

To ensure that the identified contigs represent genuine viral sequences, a functional annotation of the viral proteins for the ten longest contigs originating from different RNA virus phyla was performed. The lengths of the top ten longest contigs varied from 3,569 to 26,479 nucleotides. Remarkably, all contigs aligned to different virus families, and with the exception of one, each exhibited a protein sequence identity below 90%.

Bild.

The longest contig corresponds to a novel alphacoronavirus identified in a sample from Harrison's large-eared giant mastiff bat (*Otomops harrisoni*). This sample was obtained during a metatranscriptomic analysis of oral and rectal swabs collected from bats inhabiting caves in Kenya (Quelle zu SRA project Seite). The second largest contig, measuring 18,735 nucleotides, was one of the few identified novel sequences of Porcine reproductive and respiratory syndrome virus 2 found in wild boars (*Sus scrofa*), rather than in green monkeys (*Chlorocebus sabaeus*). Typical proteins of Coronaviridae and Arteriviridae were identified in both contigs, aligning with the correct open reading frames (ORF) and arranged in the expected order, with only previously known proteins found. The longest contig identified ORF1a along with its known non-structural proteins (NSP), while ORF1b included the RNA-dependent RNA polymerase (RdRp), non-structural protein 14 (NSP14), and helicase; the subsequent open reading frames (ORFs) also contained the structural proteins. Similarly, the contig aligning to the family Arteriviridae identified ORF1a and ORF1b with their associated proteins, as well as envelope glycoproteins 2-5 arranged (EnvGP2a,3 and 5) in the correct order.