

Global mining on filamentous elements in insect genomes

1 Input

Diptera + Hymenoptera + Lepidoptera

358 sp

326 sp

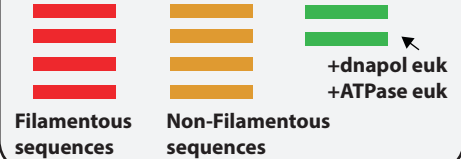
892 sp

NCBI assemblies
(n=2815 assemblies)
DATABASE



**ALL filamentous core genes
from 25 Naldaviricetes species**

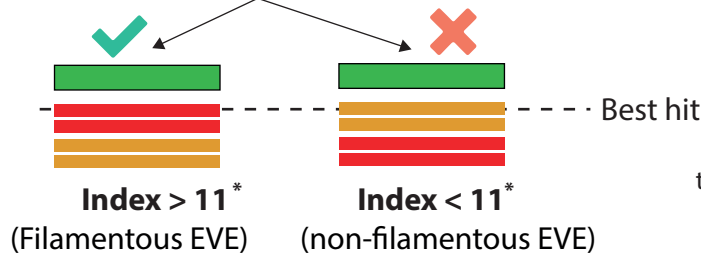
QUERY



2 tBLASTn



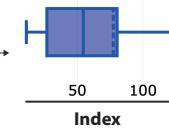
3 Filtering



$$\text{Index} = \log_{10}(\text{Best non-Filamentous E-value}) - \log_{10}(\text{Best Filamentous E-value})$$

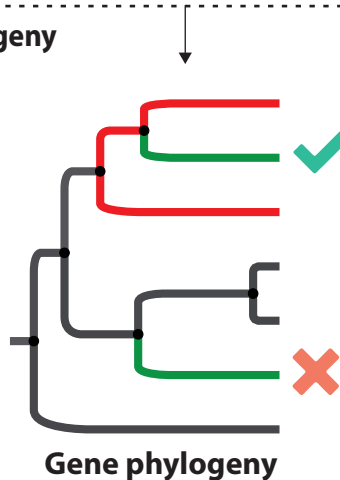
* Index distribution on EVEs controls [1]

threshold = 11.36



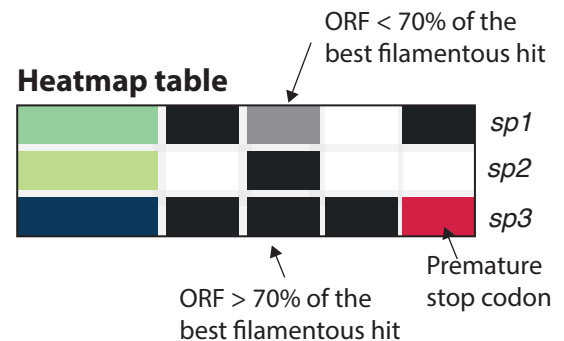
[1] Di Giovanni et al. 2020, MBE

4 Phylogeny



5 Output

Very likely endogenized
Likely endogenized
Likely free-living virus



- Nb hit < 13 & cumulative scaffold length > 300,000pb
- Nb hit > 13 & Cumulative scaffold length > 300,000pb
- Nb hit > 13 & cumulative scaffold length < 300,000pb

Filamentous clade
non-Filamentous clade
Candidate loci