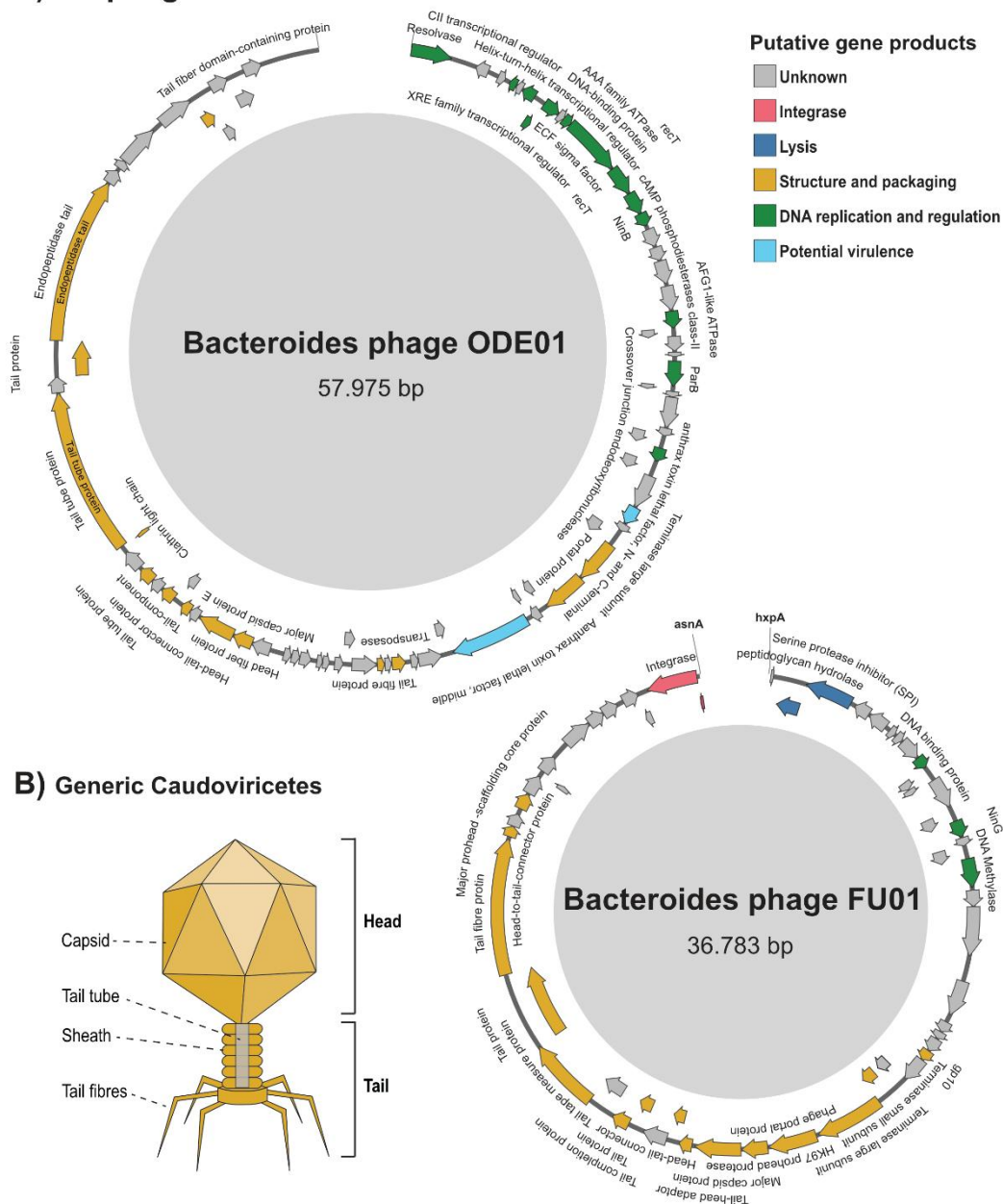
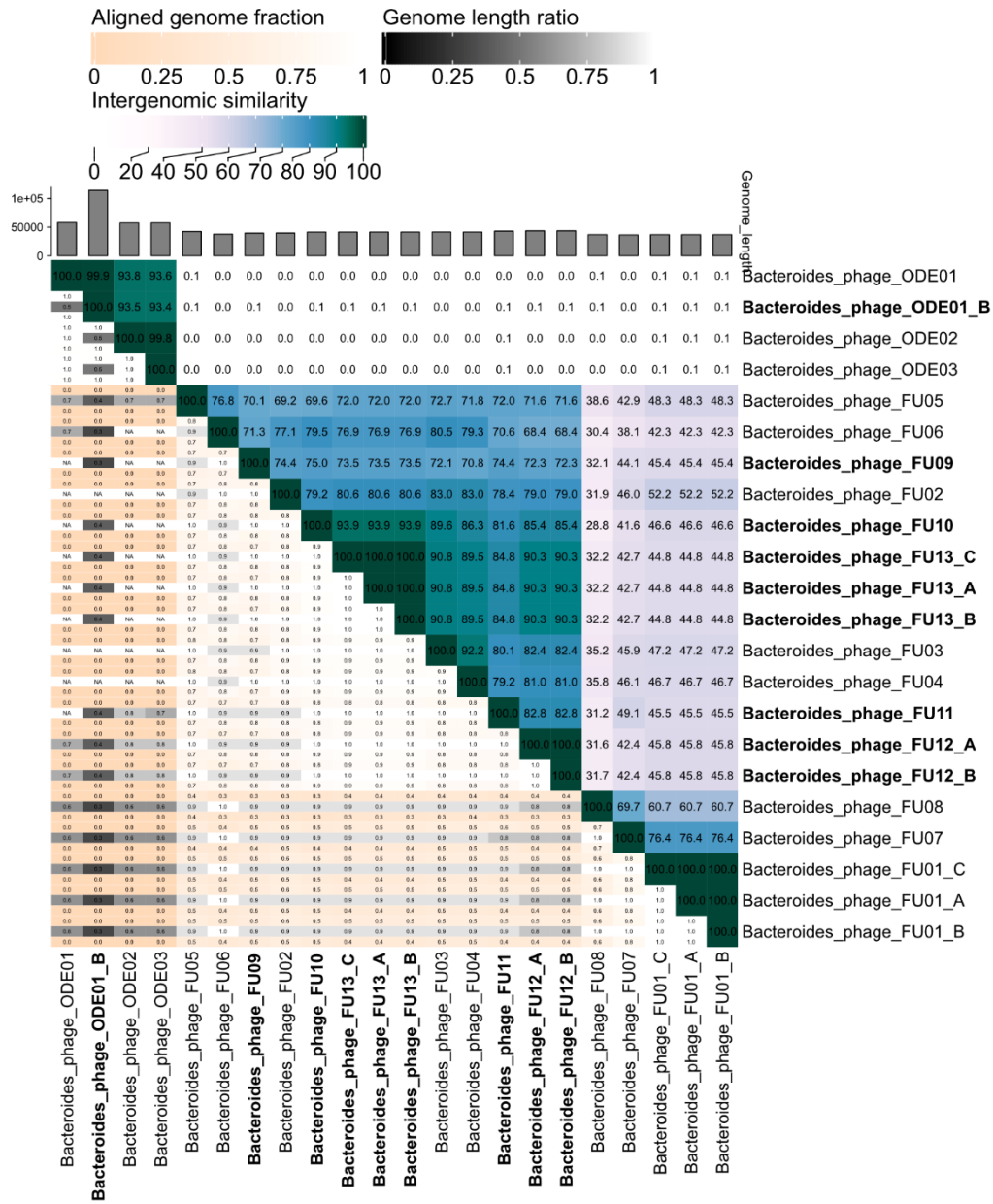


A) Prophage Genomes





7

8 Figure S2. Intergenomic sequence identity matrix for all prophages identified in this study,
 9 both from isolate and metagenomic sequencing data. The sequence identity is stated
 10 specifically for each alignment in the matrix and shaded in blue according to the degree of
 11 similarity. Phages identified in the metagenomic samples are marked in bold text.

12

13



14

15 *Figure S3. Coverage of the Bacteroides phage genomes from the metagenomic cohort*
 16 *screening on patients with colorectal cancer (CRC, red, N = 434) and controls (CTR, blue,*
 17 *N = 443). Assembled metagenomic contigs were screened with MASH and BLASTN to*
 18 *identify matching sequence fragments as described in methods. Coverage is only counted*
 19 *once for each matching position in every patient.*



20

21 *Figure S4. Coverage of the Bacteroides phage genomes from the metagenomic cohort*
 22 *screening on patients with colorectal cancer (CRC, red, N = 434) and controls (CTR, blue,*
 23 *N = 443). Assembled metagenomic contigs were screened with MASH and BLASTN to*
 24 *identify matching sequence fragments as described in methods. Coverage is only counted*
 25 *once for each matching position in every patient and hits were filtered to only include high*
 26 *quality BLASTN matches (bitscore >50).*

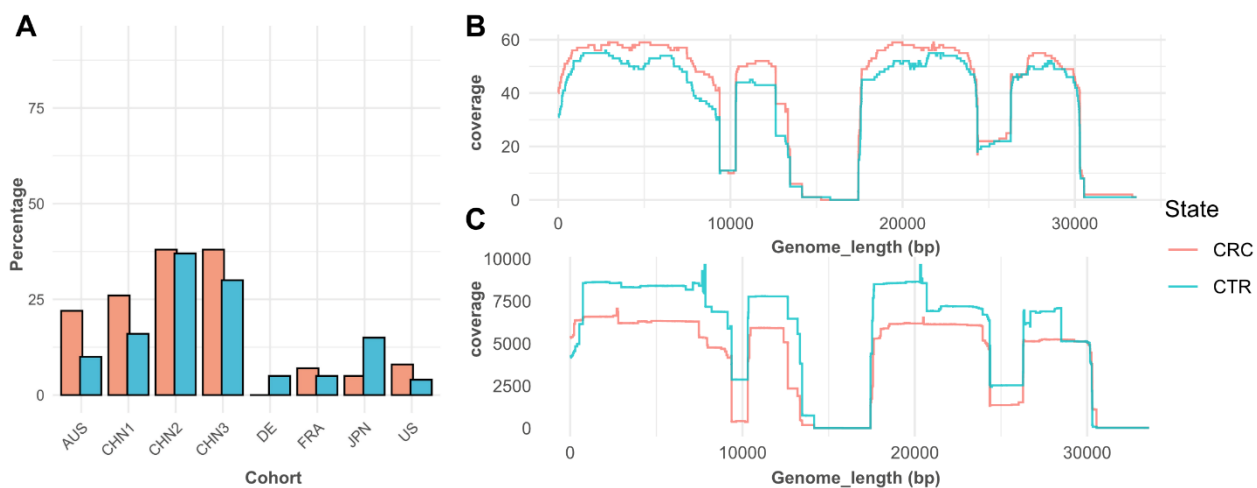
27

28



29

30 *Figure S5. Weighed coverage of the Bacteroides phage genomes from the metagenomic*
 31 *cohort screening on patients with colorectal cancer (CRC, red, N = 434) and controls*
 32 *(CTR, blue, N = 443). Assembled metagenomic contigs were screened with MASH and*
 33 *BLASTN to identify matching sequence fragments as described in methods. Weighed*
 34 *coverage is given by the accumulated coverage of all high quality BLASTN matches*
 35 *(bitscore >50).*



36

37 *Figure S6. Metagenomic screening results from the Escherichia phage P2 (NC_001895.1),*
 38 *according to patient colorectal cancer (CRC) status with CRC positives in red and controls*
 39 *(CTR) in blue (n = 877). Assembled metagenomic contigs were screened with MASH and*
 40 *BLASTN to identify matching sequence fragments as described in methods. A) Percentage*
 41 *of patients with two or more matching sequence fragments of >2000bp in length. B)*
 42 *Coverage of the P2 phage genome, counted once for each matching position in every*
 43 *patient. C) Weighed coverage of the P2 phage genome.*