Lineage-specific evolution of SVA and LAVA retrotransposons

Annette Damert*

Department of Primate Genetics, German Primate Center, Leibniz Institute for Primate Research, Göttingen, Germany

Abstract

SVA (SINE-R-VNTR-*Alu*) and LAVA (L1-Alu-VNTR-*Alu*) are composite non-autonomous non-LTR retrotransposons specific to hominoid primates. They are mobilized by the LINE-1-encoded protein machinery in *trans*. SVA insertions can cause disease in humans. Both SVA and LAVA contribute to genome evolution by co-mobilization of non-retrotransposon sequences at their 5' and 3' ends.

The origin of SVA and LAVA can be traced back to a common ancestor present in the genomes of old world monkeys – SVA2. The shared 5' *Alu*-like domain and the variant 3' ends have been acquired through splicing. The evolutionary most recent addition to the family of VNTR (Variable Number of Tandem Repeats) – containing retrotransposons are human SVAs that have co-opted an exon of a host gene as their 5' end.

Here I review current knowledge about the assembly of SVA/LAVA and their evolution in hominoid primates. Lineage- and family-specific structural differences will be discussed in the context of mobilization and host repression of expression.

Keywords: retrotransposon, SVA, LAVA, evolution

Nomenclature, origin and distribution

VNTR (Variable Number of Tandem Repeats) composite retrotransposons – SVA (SINE-R-VNTR-*Alu*), LAVA (L1-Alu-VNTR-*Alu*), PVA (PTGR2-VNTR-*Alu*) and FVA (FRAM-VNTR-*Alu*) – represent a peculiarity in terminology: disregarding all conventions in molecular biology they are named from their 3' ends. The first report of part of a VNTR composite dates back to 1986, when Ono and colleagues described repetitive sequences displaying homology to the endogenous retrovirus HERV-K in the human genome. The retroposons were around 630bp long and classified as "SINEs". They were found to be composed of 5' 40bp GC-rich repeats (later termed VNTR), followed by a sequence originating from the HERV-K 3' part of the *env* gene and 3' LTR. Based on their derivation from HERV-K Ono et al. termed them "SINE-R" [SINE of retroviral origin (1)]. In 1994, a report by Shen and colleagues characterized a SINE-R-containing retrotransposon in the *RP1* gene. They found that in the TSD (Target site duplication)-flanked structure, the SINE-R was preceded by a domain displaying homology to *Alu* elements and TCTCCC hexameric repeats at its 5' end. To give credit to those who had first discovered the elements' 3' part and to reflect the composite nature of the retrotransposons in detail, Shen and colleagues suggested the name SVA – SINE-R-VNTR-*Alu* (2). The same nomenclature approach (*i.e.*, from the 3' end) has been adopted for the additional families of VNTR composite retrotransposons discovered more recently in gibbons: LAVA (3), PVA (4, 5) and FVA (4).

SVA and LAVA are shared by gibbons and great apes (5' truncated LAVA copies only), but are absent from the genomes of old world monkeys. They must, therefore, have originated after the split of the gibbon/great ape lineage from old world monkeys. SVAs amplified and are still active in great apes; the *Nomascus leucogenys* (Northern white cheeked gibbon) genome harbours 29 SVA copies (4). LAVA reached high copy numbers only in gibbons. Presence/absence polymorphisms suggest that they are still active in this hominoid family (6).

Assembly of the prototype SVA and LAVA elements

The central VNTR domain of SVA and LAVA can be traced back to SVA2, a low copy repeat first discovered in the genome of the Rhesus macaque [*Macaca mulatta*, old world monkeys (7)]. The SVA2 VNTR is composed of 39 and 40bp repeat units (8). At its 3' end it is fused to a unique sequence, part of which has been preserved

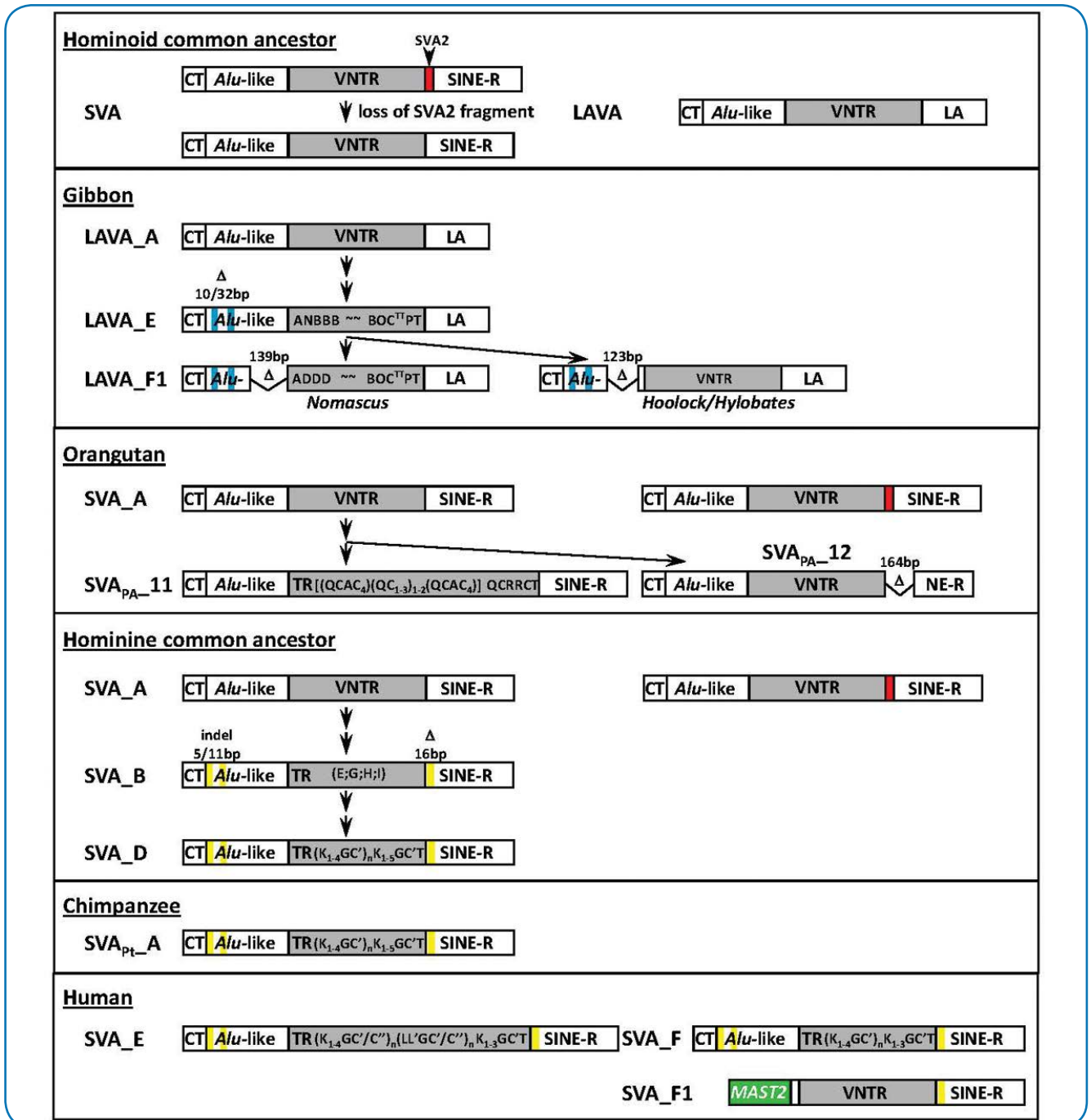*Corresponding author (E-mail: adamert@dpz.eu)

Fig. 1. Evolution of SVA and LAVA elements in hominoid primates. From top to bottom: the fragment of the SVA2 3' unique sequence (red box) remaining after acquisition of the SINE-R in SVA is eliminated through splicing. Both types of the ancestral element – with and without the SVA2 fragment – can be found in all extant hominoids. In gibbons, LAVA evolution involved small deletions in the *Alu*-like domain (blue bars) and the emergence of fixed arrays of repeat units at the VNTR 5' and 3' ends. Further on genus-specific splicing events led to the loss of the 3' part of the *Alu*-like domain in the youngest LAVA subfamilies. In orangutan, SVA subfamilies evolved from the ancestral element lacking the SVA2 fragment. The 5' and 3' domains remained largely unchanged (nucleotide substitutions only), with the exception of SVAPA_12, where splicing between the VNTR and SINE-R resulted in the loss of the 5' part of the SINE-R. The VNTR of the younger subfamilies is characterized by a constant 5' array of repeat units (TR), followed by variable arrays of repeat units (RUs) differing in both their number and the number of their constituent RUs. In the hominine common ancestor, SVA_B evolved from the ancestral element lacking the SVA2 fragment. The subfamily is characterized by indels in the 5' part of the *Alu*-like domain and a 16bp deletion in the SINE-R (yellow bars). Longer VNTR repeat units (E, G, H and I) emerged in the transition from SVA_A to SVA_B. SVA_D, the last subfamily with elements shared by all hominines, displays an organized VNTR structure with a fixed 5' array of RUs (TR) and tandemly repeated variable arrays in the 3' part of the domain. Species-specific SVA subfamilies comprise SVAPt_A in chimpanzees and SVA_E and SVA_F in humans. In the non-canonical subfamily SVA_F1, the larger part of the 5' hexameric repeat/*Alu*-like domain has been replaced by the first exon of the *MAST2* gene. VNTR repeat units are encoded as described in Lupan *et al.* (8). The tilde represents the random sequence of RUs in the interior of LAVA VNTRs. LA – LAVA 3' domain consisting of L1 and *Alu* fragments as well as two unique sequences.

in a subgroup of SVAs in all hominoid primates (red box in Fig. 1). The specific combination of antisense *Alu* fragments present in the VNTR composites' 5' domain cannot be found anywhere outside these elements. Hancks and Kazazian suggested splicing of two antisense *Alu* copies with subsequent deletion of part of the second copy as the mechanism for its assembly and fusion to the SVA2 VNTR (9). The resulting *Alu*-SVA2 – precursor of all VNTR composite families – is no longer detectable in hominoid genomes. The variant 3' ends of SVA and LAVA have been acquired by splicing: between the SVA2 3' unique sequence and the *env* gene of the endogenous retrovirus HERV-K in case of SVA and the VNTR and a splice acceptor found in the *HSD17B3* intron 2 in case of LAVA (4).

Evolution of LAVA elements in gibbons

In *Nomascus leucogenys*, between 22 and 24 subfamilies of LAVA can be distinguished (6, 8). Their sequence and structure reflect the evolution of the family over time. In the 3' LA (L1-*Alu*) domain, evolution occurred mainly at the sequence level; small deletions comprise maximal 4bp (8). By contrast to SVA (see below), the LAVA central VNTR domain did not evolve a specific interior structural organization – at least as far as can be told from currently available sequence information (the 80% GC highly repetitive region being among the most difficult to sequence and assemble). We found at most five repeat units to be conserved at the 5' and 3' ends of the domain, respectively [Fig. 1 (8)]. When compared to the 3' LA domain, more extensive changes have taken place in the 5' *Alu*-like region. They range from comparatively small deletions (10bp from LAVA_A2 to LAVA_A1 and 32bp from LAVA_B to LAVA_C) to the loss of the entire 3' part of the domain in LAVA_F1/F2 (4). As splice site prediction detects both the 5' and 3' boundaries of the deletion, it most likely occurred by splicing. The acceptor coincides exactly with the 5' end of the first VNTR repeat unit (RU). Thus, the "A" RU conserved across all VNTR composite families [with the exception of a small group of LAVA elements (8)] remained intact. Interestingly, a similar splicing event has occurred in *Hylobates* and *Hoolock* LAVA elements: in this case, however, an acceptor 16bp upstream of the *Alu*-like/VNTR boundary has been used (Fig. 1). Elements containing this particular *Alu*-like domain are absent from the *Nomascus* genome (A. Damert, unpublished observations). A detailed analysis of the 5' CT-rich region of LAVA elements is still missing. However, an enrichment of Ts at the expense of Cs has been observed (3).

Lineage-specific evolution of SVA elements

As outlined above, the prototype SVA element originated from an "*Alu*-SVA2" through splicing of a HERV-K transcript to the SVA2 3' unique sequence.

Elements of this type can still be found in all hominoid lineages [(4), A. Damert, unpublished observations]. Subsequently, the SVA2-derived sequence was eliminated (Fig. 1). Again, splicing – between a VNTR repeat unit and the HERV-K derived sequence – appears to be the most likely mechanism. The loss of the SVA2 3' unique sequence must have predated the gibbon – great ape split, as one of the *Nomascus leucogenys* SVA copies already presents the direct VNTR-SINE-R fusion.

Based on the presence of lineage-specific insertions in both orangutan and gorilla/chimpanzee/human (GCH), it can be concluded that both element types, containing and lacking the SVA2-derived sequence, have been mobilized until after the split in both lineages. In humans, approximately 1/3 of the members of the oldest subfamily, SVA_A, contain the SVA2-derived sequence. Intriguingly, further evolution in both branches, orangutan and GCH, started from the elements lacking the SVA2 fragment. In orangutans, the structure of the SINE-R remained unchanged, with the exception of SVA$_{PA}$_12, where its 5'-most 164bp were lost as a consequence of a splicing event between the VNTR and the SINE-R. VNTR evolution occurred at the level of structural organization: in the evolutionary younger orangutan SVA subfamilies the VNTRs are characterized by a fixed 5' array of repeat units followed by tandemly repeated arrays of the Q and C units (8). A preliminary analysis suggests that the *Alu*-like domains of orangutan SVAs do not display indels larger than 1bp.

In the lineage leading to gorilla/chimpanzee/humans, evolution of SVA elements is characterized by a 16bp deletion at the 5' end of the SINE-R and a 5bp insertion and a 11bp deletion in the 5' part of the *Alu*-like region (10). In the VNTR domain, repeat units longer than 40bp arose and the specific organization with an invariant 5' end (TR) and tandem arrays in the 3' part evolved (8). The resulting subfamilies SVA_B to SVA_D are shared by gorillas, chimpanzees and humans (10). After the split of the three genera, additional genus-specific subfamilies appeared: SVA_E and SVA_F in humans and SVA_Pt_A in chimpanzees (10). A more recent analysis suggests that, in all three genera, SVA_D comprises a number of genus-specific subgroups (11). Compared to the consensus initially reported for SVA_D and to the SVA_D VNTR structure, neither human SVA_E/SVA_F nor SVA_Pt_A and the SVA_D genus-specific subgroups show major structural changes in any of the elements' domains [(8, 10, 11), A. Damert, unpublished observations].

Non-canonical SVAs in humans

In 2009, we (12) and others (13, 14) described a family of human-specific SVA elements, SVA_F1, in which nearly the entire 5' domain (hexameric repeats/*Alu*-like) has been replaced by the first exon of a cellular gene, *MAST2* (Fig. 1). The exon had been acquired by splicing

to an acceptor in the 3' part of the *Alu*-like region. Although splicing to the *Alu*-like region could be shown in other cases as well, none of the resulting chimeric elements proliferated to the extent of SVA_F1. A total of 84 SVA_F1 elements were identified in the human genome (12).

Evolution and function

Amplification of non-autonomous non-LTR retrotransposons requires (i) their transcription and (ii) interaction with the proteins synthesized from their autonomous counterparts (LINE-1 in case of SVA and LAVA). The efficiency of these two processes, transcription and interaction with LINE-1 proteins, determines the evolutionary success of non-autonomous elements. Retrotransposon transcription is subject to repression by host surveillance factors. SVAs have been shown to be engaged in an evolutionary arms race with KRAB zinc finger (KZNF) transcription factors which are part of the host defence and determine the specificity of KAP1/TRIM28-induced repression (15). ZNF91, the KZNF specifically recognizing SVA in embryonic stem cells, acquired an additional set of seven zinc fingers after the split of the orangutan from the lineage leading to gorilla/chimpanzee/humans, between 8–12 million years ago (15). In the same period of time, SVA subfamilies SVA_B to SVA_D evolved (10). The emergence of these subfamilies is accompanied by structural changes: indel formation in the *Alu*-like and SINE-R domains (10) and the appearance of longer repeat subunits and of RU arrays in the 3' part of the VNTR (8). The VNTR domain was found to be necessary and sufficient for ZNF91-mediated repression of reporter gene expression (15). Turelli and colleagues found TRIM28 significantly more associated with the older SVA subfamilies SVA_A to SVA_D than with their human-specific counterparts SVA_E and SVA_F (16). This finding is in line with the hypothesis that

the evolutionary youngest retrotransposon subfamilies are not yet recognized by specific KZNFs (16, 17). More recently, ZNF33A has been identified as a KZNF specifically interacting with SVAs in HEK293 cells. In this case, elements of subfamilies SVA_D to SVA_F were found enriched in the protein binding fraction (18). Overall, there is no clear-cut picture emerging yet on the evolutionary relationship between KZNFs and SVA subfamilies: the exact binding site of the KZNF tethering KAP1/TRIM28 to SVAs in ES cells could not be determined. The ZNF33A binding site (18) localizes to the border between *Alu*-like domain and VNTR – a region that is highly conserved across all SVA subfamilies.

Evolutionary relationships between SVA/LAVA and their autonomous partner LINEs have not been explored in detail. We could show that elements carrying the *Alu*-like domain typical for subfamilies SVA_B to SVA_F are mobilized more efficiently than those carrying the ancestral type of the domain – by an L1 element currently active in humans (4). In orangutan, however, elements containing the ancestral type SVA_A *Alu*-like region amplified – mobilized by L1 elements belonging to the evolutionary older L1PA3 subfamily (19). Non-canonical SVA elements in which the *Alu*-like domain has been replaced by a GC-rich exon proliferate in humans – suggesting that they are a suitable substrate of the currently active L1 subfamily in this lineage. In gibbons with their own specific set of L1 subfamilies (20), SVA did not multiply (10); in the lineage leading to the great apes, only remnants of LAVA elements survived (3) (A. Damert, unpublished observations). In perspective, it will be interesting to see whether there really are "perfectly matched" L1 – VNTR composite pairs in the different lineages of hominoid primates.

References

1. Ono M, Kawakami M, Takezawa T, *A novel human nonviral retroposon derived from an endogenous retrovirus*, ***Nucleic Acids Res.***, **15**, 8725-8737 (1987).

2. Shen L, Wu LC, Sanlioglu S, Chen R, Mendoza AR, Dangel AW, et al., *Structure and genetics of the partially duplicated gene RP located immediately upstream of the complement C4A and the C4B genes in the HLA class III region. Molecular cloning, exon-intron structure, composite retroposon, and breakpoint of gene duplication*, ***J. Biol. Chem.***, **269**, 8466-8476 (1994).

3. Carbone L, Harris RA, Mootnick AR, Milosavljevic A, Martin DI, Rocchi M, et al., *Centromere remodeling in Hoolock leuconedys (Hylobatidae) by a new transposable element unique to the gibbons*, ***Genome Biol. Evol.***, **4**, 648-658 (2012).

4. Ianc B, Ochis C, Persch R, Popescu O, Damert A, *Hominoid composite non-LTR retroposons-variety, assembly, evolution, and structural determinants of mobilization*, ***Mol. Biol. Evol.***, **31**, 2847-2864 (2014).

5. Hara T, Hirai Y, Baicharoen S, Hayakawa T, Hirai H, Koga A, *A novel composite retrotransposon derived from or generated independently of the SVA (SINE/VNTR/Alu) transposon has undergone proliferation in gibbon genomes*, ***Genes Genet. Syst.***, **87**, 181-190 (2012).

6. Carbone L, Harris RA, Gnerre S, Veeramah KR, Lorente-Galdos B, Huddleston J, et al., *Gibbon genome and the fast karyotype evolution of small apes*, ***Nature***, **513**, 195-201 (2014).

7. Han K, Konkel MK, Xing J, Wang H, Lee J, Meyer TJ, et al., *Mobile DNA in Old World monkeys: a glimpse through the rhesus macaque genome*, ***Science***, **316**, 238-240 (2007).

8. Lupan I, Bulzu P, Popescu O, Damert A, *Lineage specific evolution of the VNTR composite retrotransposon central domain and its role in retrotransposition of gibbon LAVA elements*, ***BMC Genomics***, **16**, 389 (2015).

9. Hancks DC, Kazazian HH Jr., *SVA retrotransposons: Evolution and genetic instability*, ***Semin. Cancer Biol.***, **20**, 234-245 (2010).

10. Wang H, Xing J, Grover D, Hedges DJ, Han K, Walker JA,

et al., *SVA elements: a hominid-specific retroposon family*, **J. Mol. Biol.**, **354**, 994-1007 (2005).

11. Levy O, Knisbacher BA, Levanon EY, Havlin S, *Integrating networks and comparative genomics reveals retroelement proliferation dynamics in hominid genomes*, **Sci. Adv.**, **3**, e1701256 (2017).

12. Damert A, Raiz J, Horn AV, Lower J, Wang H, Xing J, *et al.*, *5'-Transducing SVA retrotransposon groups spread efficiently throughout the human genome*, **Genome Res.**, **19**, 1992-2008 (2009).

13. Bantysh OB, Buzdin AA, *Novel family of human transposable elements formed due to fusion of the first exon of gene MAST2 with retrotransposon SVA*, **Biochemistry (Mosc.)**, **74**, 1393-1399 (2009).

14. Hancks DC, Ewing AD, Chen JE, Tokunaga K, Kazazian HH Jr., *Exon-trapping mediated by the human retrotransposon SVA*, **Genome Res.**, **19**, 1983-1991 (2009).

15. Jacobs FM, Greenberg D, Nguyen N, Haeussler M, Ewing AD, Katzman S, *et al.*, *An evolutionary arms race between KRAB zinc-finger genes ZNF91/93 and SVA/L1 retrotransposons*, **Nature**, **516**, 242-245 (2014).

16. Turelli P, Castro-Diaz N, Marzetta F, Kapopoulou A, Raclot C, Duc J, *et al.*, *Interplay of TRIM28 and DNA methylation in controlling human endogenous retroelements*, **Genome Res.**, **24**, 1260-1270 (2014).

17. Castro-Diaz N, Ecco G, Coluccio A, Kapopoulou A, Yazdanpanah B, Friedli M, *et al.*, *Evolutionarily dynamic L1 regulation in embryonic stem cells*, **Genes Dev.**, **28**, 1397-1409 (2014).

18. Najafabadi HS, Mnaimneh S, Schmitges FW, Garton M, Lam KN, Yang A, *et al.*, *C2H2 zinc finger proteins greatly expand the human regulatory lexicon*, **Nat. Biotechnol.**, **33**, 555-562 (2015).

19. Gokcumen O, Tischler V, Tica J, Zhu Q, Iskow RC, Lee E, *et al.*, *Primate genome architecture influences structural variation mechanisms and functional consequences*, **Proc. Natl. Acad. Sci. U.S.A.**, **110**, 15764-15769 (2013).

20. Smit AF, Hubley R, Green P. RepeatMasker. http://repeatmasker.org.