## Fusarium Assemblies

Fusarium species assemblies in FASTA format are, as well as a list of assemblies and a *mimp* profile HMM are prepared as input.

## Mimp identification

Each Fusarium assembly is searched for mimps using a custom python script (using mimp TIRs) and NHMMER (3.3.1) (using a mimp profile-HMM) to identify mimps.

## **Sequence Expansion**

Sequences 2.5kb upstream and downstream of mimps identified using both approaches are extracted.

#### **Gene Prediction**

The extracted sequences are submitted to Augustus (3.3.3) for gene prediction with the "fusarium" species parameter selected.

## **Signal Peptide Filtering**

SignalP (4.1) is used to find predicted genes with a signal peptide.

#### **ORF** identification

Getorf from Emboss (6.6.0.0) is used to find and extract open reading frames (ORFs).

## **ORF Parsing**

Custom script is used to extract smaller ORFs from within the getorf output.

#### **Signal Peptide Filtering**

SignalP (4.1) is used to find ORFs with a signal peptide.

## **Non-redundant Protein Sets Generated**

Predicted genes and ORFs with a signal peptide are clustered using CD-HIT (4.8.1) generating a non-redundant protein set for each assembly.

#### **Putative Effectors Combined and Clustered**

The non-redundant protein sets across all assemblies are combined into one FASTA and clustered using CD-HIT (4.8.1) (90% identity).

#### **EffectorP Scan for Likely Effectors**

The longest sequence from each cluster is then submitted to EffectorP (2.0.1) for fungal effector prediction.

## **Final Candidate Effectors Set Generated**

A FASTA file containing the candidate effectors predicted across the Fusarium assemblies is produced.

## Fusarium pan-effectorome

ling

Effector

#### **Effector Search**

The candidate effector FASTA is queried against the Fusarium assemblies using TBLASTN, with a cut-off 1e-6 and a percentage identity and coverage threshold of 70% and 90%, respectively.

#### **Binary Data Matrix**

Using the TBLASTN hist from within the threshold, a binary data matrix indicating presence ("1") or absence ("0") is generated for the candidate effectors across the FOSC assemblies.

## **Effector Profile Heatmap**

An effector profile heatmap generated in R Studio (version 3.6.3), using the package Pheatmap (version 1.0.12).

# Fusarium Assembly **Effector Profiles**

