toxins

MDPI

Article

Insights into the Evolution of a Snake Venom Multi-Gene Family from the Genomic Organization of *Echis ocellatus* SVMP Genes

Libia Sanz * and Juan J. Calvete *

Laboratorio de Venómica Estructural y Funcional, Instituto de Biomedicina de Valencia, Consejo Superior de Investigaciones Científicas, Jaume Roig 11, 46010 València, Spain

* Correspondence: libia.sanz@ibv.csic.es (L.S.); jcalvete@ibv.csic.es (J.J.C.); Tel.: +34-96-339-1760 (L.S.); +34-96-339-1778 (J.J.C.)

Academic Editors: Jay Fox and José María Gutiérrez
Received: 12 June 2016; Accepted: 6 July 2016; Published: 12 July 2016

Abstract: The molecular events underlying the evolution of the Snake Venom Metalloproteinase (SVMP) family from an A Disintegrin And Metalloproteinase (ADAM) ancestor remain poorly understood. Comparative genomics may provide decisive information to reconstruct the evolutionary history of this multi-locus toxin family. Here, we report the genomic organization of *Echis ocellatus* genes encoding SVMPs from the PII and PI classes. Comparisons between them and between these genes and the genomic structures of *Anolis carolinensis* ADAM28 and *E. ocellatus* PIII-SVMP EOC00089 suggest that insertions and deletions of intronic regions played key roles along the evolutionary pathway that shaped the current diversity within the multi-locus SVMP gene family. In particular, our data suggest that emergence of EOC00028-like PI-SVMP from an ancestral PII(e/d)-type SVMP involved splicing site mutations that abolished both the 3′ splice AG acceptor site of intron 12* and the 5′ splice GT donor site of intron 13*, and resulted in the intronization of exon 13* and the consequent destruction of the structural integrity of the PII-SVMP characteristic disintegrin domain.

Keywords: Snake venom toxin multi-gene family; snake venom metalloproteinase; genomic organization of SVMP genes; PII-SVMP; PI-SVMP; gene duplication; intronic retroelements; intronization

1. Introduction

The ADAM (A Disintegrin-like And Metalloproteinase) family of transmembrane type 1 proteins belongs to the MEROP database M12 family of Zn^{2+}-dependent metalloendopeptidases [1] and PFAM family PF01421 [2]. Members of the ADAM family play important roles in cell signaling and in regulating cell-cell and cell-matrix interactions [3,4]. The ADAM family comprises ancient proteins whose origin extends back >750 My [5,6]. To date, close to 40 ADAM genes have been identified in vertebrate and invertebrate bilaterian animals, both in deuterostomes, from the basal chordate, *Ciona intestinalis*, to higher vertebrates, and in protostome, such as arthropods, nematodes, platyhelminths, rotifers, molluscs, and annelids. The evolutionary history of vertebrate ADAM genes is punctuated by gene duplication and retroposition events [7,8], followed by neo- or subfunctionalization [7]. Gene duplications are an essential source of genetic novelty that can lead to evolutionary innovation if the new function has no deleterious effects to its host organism or provides selective advantages. For example, in mammalian species, including marsupials and monotremes, except the platypus, ADAM28, ADAMDEC1 (decysin, a soluble ADAM-like protein), and ADAM7 form a cluster, likely as a result of tandem duplication of ADAM28 [9]. Instead, in most non-mammalian vertebrate genomes investigated, including those of aves, reptiles, and fishes, a single ADAM28 locus is present in this region [7,10]. The data suggest that ADAM7 and ADAMDEC1 were duplicated from

ADAM28, probably only in mammals [7]. On the other hand, as described below in more detail, it is thought that ADAM28 played a starring role in the emergence of toxic metalloproteinases in the superfamily Colubroidea of Caenophidian snakes (viperids, elapids, and colubrids).

The concept that gene duplication plays a major role in evolution has been around for over a century [11]. In his classic and influential book "*Evolution by Gene Duplication*" [12] Susumo Ohno argued that gene duplication is the most important evolutionary force since the emergence of the universal common ancestor. Common sources of gene duplications include ectopic homologous recombination, retrotransposition event, aneuploidy, polyploidy, and replication slippage [13]. Duplication creates genetic redundancy, where the second copy of the gene is often free from selective pressure. Thus, over generations of the organism, duplicate genes accumulate mutations faster than a functional single-copy gene, making it possible for one of the two copies to develop a new and different function. Duplicated genes may switch their transcription to other tissues by localizing closely to, and utilizing the regulatory elements of, a neighboring gene [14–16]. Examples of this are (i) the formation of toxin gene families during the evolution of the venom system of advanced snakes by co-option, multiplication, and weaponization in the venom gland of paralogs of genes encoding for normal body proteins [17–20], and (ii) the finding of 309 distinct widow spider genes exhibiting venom gland biased expression [21], suggesting that the switching of genes to venom gland expression in numerous unrelated gene families has been a dominant mode of evolution [21–23].

Because of its functional importance for prey capture, predator defense, and competitor deterrence, venom represented a key innovation that has underpinned the explosive radiation of toxicoferan reptiles in the Late Jurassic period of the Mesozoic era, ~150 million years before present (MYBP) [24–28]. Toxicofera [18] (Greek for "those who bear toxins") is the term coined for the clade of squamate reptiles that includes the Serpentes (snakes), Anguimorpha (monitor lizards, gila monster, and alligator lizards) ,and Iguania (iguanas, agamas, and chameleons) lizards. One of the founding families of advanced snake venom comprises the Zn^{2+}-dependent metalloendopeptidases (SVMPs) [17–19,29–32]. SVMPs are key enzymes contributing to toxicity of vipers and pitvipers venoms. Hemorrhage is one of the most significant effects in envenomings induced by viperid and crotalid snakebites. Damage to the microvasculature, induced by SVMPs, is the main event responsible for this effect. In addition to hemorrhagic activity, members of the SVMP family also have fibrin(ogen)olytic activity, act as prothrombin activators, activate blood coagulation factor X, possess apoptotic activity, inhibit platelet aggregation, are proinflammatory, and inactivate blood serine proteinase inhibitors [33–36].

The closest non-venom ancestors of SVMPs was likely an ADAM28 precursor gene [37]. The origin of SVMPs has been inferred to have occurred following the split of the Pareatidae from the remaining Caenophidians, approximately 60 MYBP around the Cretaceous–Paleocene boundary of the Cenozoic Era [18,19,29,31,38]. SVMPs are found in the venoms of all advanced snakes and are classified into different classes depending upon their domain structure [39–41]. The ancestral multidomain PIII form, which is found in all snake venoms, derives from the extracellular region (metalloproteinase domain with disintegrin-like and cysteine-rich domains at the *C*-terminus) of a duplicated ADAM28 precursor gene that lost the *C*-terminal epidermal-growth-factor (EGF-)-like, transmembrane, and cytoplasmic domains [31,32,41–43]. On the other hand, the derived PII-SVMPs, comprising the metalloproteinase and *C*-terminal disintegrin domain, have been only found in venoms of vipers and rattlesnakes (Viperidae). This strongly suggests that they emerged, subsequently to the separation of Viperidae and Elapidae, ~37 million years ago, in the Eocene epoch of the Cenozoic era, but before the separation of the Viperidae subfamilies Viperinae and Crotalinae 12–20 MYBP, from a duplicated PIII-SVMP gene that lost its cysteine-rich domain (see Figures 1 and 8 in [43] and Figure 18.1 in [44]). The disintegrin domain has been lost from the PII-SVMP structure on multiple occasions, resulting in the formation of the PI class of SVMPs [45] made only by the catalytic Zn^{2+}-metalloproteinase domain [39–41].

Details on the mechanisms of co-option and the molecular events underlying the transformation of an ADAM28 precursor gene copy into the SVMP multi-gene family of extant snake venoms

remain elusive. In previous works, we described a family of RPTLN genes that exhibit a broad and reptile-specific distribution, for which we hypothesize may have played a key role in the recruitment and restricted expression of SVMP genes in the venom gland of Caenophidian snakes [46]. We have also reported the genomic organization of *Echis ocellatus* PIII-SVMP gene EOC00089, and compared it to those of its closest orthologs from *Homo sapiens* and the lizard, *Anolis carolinensis* [47]. Now, we fit two new pieces in the puzzle: the genomic structures of *E. ocellatus* PII—(EOC00006-like) and PI—(EOC00028-like) SVMP genes. Insights into post-duplication events gained from the structural comparison of the three classes of SVMP genes are discussed.

2. Results and Discussion

2.1. The Genomic Structure of Pre-Pro EOC00006-Like PII-SVMP and Pre-Pro EOC00028-Like Genes

Genomic sequences encoding full-length pre-pro EOC00006-like PII-SVMP (17828 nt) [KX219964] (Figure A1) and EOC00028-like PI-SVMP (21605 nt) [KX219965] (Figure A2) genes were assembled from overlapping PCR-amplified fragments (Appendix A, Figures A1 and A2). The pre-pro PII-SVMP gene consists of 15 exons interrupted by 14 introns (Figure 1A), whereas the pre-pro PI-SVMP gene contains 13 exons and 12 introns (Figure 1B).

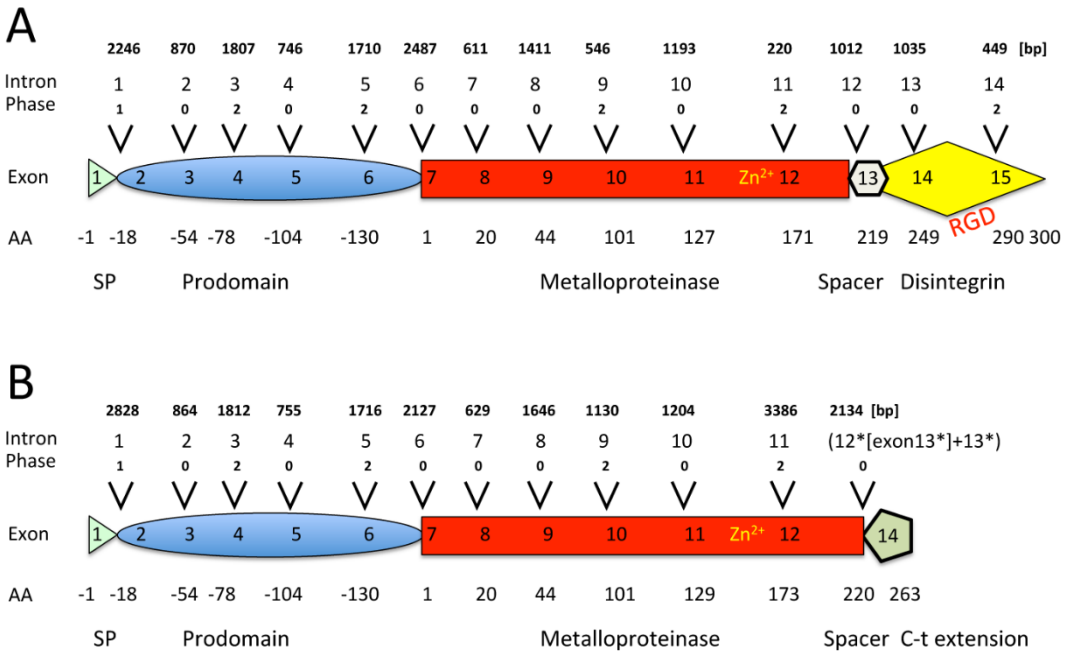

Figure 1. Scheme of the genomic organization of pre-pro EOC00006-like PII-SVMP (**A**) and pre-pro EOC00028-like PI-SVMP (**B**) genes. The distribution, phase, and size of the 14 (PII) and 12 (PI) introns and the boundaries of the protein-coding regions are highlighted. SP, signal peptide. Homologous exons and introns have identical numbering. Intron 12 of the PI-SVMP gene corresponds to the fusion of the genomic segment spanning intron12*-exon13*-intron13*. Mature PII- and PI-SVMP amino acid sequences span 299 and 263 amino acid residues, respectively. Zn^{2+}, relative location of the catalytic Zn^{2+}-binding environment; RGD, integrin-binding arginine-glycine-aspartic acid tripeptide motif.

The translated 494 (PII) and 457 (PI) pre-pro-SVMP amino acid sequences exhibit identical distribution and features (in terms of codon location and phase) for their first 11 introns and 12 exons, which code for the signal peptide (SP), prodomain (PD), metalloproteinase (MP) domain, and the short tetrapeptide (ELLQ) "spacer" sequence (Appendix A, Figures A1 and A2). These 413 (PII)/414 (PI) amino acid sequences show 85% identity, strongly suggesting that both SVMPs have a shared ancestry. It is also worth noting that the protein-coding positions interrupted by each of the introns

of the PII- and PI-SVMP genes are entirely conserved in *Anolis carolinensis* [XP_008118058] (and also in human [NG_029394]) ADAM28 gene. Introns are inserted after or between secondary structure elements, supporting the "introns-added-late" model, which proposes that during the evolution of the eukaryotic branch, introns were added at the boundaries of structural modules coded for by ancestral continuous genes [48]. In addition, as will be analyzed in detail below, pairwise alignment of topologically equivalent PII- and PI-SVMP introns show that homologous intronic nucleic acid sequences share 88%–99% identity (Figure 2). This clearly indicates that EOC00006-like PII-SVMP and EOC00028-like PI-SVMP represent paralog genes.

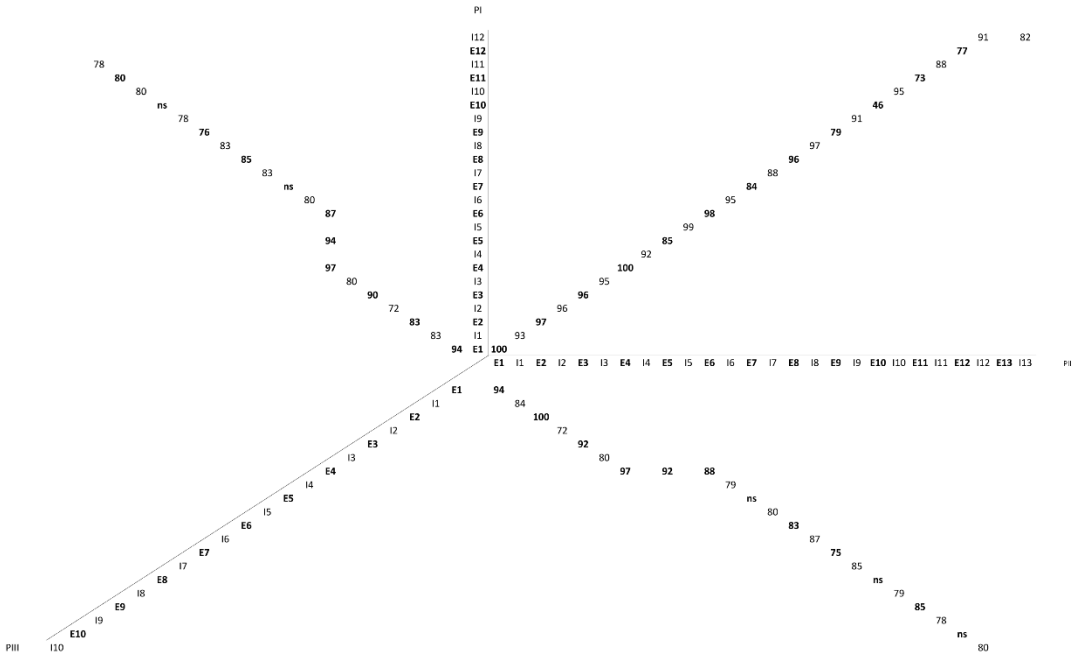

Figure 2. Pairwise comparisons of the sequence identities between the exonic and intronic nucleic acid sequences of pre-pro EOC00089-like PIII-SVMP, EOC00006-like PII-SVMP, and EOC00028-like PI-SVMP genes.

New genes can arise through four mechanisms: gene duplication, retroposition, horizontal gene transfer, and de novo origination from non-coding sequences [49]. Available evidence strongly suggests that gene duplication has played a pivotal role in the origin of venom multi-gene families [20–23,50,51]. Although the fate of many new genes may be to lose their function and become pseudogenes, some can be fixed through evolution of redundancy, subfunctionalization, or neofunctionlization. Several models have been proposed to explain functional divergence following venom toxin gene duplications [52–55]. However, this issue remains controversial and is the subject of vivid debates. The family portrait of SVMPs shows a complicated picture. SVMPs belong to different "generations", that in the canonical model for the evolutionary expansion of this multi-gene family are hierarchically related, being PIII-SVMPs the most ancient and the PII- and PI- SVMPs the succesively most recently derived family members [31,32,42]. However, due to the limited genomic information available, this model can be confounded by high rates of protein amino acid sequence divergence [56], and the occurrence of alternative routes (e.g., PIII > PI) can not be presently ruled out. The only other full-length viperid SVMP gene sequenced to date is *E. ocellatus* EOC00089-like PIII-SVMP [47] [KX219963]. The ORF encoding the pre-pro-metalloproteinase domains of this gene exhibits 63% amino acid sequence identity with the homologous coding regions of the PII- and PI-SVMPs here reported, and 72%–83% nucleotide sequence identity between topologically equivalent PIII-, PII-, and PI-SVMP introns (Figure 2). Although these figures clearly point to a common origin, it is not possible to infer whether they belong to the same or to a different PIII > PII > PI hierarchical lineage. Nonetheless, the

fact that the PIII-SVMP gene has lost introns 5 and 6 (ADAM28 numbering), with the consequence that exons 4, 5, and 6 have merged into a single exon, suggests that either these events occurred after the duplication that gave rise to the PII-SVMP ancestor, or that the PIII-SVMP EOC00089-like gene does not lay in the direct line of descent of the EOC00006-like PII-SVMP and EOC00028-like PI-SVMP genes. Refinement of the family tree of the multi-gene family of *E. ocellatus* SVMPs will surely emerge from future comparative genomic analysis of the carpet viper and other viperid species.

2.2. Role of Introns in the Evolution of the SVMP Multi-Gene Family

Since their discovery in 1977 [57,58], introns have been the subject of considerable debate. It is now generally accepted that introns represent more than merely junk DNA that must be pruned from pre-mRNAs to yield mature, functional mRNAs prior to their translation. Mounting evidence indicates that while introns do not encode protein products, they play essential roles in a wide range of gene expression regulatory functions such as non-sense mediated decay [59], mRNA export [60], and regulation of the amount of recombination between the flanking exons [61], or they serve as locations for nonhomologous recombination that would allow for exon shuffling [62,63]. As discussed below, most of the structural divergence between the EOC00006-like PII-SVMP and EOC00028-like PI-SVMP genes is due to the different size of their topologically equivalent eleven (1–11) introns (Supplementary Figure S1). The role of introns in the evolution of snake venom gene families remains elusive. However, in other biological systems, i.e., *Arabidopsis* and *Drosophila*, intron features, such as sequence and length, have been shown to function in maintaining pre-mRNA secondary structure, thus influencing temporal and spatial patterns of gene expression by modulating transcription efficiency and splicing accuracy [64–67].

Most PII- and PI-SVMP introns belong to phase 0, followed by phase 2; and, in both genes, only intron 1, separating the monoexonic signal peptide from the start of the prodomain, is a phase 1 intron (Figure 1). Analysis of the exon–intron structures of a large number of human genes has revealed a statistically highly significant enrichment of phase 1 introns flanking signal peptide cleavage sites [68]. Phase 1 introns most frequently split the four GGN codons encoding glycine. A plausible explanation for the correlation between signal peptide domains and the intron phase is that the base preferences of proto-splice sites [69,70] mirrors the amino acid preference for glycine in the signal peptidase consensus cleavage site [71].

The signal peptide is the most conserved structural element between pre-pro EOC00006-like PII-SVMP and EOC00028-like PI-SVMP is (Figure 2). In both genes, it is encoded by identical exon 1 amino acid sequences (Figures A1 and A2), which is also highly conserved in present-day SVMPs [46]. These findings support the view that co-option of this signal peptide may have played a role in the restricted expression of SVMP genes in the venom gland of Caenophidian snakes, some 60–50 Mya [46].

Nucleotide sequence comparison of the topologically equivalent introns of the *E. ocellatus* PII- and PI-SVMPs (Supplementary Figure S1) provide insights into the events underlying the conversion of a PII-SVMP into a PI-SVMP gene. In this regard, some introns differ in the number and location of intronic retroelements (Table 1). Thus, insertions in introns PI-SVMP 1 and 9 introduced complete and truncated SINE/Sauria elements in positions 1764–2101 (Figure S1, panel A) and 321–502 (Figure S1, panel I), respectively. The inserted nucleic acid sequence in intron 9 retains the GT-AG splicing sites, indicating that this insertion event created a twintron, an intron within an intron. PII-SVMP intron 6 (Figure S1, panel F) and PI-SVMP introns 11 (Figure S1, panel K) and 12 (Figure S1, panel L) are also twintrons. Compared to its topologically equivalent PII-SVMP intron, a large insertion in intron 11 of the EOC00028-like PI-SVMP gene replaced the first 66 nucleotides for a longer stretch of 3281 nucleotides; region 2461–2561 of the inserted nucleic acid sequence is 97% identical to *Hyla tsinlingensis* Hts-35 [KP204922], a microsatellite sequence that is also partly present in intron 61 of Podarcis reelin (RELN) genes [GU181006-13] (positions 554–623) [72]. Microsatellites are simple nucleotide sequence repeats (SSR) ranging in length from two to five base pairs that are tandemly repeated, typically 5–50 times (reviewed in [73]). These non-coding elements are abundant in major

lineages of vertebrates. Mammalian, fish, and squamate reptile genomes appear to be relatively microsatellite rich [74]. However, besides Hts35, RepeatMasker only identified few SSR tracks in introns 1 (5× GTTT; 28× TC) and 2 (13× ATTT; 4× TAA) of the PII-SVMP gene (Figure A1), and introns 1 (11× GTTT; 21× AG) and 2 (9× GTTT; 4× TAA) of the PI-SVMP gene (Figure A2).

Table 1. Comparison of type and location of retroelements identified in introns of *E. ocellatus* PII-SVMP EOC00006-like and PI-SVMP EOC00028-like genes.

Intron	PII-SVMP	PI-SVMP
	Inserted Retroelement	
1	SINE/Sauria	2 SINE/Sauria, LTR/ERV1, DNA/hAT-Ac
3	LINE/L2/CR1	LINE/L2/CR1
5	LINE/L2/CR1	LINE/L2/CR1
6	SINE/Sauria	-
8	LINE/L2/CR1	-
9	-	SINE/Sauria
10	DNA transposon	DNA transposon

Growing evidence supports that repetitive intronic elements, such as the long interspersed elements (LINEs) and the short interspersed elements (SINEs) contained in several introns of both PII- and PI-SVMP genes (Table 1) can influence genome stability and gene expression (reviewed in [75]). Thus, these interspersed repeats may alter genome recombination structure and rates, through a number of mechanisms, including replication slippage and unequal crossover [76,77], potentially impacting regulation of gene expression [78], recombination events leading to tandem duplication of segments of the genome [79,80], gene conversion [81], and chromosomal organization [79]. Moreover, the insertion of interspersed repeats into a new genomic position may introduce promoter or enhancer sequence motifs for transcription of nearby genes [82,83], and alternative splicing sites or polyadenylation sites [84], thereby resulting in a change of overall level of gene expression. Interspersed repetitive elements have also played an important role in expanding the repertoire of transcription factor binding sites in eukaryotic genomes [85]. However, whether these elements have contributed to the genomic context that facilitated the evolution and radiation of venom *loci* in snakes deserves future detailed comparative genomic studies.

2.3. A Fusion Event Led to the Conversion of a PII(e/d)-Type SVMP into EOC00028-like PI-SVMP

PI-SVMP intron 12 is a twintron resulting from the fusion of the genomic region spanning ancestral introns 12* and 13* and exon 13* (homologous to identical numbered elements in the genomic structure PII) (Figures 1 and 3A). Splicing site mutations affecting both the 3′ splice AG acceptor site of intron 12* and the 5′ splice GT donor site of intron 13* led to the retention, and subsequent intronization, of exon 13* within a fused (12* + 13*) twintron (Figure 3A). Intronization of exon-coding nucleic acid sequences has been proposed as a major contributor to intron creation [86]. Intron 13* encoded part of the *N*-terminal region of a disintegrin domain, most likely, as discussed below, an eventual subunit of dimeric disintegrin. In addition to the disruption of the structural integrity of the disintegrin domain, a stop codon after exon 14 removed intron 14 and exon 15 from the PII(e/d)-type SVMP (Fox & Serrano's nomenclature [40]) precursor gene structure, thereby completing the conversion of the PII-SVMP into present EOC00028-like PI-SVMP gene (Figure 3A).

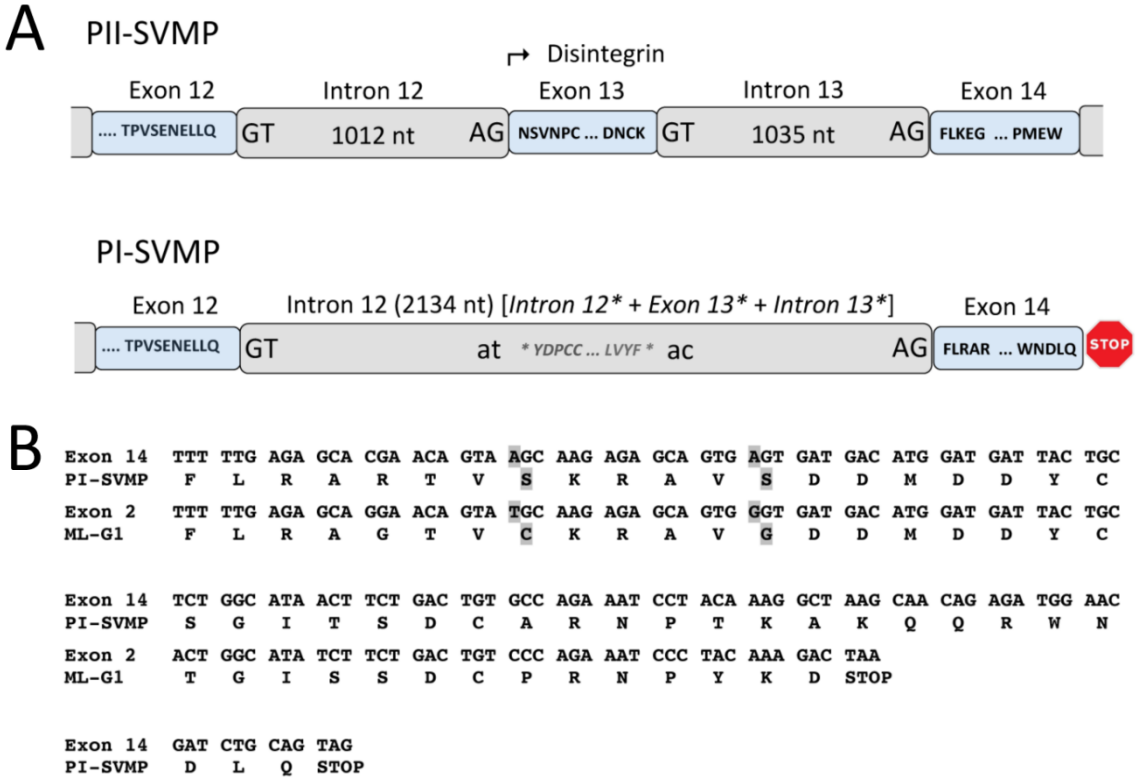

Figure 3. Panel **A**, cartoon comparing the 3′ regions of the PII-SVMP and PI-SVMP genes and highlighting the processes (intronization of ancestral exon 13* inside twintron 12 resulting from the fusion of introns 12* and 13*, and creation of a stop codon after exon 14) that destroyed the integrity of the disintegrin domain, converting an ancestral PII(e/d)-type SVMP into extant EOC00028-like PI-SVMP. Panel **B**, alignment of the amino acid sequences encoded by exon 14 of EOC00028-like PI-SVMP and exon 2 of the dimeric disintegrin subunit ML-G1 [AM261811] [87]. Degeneration of PI-SVMP's conserved functional and structural amino acid residues in dimeric disintegrins are highlighted in boldface and grey background.

Region 1013–2134 of PI-SVMP intron 12 exhibits 91% nucleotide sequence identity with range 14 to 1135 of *Macrovipera lebetina* gene encoding part of exon 1 and full-length intron 1 of the VGD-containing dimeric disintegrin subunit precursor, ML-G1 [AM261811] [87]. PI-SVMP exon 14 (mature protein amino acid residues 221–263, Figure A2) exhibits strong homology (79% identity) to exon 2 of the same VGD-bearing dimeric disintegrin subunit. The PI-SVMP exon 14 shows the consequences of genetic drift (Figure 3B): the conseved $\alpha_5\beta_1$ integrin-inhibitory VGD tripeptide motif [44] of the PII-SVMP precursor gene has been replaced by a VSD motif (generated by a G > A mutation: GTG AGT GAT > GTG GGT GAT), and the absolutely conserved tenth cysteine residue of dimeric disintegrin subunits has degenerated (TGC) to a serine residue (AGC) (Figure 3B).

3. Concluding Remarks and Perspectives

The event that gave birth to the family of SVMPs was the generation of a STOP codon at the 3′ end of exon 16 of a duplicated ADAM28 gene (Figure 4). This mutation produced an ORF truncated at the *N*-terminal part of the EGF-like domain, which encoded a precursor of an ancestral PIII-SVMP lacking this domain and the *C*-terminal membrane anchoring and cytoplasmic polypeptides (Figure 4). On the other hand, our results comparing the available genomic structures of SVMP genes, e.g., EOC00089-like PIII-SVMP [47] [KX219963], EOC00006-like PII-SVMP [KX219964], and EOC00028-like PI-SVMP [KX219965] (this work), suggest that the evolutionary history of SVMPs is marked with events of insertions and deletions of intronic regions. This scenario points to introns as key players in

the formation of the multi-locus SVMP gene multifamily. Thus, comparison of the genomic structures of EOC00089-like PIII-SVMP and EOC00006-like PII-SVMP (Figure 5) indicates that replacement of the PIII-specific cysteine-rich domain by a non-homologous region encoding intron 14-exon 15 followed by a STOP codon may represent a step in the conversion of a PIII-SVMP into a PII-SVMP.

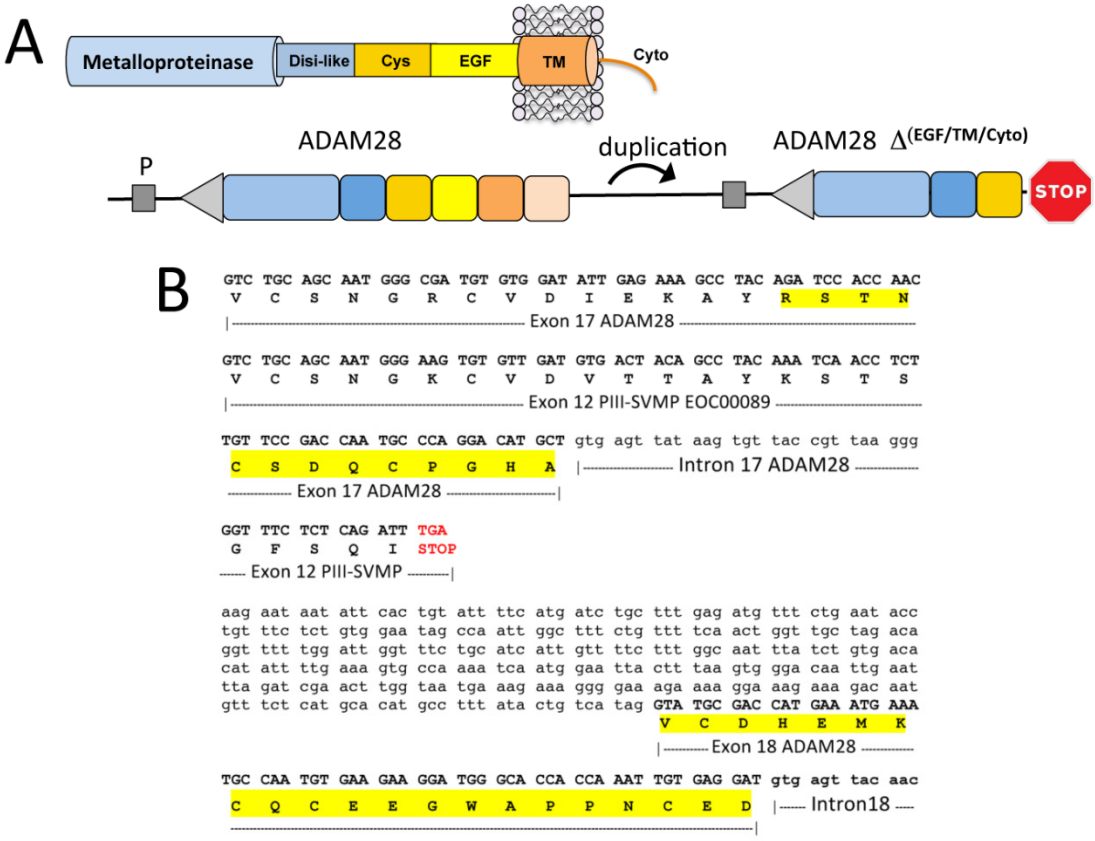

Figure 4. Comparison of the genomic region encompassing exons 17 through 18 of *Anolis carolinensis* ADAM28 [XP_003226913] and the homologous amino acid sequence of *E. ocellatus* SVMP EOC00089 [ADW54351], highlighting the STOP codon after exon 12 of the latter generating a *C*-terminally truncated molecule, which eventually gave rise to the ancestor of the PIII-SVMPs.

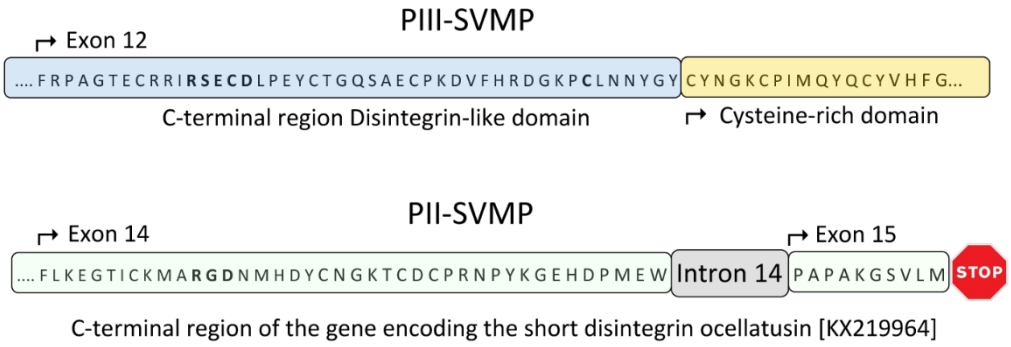

Figure 5. Comparison of the genomic region encoding the *C*-terminal domains of *E. ocellatus* EOC00089-like PIII-SVMP and EOC00006-like PII-SVMP, suggesting that 3′ genomic remodeling represents a seminal step in the generation of PII-SVMPs.

This view is consistent with structural evidence suggesting that the loss of the cysteine-rich domain represents an early seminal event that facilitated the formation of PII class SVMPs [43].

The PII subfamily of SVMPs is characterized by the diversity of disintegrin domains exhibited by different family members [39,40], ranging from the more ancestral long disintegrin domains (~84 amino-acid-residue polypeptide cross-linked by 7 disulfide linkages) to the more recently evolved short disintegrin (41–51 amino-acid-residues crosslinked by 4 disulfide bonds) [42]; for a scheme of the evolutionary path of the disintegrin domains, see Figure 1 in [43]. EOC00006-like is an example of a PII-SVMP with short disintegrin domain. Given the structural diversity of PII-SVMPs, genomic sequences from the different members of the subfamily are required for a more accurate glimpse of the genomic mechanisms operating in the generation and subsequent diversification of PII-SVMPs.

Comparison of the EOC00006-like PII-SVMP and EOC00028-like PI-SVMP gene structures also points to genomic remodeling of the 3′ region of a PII(e/d)-type SVMP precursor gene [39,40] as the EOC00028-like PI-SVMP gene generator mechanism. The PII > PI conversion involved the generation of twintron 12 (by fusion of introns 12* and 13*) and the loss, by intronization, of exon 13*, thereby destroying the consistency of the region coding for the disintegrin domain. This elaborated mechanism indicates that the structural diversification of SVMPs is not due to a random mutation generating a STOP codon before the disintegrin domain, but follows a well orchestrated sequence of events imprinted in the genome of snake species sometime after the split of Viperidae and Elapidae, 37 million years ago, but before the separation of the Viperidae subfamilies Viperinae and Crotalinae 12–20 MYBP. The mechanisms underlying loss or gain of spliceosomal introns are still poorly understood. The most widely accepted hypothesis is that intron insertion may occur via a process similar to group II intron retrotransposition [88,89]. According to this view, the spliceosomal components remain transiently associated with a recently excised intron and then attach at a potential splice site of a non-homologous pre-mRNA, where they catalyze the reverse reaction [90,91]. The modified pre-mRNA is reverse-transcribed and the resulting cDNA participates in a recombination with its parent gene, thereby inserting a novel intron into the target gene [90–93]. An attractive feature of this mechanism is that it ensures that the inserted nucleic acid sequence has the full complement of intron signature sequences required for efficient splicing [94].

Studies of multi-gene protein families are crucial for understanding the role of gene duplication and genomic exon-intron organization in generating protein diversity. For example, full-length genomic sequences of Crotalinae group II PLA_2 isogenes from *P. flavoviridis* (Tokunoshima and Amami-Oshima islands, Japan) [95], and *T. gramineus* (Taiwan) [96] have been reported. All these genes exhibit four coding regions and conserved exon-intron structures spanning about 1.9 kb. A cluster of five tandemly arranged PLA_2 genes have been located in a 25 kb 3′ segment of a 31 kb fragment of the Amami-Oshima *P. flavoviridis* genome [97], which in addition harbors a PLA_2 pseudogene in its 6 kb 5′ region [98]. Genomic sequence comparisons between the pancreatic PLA_2 gene of *P. elegans*, group IB pancreatic PLA_2 gene of *L. semifasciata*, and the *L. semifasciata* group IA venom PLA_2 gene, suggest that Crotalinae group II venom PLA_2 genes emerged before the divergence of Elapinae and Crotalinae, whereas groups of IB and IA PLA_2 genes appeared after Elapinae was established as a taxonomic lineage [99].

Duplicated structures found in eukaryotic genomes may result from complex interplays between different mechanisms [100]. Mitotic and meiotic non-allelic homologous recombination (NAHR) events, resolved as unequal crossing-over, have been traditionally invoked to account for segmental duplications within genomes [101,102]. Duplicated regions can be organized as direct tandems (e.g., the cluster of tandem snake venom PLA_2 genes), but also be separated by hundreds of kb [100]. Our present and previous work [47] inaugurate a line of research that will allow the depiction of a more precise characterization of the genomic context in which the SVMP multi-gene family has emerged. This goal demands populating the current databases with genomic sequences of genes representing the different members of the SVMPs. Although the variety of structural forms comprising the PII family may be considered a challenge for this purpose, this circumstance can be also regarded as a valuable opportunity for the step-by-step description of the molecular pathways that led to the formation of this multi-gene family. Without a doubt, ongoing Viperidae snake genome sequencing projects will mark

the beginning of comparative snake genomics, and will be key to revealing not only the topology and copy number of the genes encoding SVMPs, but also to provide decisive information to reconstruct the evolutionary history of this multilocus gene family.

4. Materials and Methods

4.1. Genomic DNA

Genomic DNA was extracted from the fresh liver of *E. ocellatus* (Kaltungo, Nigeria) maintained at the herpetarium of the Liverpool School of Tropical Medicine. *Echis ocellatus* liver was ground to a fine powder under liquid nitrogen and the genomic DNA extracted using a Roche DNA isolation kit for cells and tissue containing SDS (2% final concentration) and proteinase K (400 μg/mL final concentration). The homogenates were incubated at 55 °C overnight. Thereafter, 300 μL of 6 M NaCl (NaCl-saturated H_2O) was added to each sample, and the mixture was vortexed for 30 s at maximum speed and centrifuged for 30 min at 10,000 *g*. An equal volume of isopropanol was added to each supernatant, and the sample mixed, incubated at −20 °C for 1 h, and centrifuged for 20 min at 4 °C and 10,000 *g*. The resulting pellets were washed with 70% ethanol, dried, and, finally, resuspended in 300–500 μL sterile distilled H_2O.

4.2. Strategy for PCR Amplification of Overlapping Genomic DNA Fragments

For sequencing *E. ocellatus* genes encoding PII-SVMP EOC00006 [Q14FJ4] and PI-SVMP EOC00028 [Q2UXQ3] we employed a similar iterative process as described in [47]. Full-length cDNA-deduced amino acid sequences of disintegrin domains [103] and of the genomic organization of dimeric disintegrin domains [AM286800] [87] and PIII-SVMP EOC00089 [47] from the same species were used as templates to design primers for the PCR-amplification of protein-specific genomic sequences (Table 2).

PI-SVMP stretch ^{72}AREILNS.....QRWNDLQ^{263} was amplified on an Eppendorf Mastercycle® epgradient S instrument in a 50 μL reaction mixture containing 17.5 μL of H_2O, 25 μL Master-Mix (Thermo Scientific, Waltham, MA USA) including buffer, dNTPs, and Phusion High-Fidelity DNA polymerase, 2.5 μL of each primer (10 μM) Met1PIRv and Met5PIFw, 1.5 μL of DMSO (100%), and 1 μL of genomic DNA (50 ng/μL). PCR conditions included an initial denaturation step at 98 °C for 30 s followed by 35 cycles of denaturation (20 s at 98 °C), annealing (15 s at 63 °C), extension (300 s at 72 °C), and a final extension for 5 min at 72 °C. All other PCR amplifications were carried out in the same thermocycler using iProof High Fidelity polymerase (BioRad, Hercules, CA, USA). The 50 μL reaction mixture contained 10 μL of 5×buffer, 1 μL of 10 mM (each) dNTPs, 2 μL of $MgCl_2$ 50 mM, 1.5 μL of DMSO (100%), 1 μL of each Fw and Rv primer (10 μM), 1 μL of genomic DNA (50 ng/μL), and 32.5 μL of water. PCR conditions included an initial denaturation step at 98 °C for 120 s followed by 35 cycles of denaturation (10 s at 98 °C), annealing (15 s at the lower melting temperature of the primers), extension (60 s per Kb at 72 °C), and a final extension for 5 min at 72 °C.

4.3. Purification and Cloning of PCR Products

PCR-amplified DNA fragments were purified from agarose electrophoretic bands using the GENECLEAN Turbo kit (MP Biomedicals). The purified fragments were inserted into pJET_1.2 (Thermo Scientific, Waltham, MA USA) using phage T4 ligase and cloned into *E. coli* DH5α by electroporation at 1700 V. Transformed cells, resuspended in 200 μL LB medium, were incubated at 37 °C for 1 h, and were subsequently plated on LB agar/ampicilline to select positive clones. The presence of the inserted DNA fragments was verified by PCR amplification or digestion of the expression vector with the restriction enzyme Bgl II. The inserted DNA fragments were sequenced in-house on an Applied Biosystems model 377 DNA sequencing system (Foster City, CA, USA) using pJETFw and pJETRv primers.

Table 2. Forward (Fw) and reverse (Rv) primers used to PCR-amplify genomic DNA stretches from *E. ocellatus* PII-SVMP EOC00006-like (**right**) and PI-SVMP EOC00028-like (**left**) genes.

Primer	DNA sequence	Primer	DNA sequence
Sp35_Eo Fw	ATGATCCAAGTTCTCTTGGTAACTATATGCTTAGC	5′ PS-Disi Fw	ATGATCCAAGTTCTCTTGG
Met14PI Fw	CTATATGCTTAGCAGTTTTTCCATATC	Intr4 Fw	ATGACACTGACCTCTAGAGTTGG
Intr1F1PI Fw	CTAGTCATTCCGGCCATATGAC	IntrB9_4-2 Fw	AAGCTTGCTTGCTAGTAGGTGG
Intr2F1PI Fw	ATCAGTCTGAGAGGATGCATTTCC	Intr4 Rv	TGGACATTGTATGGTCACCTG
Intr3F1PI Fw	GTGACCATGCAATGTCCATATG	Prodom 3 Fw	GGAGCTTTTAAGCAGCCAGAG
Met15PI Fw	GTTGCCTGTAGGAGCTGTTAAG	Prodom 3 Rv	CTCTGGCTGCTTAAAAGCTCC
Prodom 2 Fw	GACGCTGTGCAATATGAATTTG	Prodom 2 Fw	GACGCTGTGCAATATGAATTTG
Prodom 2 Rv	CAAATTCATATTGCACAGCGTC	Prodom 2 Rv	CAAATTCATATTGCACAGCGTC
Intr3 Rv	GCACCAACTCTGTATCTCAGTC	Intr3 Fw	CACAGGTAAATAAGCCACAAACACC
Pro2 Fw	CAGTGAGACTCATTATTCCCCTGATGGCAG	Intr3 Rv	GCACCAACTCTGTATCTCAGTC
Pro3 Rv	CTGCCATCAGGGGAATAATGAGTCTCACTG	Pro2-SVMP_Fw	CAGAAGATTACAGTGAGACTCATTA TTCCCWGATGG
IntrB13-1 Fw	CTTGCCTCCCTATAGGATCACTGC	Pro3-SVMP_Rv	CTGCCATCAGGGGAATAATGAGTCTCACT
Met16PI Rv	GATGCGTCCATAATAATAGCAGTG	IntrB13-1 Fw	CTTGCCTCCCTATAGGATCACTGC
Prodom 1 Fw	GATGCCAAAAAAAAGGATGAGG	Prodom 1 Fw	GATGCCAAAAAAAAGGATGAGG
Prodom 1 Rv	CCTCATCCTTTTTTTTGGCATC	Prodom 1 Rv	CCTCATCCTTTTTTTTGGCATC
IntronB7PI Fw	TGGAACAACAGCTGTTGTTATGACG	Intr2 Fw	ACAATGGGAAACTGAGGAACAG
IntronB7PI Rv	TGAGAGACATGCTGATGTGGTC	Intr2 Rv	GGGAACTCTGACTTAGAGAAAGTC
Met4 PI Fw	GACCCAAGATACATTCAGCTTGTC	Met1PII Fw	CAACAGCATTTTCACCCAAGATAC
Met4 PI Rv	GACAAGCTGAATGTATCTTGGGTC	Met1PII Rv	GTATCTTGGGTGAAAATGCTGTTG
Met8PI Rv	TATCCATGTTGTTATAGCAGTTAAATC	Met 1-2 Fw	CATGGATACATCAAATTGTCAACG
Intron B16 Fw	TGTGCTTACCCAACACTGAGCC	Met 1-3 Rv	TGTACATCTGTCAGGTGGACATG
Met5 PI Fw	GCACGTGAAATTTTGAACTCA	Met2PII Fw	GCCGTTCACCTTGATAACCTTATAGG
Met5PI Rv	GAGTTCAAAATTTCACGTGCTG	Met2PII Rv	CCTATAAGGTTATcAAGGTGAACGGC
Met9PI Rv	AGCATTATCATGCGTTATGCG	Met 6 PII Fw	CCACAATCGTCTGTAGCAATTACTGA
Met3 PI Fw	GGAAGAGCTTACATGGAGAG	Met 6 PII Rv	TCAGTAATTGCTACAGACGATTGTGG
Met3PI Rv	CTCTCCATGTAAGCTCTTCC	Met3 PII Fw	GATCATAGCACAGATCATCTTTGG
Met2PI Rv	GCTCCCCAGACATAACGCATC	Met3PII Rv	CCAAAGATGATCTGTGCTATGATCc
IntrB23PI Fw	CTGACTATGACTCACTTAACAACTGG	Met 4 Fw	ATGATCCAGGTTCTCTTGGTAACTATATG
IntrF2PI Fw	GGCCGCGTGAATGCATCTGCTTC	Met 4 Rv	TGAACTGATAGGAACGGTATTGTG
Intr2F2PI Fw	GCATCAGTTTGTTCGCACTCAATAAAG	Fw_Ocella NcoI	ATCCATGGTAGACTGTGAATCTGGACC
Intr3F2PI Fw	GAGCATAATCTGGAACTAAGATCAAG	IntrDis1 Rv	ATACGGCTAGTATGGAGCAGG
Met7PI Fw	GCACAAGATTCCTATCACTTCAG	Dis PII Rv	TCACATCAACACACTGCCTTTTGC
Met13PI Rv	TCCTACCTGCAAAAGTTCATTTTC	-	-
Intron B10PI Rv	CTGACTCAGGGCACCAATCTC	-	-
Met1PI Rv	CTACTGCAGATCGTTCCATCTCTG	-	-

4.4. Sequence Analysis

Exon-intron boundaries were localized by visual inspection and corroborated using Wise2 [104]. Amino acid and nucleotide sequence similarity searches were done using BLAST [105]. Multiple sequence alignments were performed using ClustalW2 [106]. The occurrence of retrotransposable elements and simple nucleotide sequence repeats (SSRs) were assessed using RepeatMasker (version rm-20110920) [107], a program that screens DNA sequences for interspersed repeats and low complexity DNA sequences included in the Repbase database [108].

4.5. Sequence Availability

Pre-pro EOC00006-like PII-SVMP and EOC00028-like PI-SVMP gene sequences have been deposited with the NCBI GeneBank [109] and are accessible under accession codes KX219964 and KX219965, respectively.

Supplementary Materials: The following are available online at www.mdpi.com/2072-6651/8/7/216/s1, Figure S1: Pairwise nucleotide sequence alignments of topologically equivalent paralog introns 1–12 from Pre-pro EOC00006-like PII-SVMP and 1–13 from Pre-pro EOC00028-like PI-SVMP gene sequences.

Acknowledgments: This work has been financed by Grant BFU2013-42833-P from the Ministerio de Economía y Competitividad, Madrid (Spain).

Author Contributions: J.J.C. and L.S. conceived and designed the experiments; L.S. performed the experiments; J.J.C. and L.S. analyzed the data; J.J.C. wrote the paper.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Genomic sequence of *E. ocellatus* EOC00006-like PII-SVMP gene. The locations and identities of the primers used to PCR-amplify genomic sequences (listed in Table 2) are indicated. Protein-coding DNA regions are in upper letters and boldface, and the encoded amino acid sequence is displayed below the DNA sequence. Start of introns are labelled EoPII-X, where "X" corresponds to intron number. The beginning and the signal peptide, propeptide, metalloproteinase and the short-disintegrin domains are specified. Numbers at the right correspond to amino acid numbering of the DNA-deduced pre-pro-PII-SVMP relative to the mature SVMP. The *N*-terminal glutamine of the metalloproteinase domain has been assigned residue 1. The extended Zn^{2+}-binding environment (HEXXHXXGXXH) and the RGD integrin inhibitory motif stand on yellow background. The only two amino acids (−70I/V, and −111H/R) that distinguish this sequence from that of PII-SVMP EOC00006 (Q14FJ4) are shown in bold and red. The remains of a disintegrin-like domain transformed into intron EoPII-12 are underscored in italics and on cyan background. SINE/Sauria, LINE/L2/CR1 and DNA transposon retroelements are highlighted on a gray background. Simple sequence repeats (SSR, microsatellites) are shown in light green background.

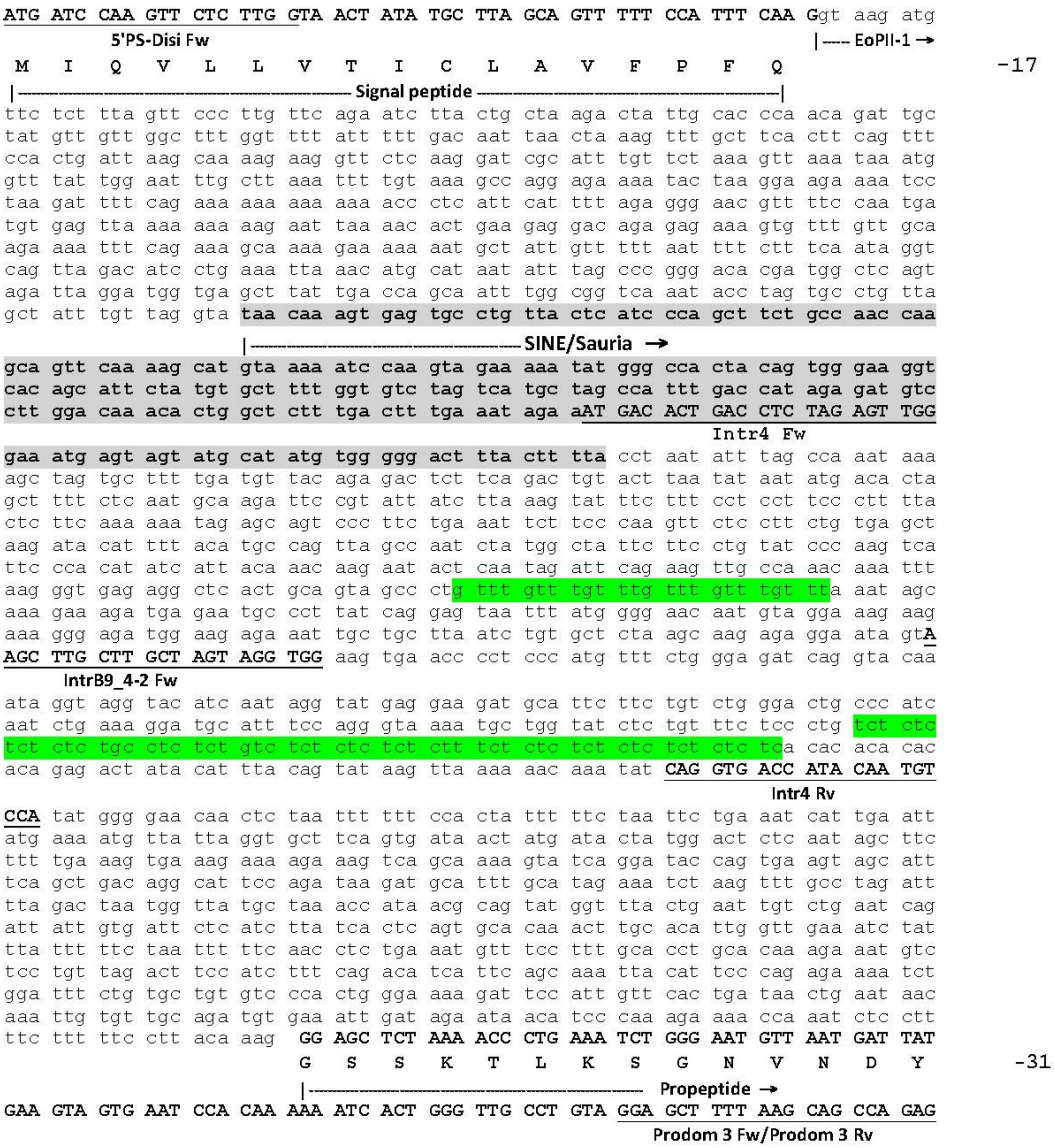

Figure A1. *Cont.*

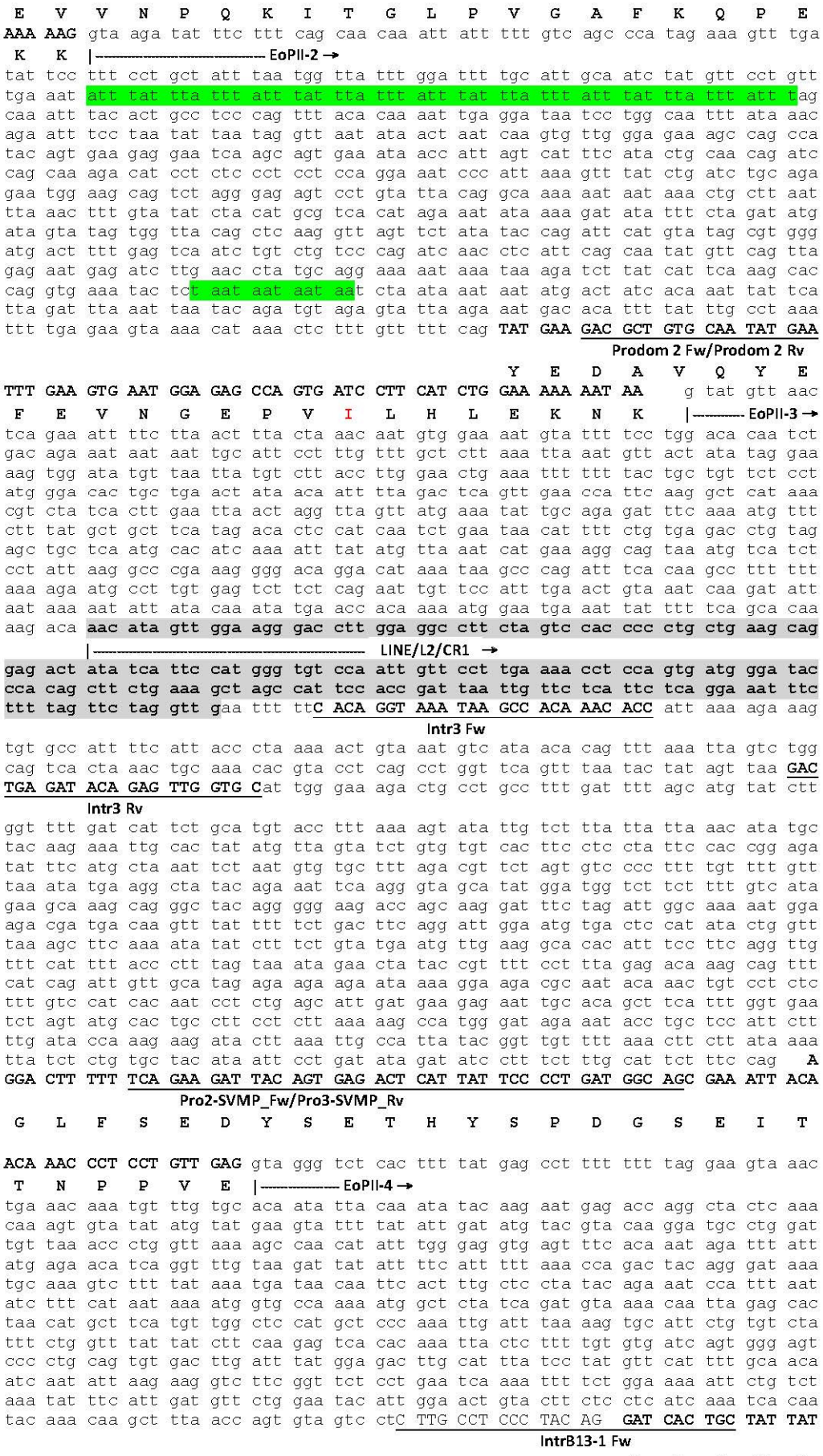

Figure A1. *Cont.*

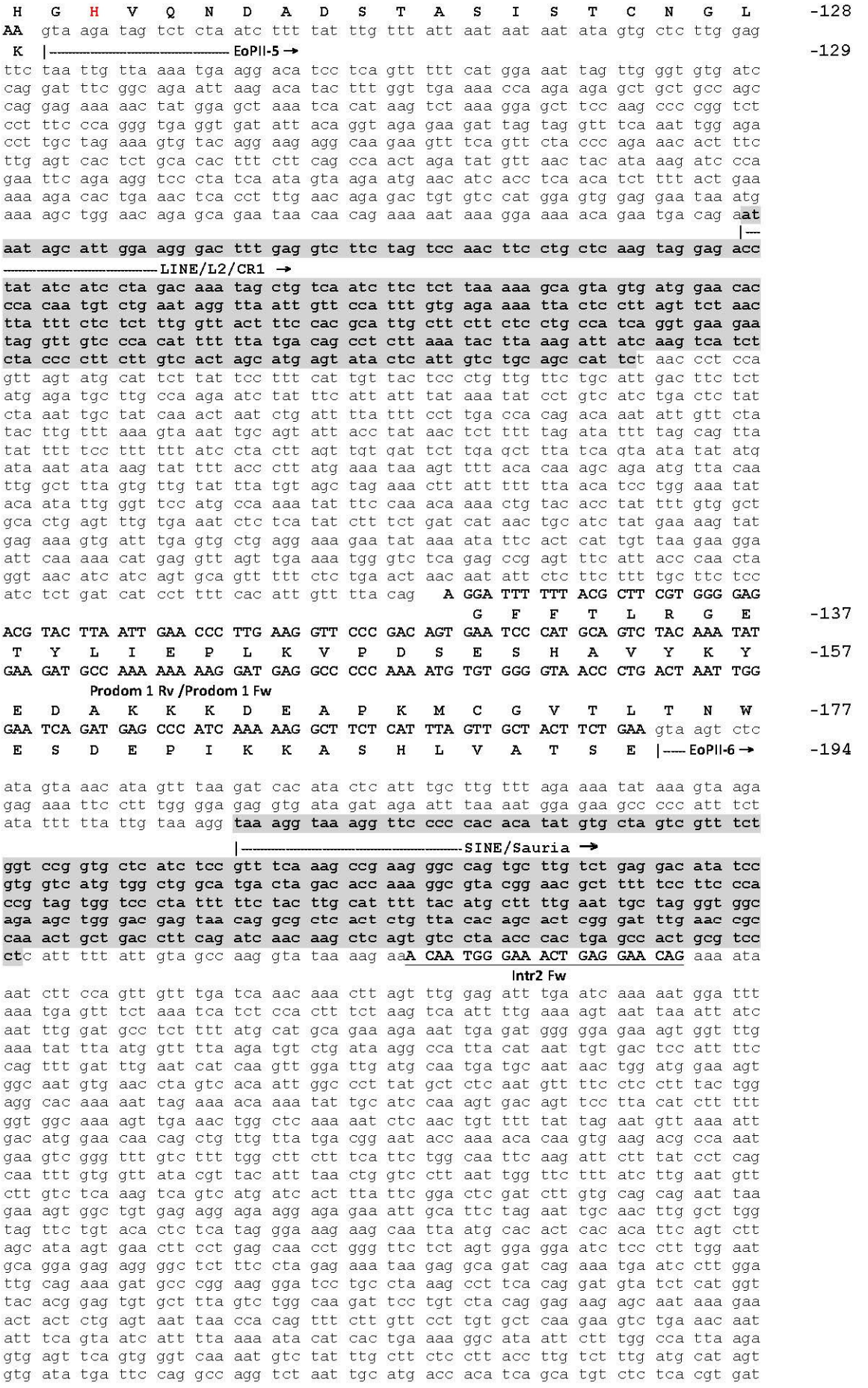

Figure A1. *Cont.*

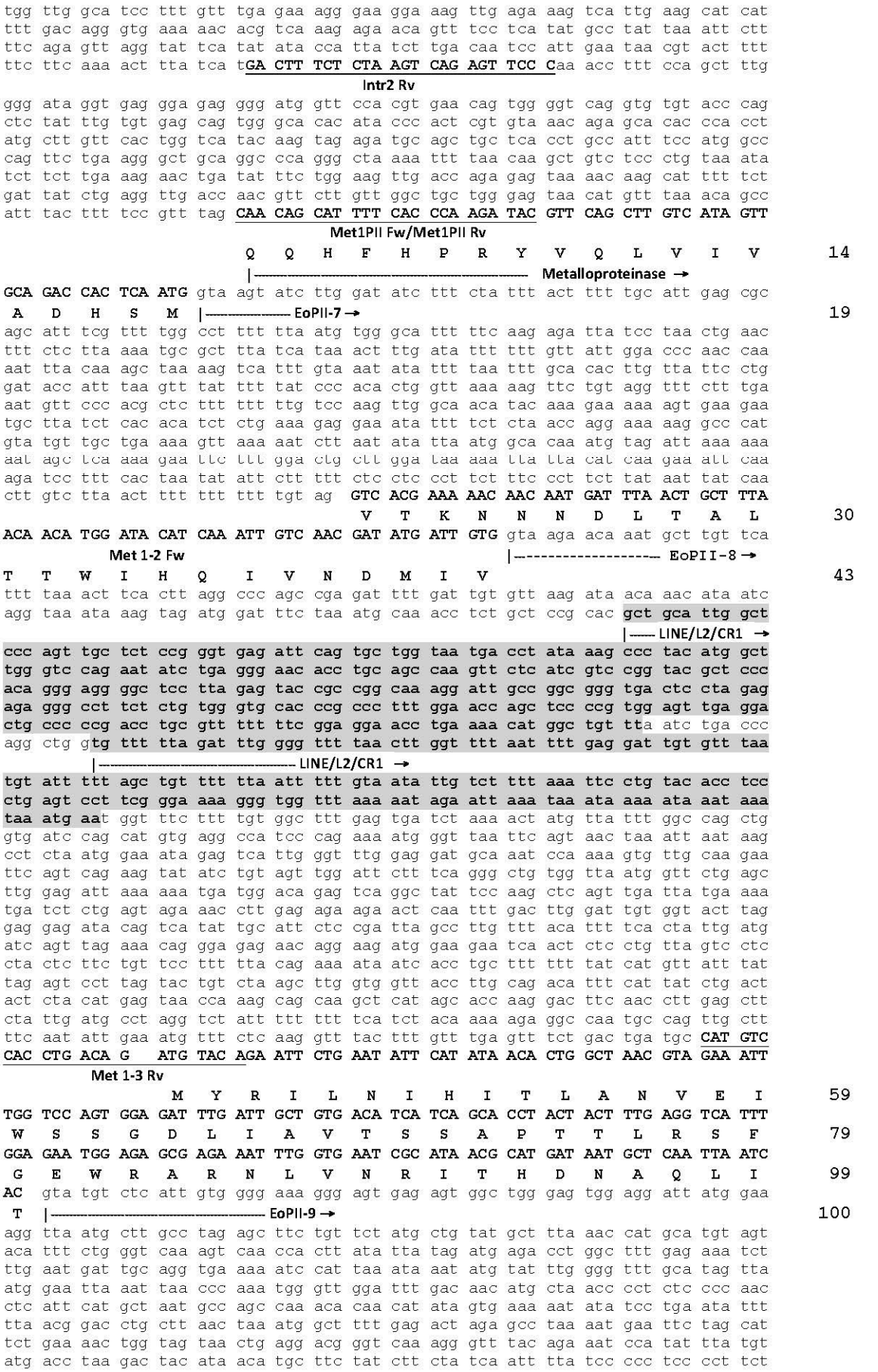

Figure A1. *Cont.*

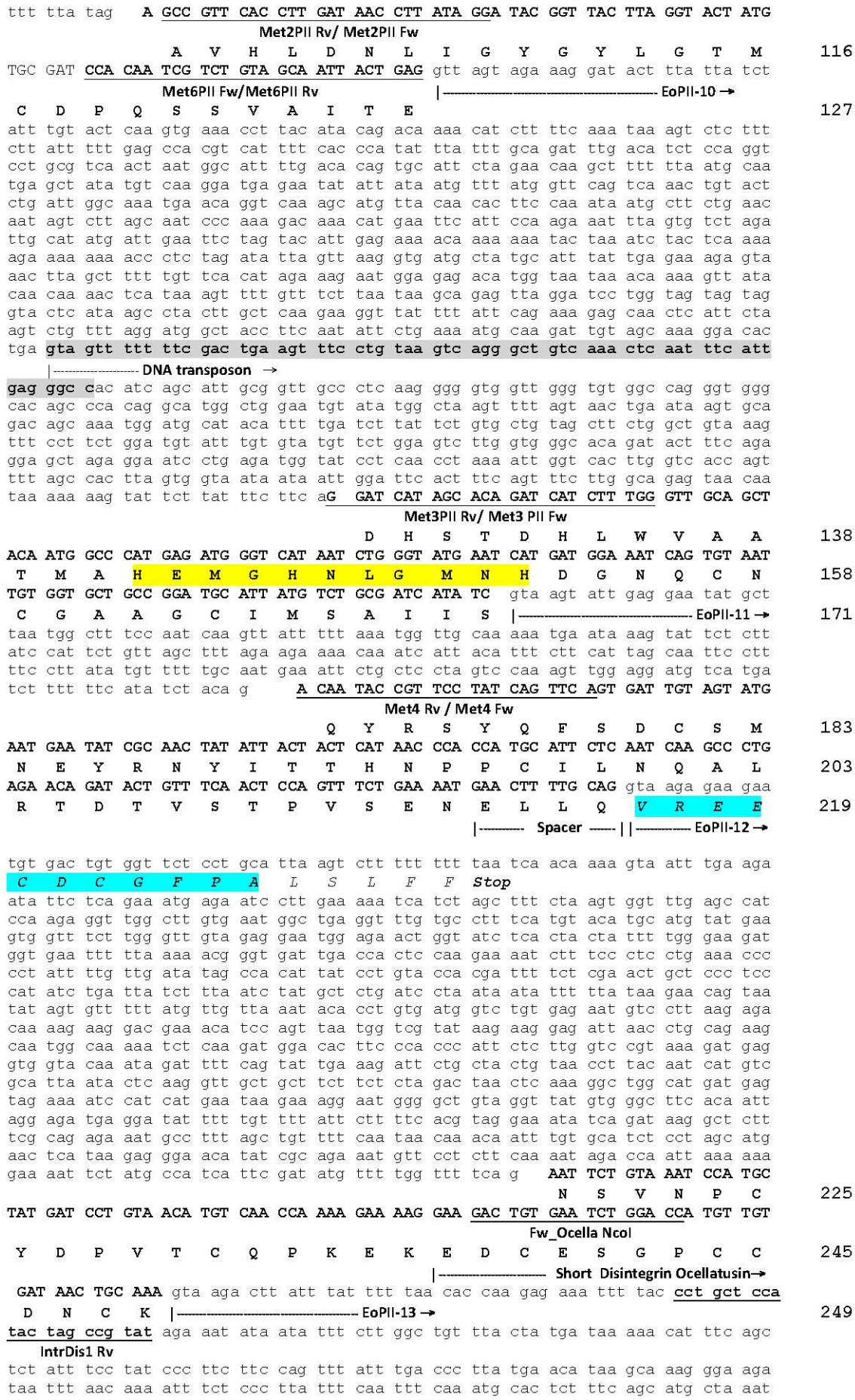

Figure A1. *Cont.*

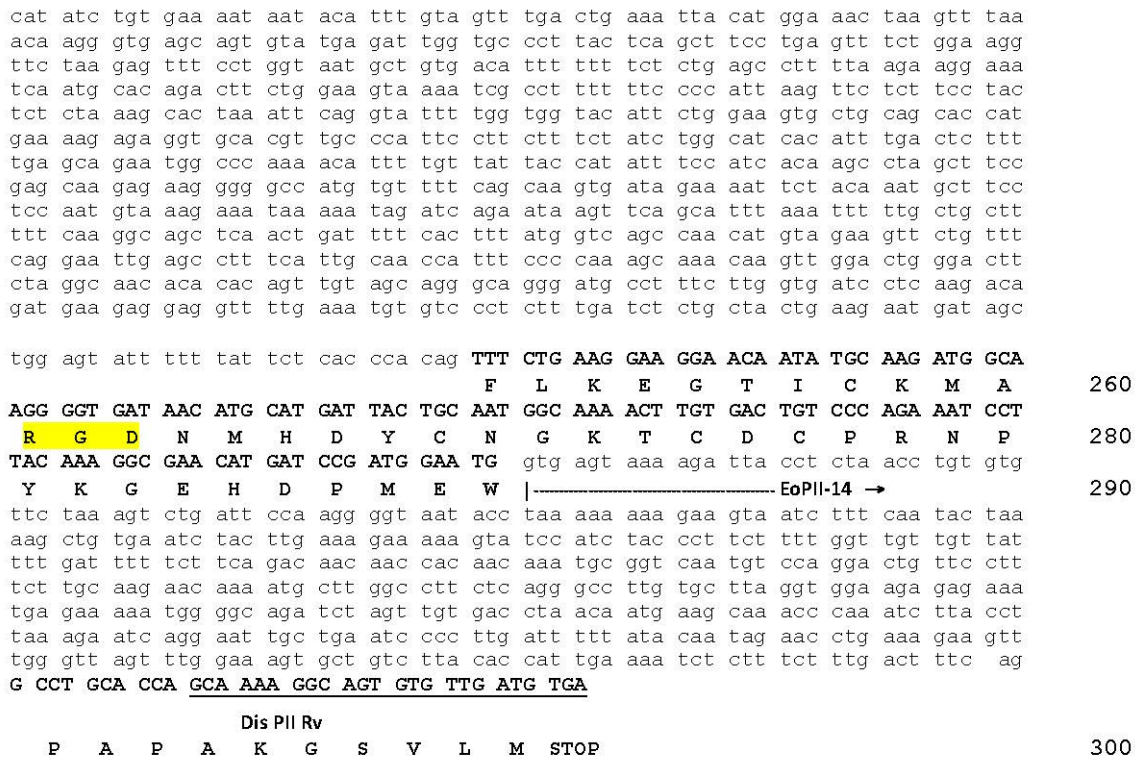

Figure A1. Genomic organization of *E. ocellatus* EOC00006-like PII-SVMP gene.

Genomic sequence of *E. ocellatus* EOC00028-like PI-SVMP gene. The locations and identities of the primers used to PCR-amplify genomic sequences (listed in Table 2) are indicated. Protein-coding DNA regions are in upper letters and boldface, and the encoded amino acid sequence is displayed below the DNA sequence. Start of introns are labelled EoPI-X, where "X" corresponds to intron number. The beginning and the signal peptide, propeptide, metalloproteinase domains and the *C*-terminal extension are specified. Numbers at the right correspond to amino acid numbering of the DNA-deduced pre-pro-PII-SVMP relative to the mature SVMP. The extended Zn^{2+}-binding environment (HEXXHXXGXXH) stands on yellow background. The only two amino acids (-124T/A and 15T/A) that distinguish this sequence from that of PI-SVMP EOC00028 (Q2UXQ3) are shown in bold and red. The remains of a disintegrin-like domain and a dimeric disintegrin domain transformed into intron EoPI-12 are underscored in italics and on cyan background. The *N*-terminal glutamine of the metalloproteinase domain has been assigned residue 1. SINE/Sauria, LINE/L2/CR1, LTR/ERV1, DNA/hAT-Ac and DNA transposon retroelements are highlighted on a gray background. Inserted nucleotide sequences in introns 1 (582 nucleotides between positions 1534–1582, including a SINE/Sauria element); 9 (between nucleotides 194–195 of the topologically equivalent intron of PII); 11 (replacing nucleotides 1-66 of PII intron 11 for a stretch of 3281 nucleotides); and 12 (after nucleotide 999 of the homologous PII intron) are underlined. Simple sequence repeats (SSR, microsatellites) are shown in light green background.

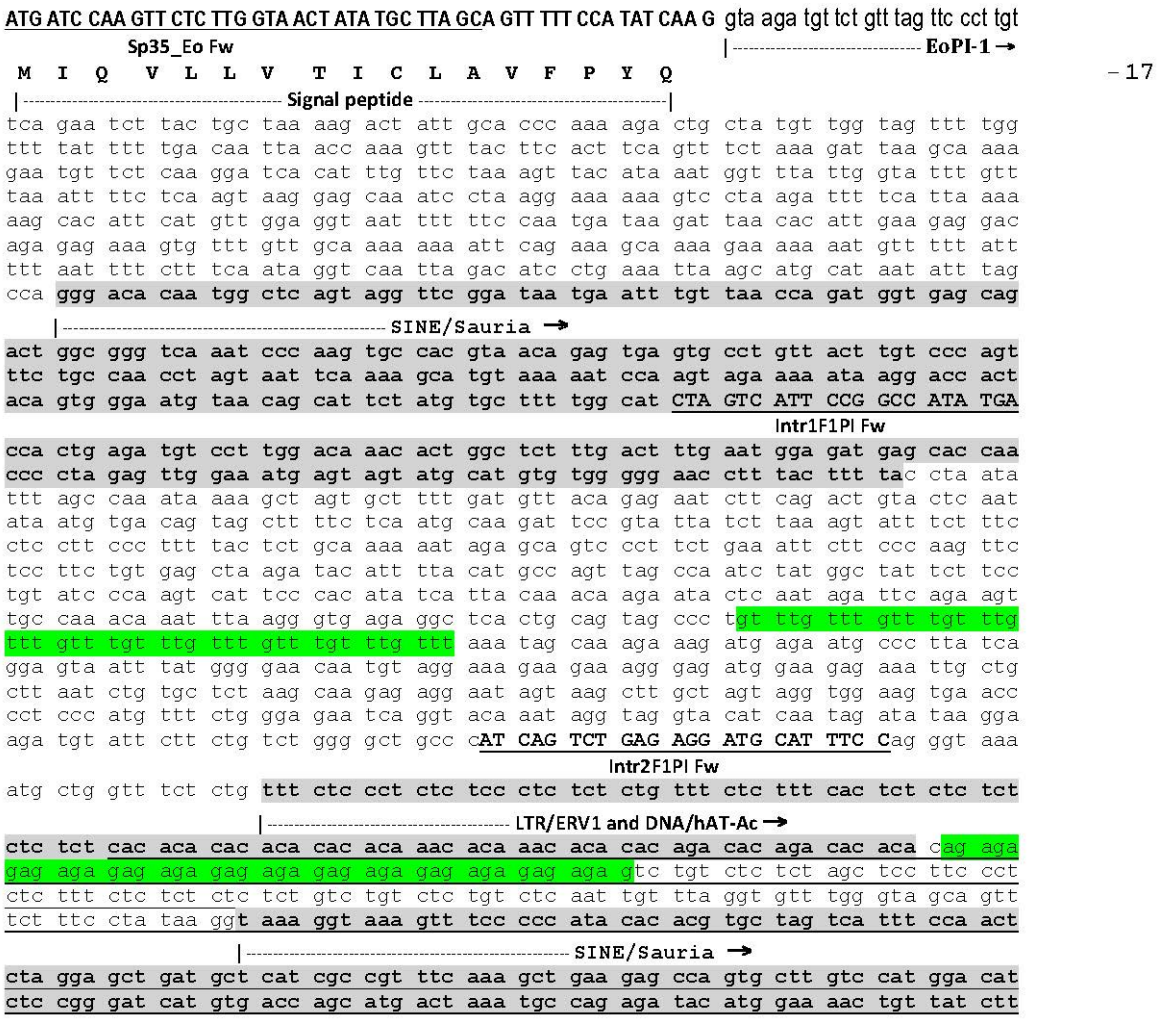

Figure A2. *Cont*.

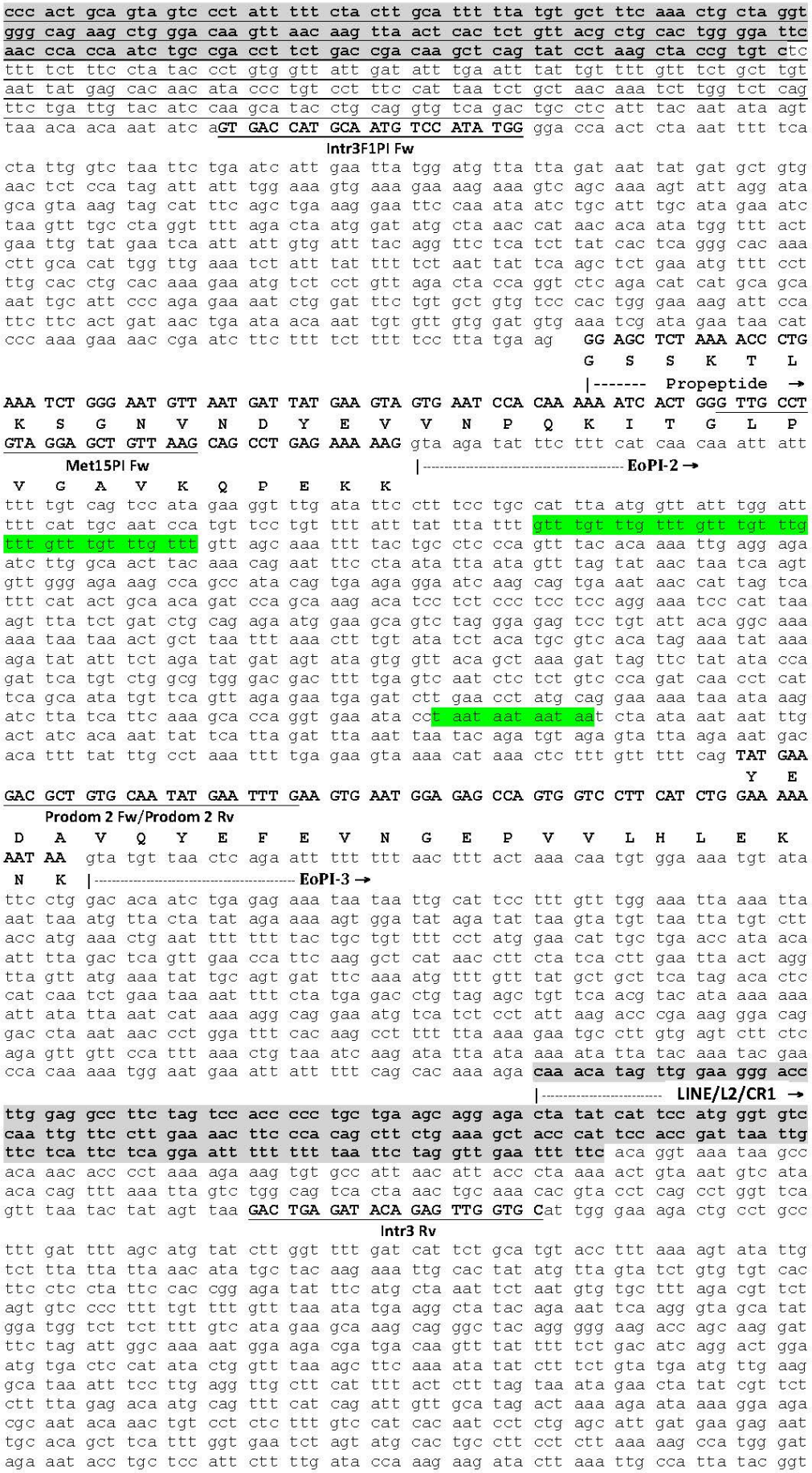

Figure A2. *Cont.*

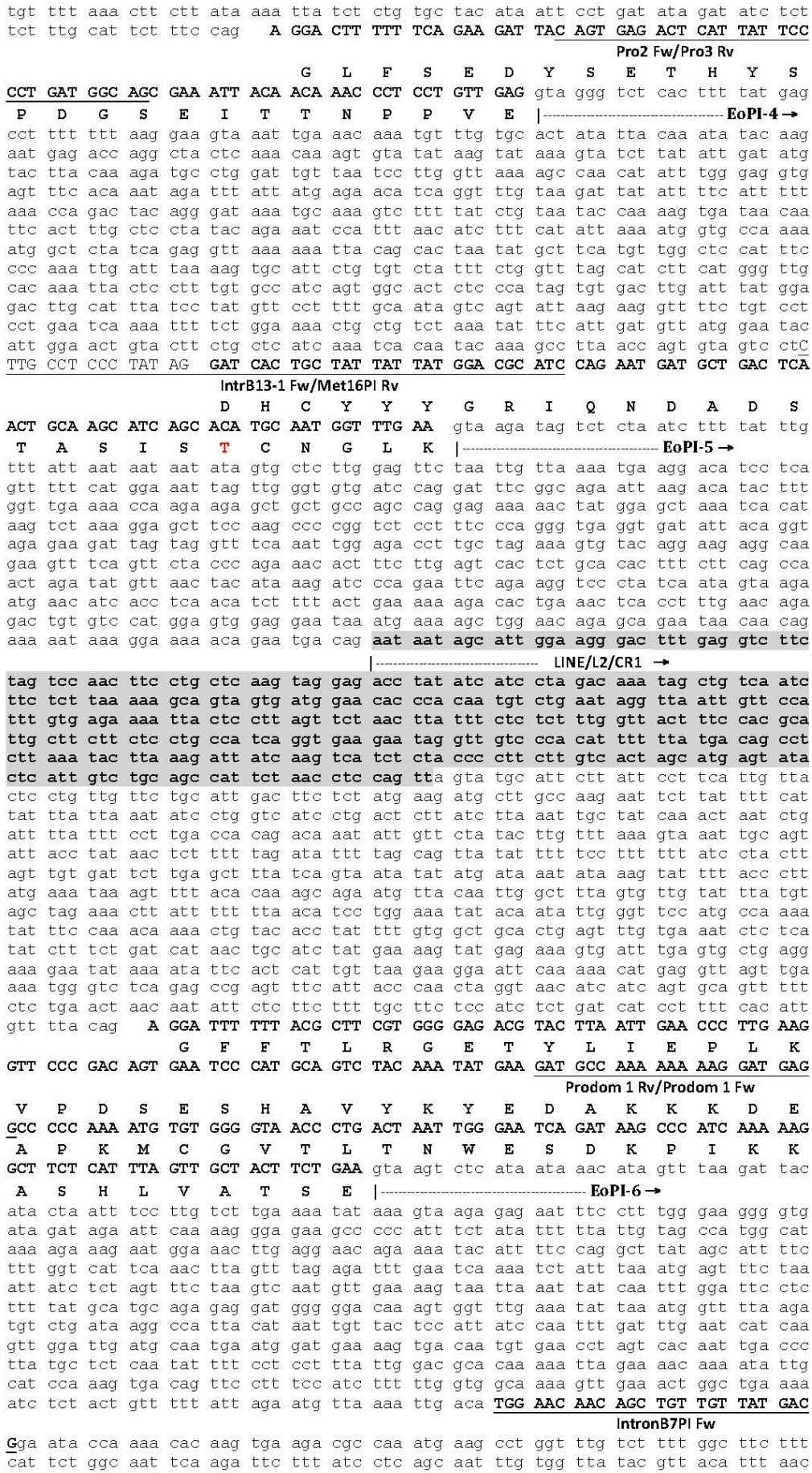

Figure A2. *Cont.*

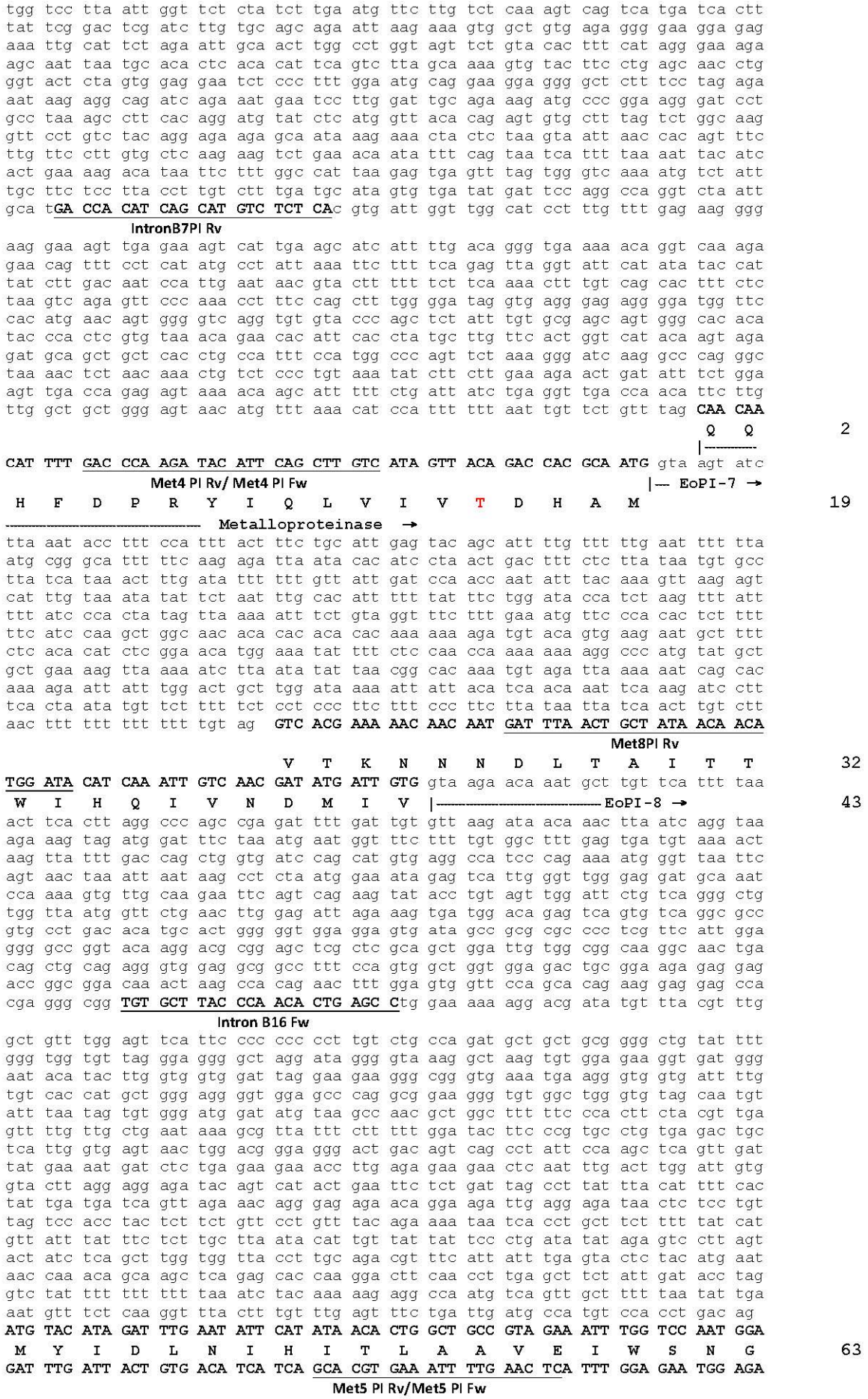

Figure A2. *Cont.*

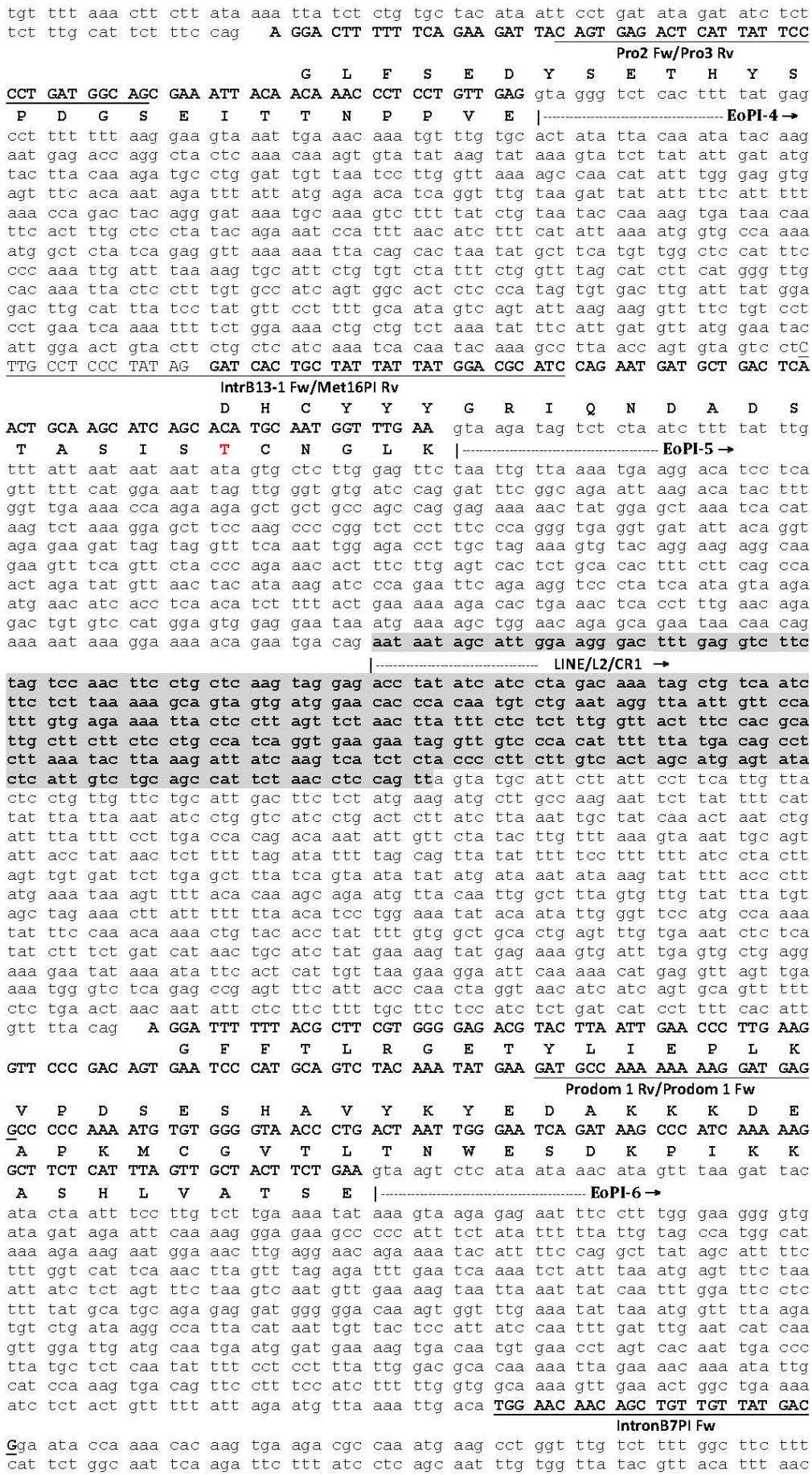

Figure A2. *Cont.*

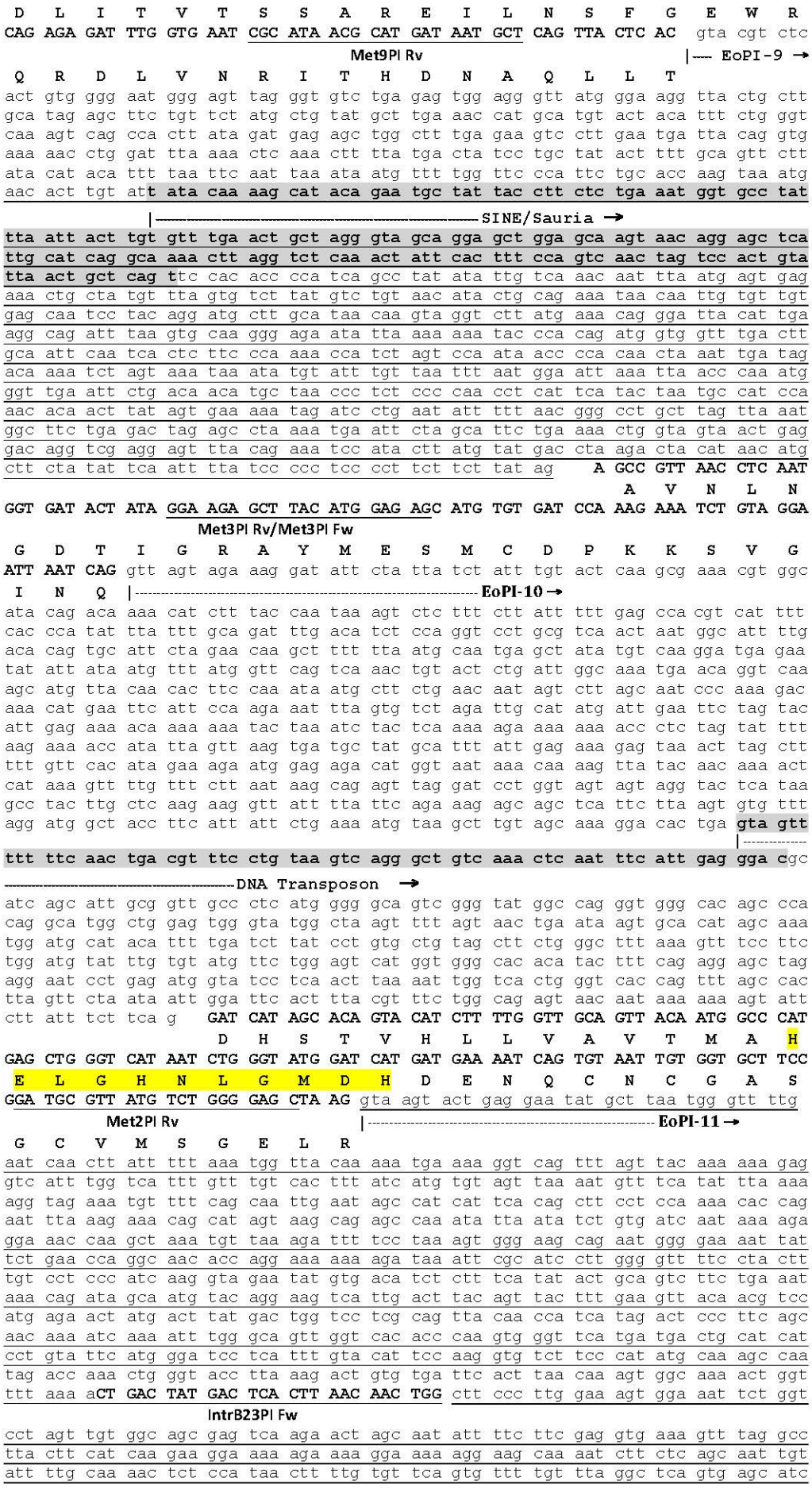

Figure A2. *Cont.*

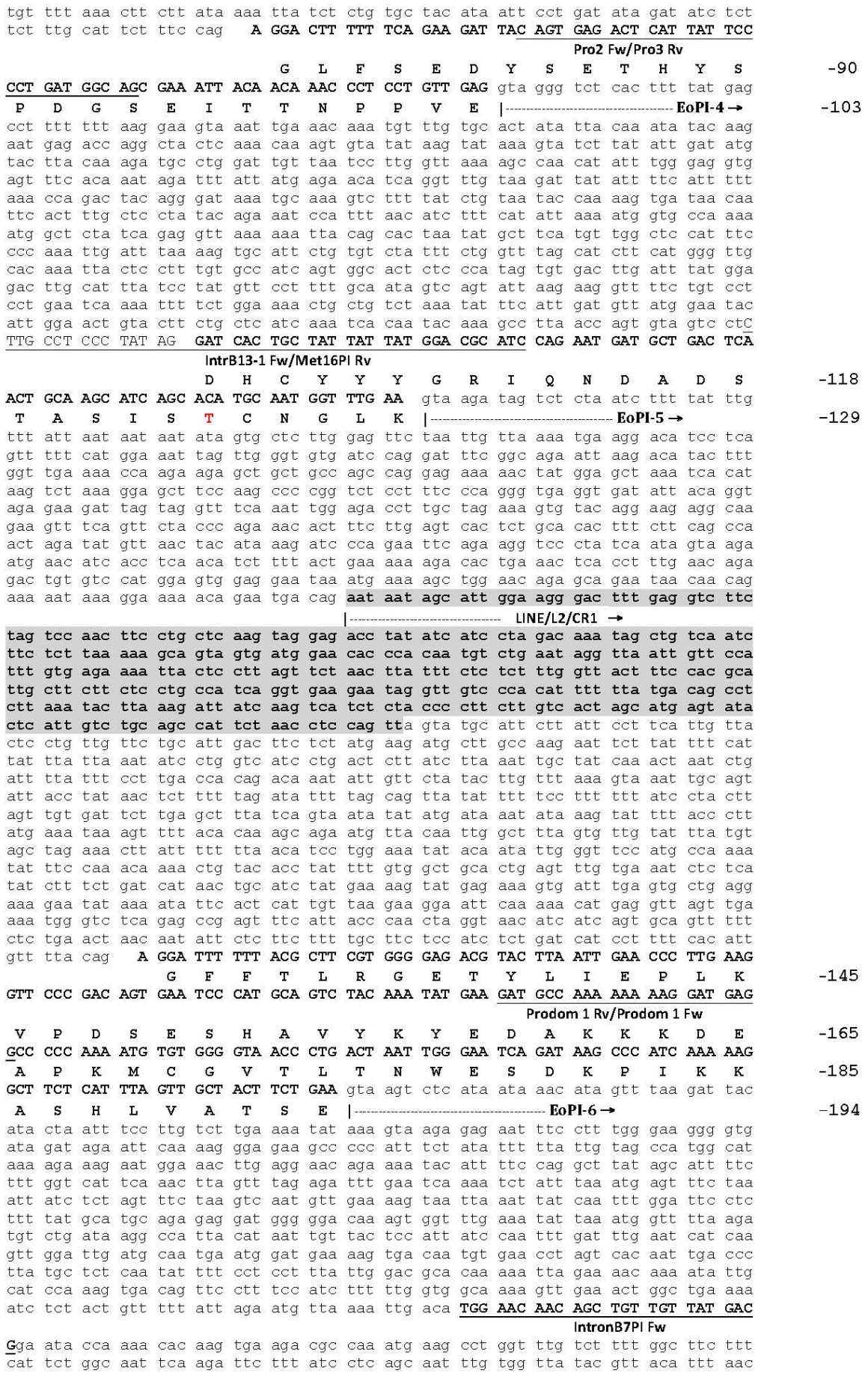

Figure A2. *Cont.*

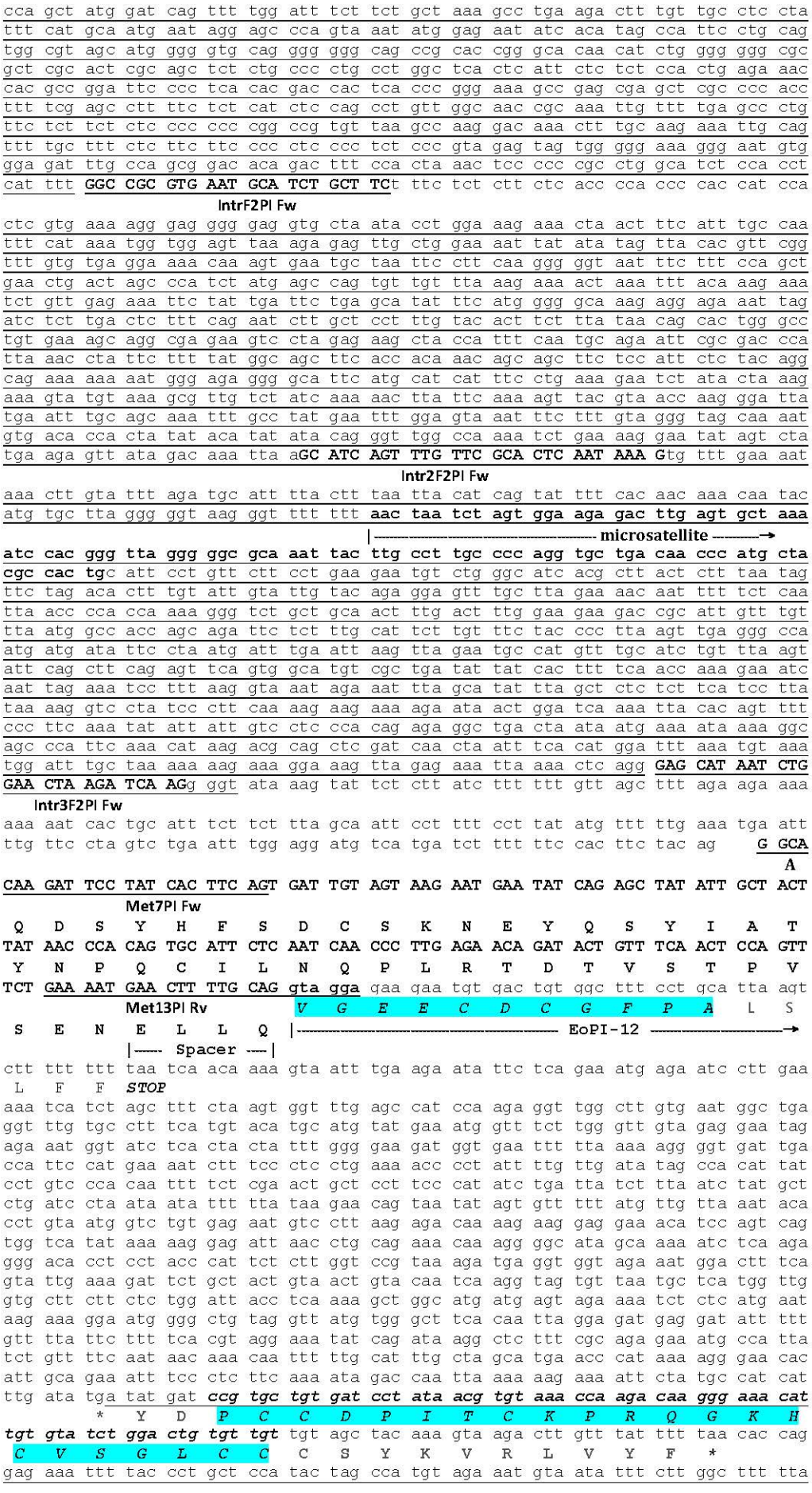

Figure A2. *Cont.*

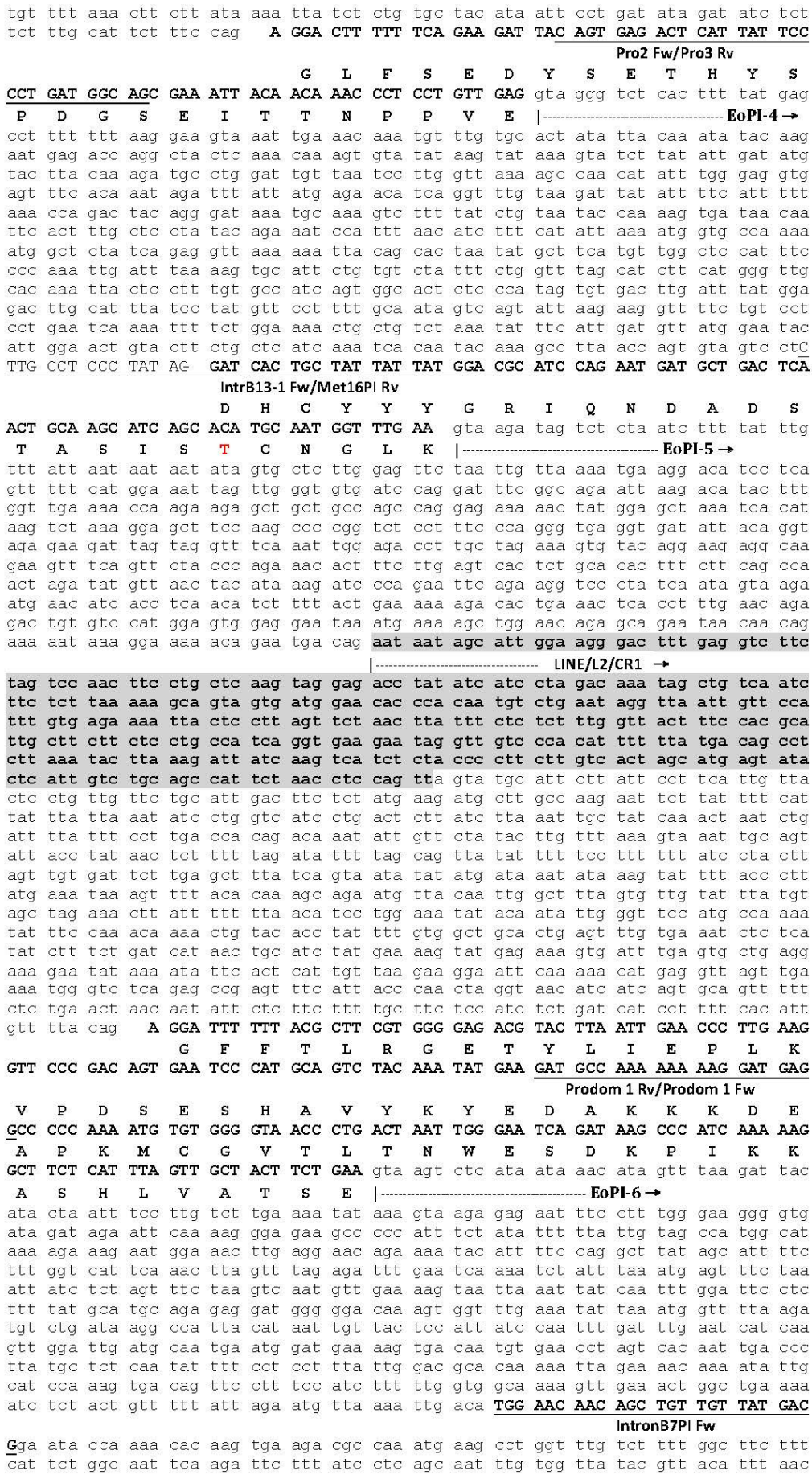

Figure A2. *Cont.*

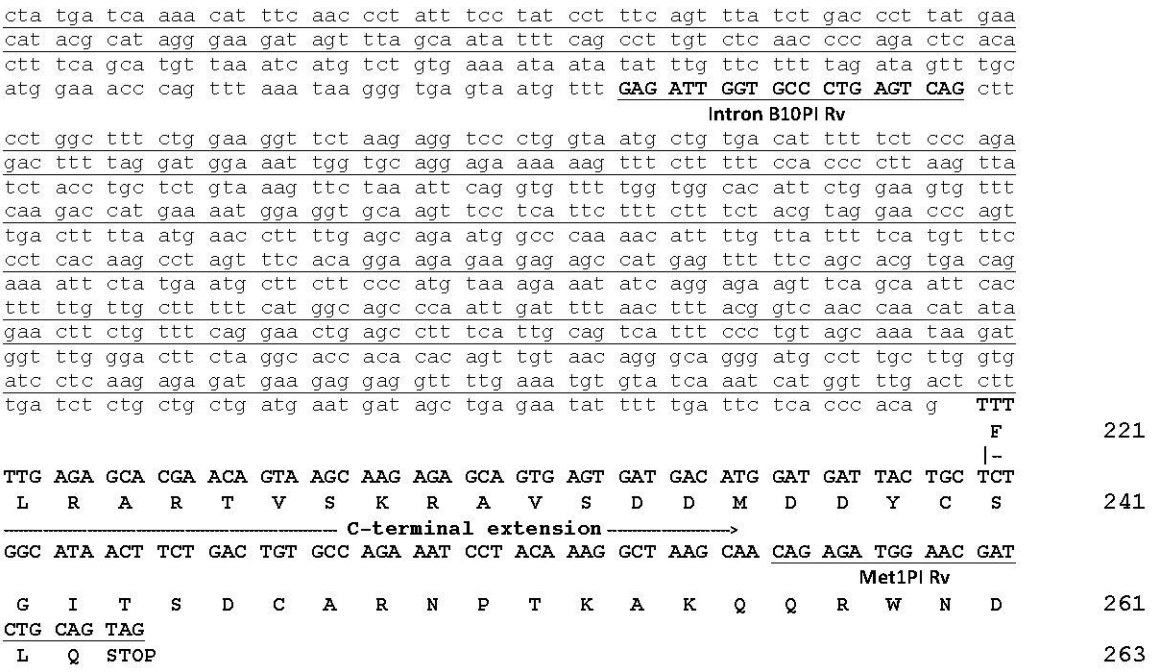

Figure A2. Genomic organization of *E. ocellatus* EOC00028-like PI-SVMP gene.

References

1. MEROPS- the Peptidase Database: Family M72. Available online: https://merops.sanger.ac.uk/cgi-bin/famsum?family=M12 (accessed on 8 May 2016).
2. PFAM Family: REprolysin. Available online: http://pfam.xfam.org/family/PF01421 (accessed on 8 May 2016).
3. Seals, D.F.; Courtneidge, S.A. The ADAMs family of metalloproteases: Multidomain proteins with multiple functions. *Genes Dev.* **2003**, *17*, 7–30. [CrossRef] [PubMed]
4. Giebeler, N.; Zigrino, P. A disintegrin and metalloprotease (ADAM): Historical overview of their functions. *Toxins* **2016**, *8*. [CrossRef] [PubMed]
5. Arendt, D.; Technau, U.; Wittbrodt, J. Evolution of the bilaterian larval foregut. *Nature* **2001**, *409*, 81–85. [CrossRef] [PubMed]
6. Tucker, R.P.; Adams, J.C. Adhesion Networks of Cnidarians: A Postgenomic View. *Int. Rev. Cell Mol. Biol.* **2014**, *308*, 323–377. [PubMed]
7. Bahudhanapati, H.; Bhattacharya, S.; Wei, S. Evolution of Vertebrate Adam Genes; Duplication of Testicular Adams from Ancient Adam9/9-like Loci. *PLoS ONE* **2015**, *10*, e0136281. [CrossRef] [PubMed]
8. Cho, C. Testicular and epididymal ADAMs: Expression and function during fertilization. *Nat. Rev. Urol.* **2012**, *9*, 550–560. [CrossRef] [PubMed]
9. Bates, E.E.; Fridman, W.H.; Mueller, C.G. The ADAMDEC1 (decysin) gene structure: Evolution by duplication in a metalloprotease gene cluster on chromosome 8p12. *Immunogenetics* **2002**, *54*, 96–105. [CrossRef] [PubMed]
10. Wei, S.; Whittaker, C.A.; Xu, G.; Bridges, L.C.; Shah, A.; White, J.M.; Desimone, D.W. Conservation and divergence of ADAM family proteins in the Xenopus genome. *BMC Evol. Biol.* **2010**, *10*. [CrossRef] [PubMed]
11. Taylor, J.S.; Raes, J. Duplication and divergence: The evolution of new genes and old ideas. *Annu. Rev. Genet.* **2004**, *38*, 615–643. [CrossRef] [PubMed]
12. Ohno, S. *Evolution by Gene Duplication*; Springer-Verlag: Berlin-Heidelberg, Germany, 1970.
13. Zhang, J. Evolution by gene duplication: An update. *Trends Ecol. Evol.* **2003**, *18*, 292–298. [CrossRef]
14. True, J.R.; Carroll, S.B. Gene co-option in physiological and morphological evolution. *Annu. Rev. Cell Dev. Biol.* **2002**, *18*, 53–80. [CrossRef] [PubMed]

15. Kaessmann, H.; Vinckenbosch, N.; Long, M. RNA-based gene duplication: Mechanistic and evolutionary insights. *Nat. Rev. Genet.* **2009**, *10*, 19–31. [CrossRef] [PubMed]
16. Vinckenbosch, N.; Dupanloup, I.; Kaessmann, H. Evolutionary fate of retroposed gene copies in the human genome. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 3220–3225. [CrossRef] [PubMed]
17. Fry, B.; Wüster, W. Assembling an Arsenal: Origin and evolution of the snake venom proteome inferred from phylogenetic analysis of toxin sequences. *Mol. Biol. Evol.* **2004**, *21*, 870–883. [CrossRef] [PubMed]
18. Fry, B.G.; Vidal, N.; Norman, J.A.; Vonk, F.J.; Scheib, H.; Ramjan, S.F.; Kuruppu, S.; Fung, K.; Hedges, S.B.; Richardson, M.K.; et al. Early evolution of the venom system in lizards and snakes. *Nature* **2006**, *439*, 584–588. [CrossRef] [PubMed]
19. Fry, B.G.; Casewell, N.R.; Wüster, W.; Vidal, N.; Young, B.; Jackson, N. The structural and functional diversification of the Toxicofera reptile venom system. *Toxicon* **2012**, *60*, 434–448. [CrossRef] [PubMed]
20. Casewell, N.R.; Wüster, W.; Vonk, F.J.; Harrison, R.A.; Fry, B.G. Complex cocktails: The evolutionary novelty of venoms. *Trends Ecol. Evol.* **2013**, *28*, 219–229. [CrossRef] [PubMed]
21. Haney, R.A.; Clarke, T.H.; Gadgil, R.; Fitzpatrick, R.; Hayashi, C.Y.; Ayoub, N.A.; Garb, J.E. Effects of gene duplication, positive selection, and shifts in gene expression on the evolution of the venom gland transcriptome in widow spiders. *Genome Biol. Evol.* **2016**, *8*, 228–242. [CrossRef] [PubMed]
22. Wong, E.S.; Belov, K. Venom evolution through gene duplications. *Gene* **2012**, *496*, 1–7. [CrossRef] [PubMed]
23. Vonk, F.J.; Casewell, N.R.; Henkel, C.V.; Heimberg, A.M.; Jansen, H.J.; McCleary, R.J.; Kerkkamp, H.M.; Vos, R.A.; Guerreiro, I.; Calvete, J.J.; et al. The king cobra genome reveals dynamic gene evolution and adaptation in the snake venom system. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 20651–20656. [CrossRef] [PubMed]
24. Hedges, S.B.; Vidal, N. Lizards, snakes, and amphisbaenians (Squamata). In *The Timetree of Life*; Hedges, S.B., Kumar, S., Eds.; Oxford University Press: Oxford, UK, 2009; pp. 383–389.
25. Jones, M.E.; Anderson, C.L.; Hipsley, C.A.; Müller, J.; Evans, S.E.; Schoch, R.R. Integration of molecules and new fossils supports a Triassic origin for Lepidosauria (lizards, snakes, and tuatara). *BMC Evol. Biol.* **2013**, *13*. [CrossRef] [PubMed]
26. Pyron, R.A.; Burbrink, F.T.; Wiens, J.J. A phylogeny and revised classification of Squamata, including 4161 species of lizards and snakes. *BMC Evol. Biol.* **2013**, *13*. [CrossRef] [PubMed]
27. Reeder, T.W.; Townsend, T.M.; Mulcahy, D.G.; Noonan, B.P.; Wood, P.L., Jr.; Sites, J.W., Jr.; Wiens, J.J. Integrated analyses resolve conflicts over squamate reptile phylogeny and reveal unexpected placements for fossil. *PLoS ONE* **2015**, *10*, e0118199. [CrossRef] [PubMed]
28. Hsiang, A.Y.; Field, D.J.; Webster, T.H.; Behlke, A.D.; Davis, M.B.; Racicot, R.A.; Gauthier, J.A. The origin of snakes: Revealing the ecology, behavior, and evolutionary history of early snakes using genomics, phenomics, and the fossil record. *BMC Evol. Biol.* **2015**, *15*. [CrossRef] [PubMed]
29. Reeks, T.A.; Fry, B.G.; Alewood, P.F. Privileged frameworks from snake venom. *Cell. Mol. Life Sci.* **2015**, *72*, 1939–1958. [CrossRef] [PubMed]
30. Hite, L.A.; Jia, L.G.; Bjarnason, J.B.; Fox, J.W. cDNA sequences for four snake venom metalloproteinases: Structure, classification, and their relationship to mammalian reproductive proteins. *Arch. Biochem. Biophys.* **1994**, *308*, 182–191. [CrossRef] [PubMed]
31. Moura-da-Silva, A.M.; Theakston, R.D.; Crampton, J.M. Evolution of disintegrin cysteine-rich and mammalian matrix-degrading metalloproteinases: Gene duplication and divergence of a common ancestor rather than convergent evolution. *J. Mol. Evol.* **1996**, *43*, 263–269. [CrossRef] [PubMed]
32. Casewell, N.R. On the ancestral recruitment of metalloproteinases into the venom of snakes. *Toxicon* **2012**, *60*, 449–454. [CrossRef] [PubMed]
33. Escalante, T.; Rucavado, A.; Fox, J.W.; Gutiérrez, J.M. Key events in microvascular damage induced by snake venom hemorrhagic metalloproteinases. *J. Proteomics* **2011**, *74*, 1781–1794. [CrossRef] [PubMed]
34. Markland, F.S., Jr.; Swenson, S. Snake venom metalloproteinases. *Toxicon* **2013**, *62*, 3–18. [CrossRef] [PubMed]
35. Herrera, C.; Escalante, T.; Voisin, M.B.; Rucavado, A.; Morazán, D.; Macêdo, J.K.; Calvete, J.J.; Sanz, L.; Nourshargh, S.; Gutiérrez, J.M.; et al. Tissue localization and extracellular matrix degradation by PI, PII and PIII snake venom metalloproteinases: Clues on the mechanisms of venom-induced hemorrhage. *PLoS Negl. Trop. Dis.* **2015**, *9*, e0003731. [CrossRef] [PubMed]
36. Gutiérrez, J.M.; Escalante, T.; Rucavado, A.; Herrera, C. Hemorrhage Caused by Snake Venom Metalloproteinases: A Journey of Discovery and Understanding. *Toxins* **2016**, *8*. [CrossRef] [PubMed]

37. Jia, L.G.; Shimokawa, K.; Bjarnason, J.B.; Fox, J.W. Snake venom metalloproteinases: Structure, function and relationship to the ADAMs family of proteins. *Toxicon* **1996**, *34*, 1269–1276. [CrossRef]
38. Pyron, R.A.; Burnbrink, F.T. Extinction ecological opportunity and the origins of global snake diversity. *Evolution* **2012**, *66*, 163–178. [CrossRef] [PubMed]
39. Fox, J.W.; Serrano, S.M. Structural considerations of the snake venom metalloproteinases, key members of the M12 reprolysin family of metalloproteinases. *Toxicon* **2005**, *45*, 969–985. [CrossRef] [PubMed]
40. Fox, J.W.; Serrano, S.M. Insights into and speculations about snake venom metalloproteinase (SVMP) synthesis, folding and disulfide bond formation and their contribution to venom complexity. *FEBS J.* **2008**, *275*, 3016–3030. [CrossRef] [PubMed]
41. Casewell, N.R.; Sunagar, K.; Takacs, Z.; Calvete, J.J.; Jackson, T.N.W.; Fry, B.G. Snake venom metalloprotease enzymes. In *Venomous Reptiles and Their Toxins: Evolution, Pathophysiology and Biodiscovery*; ISBN: 978-0-19-930939-9. Fry, B.G., Ed.; Oxford University Press: Oxford, UK, 2015; Chapter 23; pp. 347–363.
42. Juárez, P.; Comas, I.; González-Candelas, F.; Calvete, J.J. Evolution of snake venom disintegrins by positive Darwinian selection. *Mol. Biol. Evol.* **2008**, *25*, 2391–2407. [CrossRef] [PubMed]
43. Carbajo, R.J.; Sanz, L.; Pérez, A.; Calvete, J.J. NMR structure of bitistatin—A missing piece in the evolutionary pathway of snake venom disintegrins. *FEBS J.* **2015**, *282*, 341–360. [CrossRef] [PubMed]
44. Calvete, J.J. Brief History and Molecular Determinants of Snake Venom Disintegrin Evolution. In *Toxins and Hemostasis. From Bench to Bedside*; Kini, R.M., Markland, F., McLane, M.A., Morita, T., Eds.; Springer Science+Business Media B.V.: Amsterdam, The Netherlands, 2010; pp. 285–300.
45. Casewell, N.R.; Wagstaff, S.C.; Harrison, R.A.; Renjifo, C.; Wüster, W. Domain loss facilitates accelerated evolution and neofunctionalization of duplicate snake venom metalloproteinase toxin genes. *Mol. Biol. Evol.* **2011**, *28*, 2637–2649. [CrossRef] [PubMed]
46. Sanz-Soler, R.; Sanz, L.; Calvete, J.J. Distribution of *RPTLN* genes across Reptilia. Hypothesized role for *RPTLN* in the evolution of SVMPs. *Integr. Compar. Biol.* **2016**. [CrossRef] [PubMed]
47. Sanz, L.; Harrison, R.A.; Calvete, J.J. First draft of the genomic organization of a PIII-SVMP gene. *Toxicon* **2012**, *60*, 455–469. [CrossRef] [PubMed]
48. Endo, T.; Fedorov, A.; de Souza, S.J.; Gilbert, W. Do Introns Favor or Avoid Regions of Amino Acid Conservation? *Mol. Biol. Evol.* **2002**, *19*, 521–525. [CrossRef] [PubMed]
49. Zhou, Q.; Wang, W. On the origin and evolution of new genes-a genomic and experimental perspective. *J. Genet. Genomics* **2008**, *35*, 639–648. [CrossRef]
50. Kordis, D.; Gubensek, F. Adaptive evolution of animal toxin multigene families. *Gene* **2000**, *261*, 43–52. [CrossRef]
51. Cao, Z.; Yu, Y.; Wu, Y.; Hao, P.; Di, Z.; He, Y.; Chen, Z.; Yang, W.; Shen, Z.; He, X.; et al. The genome of Mesobuthus martensii reveals a unique adaptation model of arthropods. *Nat. Commun.* **2013**, *4*. [CrossRef] [PubMed]
52. Fry, B.G.; Wüster, W.; Kini, R.M.; Brusic, V.; Khan, A.; Venkataraman, D.; Rooney, A.P. Molecular evolution and phylogeny of elapid snake venom three-finger toxins. *J. Mol. Evol.* **2003**, *57*, 110–129. [CrossRef] [PubMed]
53. Reyes-Velasco, J.; Card, D.C.; Andrew, A.L.; Shaney, K.J.; Adams, R.H.; Schield, D.R.; Casewell, N.R.; Mackessy, S.P.; Castoe, T.A. Expression of venom gene homologs in diverse python tissues suggests a new model for the evolution of snake venom. *Mol. Biol. Evol.* **2015**, *32*, 173–183. [CrossRef] [PubMed]
54. Hargreaves, A.D.; Swain, M.T.; Logan, D.W.; Mulley, J.F. Testing the Toxicofera: Comparative transcriptomics casts doubt on the single, early evolution of the reptile venom system. *Toxicon* **2014**, *92*, 140–156. [CrossRef] [PubMed]
55. Sunagar, K.; Jackson, T.N.; Undheim, E.A.; Ali, S.A.; Antunes, A.; Fry, B.G. Three-fingered RAVERs: Rapid Accumulation of Variations in Exposed Residues of snake venom toxins. *Toxins* **2013**, *5*, 2172–2208. [CrossRef] [PubMed]
56. Chang, D.; Duda, T.F. Extensive and continuous duplication facilitates rapid evolution and diversification of gene families. *Mol. Biol. Evol.* **2012**, *29*, 2019–2029. [CrossRef] [PubMed]
57. Chow, L.T.; Gelinas, R.E.; Broker, T.R.; Roberts, R.J. An amazing sequence arrangement at the 5′ ends of adenovirus 2 messenger RNA. *Cell* **1977**, *12*, 1–8. [CrossRef]
58. Berget, S.M.; Moore, C.; Sharp, P.A. Spliced segments at the 5′ terminus of adenovirus 2 late mRNA. *Proc. Natl. Acad. Sci. USA* **1977**, *74*, 3171–3175. [CrossRef] [PubMed]

59. Bicknell, A.A.; Cenik, C.; Chua, H.N.; Roth, F.P.; Moore, M.J. Introns in UTRs: Why we should stop ignoring them. *BioEssays* **2012**, *34*, 1025–1034. [CrossRef] [PubMed]
60. Cenik, C.; Chua, H.N.; Zhang, H.; Tarnawsky, S.P.; Akef, A.; Derti, A.; Tasan, M.; Moore, M.J.; Palazzo, A.F.; Roth, F.P. Genome analysis reveals interplay between 5′-UTR introns and nuclear mRNA export for secretory and mitochondrial genes. *PLoS Genet.* **2011**, *7*, e1001366. [CrossRef] [PubMed]
61. Comeron, J.M.; Kreitman, M. The correlation between intron length and recombination in Drosophila: Dynamic equilibrium between mutational and selective forces. *Genetics* **2000**, *156*, 1175–1190. [PubMed]
62. De Souza, S.J.; Long, M.; Gilbert, W. Introns and gene evolution. *Genes Cells* **1996**, *1*, 493–505. [CrossRef] [PubMed]
63. Patthy, L. Exon shuffling and other ways of module exchange. *Matrix Biol.* **1996**, *15*, 301–310. [CrossRef]
64. Hughes, A.L.; Hughes, M.K. Small genomes for better flyers. *Nature* **1995**, *377*, 391. [CrossRef] [PubMed]
65. Lynch, M. Intron evolution as a population-genetic process. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 6118–6123. [CrossRef] [PubMed]
66. Haddrill, P.R.; Charlesworth, B.; Halligan, D.L.; Andolfatto, P. Patterns of intron sequence evolution in Drosophila are dependent upon length and GC content. *Genome Biol.* **2005**, *6*, R67. [CrossRef] [PubMed]
67. Zhao, M.; He, L.; Gu, Y.; Wang, Y.; Chen, Q.; He, C. Genome-wide analyses of a plant-specific LIM-domain gene family implicate its evolutionary role in plant diversification. *Genome Biol. Evol.* **2014**, *6*, 1000–1012. [CrossRef] [PubMed]
68. Tordai, H.; Patthy, L. Insertion of spliceosomal introns in proto-splice sites: The case of secretory signal peptides. *FEBS Lett.* **2004**, *575*, 109–111. [CrossRef] [PubMed]
69. Tomita, M.; Shimizu, N.; Brutlag, D.L. Introns and reading frames: Correlation between splicing sites and their codon positions. *Mol. Biol. Evol.* **1996**, *13*, 1219–1223. [CrossRef] [PubMed]
70. Long, M.; de Souza, S.J.; Rosenberg, C.; Gilbert, W. Relationship between proto-splice sites and intron phases: Evidence from dicodon analysis. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 219–223. [CrossRef] [PubMed]
71. Von Heijne, G. Patterns of amino acids near signal-sequence cleavage sites. *Eur. J. Biochem.* **1983**, *133*, 17–21. [CrossRef] [PubMed]
72. Pinho, C.; Rocha, S.; Carvalho, B.M.; Lopes, S.; Mourão, S.; Vallinoto, M.; Brunes, T.O.; Haddad, C.F.B.; Gonçalves, H.; Sequeira, F.; et al. New primers for the amplification and sequencing of nuclear *loci* in a taxonomically wide set of reptiles and amphibians. *Conserv. Genet. Resour.* **2010**, *2*, 181–185. [CrossRef]
73. Ellegren, H. Microsatellites: Simple sequences with complex evolution. *Nat. Rev. Genet.* **2004**, *5*, 435–445. [CrossRef] [PubMed]
74. Adams, R.H.; Blackmon, H.; Reyes-Velasco, J.; Schield, D.R.; Card, D.C.; Andrew, A.L.; Waynewood, N.; Castoe, T.A. Microsatellite landscape evolutionary dynamics across 450 million years of vertebrate genome evolution. *Genome* **2016**, *59*, 295–310. [CrossRef] [PubMed]
75. Liang, K.C.; Tseng, J.T.; Tsai, S.J.; Sun, H.S. Characterization and distribution of repetitive elements in association with genes in the human genome. *Comput. Biol. Chem.* **2015**, *57*, 29–38. [CrossRef] [PubMed]
76. Schlötterer, C.; Tautz, D. Slippage synthesis of simple sequence DNA. *Nucleic Acids Res.* **1992**, *20*, 211–215. [CrossRef] [PubMed]
77. Charlesworth, B.; Sniegowski, P.; Stephan, W. The evolutionary dynamics of repetitive DNA in eukaryotes. *Nature* **1994**, *371*, 215–220. [CrossRef] [PubMed]
78. Martin, P.; Makepeace, K.; Hill, S.A.; Hood, D.W.; Moxon, E.R. Microsatellite instability regulates transcription factor binding and gene expression. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 3800–3804. [CrossRef] [PubMed]
79. Li, Y.C.; Korol, A.B.; Fahima, T.; Beiles, A.; Nevo, E. Microsatellites: Genomic distribution, putative functions and mutational mechanisms: A review. *Mol. Ecol.* **2002**, *11*, 2453–2465. [CrossRef] [PubMed]
80. Shaney, K.J.; Schield, D.R.; Card, D.C.; Ruggiero, R.P.; Pollock, D.D.; Mackessy, S.P.; Castoe, T.A. Squamate reptile genomics and evolution. In *Handbook. of Toxinology: Venom Genomics and Proteomics*; Gopalakrishnakone, P., Calvete, J.J., Eds.; Springer Science+Business Media: Dordrecht, The Netherlands, 2016; pp. 29–49.
81. Balaresque, P.; King, T.E.; Parkin, E.J.; Heyer, E.; Carvalho-Silva, D.; Kraaijenbrink, T.; de Knijff, P.; Tyler-Smith, C.; Jobling, M.A. Gene conversion violates the stepwise mutation model for microsatellites in Y-chromosomal palindromic repeats. *Hum. Mutat.* **2014**, *35*, 609–617. [CrossRef] [PubMed]

82. Eller, C.D.; Regelson, M.; Merriman, B.; Nelson, S.; Horvath, S.; Marahrens, Y. Repetitive sequence environment distinguishes housekeeping genes. *Gene* **2007**, *390*, 153–165. [CrossRef] [PubMed]
83. Sverdlov, E.D. Perpetually mobile footprints of ancient infections in human genome. *FEBS Lett.* **1998**, *428*, 1–6. [CrossRef]
84. Makalowski, W. Genomic scrap yard: How genomes utilize all that junk. *Gene* **2000**, *259*, 61–67. [CrossRef]
85. Bourque, G.; Leong, B.; Vega, V.B.; Chen, X.; Lee, Y.L.; Srinivasan, K.G.; Chew, J.L.; Ruan, Y.; Wei, C.L.; Ng, H.H.; et al. Evolution of the mammalian transcription factor binding repertoire via transposable elements. *Genome. Res.* **2008**, *18*, 1752–1762. [CrossRef] [PubMed]
86. Irimía, M.; Rukov, J.L.; Penny, D.; Vinther, J.; García-Fernández, J.; Roy, S.W. Origin of introns by 'intronization' of exonic sequences. *Trends Genet.* **2008**, *24*, 378–381. [CrossRef] [PubMed]
87. Bazaa, A.; Juarez, P.; Marrakchi, N.; Bel Lasfer, Z.; El Ayeb, M.; Harrison, R.A.; Calvete, J.J.; Sanz, L. Loss of introns along the evolutionary diversification pathway of snake venom disintegrins evidenced by sequence analysis of genomic DNA from *Macrovipera lebetina transmediterranea* and *Echis ocellatus*. *J. Mol. Evol.* **2007**, *64*, 261–271. [CrossRef] [PubMed]
88. Bonen, L.; Vogel, J. The ins and outs of group II introns. *Trends Genet.* **2001**, *17*, 322–331. [CrossRef]
89. Dibb, N.J.; Newman, A.J. Evidence that introns arose at proto-splice sites. *EMBO J.* **1989**, *8*, 2015–2021. [PubMed]
90. Logsdon, J.M., Jr. The recent origins of spliceosomal introns revisited. *Curr. Opin. Genet. Dev.* **1998**, *8*, 637–648. [CrossRef]
91. Lynch, M.; Richardson, A.O. The evolution of spliceosomal introns. *Curr. Opin. Genet. Dev.* **2002**, *12*, 701–710. [CrossRef]
92. Patel, A.A.; Steitz, J.A. Splicing double: Insights from the second spliceosome. *Nat. Rev. Mol. Cell Biol.* **2003**, *4*, 960–970. [CrossRef] [PubMed]
93. Rodríguez-Trelles, F.; Tarrío, R.; Ayala, F.J. Origins and evolution of spliceosomal introns. *Annu. Rev. Genet.* **2006**, *40*, 47–76. [CrossRef] [PubMed]
94. Luo, Y.; Li, C.; Gong, X.; Wang, Y.; Zhang, K.; Cui, Y.; Sun, Y.E.; Li, S. Splicing-related features of introns serve to propel evolution. *PLoS ONE* **2013**, *8*, e58547. [CrossRef] [PubMed]
95. Nakashima, K.; Ogawa, T.; Oda, N.; Hattori, M.; Sakaki, Y.; Kihara, H.; Ohno, M. Accelerated evolution of *Trimeresurus. flavoviridis* venom gland phospholipase A_2 isozymes. *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 5964–5968. [CrossRef] [PubMed]
96. Nakashima, K.; Nobuhisa, I.; Deshimaru, M.; Nakai, M.; Ogawa, T.; Shimohigashi, Y.; Fukumaki, Y.; Hattori, M.; Sakaki, Y.; Hattori, S.; et al. Accelerated evolution in the protein-coding regions is universal in crotalinae snake venom gland phospholipase A_2 isozyme genes. *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 5605–5609. [CrossRef] [PubMed]
97. Ikeda, N.; Chijiwa, T.; Matsubara, K.; Oda-Ueda, N.; Hattori, S.; Matsuda, Y.; Ohno, M. Unique structural characteristics and evolution of a cluster of venom phospholipase A_2 isozyme genes of *Protobothrops flavoviridis* snake. *Gene* **2010**, *461*, 15–25. [CrossRef] [PubMed]
98. Chijiwa, T.; Ikeda, N.; Masuda, H.; Hara, H.; Oda-Ueda, N.; Hattori, S.; Ohno, M. Structural characteristics and evolution of a novel venom phospholipase A_2 gene from *Protobothrops flavoviridis*. *Biosci. Biotechnol. Biochem.* **2012**, *76*, 551–558. [CrossRef] [PubMed]
99. Chijiwa, T.; Nakasone, H.; Irie, S.; Ikeda, N.; Tomoda, K.; Oda-Ueda, N.; Hattori, S.; Ohno, M. Structural characteristics and evolution of the *Protobothrops elegans* pancreatic phospholipase A_2 gene in contrast with those of *Protobothrops* genus venom phospholipase A_2 genes. *Biosci. Biotechnol. Biochem.* **2013**, *77*, 97–102. [CrossRef] [PubMed]
100. Koszul, R.; Fischer, G. A prominent role for segmental duplications in modeling eukaryotic genomes. *C. R. Biol.* **2009**, *332*, 254–266. [CrossRef] [PubMed]
101. Ohta, T. Simple model for treating evolution of multigene families. *Nature* **1976**, *263*, 74–76. [CrossRef] [PubMed]
102. Smith, G.P. Evolution of repeated DNA sequences by unequal crossover. *Science* **1976**, *191*, 528–535. [CrossRef] [PubMed]
103. Juárez, P.; Wagstaff, S.C.; Sanz, L.; Harrison, R.A.; Calvete, J.J. Molecular cloning of *Echis ocellatus* disintegrins reveals non-venom-secreted proteins and a pathway for the evolution of ocellatusin. *J. Mol. Evol.* **2006**, *63*, 183–193. [CrossRef] [PubMed]

104. GeneWise. Available online: http://www.ebi.ac.uk/Tools/Wise2/index.html (accessed on 31 May 2016).
105. Basic Local Alignment Search Tool (BLAST). Available online: http://blast.ncbi.nlm.nih.gov/Blast.cgi (accessed on 5 June 2016).
106. ClustalW2-Multiple Sequence Alignment. Available online: http://www.ebi.ac.uk/Tools/msa/clustalw2 (accessed on 5 June 2016).
107. RepeatMasker. Available on: http://www.repeatmasker.org (accessed on 30 December 2015).
108. Genetic Information Research Institute. Available on: http://www.girinst.org (accessed on 7 January 2016).
109. National Center for Biotechnology Information (NCBI) Database. Available on: http://www.ncbi.nlm.nih.gov (accessed on 10 January 2016).