Integrating vMAGs and Amplicon Sequencing: Bridging Gaps in Illumina-Based Assembly of Novel Large Viruses from Honey Bee Viromes

Dominika Kadlečková[1], Martina Saláková[1], Tomáš Erban[2], and Ruth Tachezy[1]
[1]Department of Genetics and Microbiology, Faculty of Science BIOCEV, Charles University, Vestec, Czechia
[2] Crop Research Institute, Prague, Czechia

Introduction

The study of viromes has been revolutionized by advanced bioinformatics methodologies that enable the precise identification of novel viruses. However, assembling complete genomes of large viruses is still challenging. To address these issues, we employed a viral binning approach that integrates fragmented sequences to reconstruct viral genomes. Our methodology focused on identifying and elongating contigs and reducing their number where feasible. Additionally, we complemented our efforts with wet laboratory techniques, specifically generating and sequencing PCR products, to ensure genome completeness. Our methodological approach revealed complete genomes of previously unidentified large DNA viruses in the honey bee population.

Methods

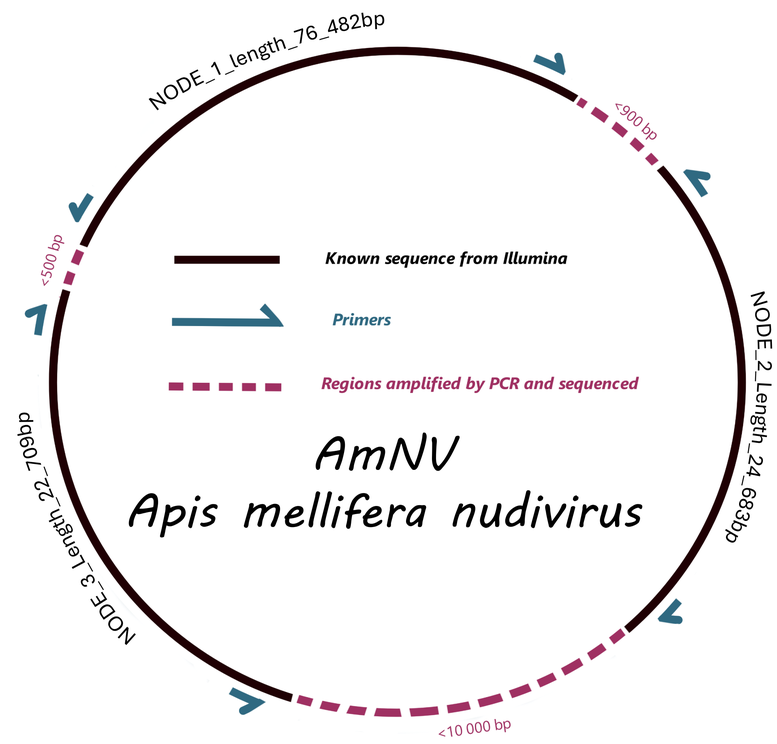
1) Viral binning to gain the majority of the genome of a large virus
2) Assembly of genome(s)
   - Co-assembly
   - Read filtering (mapped on target contigs and not on scaffold)
   - Different settings (--meta/--metaviral/--isolate)

   ⇒ reduction of the number of contigs and gaining longer contigs
   (Virify; geNomad; vRhyme; BWA-MEM2; SPAdes)

3) Design PCR primers aiming outwards from the ends
   - Similar Tm and specificity for all combinations
   - Overlaps (parts of already known sequence)
4) Perform PCR with all combinations (we initially used 5 min elongation)
5) Perform additional PCRs optimized for the selected combination of primers (longer/shorter elongation time, Tm, polymerase, and etc.)
6) Clean-up of amplified products and their sequencing (Illumina/Sanger)

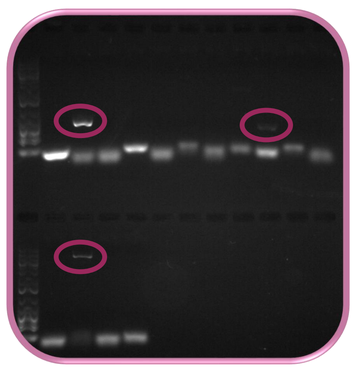
7) Complete the genome with the help of the sequenced amplicons
8) Polish the genome to correct the remaining small errors

=> COMPLETE vMAG



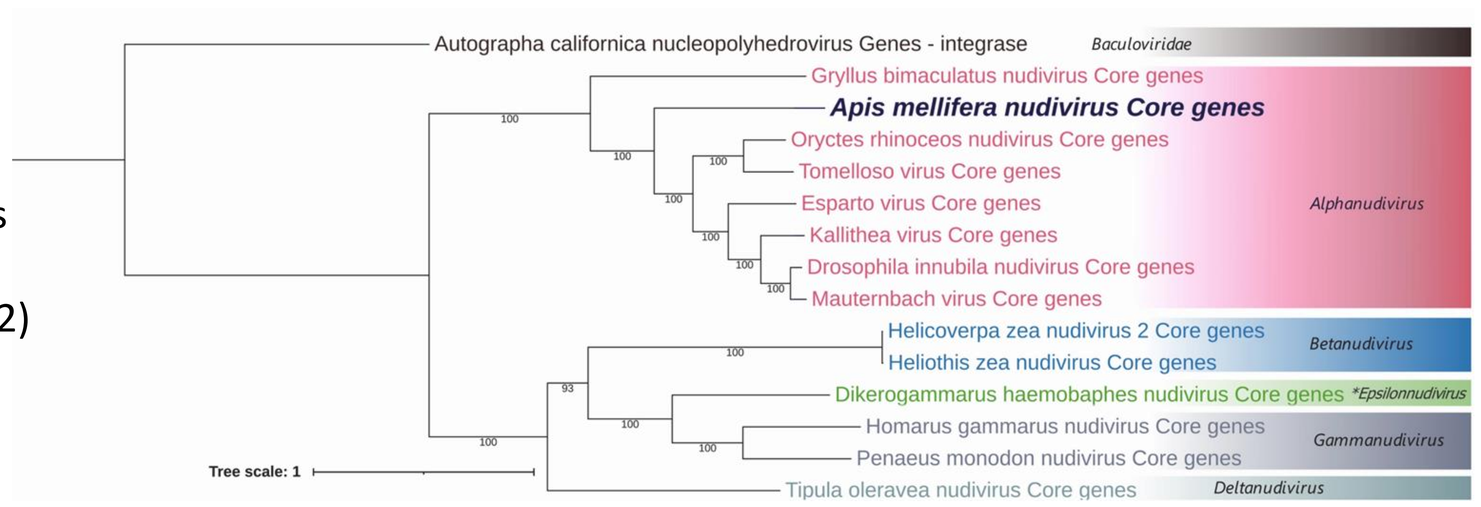
Scheme of additional post-sequencing method to finish vMAG

The initial PCR with different combinations of primers. Sequencing of the three PCR products allowed for the completion of the genome.

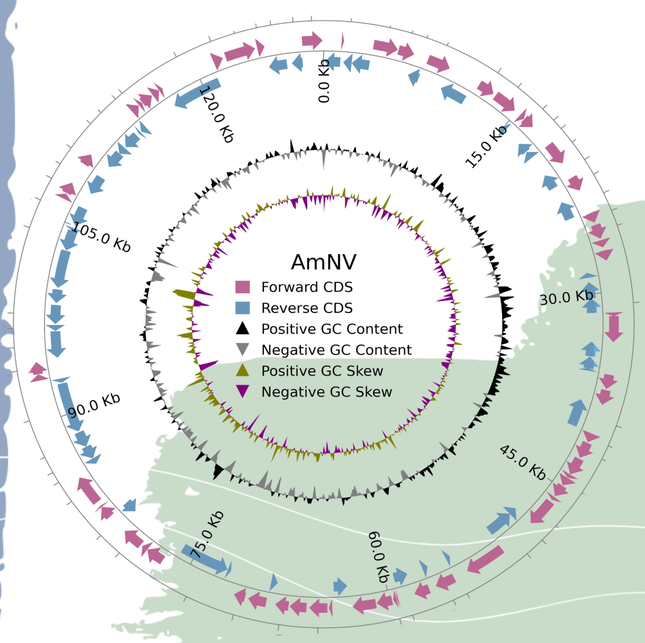


Results

- 12 => 3 contigs: 76,482, 24,683 and 22,709 bp
  = ~96% of the whole genome of 129,467bp
- Alphanudivirus; 106 proteins; 75% has alignments
- 25/29 *Nudiviridae* core genes (for example Integrase/Recombinase; pif 1,2,3; DdDp-B; Helicase2)



Phylogenetic analysis of the concatenated core genes (Baculoviridae used as outgroup and lack integrase)
*Epsilonnudivirus = suggested new group



Scheme of the complete circular genome of AmNV with ORF

Conclusions

Our study underscores the synergistic advantages of integrating computational techniques with laboratory methodologies for the thorough characterization of large viruses. By combining viral binning with PCR, we achieved complete genomes. This combined approach not only confirms the accuracy of assembly and binning but also ensures the completeness of viral genomes.
Furthermore, our results expanded the repertoire of known viruses within the honey bee population, highlighting the importance of ongoing surveillance efforts underlining further research focused on the role of these viruses in bees. This comprehensive understanding of viromes is crucial for ecological conservation and public health management.

Contact me:
dom.Kadleckova@gmail.com
kadlecdo@natur.cuni.cz
+420 607 123 905

Funded by the European Union
NextGenerationEU
CZECH RECOVERY PLAN
MINISTRY OF EDUCATION, YOUTH AND SPORTS
MINISTRY OF AGRICULTURE OF THE CZECH REPUBLIC
DURABLE Research Network against Epidemics