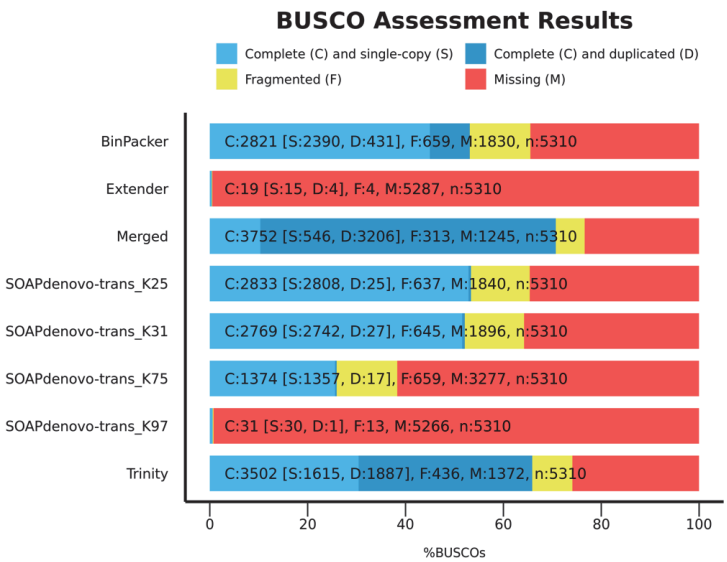
Convergent molecular evolution of toxins in the venom of advanced snakes (Colubroidea)

Xie, B.

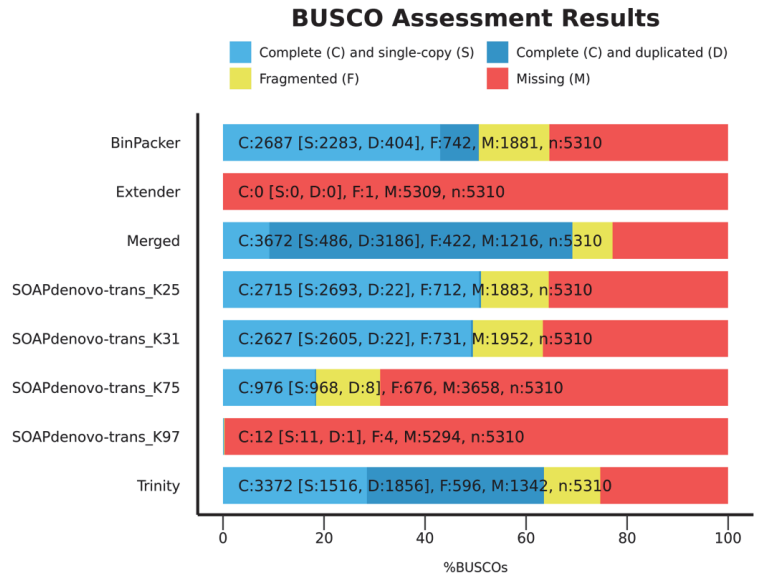
Chapter 7. Supplementary Materials

This chapter is published as Supplementary Material of: Bing Xie, Daniel Dashevsky, Darin Rokyta, Parviz Ghezellou, Behzad Fathinia, Qiong Shi, Michael K. Richardson and Bryan G. Fry. Dynamic genetic differentiation drives the widespread structural and functional convergent evolution of snake venom proteinaceous toxins. *BMC Biology*, 2022, 20:4.

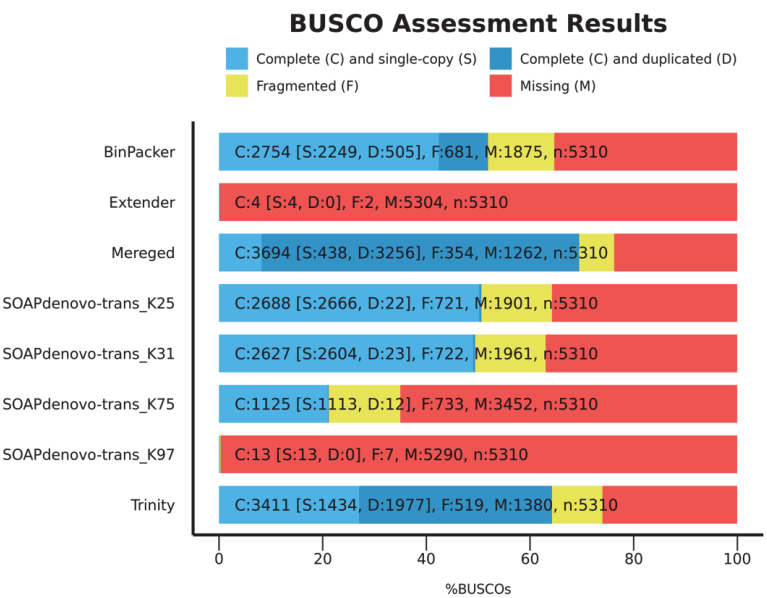
https://doi.org/10.1186/s12915-021-01208-9



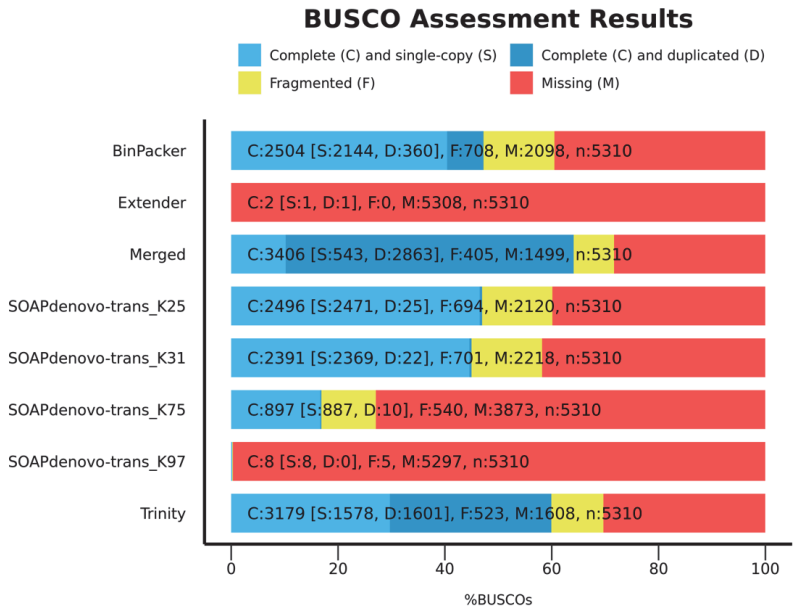
Supplementary Figure 1A: BUSCO completeness analyses of venom gland transcriptome assemblies for *Helicops leopardinus*. Snake contigs were matched against 5310 orthologous loci defined within Tetrapoda.



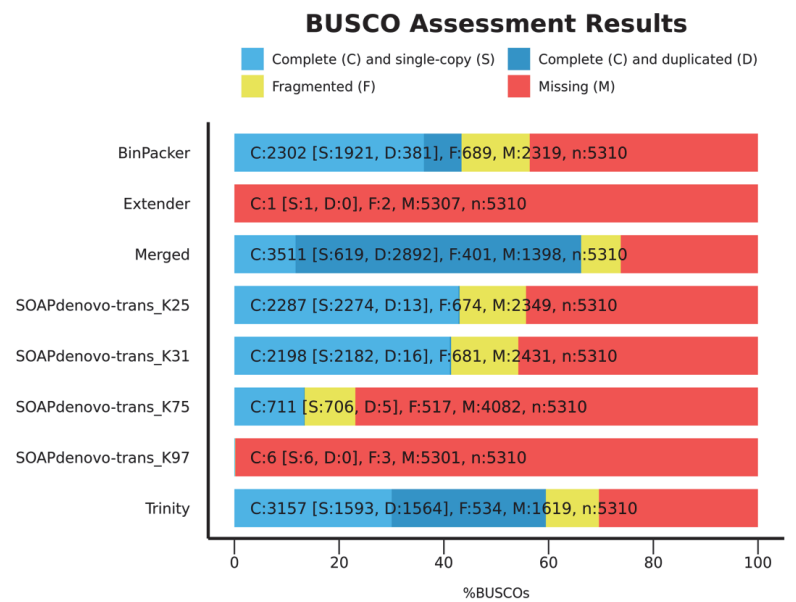
Supplementary Figure 1B: BUSCO completeness analyses of venom gland transcriptome assemblies for *Rhabdophis subminiatus*. Snake contigs were matched against 5310 orthologous loci defined within Tetrapoda.



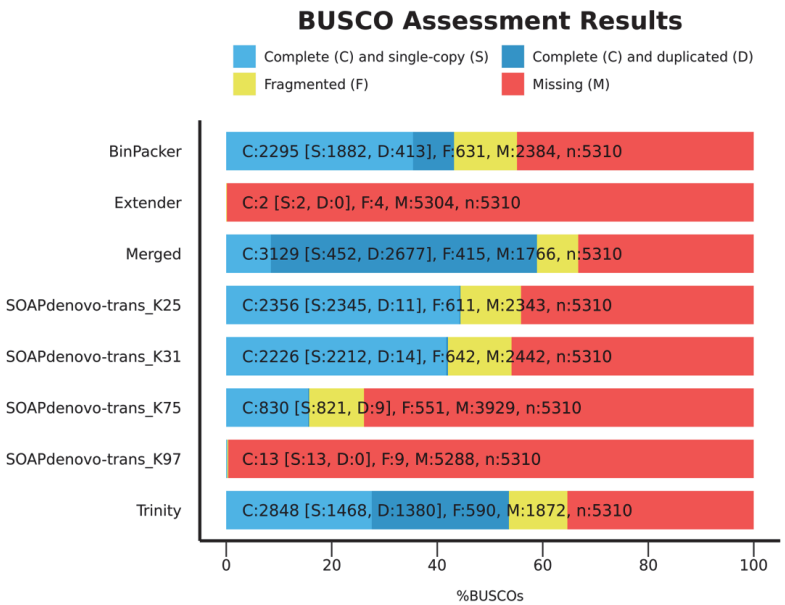
Supplementary Figure 1C: BUSCO completeness analyses of venom gland transcriptome assemblies for *Heterodon nasicus*. Snake contigs were matched against 5310 orthologous loci defined within Tetrapoda.



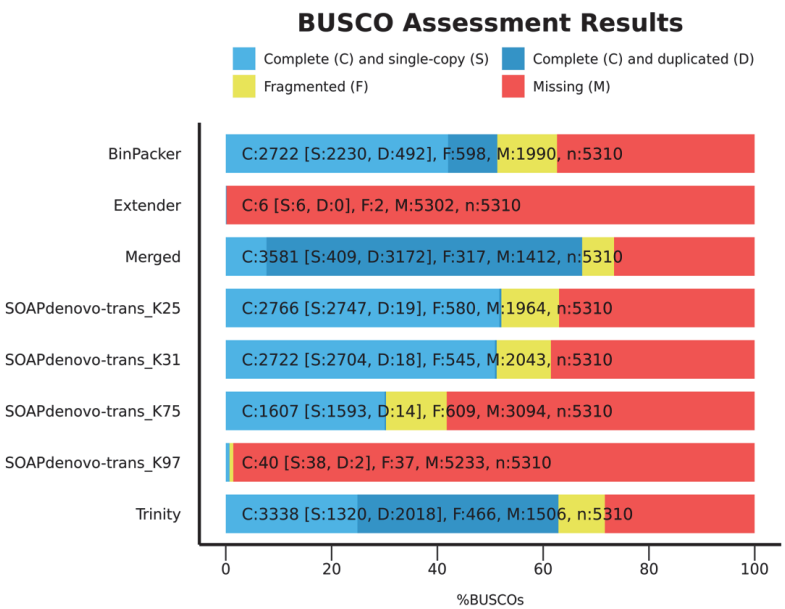
Supplementary Figure 1D: BUSCO completeness analyses of venom gland transcriptome assemblies for *Malpolon monspessulanus*. Snake contigs were matched against 5310 orthologous loci defined within Tetrapoda.



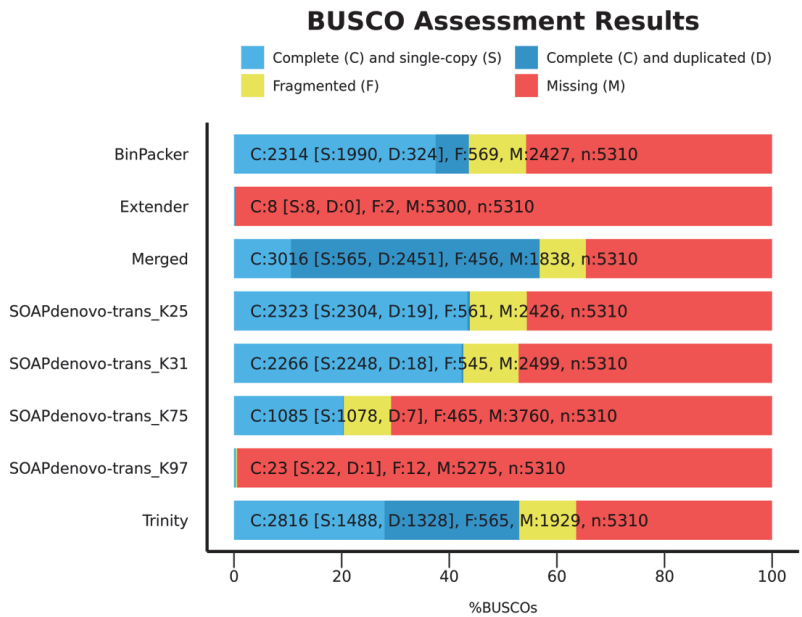
Supplementary Figure 1E: BUSCO completeness analyses of venom gland transcriptome assemblies for *Psammophis schokari*. Snake contigs were matched against 5310 orthologous loci defined within Tetrapoda.



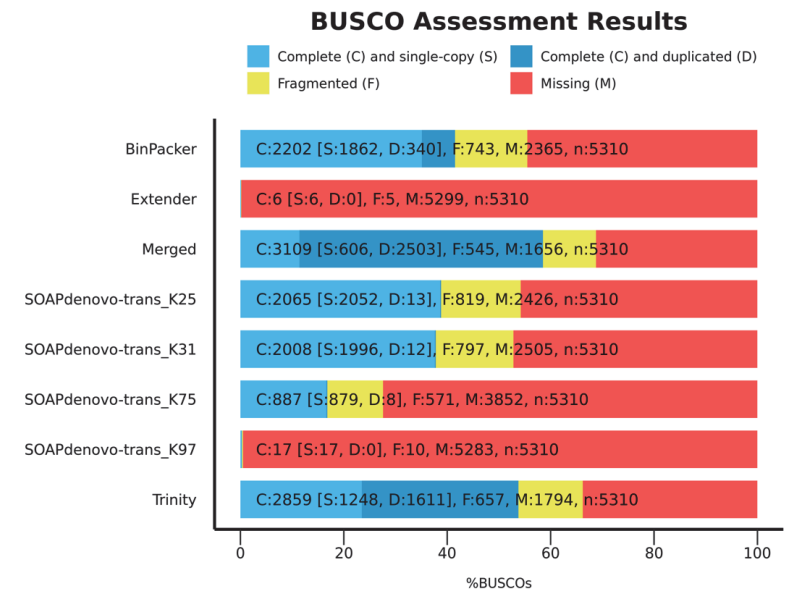
Supplementary Figure 1F: BUSCO completeness analyses of venom gland transcriptome assemblies for *Psammophis subtaeniatus*. Snake contigs were matched against 5310 orthologous loci defined within Tetrapoda.



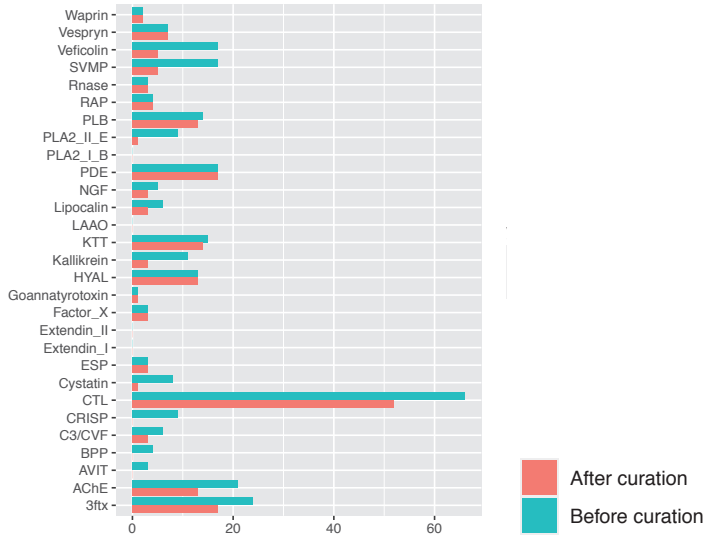
Supplementary Figure 1G: BUSCO completeness analyses of venom gland transcriptome assemblies for *Homalopsis buccata*. Snake contigs were matched against 5310 orthologous loci defined within Tetrapoda.



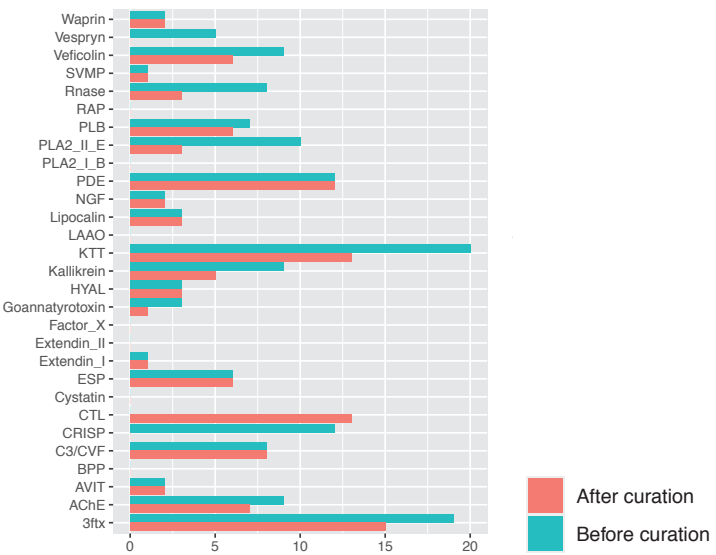
Supplementary Figure 1H: BUSCO completeness analyses of venom gland transcriptome assemblies for *Pseudocerastes urarachnoides*. Snake contigs were matched against 5310 orthologous loci defined within Tetrapoda.



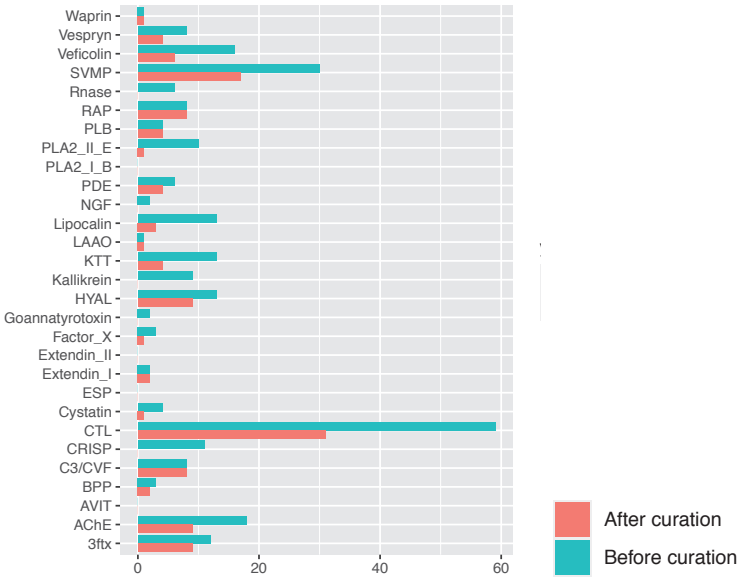
Supplementary Figure 1I: BUSCO completeness analyses of venom gland transcriptome assemblies for *Vipera transcaucasiana*. Snake contigs were matched against 5310 orthologous loci defined within Tetrapoda.



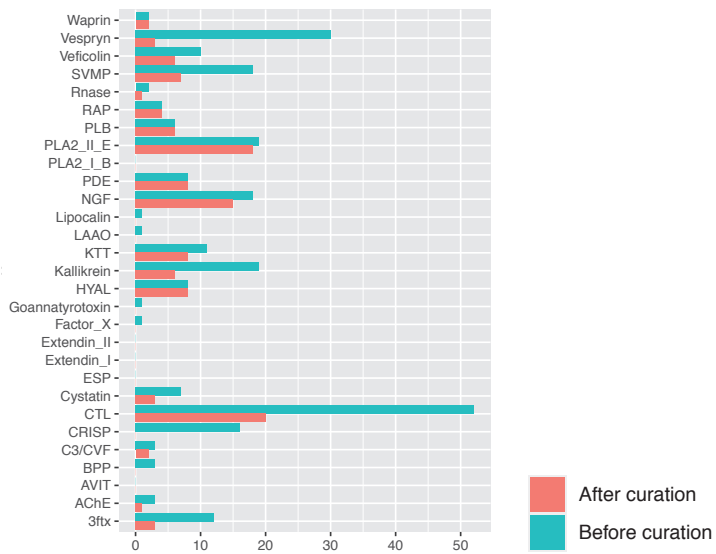
Supplementary Figure 2A: The statistics of toxins from *Helicops leopardinus* before and after curation. As can be seen, curation leads to a huge decline in the number of both toxin families diversities and toxin transcripts. Some toxin families are discarded as a whole. X axis is toxin family name and Y axis is the number of the corresponding toxin transcripts.



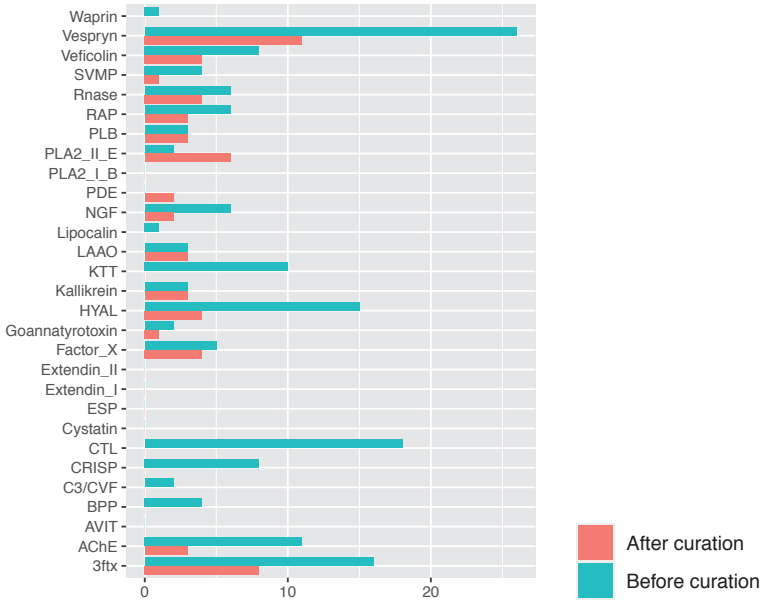
Supplementary Figure 2B: The statistics of toxins from *Rhabdophis subminiatus* before and after curation. As can be seen, curation leads to a huge decline in the number of both toxin families diversities and toxin transcripts. Some toxin families are discarded as a whole. X axis is toxin family name and Y axis is the number of the corresponding toxin transcripts.



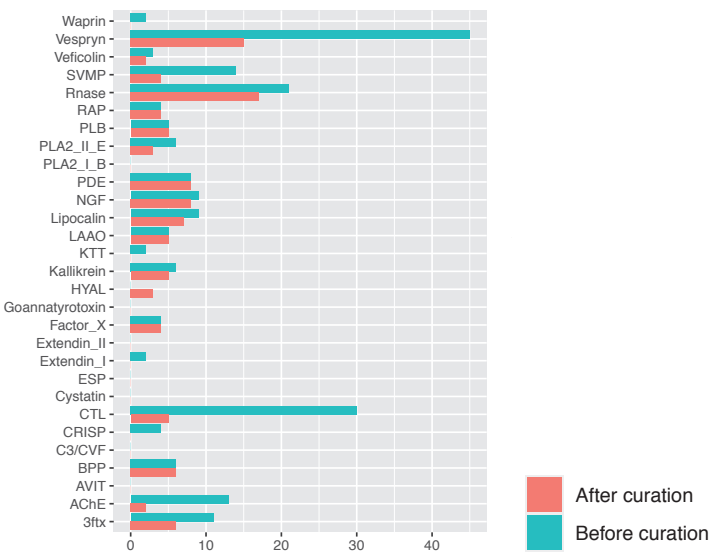
Supplementary Figure 2C: The statistics of toxins from *Heterodon nasicus* before and after curation. As can be seen, curation leads to a huge decline in the number of both toxin families diversities and toxin transcripts. Some toxin families are discarded as a whole. X axis is toxin family name and Y axis is the number of the corresponding toxin transcripts.



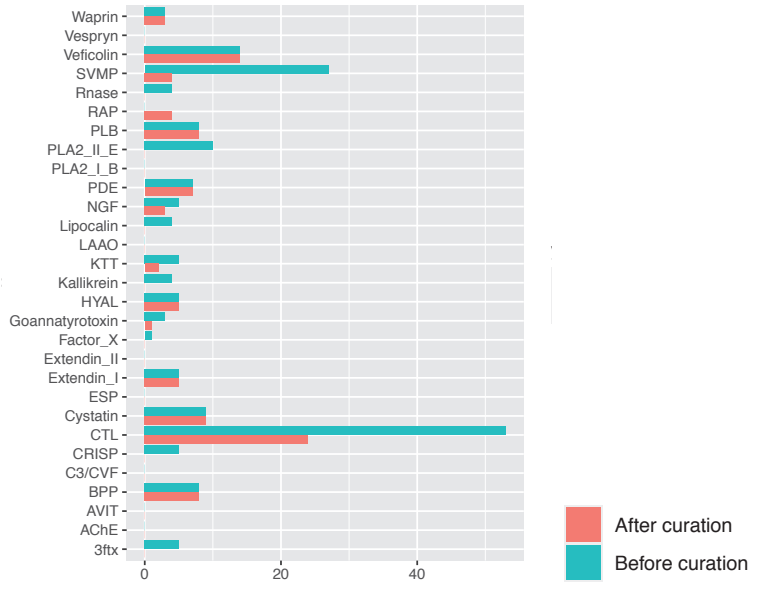
Supplementary Figure 2D: The statistics of toxins from *Malpolon monspessulanus* before and after curation. As can be seen, curation leads to a huge decline in the number of both toxin families diversities and toxin transcripts. Some toxin families are discarded as a whole. X axis is toxin family name and Y axis is the number of the corresponding toxin transcripts.



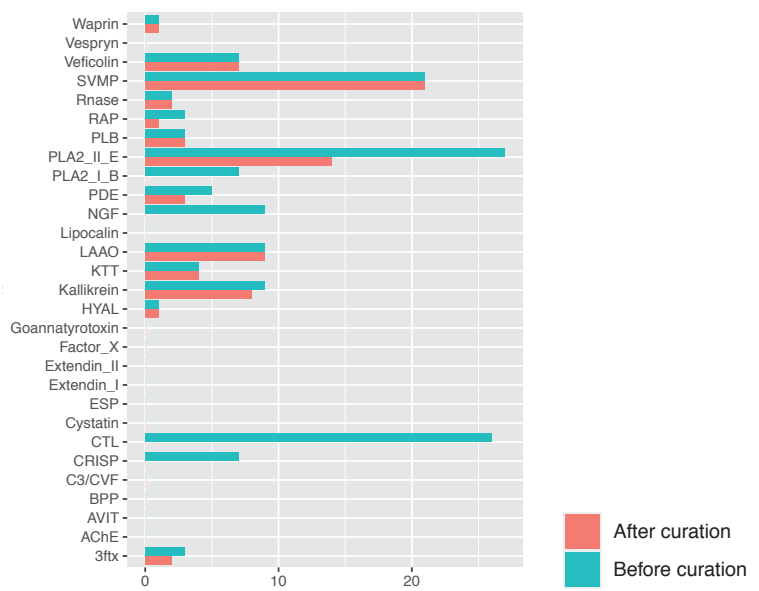
Supplementary Figure 2E: The statistics of toxins from *Psammophis schokari* before and after curation. As can be seen, curation leads to a huge decline in the number of both toxin families diversities and toxin transcripts. Some toxin families are discarded as a whole. X axis is toxin family name and Y axis is the number of the corresponding toxin transcripts.



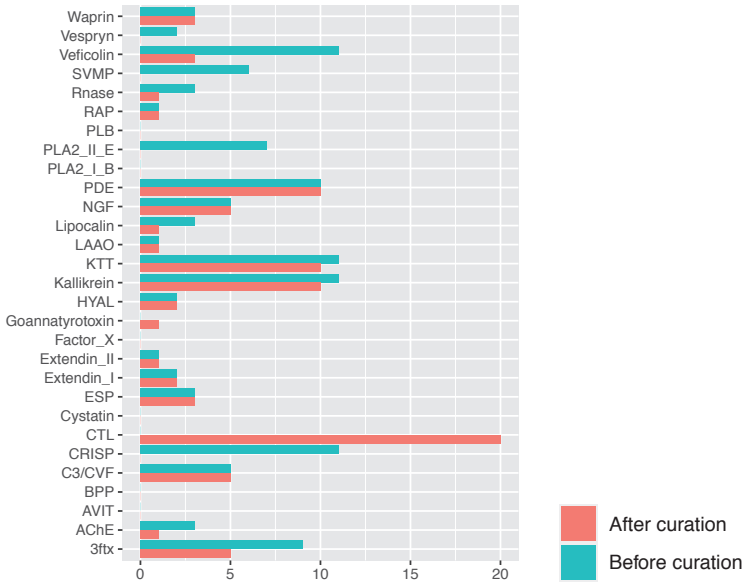
Supplementary Figure 2F: The statistics of toxins from *Psammophis subtaeniatus* before and after curation. As can be seen, curation leads to a huge decline in the number of both toxin families diversities and toxin transcripts. Some toxin families are discarded as a whole. X axis is toxin family name and Y axis is the number of the corresponding toxin transcripts.



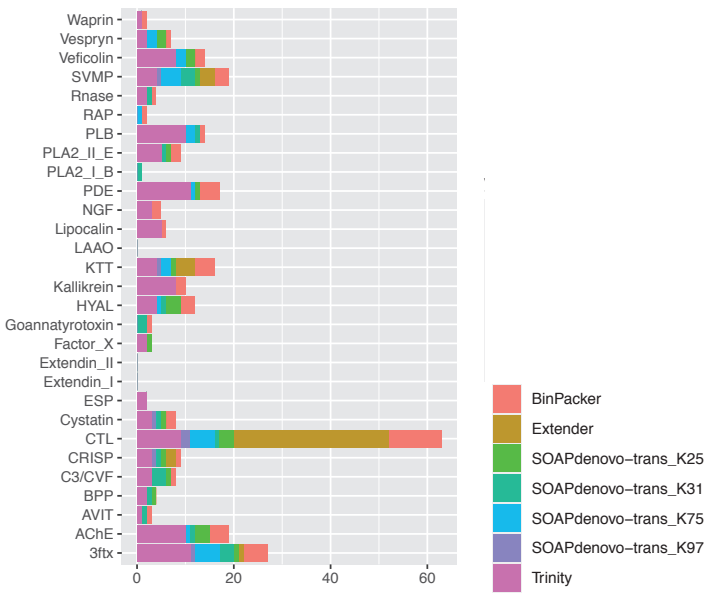
Supplementary Figure 2G: The statistics of toxins from *Homalopsos buccata* before and after curation. As can be seen, curation leads to a huge decline in the number of both toxin families diversities and toxin transcripts. Some toxin families are discarded as a whole. X axis is toxin family name and Y axis is the number of the corresponding toxin transcripts.



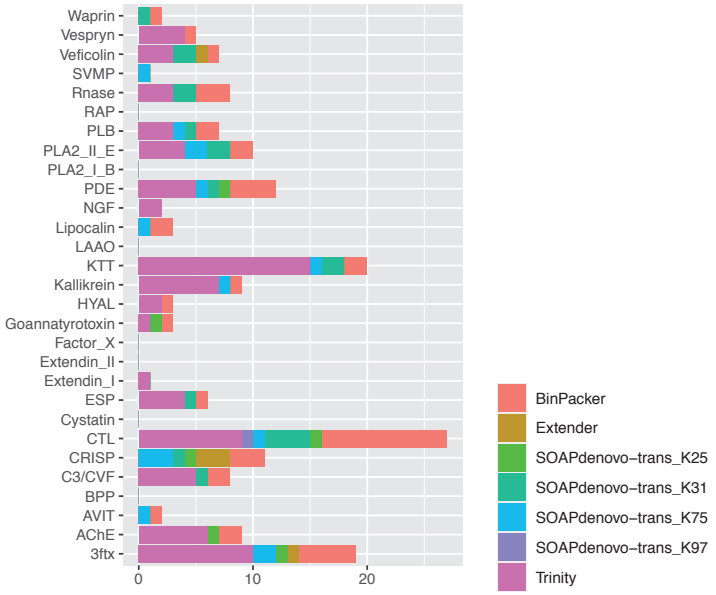
Supplementary Figure 2H: The statistics of toxins before and after curation. As can be seen, curation leads to a huge decline in the number of both toxin families diversities and toxin transcripts. Some toxin families are discarded as a whole. X axis is toxin family name and Y axis is the number of the corresponding toxin transcripts.



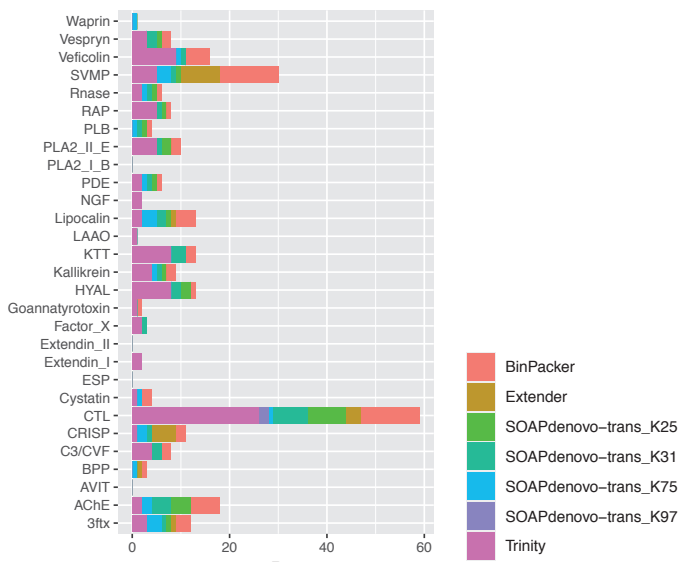
Supplementary Figure 2I: The statistics of toxins before and after curation. As can be seen, curation leads to a huge decline in the number of both toxin families diversities and toxin transcripts. Some toxin families are discarded as a whole. X axis is toxin family name and Y axis is the number of the corresponding toxin transcripts.



Supplementary Figure 3A: Toxin distribution of *Helicops leopardinus* before curation. For each toxin family, the numbers of toxin transcripts recovered by different assembly methods are indicated by different colours.

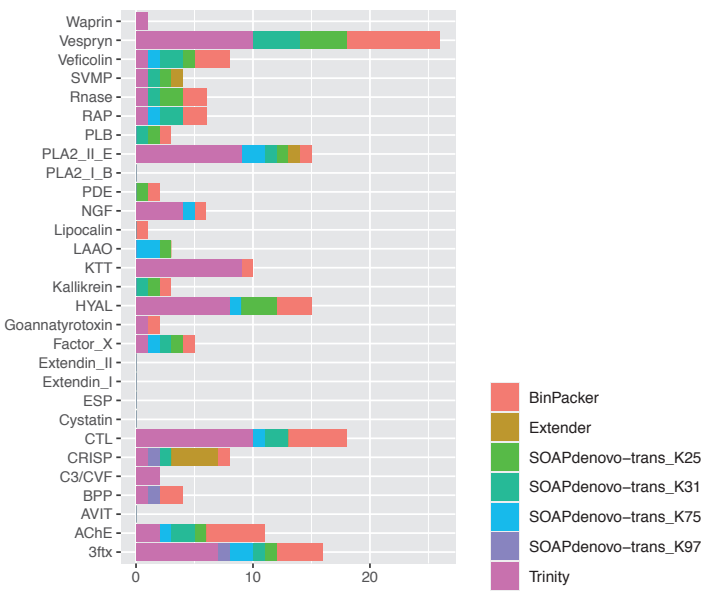


Supplementary Figure 3B: Toxin distribution of *Rhabdophis subminiatus* before curation. For each toxin family, the numbers of toxin transcripts recovered by different assembly methods are indicated by different colours.

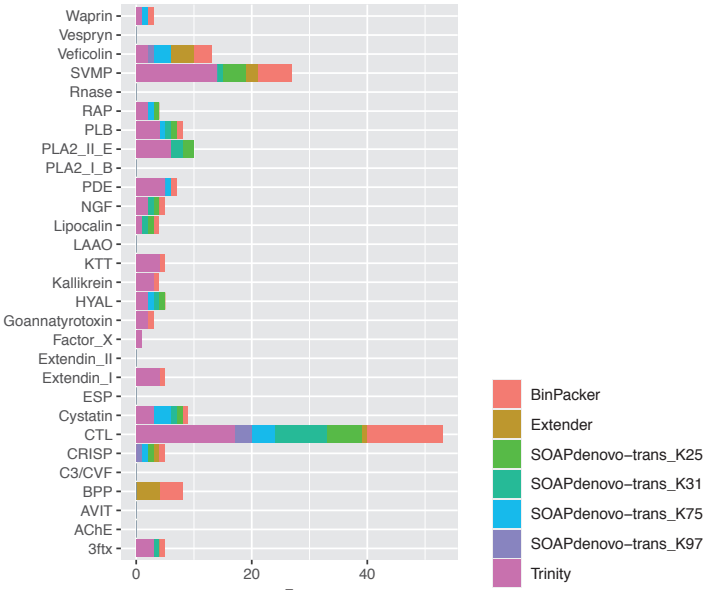


Supplementary Figure 3C: Toxin distribution of *Heterodon nasicus* before curation. For each toxin family, the numbers of toxin transcripts recovered by different assembly methods are indicated by different colours.

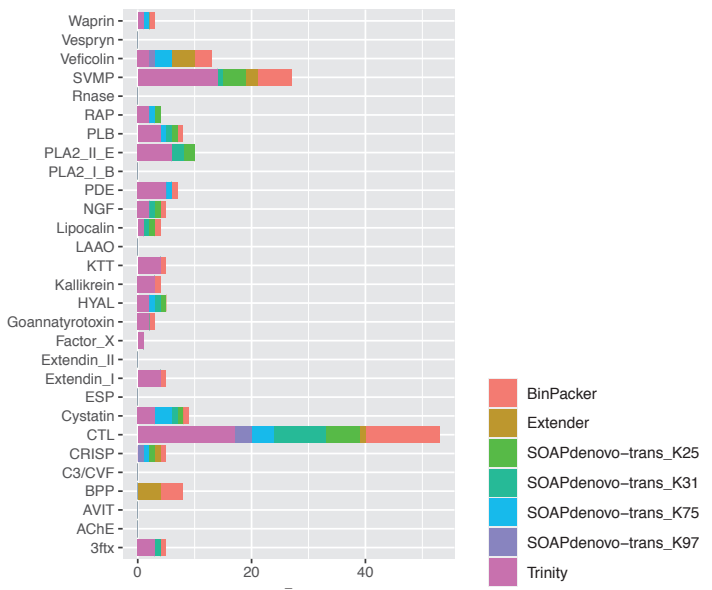
Supplementary Figure 3D: Toxin distribution of *Malpolon monspessulanus* before curation. For each toxin family, the numbers of toxin transcripts recovered by different assembly methods are indicated by different colours.



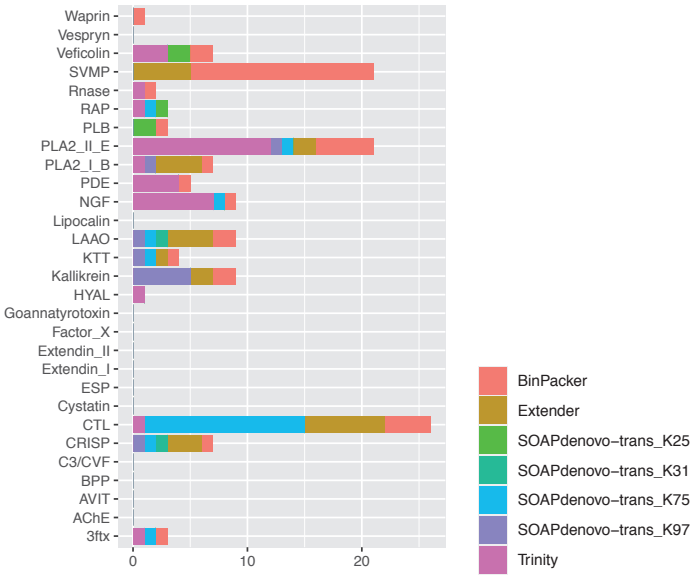
Supplementary Figure 3E: Toxin distribution of *Psammophis schokari* before curation. For each toxin family, the numbers of toxin transcripts recovered by different assembly methods are indicated by different colours.



Supplementary Figure 3F: Toxin distribution of *Psammophis subtaeniatus* before curation. For each toxin family, the numbers of toxin transcripts recovered by different assembly methods are indicated by different colours.

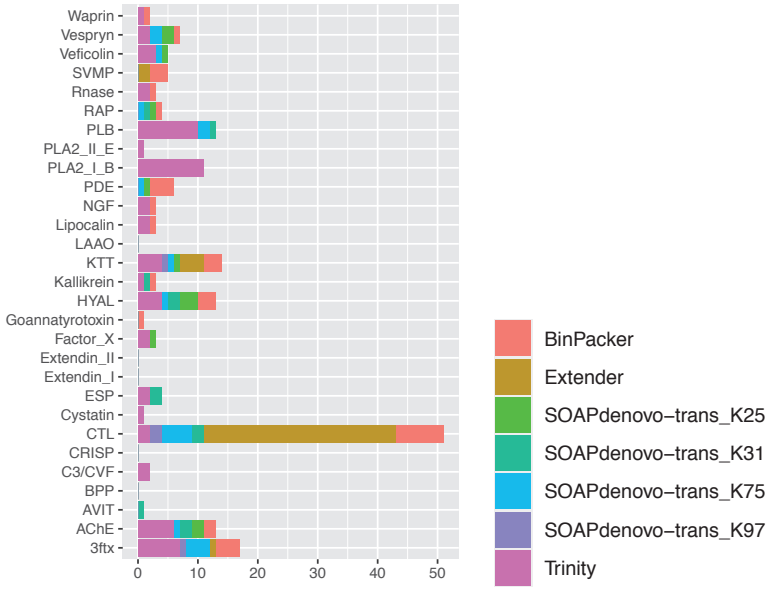


Supplementary Figure 3G: Toxin distribution of *Homalopsis buccata* before curation. For each toxin family, the numbers of toxin transcripts recovered by different assembly methods are indicated by different colours.

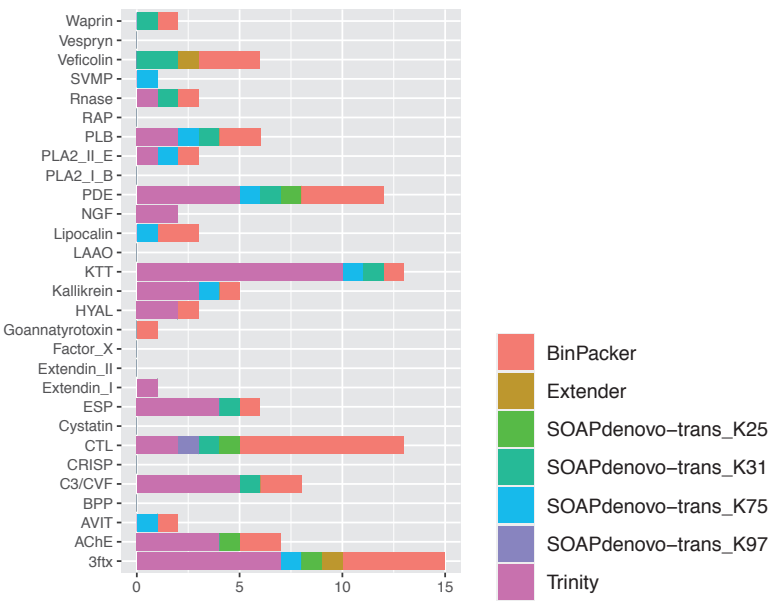


Supplementary Figure 3H: Toxin distribution of *Pseudocerastes urarachnoides* before curation. For each toxin family, the numbers of toxin transcripts recovered by different assembly methods are indicated by different colours.

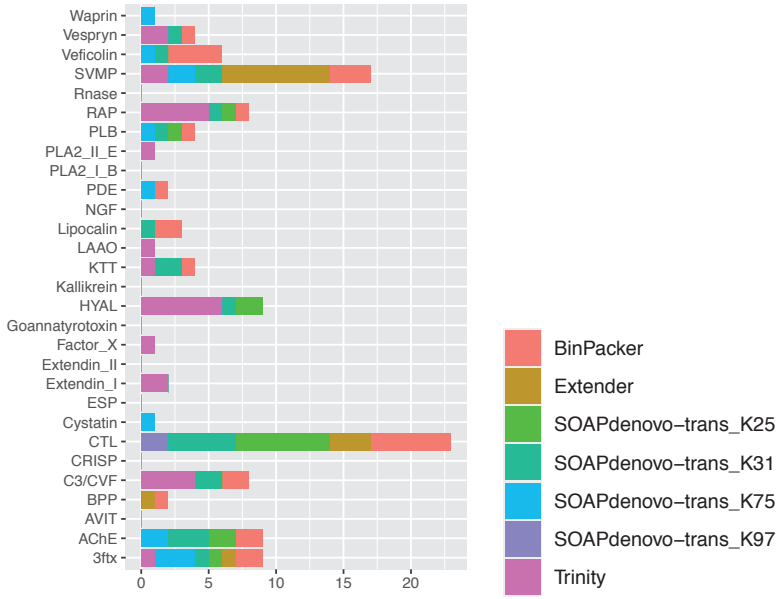
Supplementary Figure 3I: Toxin distribution of *Vipera transcaucasiana* before curation. For each toxin family, the numbers of toxin transcripts recovered by different assembly methods are indicated by different colours.



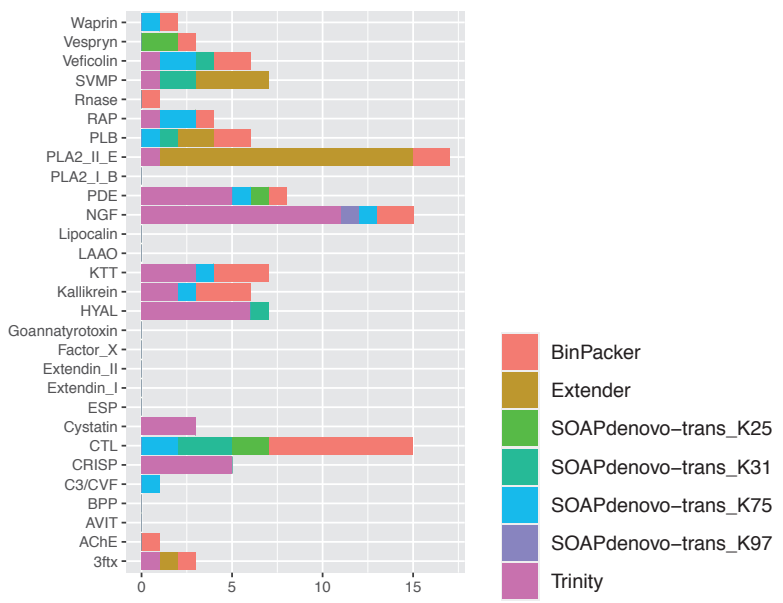
Supplementary Figure 4A: Toxin distribution of *Helicops leopardinus* after curation. For each toxin family, the numbers of toxin transcripts recovered by different assembly methods are indicated by different colours.



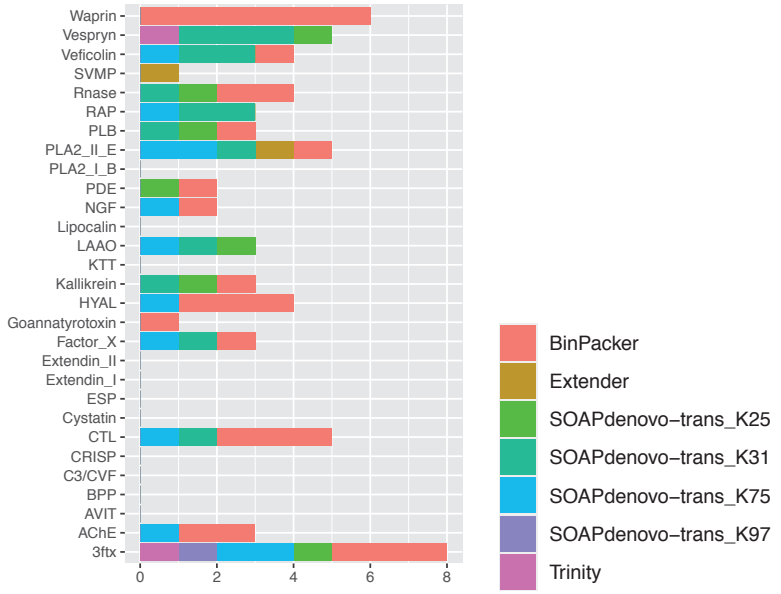
Supplementary Figure 4B: Toxin distribution of *Rhabdophis subminiatus* after curation. For each toxin family, the numbers of toxin transcripts recovered by different assembly methods are indicated by different colours.



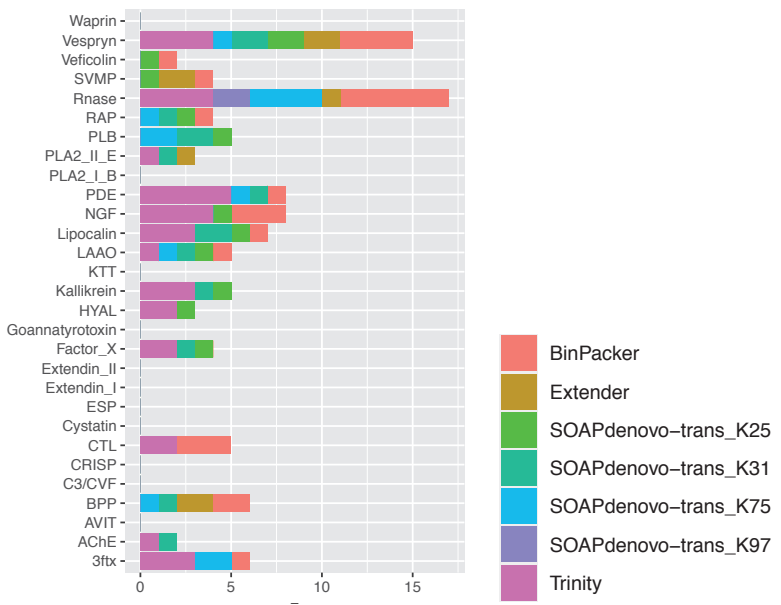
Supplementary Figure 4C: Toxin distribution of *Heterodon nasicus* after curation. For each toxin family, the numbers of toxin transcripts recovered by different assembly methods are indicated by different colours.



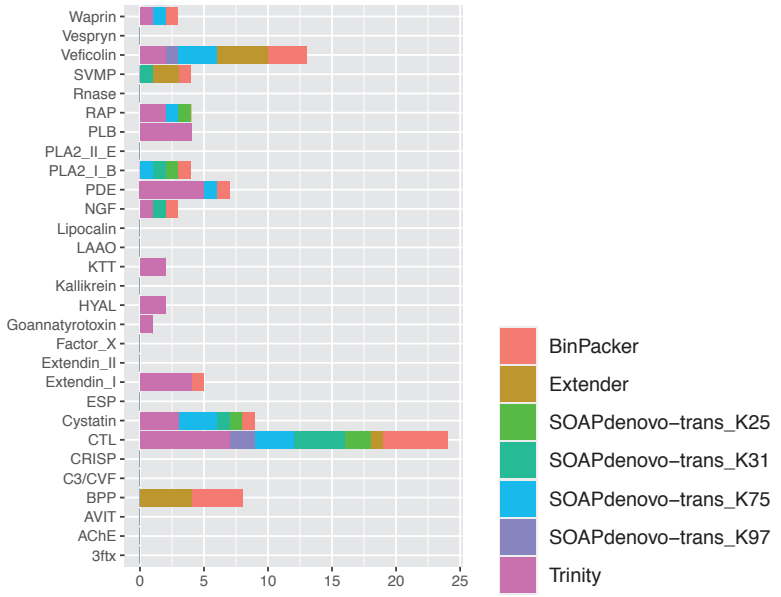
Supplementary Figure 4D: Toxin distribution of *Malpolon monspessulanus* after curation. For each toxin family, the numbers of toxin transcripts recovered by different assembly methods are indicated by different colours.



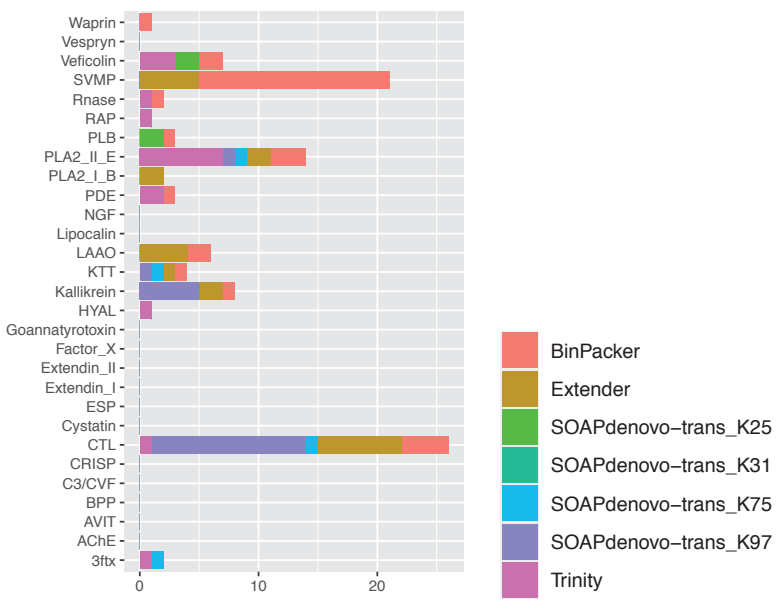
Supplementary Figure 4E: Toxin distribution of *Psammophis schokari* after curation. For each toxin family, the numbers of toxin transcripts recovered by different assembly methods are indicated by different colours.



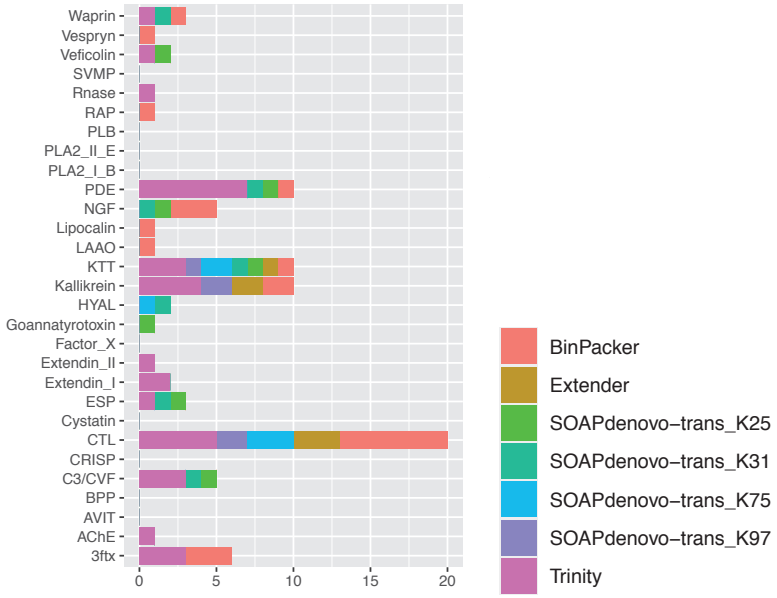
Supplementary Figure 4F: Toxin distribution of *Psammophis subtarniatus* after curation. For each toxin family, the numbers of toxin transcripts recovered by different assembly methods are indicated by different colours.



Supplementary Figure 4G: Toxin distribution of *Homalopsis buccata* after curation. For each toxin family, the numbers of toxin transcripts recovered by different assembly methods are indicated by different colours.



Supplementary Figure 4H: Toxin distribution of *Pseudocerastes urarachnoides* after curation. For each toxin family, the numbers of toxin transcripts recovered by different assembly methods are indicated by different colours.



Supplementary Figure 4I: Toxin distribution of *Vipera transcaucasiana* after curation. For each toxin family, the numbers of toxin transcripts recovered by different assembly methods are indicated by different colours.

Supplementary File 2: The settings for MrBayes

begin mrbayes;

log start replace;

set autoclose = no nowarn=no;

lset applyto = (all) nst = 6 rates = invgamma;

prset applyto = (all) aamodelpr = mixed;

unlink revmat = (all) shape = (all) pinvar = (all) statefreq = (all) tratio = (all);

showmodel;

mcmc ngen = 15000000 printfreq = 1000 samplefreq = 100 nchains = 4 temp = 0.2 checkfreq = 50000 diagnfreq = 1000 stopval = 0.01 stoprule = yes;

sumt relburnin = yes burninfrac = 0.25 contype = halfcompat;

sump relburnin = yes burninfrac =0.25;

outgroup 1;

log stop;

end;

For the following **Supplementary Files**, they can be viewed via

https://figshare.com/articles/dataset/Supplementary_Files/15085353

Supplementary_File_1_inhouse_toxin_database.fasta

Supplementary_File_3_kunitz_alignment.fasta

Supplementary_File_4_kunitz_alignment.con.tre

Supplementary_File_5_lectin_alignment.fasta

Supplementary_File_6_lectin_tree.con.tre

Supplementary_File_7_SVMP_alignment.fasta

Supplementary_File_8_SVMP_tre.con.tre

Supplementary_File_9_SVMP_propeptide_alignment.fasta

Supplementary_File_10_SVMP_propeptide_tree.con.tre