



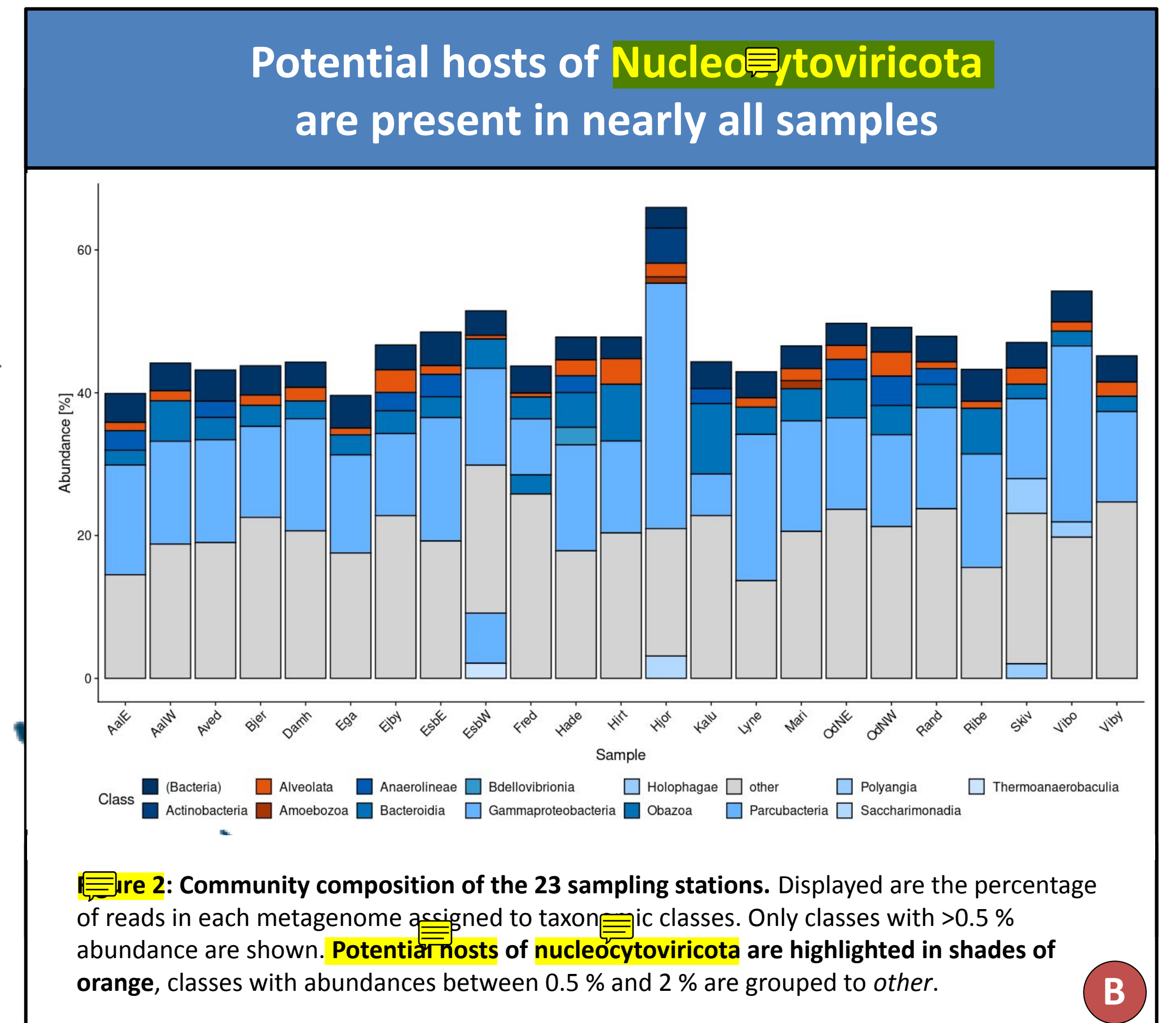
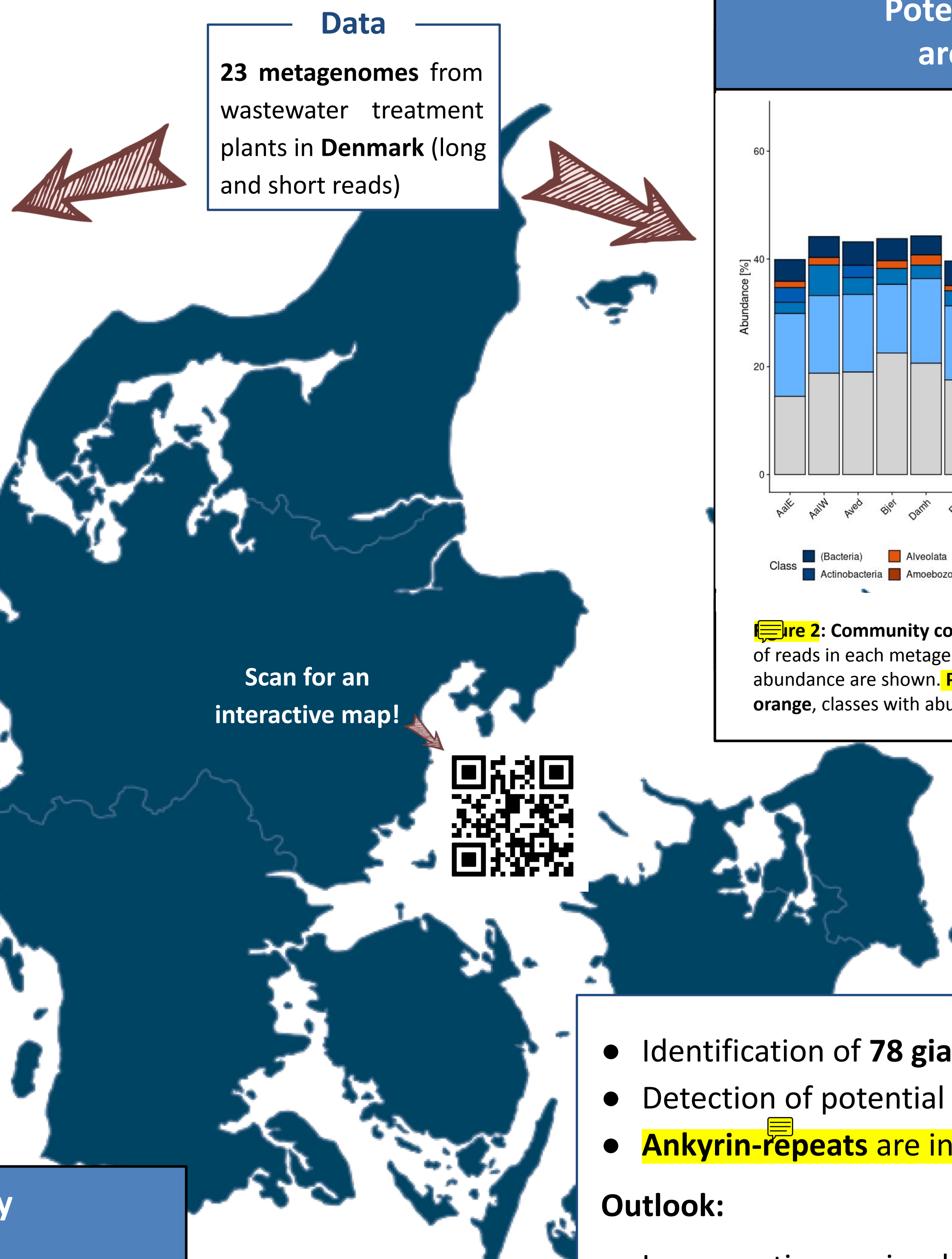
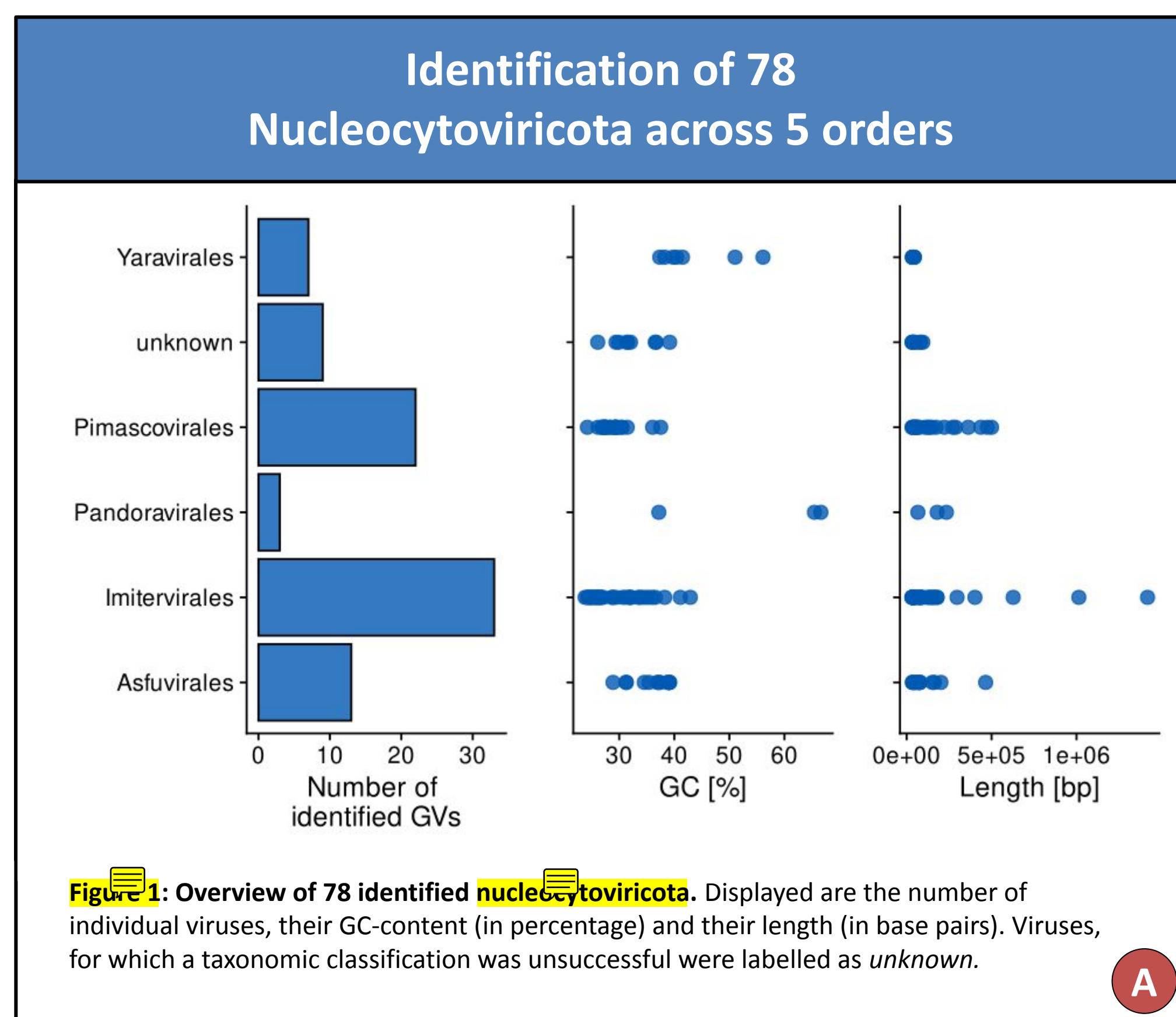
# Distribution of Giant Viruses and Mobile Genetic Elements in Selected Habitats

Dominik Lücking<sup>1</sup>, Alejandro Manzano-Marín<sup>1</sup>, Anouk Willemsen<sup>1</sup>

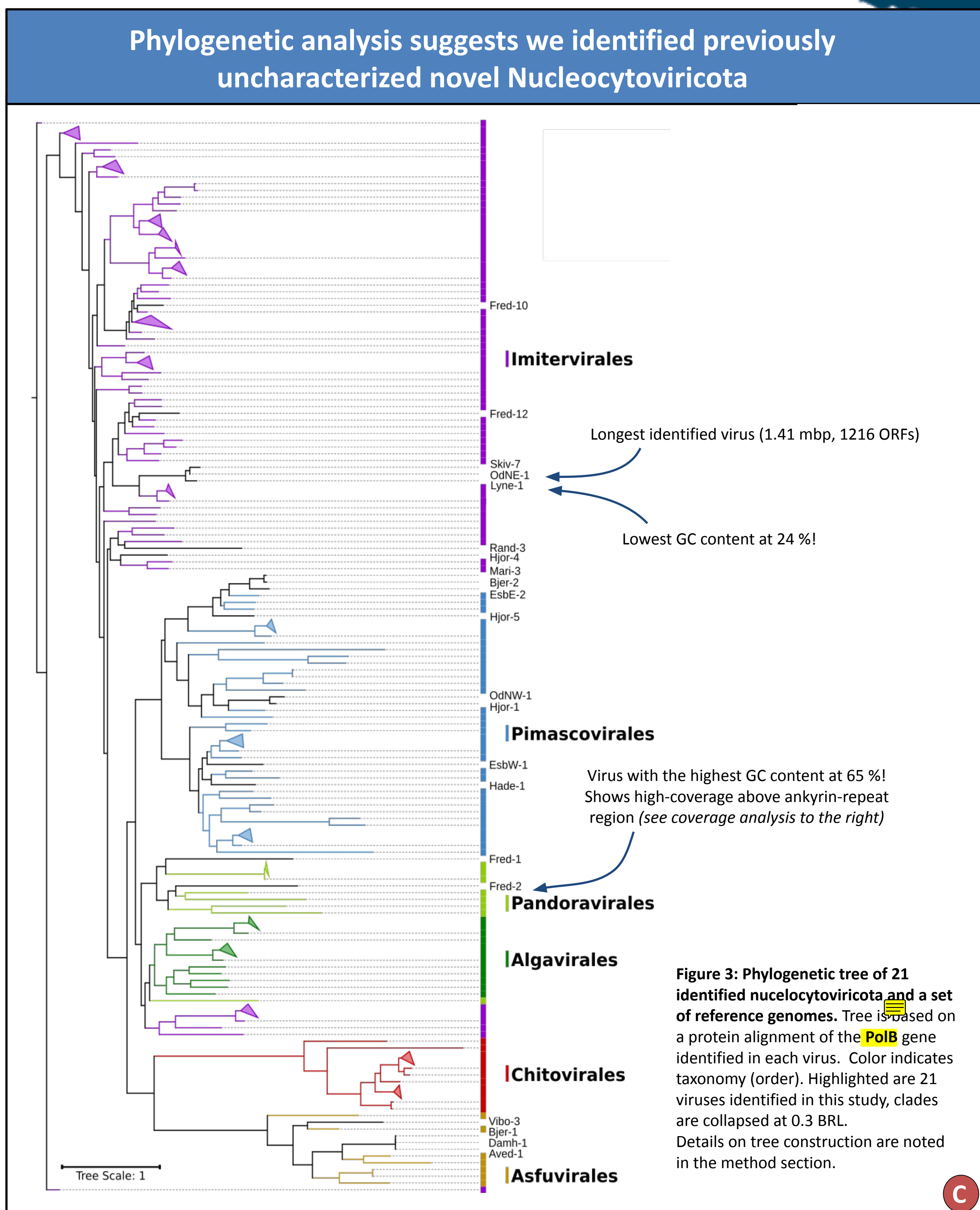
<sup>1</sup> Centre for Microbiology and Environmental Systems Science, University of Vienna, 1030 Vienna, Austria

## ABSTRACT

Giant viruses (GVs) play crucial ecological roles in oceans, freshwater and soil, yet their temporal and spatial distributions, as well as their interactions within complex systems, remain largely unexplored. In this study, we investigate the use of mobile genetic elements (MGEs) identified on GV genomes to trace potential interaction partners and unravel the evolutionary history of these viruses. Combined with viral, bacterial, and host diversity, we aim to provide a comprehensive understanding of the ecological functions of GV in both natural and biotechnological ecosystems. Samples from Danish wastewater treatment plants were sequenced and scrutinized for the presence of potential GV and their hosts. Within the genomes of the identified novel GV, we detect Ankyrin repeat-containing genes, inserted in the replication machinery of the host, suggesting a complex interaction network of viruses and hosts. Future work aims to unravel this interaction network, by integrating available time series data and tracing mobile genetic elements.



Are the identified giant viruses novel?



## CONCLUSION AND OUTLOOK

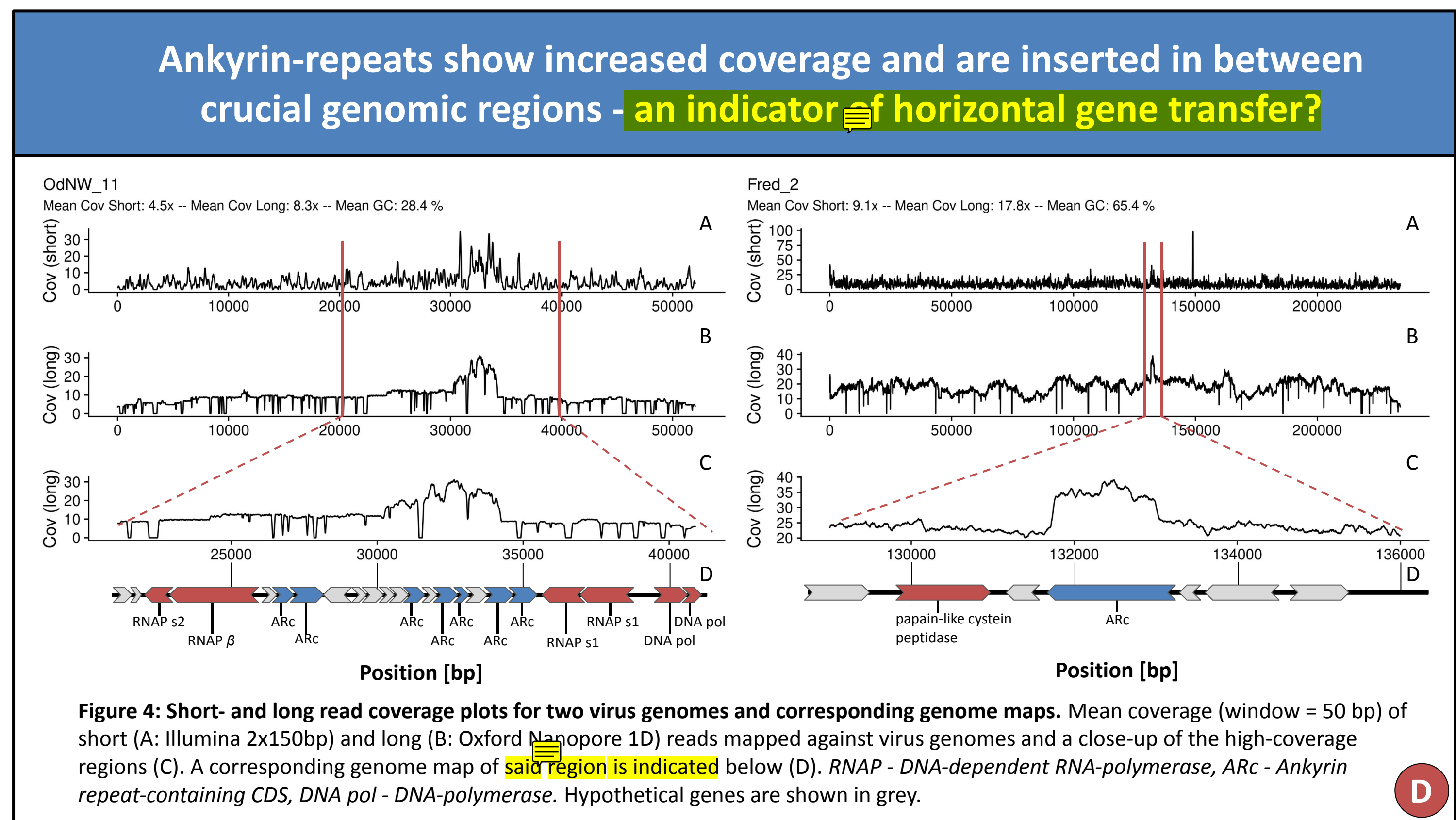
- Identification of **78 giant viruses** across 4 orders
- Detection of potential giant virus hosts
- **Ankyrin-repeats** are inserted in virus replication machinery

### Outlook:

- Leverage time series data in order to establish virus-host interactions
- Connect the presence of **mobile genetic elements** to virus-host pairs in order to entangle their complex ecological and evolutionary interaction network

Are there traces of mobile genetic elements?

Can mobile genetic elements help us entangle evolutionary history?



## Methods

- Identification and analysis of potential nucleocytoviricota.** Assembled contigs were screened with *ViralRecall*, *kraken*, *ncldv\_markersearch*, and *diamond blastp*. rRNA genes were predicted using *rnammer*-1.2. Promising contigs were manually searched against *nr* using *blastn* and *blastp* and a final score was given to each contig ([https://github.com/dluecking/wwwtp\\_denmark](https://github.com/dluecking/wwwtp_denmark)). This yielded 78 potential nucleocytoviricota which were subsequently analysed. GC content was calculated using *seqkit*.
- Assessing the community composition of each sampling site.** For each sampling site, the *illumina* reads were analysed using *phyloFlash*.
- Phylogenetic tree construction.** First, *polB* proteins were identified in each contig using *ncldv\_markersearch*. This yielded strong signals in 21 out of 78 virus contigs, which were subsequently aligned using *muscle-align* with default parameters. The alignment was cleaned using *trimal* with parameter setting 'automated1' and finally, a maximum-likelihood (ML) tree was calculated using *iqtree*, with default bootstrap values and the LG+R7 model. Clades are collapsed by 0.3 BRL.
- Coverage maps.** Mapping was done for all contigs (viral and non-viral) of one sample in parallel in order to avoid cross-contamination. For illumina short reads, this was done using *bowtie2* with the following parameters: --very-sensitive-local --no-discordant -1 150 -X 700. Oxford nanopore reads were mapped using *minimap2* with the --ax map-on flag. Coverage per contig was calculated with *samtools depth* and visualized using R. **Genome maps.** Potential viral contigs were annotated using *prokka* with the following parameters: --evaluate '1e-5' --coverage '60' while using a manually curated set of giant virus proteins in order to improve annotation.