Global mining on filamentous elements in insect genomes 1 Input **ALL filamentous core genes** Diptera + Hymenoptera + Lepidoptera from 25 Naldaviricetes species **OUERY NCBI** assemblies (n=2815 assemblies) +dnapol euk **DATABASE** +ATPase euk **Filamentous** Non-Filamentous seauences sequences tBLASTn bits > 50 (viral origin) No hit **False** positive qcov > 25% (euk origin) (nothing found) (3) Filtring Index= log10(Best non-Filamentous E-value) - log10(Best Filamentous E-value) * Index distribution on EVEs controls [1] Best hit threshold = 11.36Index > 11 Index < 11 (Filamentous EVE) (non-filamentous EVE) Index [1] Di Giovanni et al. 2020. MBE **5**) Output $(\mathbf{4})$ Phylogeny ORF < 70% of the best filamentous hit **Heatmap table** Very likely endogenized sp1 Kept Likely endogenized sp2 Likely free-living virus sp3 Prèmature ORF > 70% of the stop codon best filamentous hit **Gene phylogeny** Nb hit < 13 & cumulative scaffold length > 300,000pb Nb hit > 13 & Cumulative scaffold length > 300,000pb Filamentous clade non-Filamentous clade Nb hit > 13 & cumulative scaffold length < 300,000pb Candidate loci