

Transcriptome Assembly

.fasta



Pre-filter step

Blast search
($\geq 90\%$ coverage)
using the contigs
against a toxinDB

$\geq 90\%$
coverage

CDS prediction step

Prediction of CDS
using gHMM models
pre-build to each
toxin family



Toxin filter step

Blast search
($\geq 80\%$ coverage)
using the predicted CDSs
against a toxinDB

Toxin prediction step

$< 90\%$
coverage



Non-Toxins

.fasta

$< 80\%$ coverage
or no CDS predicted



Putative Toxins

.fasta

$\geq 80\%$
coverage



Toxins

.fasta

SignalP filter step

Filter CDSs based
on similarity search
against a toxinDB
that has a signalP



SignalP filter step

Keep only CDSs
with a predicted signalP
using SignalP-4.1



Redundancy filter step

Clustering of 100%
similar toxins predicted



Putative Toxins

SP_filtered
Redundancy_filtered
.fasta



Toxins

SP_filtered
Redundancy_filtered
.fasta

Post-toxin prediction step

