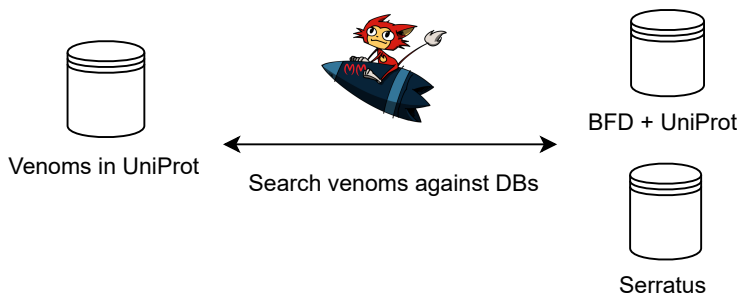


1. Search



18184892 pairs

2. Predict cleaving

Annotation: **VASKRWTRPSVNL**P

Alignment: **MPGAVASKRWTRPSV-NLP**AGTNSD <- given venom
 MPA**AVASS----**TE**ACLQ**PVGSDAS <- DB hit

Inferred: **VASSTEACLQ**P

Requires > 90% overlap

2924870 pairs
739724 metagenomic
 8202 venoms

3. Length Cutoff (100 aa)

VPRDDN**PGGASGK**FMNVLRRSGCPWEPWCG
 NTGGAS**RKFLNVP**RESECPWRPWCG
 ECPW**KPWCG**
 VPRDDY**SRRANR**

865100 pairs
333787 metagenomic
 6471 venoms

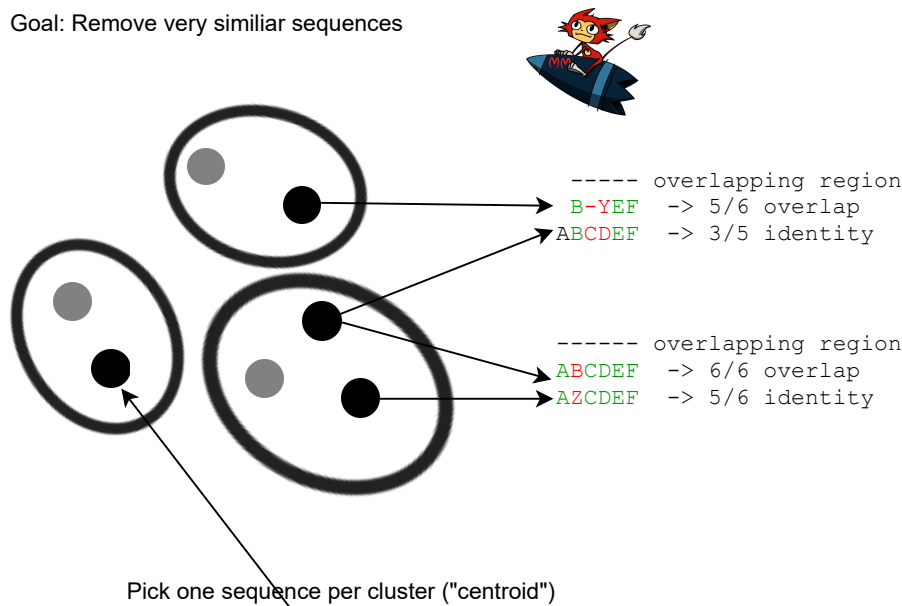
4. Remove redundant hits



1380546 pairs
381128 metagenomic
 6913 venoms

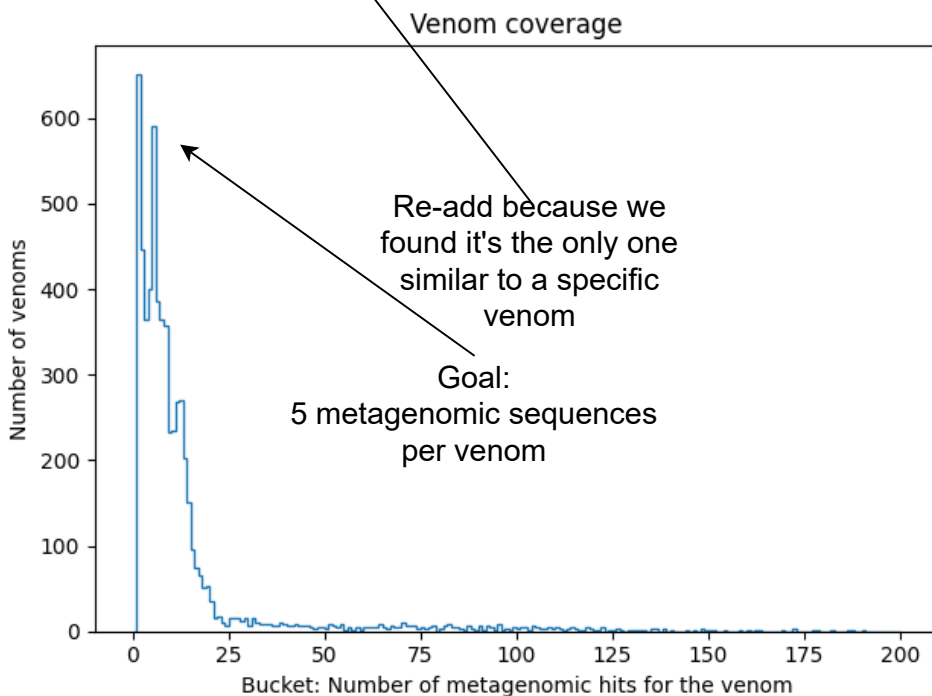
5. Cluster

Goal: Remove very similar sequences



70415 pairs
37672 metagenomic
 5424 venoms

6. Improve venom coverage



85406 pairs
39583 metagenomic
 5941 venoms

