Supplementary Table 1. Comparation of BtToxin\_Digger with existing software

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Tools** | **Available** | **methods used** | **Supported inputs** | **Target genes** | **Output files** | **Flux** |
| BtToxin\_Digger | Web services, command line | Blast, HMM, SVM | Illumina/PacBio/ONT reads, assembled genomes, protein sequences, and coding sequences | Cry, Cyt, Vip, Gpp, Mcf, Mpf, Mpp, Mtx, Pra, Prb, Spp, Tpp, Vpa, Vpb, Xpp, Sip, Chitinase, InhA, Bmp1, Enhancin, and ZwA. | Toxin list file and sequences file for each input, a matrix file describes all number of strains vs. all toxins, an integrated file contains sequences and information of all inputs | Unlimited number of inputs with a one-line command |
| BtToxin\_scanner | Web services | Blast, HMM, SVM | Assembled genomes, protein sequences， and ORFs | Cry toxins | Toxin list file and sequences file for each input | One submission at a time |
| CryProcessor | Web services (Need register and be activated by the administrator), command line | HMM | Illumina reads, protein sequences | Three-domain Cry toxins | A directory containing multiple files for each input | Unlimited number of inputs with additional shell scripts |

**Blast**: Basic Local Alignment Search Tool. **SVM**: Support Vector Machines. **HMM**: Hidden Markov Model. **Cry**: crystal proteins. **Cyt**: Primarily dipteran-active proteins. **Vip**: Vegetative Insecticidal Proteins. **Gpp**: Aegerolysin-related pesticidal proteins. **Mcf**: proteins can make caterpillars floppy. **Mpf**: membrane attack complex/ perforin proteins. **Mpp**: Etx/Mtx2-related pesticidal proteins. **Mtx**: the mosquitocidal Mtx1 protein. **Pra**: PirA-related proteins. **Prb**: PirB-related proteins. **Spp**: Sphaericolysin-related pesticidal proteins. **Tpp**: Toxin-10 pesticidal proteins. **Vpa**: the active component of the Vpa/Vpb binary toxin. **Vpb**: the binding component of the Vpa/Vpb binary toxin. **Xpp**: pesticidal proteins without structure determined. **Sip**: secreted protein with activity against coleopteran larvae. **Chitinase**: activity against *Caenorhabditis elegans*. **InhA**: a protease degrades attacins and cecropins, two classes of antibacterial proteins in insects. **Bmp1**: nematicidal virulence factor. **Enhancin**: enhances the toxicity of Cry1Ac protein to *Helicoverpa armigera* larvae. **ZwA**: an antifungal and plant protection agent.

Supplementary Table 2. The performance of the local version of BtToxin\_Digger and CryProcessor

|  |  |  |
| --- | --- | --- |
| **Timing and toxins** | **BtToxin\_Digger** | **CryProcessor**  **(“find domains” mode)** |
| Timing (min/genome) | 0.55 | 0.28 |
| Cry | 1556 (1245 with at least one domian) | 874 |
| App | 80 | 0 |
| Cyt | 96 | 0 |
| Mpp | 487 | 0 |
| Mtx | 79 | 0 |
| Spp | 632 | 0 |
| Tpp | 110 | 0 |
| Vip | 126 | 0 |
| Vpa | 277 | 0 |
| Vpb | 518 | 0 |
| Xpp | 22 | 0 |
| Bmp1 | 1840 | 0 |
| Chitinase | 600 | 0 |
| Enhancin | 361 | 0 |
| InhA | 1906 | 0 |
| Sip | 29 | 0 |
| ZwA | 3556 | 0 |
| Unnamed toxins only found by SVM | 39 | 0 |

The numbers in the two columns after the toxin represent the number of corresponding toxins found by BtToxin\_Digger and CryProcessor in 600 *Bacillus thuringiensis* genomes.