Incidence and diversity of insecticidal proteins in metagenomes from targeted environments


1Second Genome, Inc, 341 Allerton Ave, South San Francisco, CA USA, 2Bayer, 700 Chestefield Parkway West, Chestefield, MO USA

Abstract

Insecticidal microbes are an effective and environmentally safe solution against agricultural pests and insects that vector human pathogens. The increasing rate of insect resistance to commonly used toxin proteins, e.g. in Cry families, necessitates the search for novel insecticidal proteins. Historically, insecticidal bacteria have been discovered by selective isolation of insecticidal species from dead insects, grain dust, and soil. Here we used metagenomics to access genetic information from insecticidal bacteria that escaped isolation and to enable efficient mining of a wide variety of environmental sources. However, insecticidal bacteria outside of an insect host are often found at low abundances making assembly of their genes from complex metagenomes especially challenging. In the present study, we evaluated the presence and diversity of insecticidal proteins in various environments including soil, grain, and plant surfaces. To improve the recovery of assembled insecticidal genes from complex environmental samples, we developed a set of enrichment approaches facilitating the growth of insecticidal bacteria. The enriched metagenomes contained hundreds of insecticidal-like proteins identified using the Second Genome discovery platform based on both homology and machine learning methods. The results of this study showed that insecticidal genes can be successfully assembled from enriched metagenomes and the occurrence of insecticidal protein classes varies by environmental source and enrichment approach. The results have implications for agriculture and healthcare and for better understanding the ecology of insecticidal bacteria.

I. Second Genome discovery platform is based on generating and mining ‘omics data, curated metadata and machine learning

II. Biosourcing strategy: diverse environmental samples were collected and treated to enrich for insecticidal bacteria

A. Incidence of insecticidal protein classes and diversity of classes were estimated for each metagenome.
B. A biospecimen was considered a candidate insecticidal biospecimen for a class of proteins (Class 1, Class 2) if its metagenome contained at least one insecticidal protein of this class. A protein was considered present in a metagenome if reads to its gene were found in 3X coverage over the length of the protein.
C. A biospecimen was considered a candidate insecticidal biospecimen for diversity of protein classes if its metagenome contained insecticidal proteins from at least five classes.
D. Enrichment for candidate insecticidal biospecimens for each criterion (Class or Class diversity) is shown on Y axis. Counts of candidate insecticidal biospecimens and total biospecimens shown above bar for each enrichment type on X axis.

III. Statistics of metagenome libraries and assembles

- Methods
  - Metagenomes were sequenced using Illumina NexSeq (150 bp PE), targeting 10-100 Mb reads per library
  - Read QC included adapter and quality trimming, contaminant and human reads removal (Trimmomatic, Bowtie2, Kraken)
  - Assembly was done with Megahit or SPAdes depending on library complexity
  - Prodigal was used for gene calling.

- Statistics
  - Average per metagenome:
    - Number of post-QC reads
    - Bases assembled, bp
    - N50, bp
    - Contigs
    - Max contig length, bp
    - Median contig length, bp
    - Unique genes, (ftp1=1350)
    - Unique proteins, (ftp1=1350)

- Data is shown for n=1839 libraries, unless noted.

IV. Enrichments resulted in highest rate of insecticidal proteins occurrence and diversity

- Richness of insecticidal proteins varied with geographic location
- The richness of insecticidal proteins per sampling site is shown here on an example of Class 1 proteins.
- Richness was calculated as count of unique Class 1 proteins per million of reads in each metagenome.
- Rich biospecimens for Class 1 proteins occur independently of geographic location.

V. Richness of insecticidal proteins varied with geographic location

- Bioassays more frequently than from metagenomes obtained from original biospecimens.
- Nominated candidate insecticidal proteins from enriched metagenomes produced hits in insect bioassays.

VI. Candidate proteins identified in metagenomes were positive in insect bioassays

- Bioassays for selected candidate insecticidal proteins in insects showed positive results.

Conclusions

- Incidence and diversity of candidate insecticidal proteins varied by biospecimen and enrichment types and were highest in enrichments than in original biospecimens.
- Latitude and longitude had no clear effect on the rate of candidate insecticidal proteins presence in metagenomes evaluated here.
- Nominated candidate insecticidal proteins from enriched metagenomes produced hits in insect bioassays more frequently than from metagenomes obtained from original biospecimens.
- Higher number of hits from enriched metagenomes is likely due to many factors including lower number of original biospecimens, challenges in assembling genes from metagenome of original biospecimens with high microbial diversity and also biology of insecticidal bacteria.

Acknowledgments

- The MAD lab at Second Genome for sequencing >2400 metagenomes and >2200 V4 biospecimens, T.