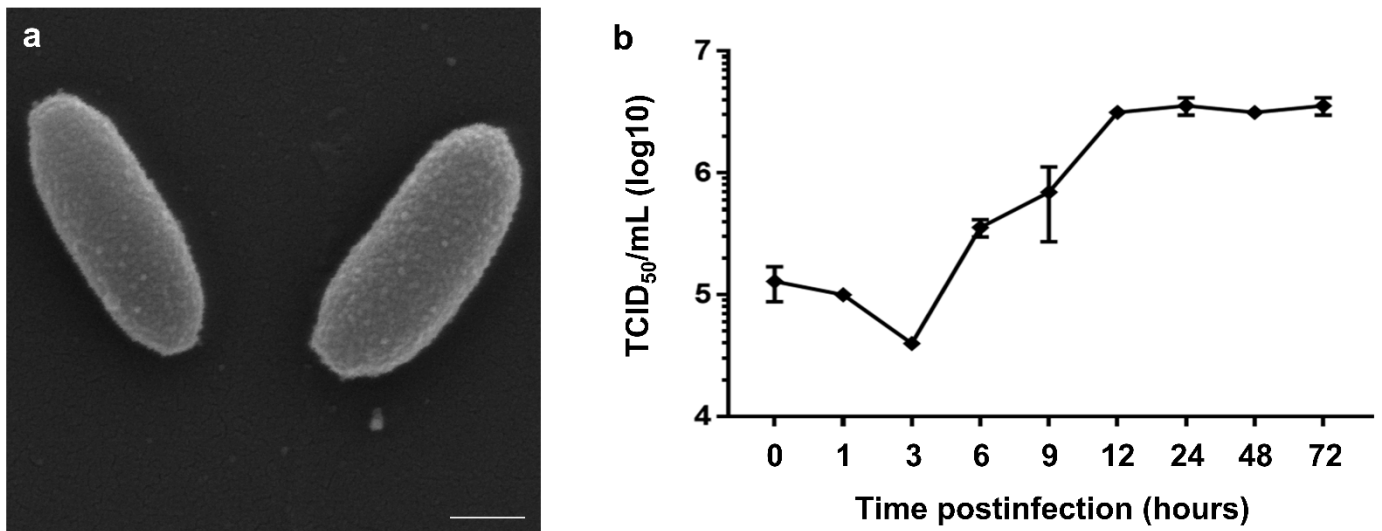
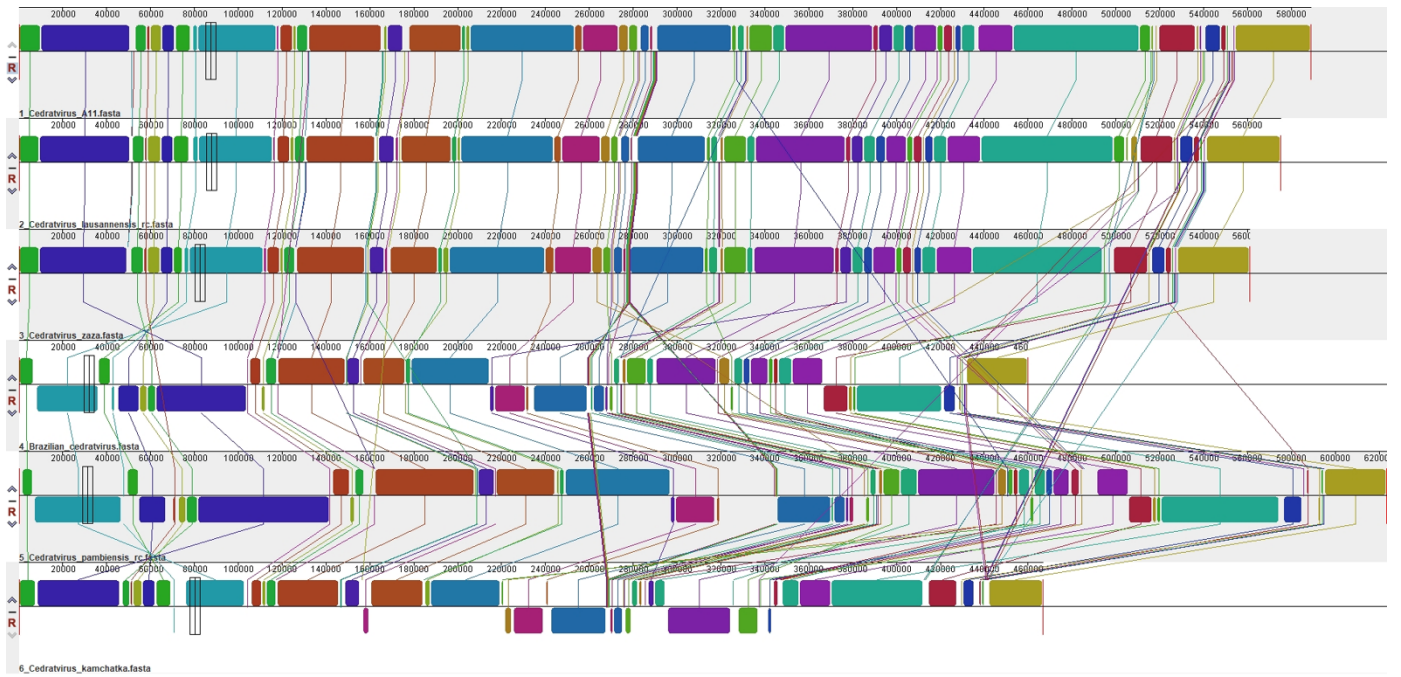


## Supplementary figures - Gene Duplication as a Major Force Driving the Genome Expansion in Some Giant Viruses

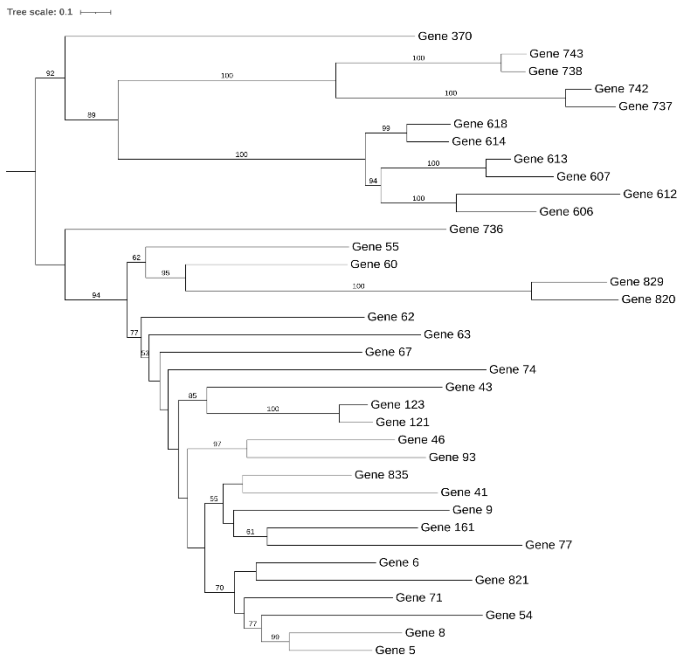


**Supplementary Fig.1: Particle and one-step growth curve graph of cedratvirus pambiensis.** **a** Viral particles visualized by scanning electron microscopy. Scale bar: 300 nm. **b** The replication cycle begins with the phagocytosis of the particles by the host cell. At 3hpi appears to be the viral eclipse phase. From 3hpi to 12hpi new particles are formed and assembled. At 12hpi the new viral progeny is already fully formed, as the viral titer maintains a plateau up to 72 hpi.

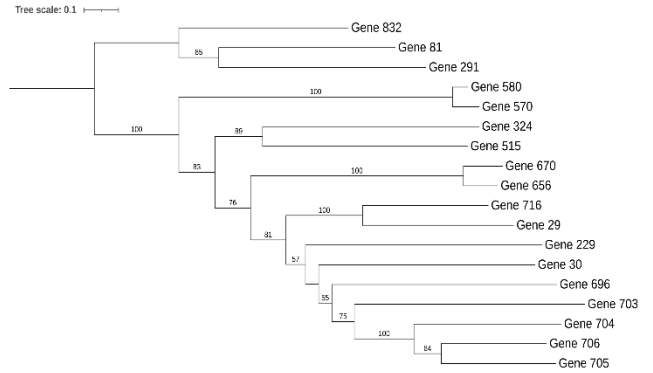


**Supplementary Fig.2: Synteny analysis of cedratviruses.** The analysis was performed using the genome of cedratvirus A11 (NC\_032108.1), cedratvirus lausannensis (LT907979.1), cedratvirus zaza (LT994652.1), brazilian cedratvirus (LT994651.1), cedratvirus kamchatka (MN873693.1) and the genome sequence of the new isolate, cedratvirus pambiensis.

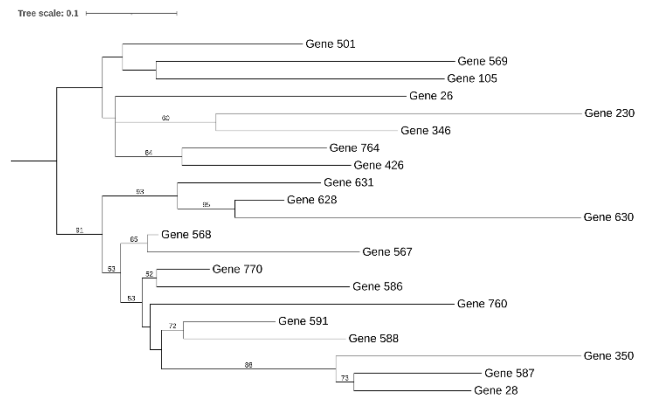
### Cluster 1



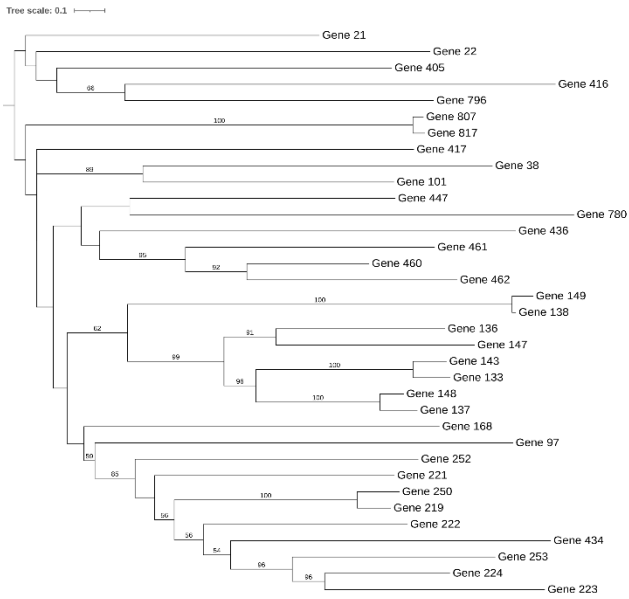
### Cluster 2



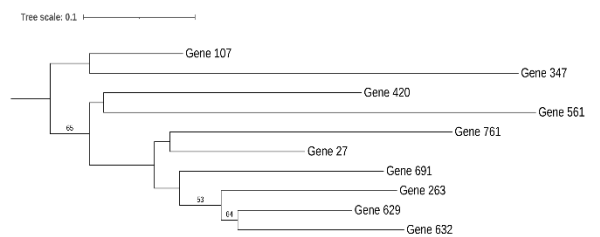
### Cluster 5



### Cluster 4

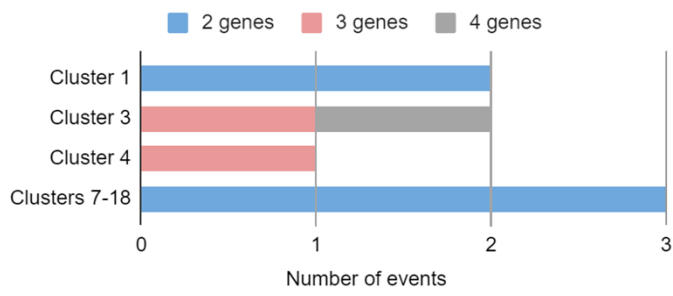


### Cluster 6

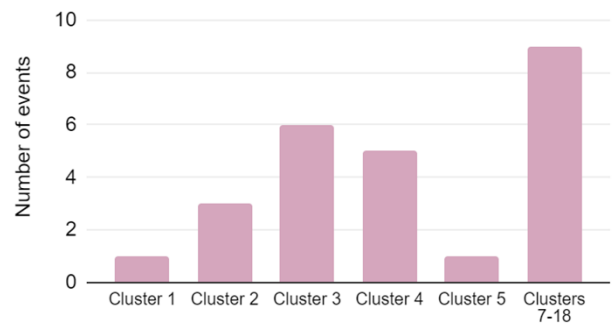


**Supplementary Fig.3: Maximum likelihood phylogenetic tree constructed with sequences of the six largest families of paralogs belonging to *c. pambiensis*. Cluster 3 has already been represented in figure 4.**

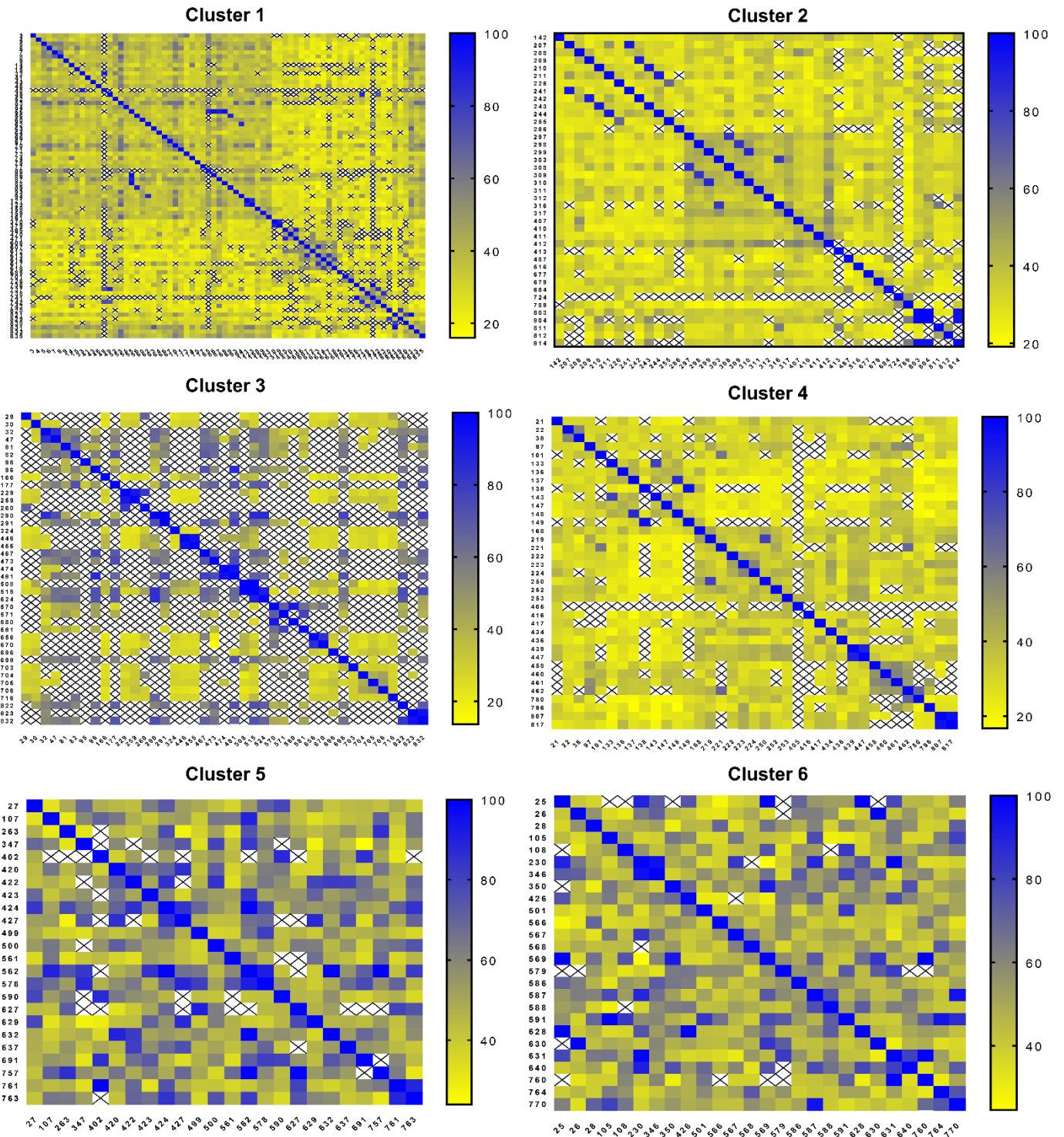
**a** Chromosomal segment duplications



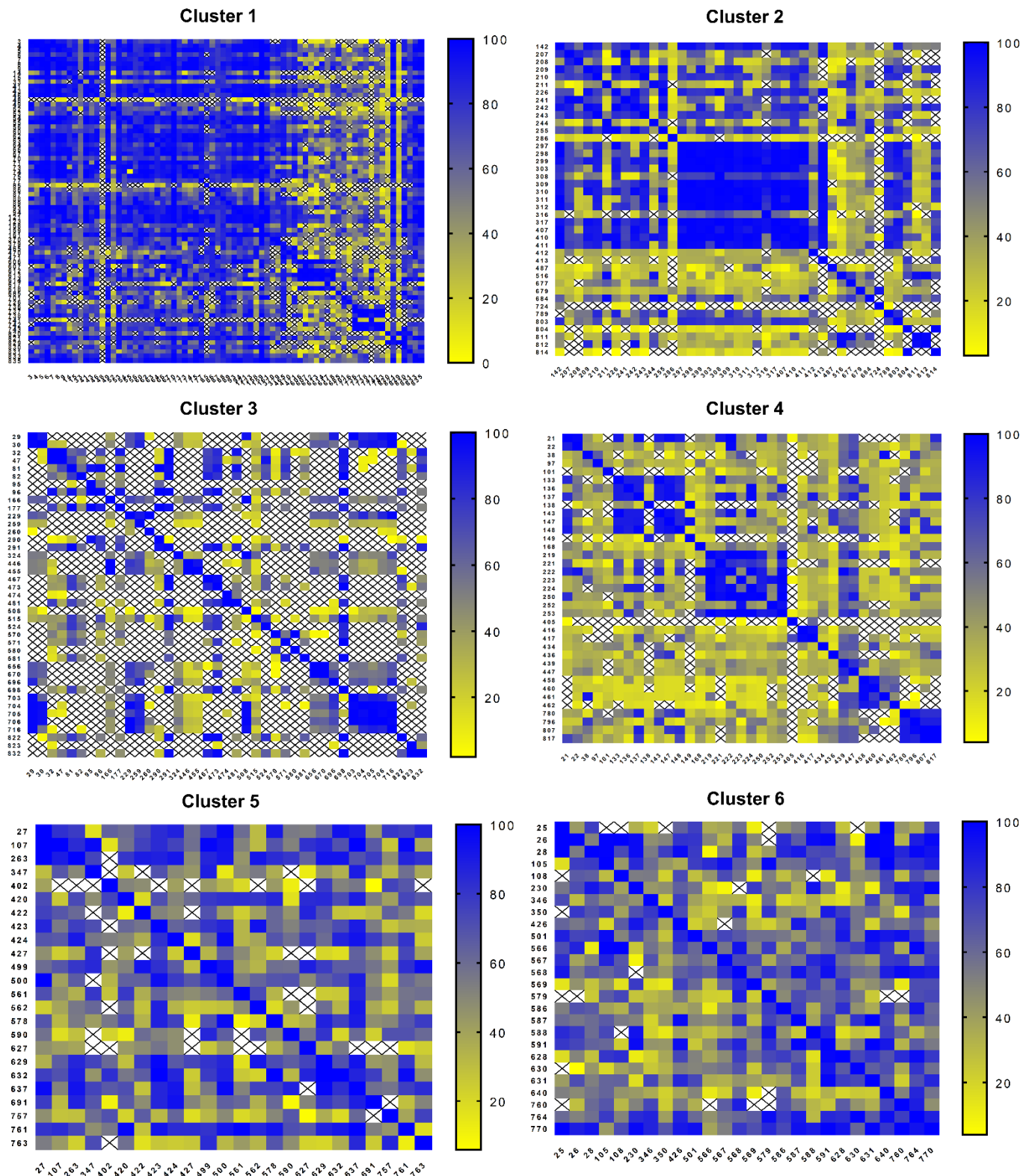
**b** Proximal tandem duplication events



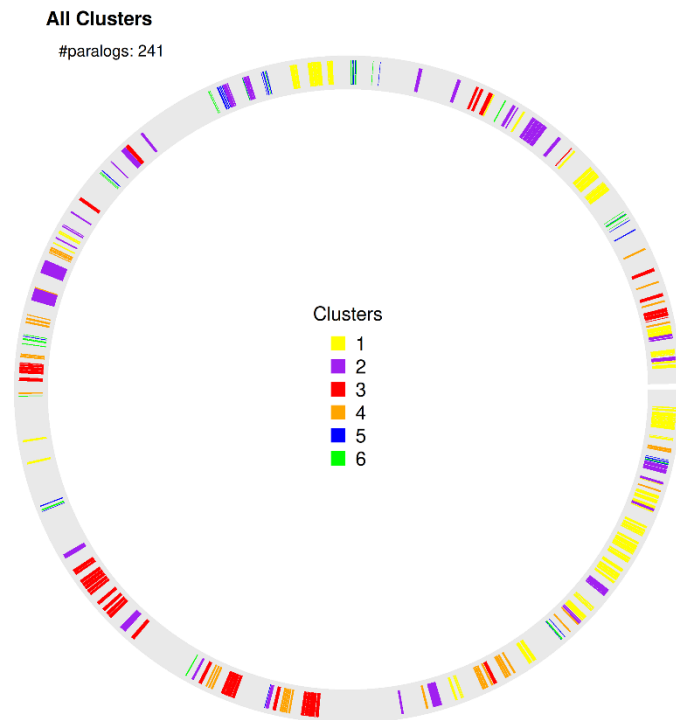
**Supplementary Fig.4: Complementary analyses of paralogy related to *c. pambiensis* genome - 18 largest gene clusters** **a** Number of chromosome segment duplication events. **b** Number of proximal tandem duplication events in the clusters.



**Supplementary Fig.5: Coverage of the genes that form each cluster.** Shown is a coverage heat map obtained from all-against-all proteins comparisons. A BLASTp was performed with an e-value of  $1e-5$  as the threshold. For each comparison the best hit was used to compose the matrix.



**Supplementary Fig.6: Identity of the genes that form each cluster.** Shown is an Identity heat map obtained from all-against-all proteins comparisons. A BLASTp was performed with an e-value of  $1e-5$  as the threshold. For each comparison the best hit was used to compose the matrix.



**Supplementary Fig.7: Location of the six largest clusters in the cedratvirus pambiensis genome.** Paralogous genes are scattered around the genome. This figure is a compilation of all graphs presented in Fig.6.

**Supplementary Table 1 (XLS file): Information about each paralog genes of cedratvirus pambiensis.** Cluster genes were arranged in ascending order. Paralogous pairs were arranged in ascending order for "gene 1" of each pair.