

REGULAR EXPRESSION MIMP SEARCH

Mimps were identified by searching for the occurrence of the TIRs CAGTGGG..GCAA[TA]AA and TT[TA]TTGC..CCCACTG within 400 nucleotides.

EXPAND HITS

Sequences 2.5kb upstream and downstream of the *mimps* hits were extracted.

GENE PREDICTION

The extracted sequences were submitted to AUGUSTUS (3.3.3) for gene prediction with the “fusarium” species parameter selected.

ORF FINDING

The getorf tool from EMBOSS (6.6.0.0) was used to find open reading frames (ORFs) within the extracted sequences.

SIGNALP 4.1

SignalP was run on the genes predicted by Augustus to identify predicted genes with a signal peptide (default parameters).

SIGNALP 4.1

SignalP was run on the ORFs identified by getorf (default parameters).

NON-REDUNDANT PROEIN SETS GENERATED

Predictions and ORFs with a signal peptide were then clustered using CD-HIT (4.8.1) to generate a non-redundant protein set for each genome.

PUTATIVE EFFECTORS COMBINED

The non-redundant protein sets were combined to generate a FASTA containing all putative effectors.

CLUSTERING

The FASTA containing the putative effectors from all genomes was then clustered using CD-HIT (4.8.1) (90% sequence identity).

EFFECTORP

The clustered effector FASTA was then submitted to EffectorP (2.0.1) for prediction and a FASTA file containing sequences predicted to be effectors was produced .