Figure S1. Schematic summary of *Dynamite* workflow used in this study.

conserved NP protein queries

local NCBI protein databank

blastp query NCBI protein databank

hits and associated metadata

sort based on GI#

gene clusters

classify by biosynthetic characteristics based on presence of specific sets of hits

gene clusters classified by biosynthetic type (e.g. type I PKS, type II PKS, NRPS)

cluster summary files

multi-fasta files of protein sequences of specific hit types

user viewing, manual analysis (e.g. sequence alignment, dendrogram)

further analysis