Development ofONT-cappable-seq to unravel the transcriptional landscape of Pseudomonas phages

Leena Putzeys a,1, Maarten Boon a,1, Eveline-Marie Lammens a, Konstantin Kuznedelov b, Konstantin Severinov b, Rob Lavigne a,b,*

a Department of Biosystems, Laboratory of Gene Technology, KU Leuven, Leuven 3001, Belgium
b Waksman Institute, Rutgers, The State University, Piscataway, NJ 08854, USA

1. Introduction

Over the last decade, high-throughput sequencing approaches have fuelled a new era of molecular research in bacteria and their viral predators, bacteriophages. The surge in microbial transcriptome studies has provided major insights in the fundamentals of transcription and paves the way for standardized, in-depth transcription studies for microbial viruses or bacteria in general.
reveal transcriptional landmarks and operon structures in bacteria [15]. However, as this method uses size selection to enrich for longer transcripts, it is less suited for the exploration of phage transcriptional landscapes, which are often hallmarkmed by high gene densities and short open reading frames. In parallel, Oxford Nanopore’s long-read sequencing technology (ONT) is rapidly improving in quality and is emerging as a cost-effective technique in microbial transcriptomics [12,16–20]. Currently, one of the main drawbacks of native RNA sequencing is the inability to sequence ~12 nucleotides at the 5’ transcript end, resulting in ambiguous TSS mapping [16,21,22]. This technical limitation, which has not been reported for (PCR)-cDNA nanopore sequencing approaches, can be attributed to a loss of control of the RNA translocation speed through the nanopore after premature detachment of the motor protein.

In this study, we present a nanopore-based cDNA Cappable-seq approach (ONT-cappable-seq), which can be especially useful for resolving regulatory features in densely coded viral genomes. The power of full-length primary transcriptome sequencing was just recently demonstrated for the eukaryotic severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), where it enabled a better understanding of its complex transcriptome [23]. The ONT-cappable-seq method outlined in this work enables end-to-end cDNA sequencing of primary prokaryotic transcripts to unravel the complex transcriptomes of microbial viruses and their bacterial hosts.

As a proof-of-concept, we elucidated the transcriptional properties of Pseudomonas aeruginosa virus LU77. Like other relatives of coliphage N4, LUZ7 has a peculiar transcriptional strategy that relies successively on three different RNA polymerases (RNAP) and single-stranded DNA-binding protein Drc activate transcription of the middle genes [26]. Finally, the host RNA polymerase takes over during the final infection stage and carries out transcription of the late genes [25,35]. All collected samples were immediately mixed with 12 nucleotides at the 5’ transcript end, resulting in ambiguous TSS mapping [16,21,22]. This technical limitation, which has not been reported for (PCR)-cDNA nanopore sequencing approaches, can be attributed to a loss of control of the RNA translocation speed through the nanopore after premature detachment of the motor protein.

In this study, we present a nanopore-based cDNA Cappable-seq approach (ONT-cappable-seq), which can be especially useful for resolving regulatory features in densely coded viral genomes. The power of full-length primary transcriptome sequencing was just recently demonstrated for the eukaryotic severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), where it enabled a better understanding of its complex transcriptome [23]. The ONT-cappable-seq method outlined in this work enables end-to-end cDNA sequencing of primary prokaryotic transcripts to unravel the complex transcriptomes of microbial viruses and their bacterial hosts.

As a proof-of-concept, we elucidated the transcriptional properties of Pseudomonas aeruginosa virus LU77. Like other relatives of coliphage N4, LUZ7 has a peculiar transcriptional strategy that relies successively on three different RNA polymerases (RNAP) [24,25]. Upon phage infection, a large virion-encapsulated RNAP (vRNAP) is co-injected with the viral DNA and initiates transcription of the early genes from single-stranded hairpin promoters. Among the early phage gene products, the heterodimeric RNAP (RNAP1) and single-stranded DNA-binding protein Drc activate transcription of the middle genes [26]. Finally, the host RNA polymerase takes over during the final infection stage and carries out transcription from late phage promoters, which weakly resemble the σ32 promoter consensus motif [27]. However, no middle and late viral promoters have been definitively identified in LU77 and its closest relatives, despite the availability of RNA-seq data for one of these phages [28]. This gap of knowledge and a complex transcriptional strategy makes LU77 an interesting model to benchmark our method and at the same time unravel the unknowns of LU77 genome transcription.

2. Materials and methods

2.1. Bacteriophage propagation

P. aeruginosa strain US449 was cultured at 37 °C in Lysogeny Broth (LB) medium. To amplify Pseudomonas phage LU77 [29], a culture of US449 was grown to an optical density at 600 nm (OD600) of 0.3 and infected with a high titre lysate of LU77, followed by overnight incubation at 37 °C. Afterwards, the phage lysate was purified with polyethylene glycol 8000 (PEG8000) and stored in phage buffer at 4 °C as described elsewhere [30].

2.2. Bacterial genome extraction, nanopore sequencing and hybrid assembly

The genome of P. aeruginosa strain US449 was previously sequenced using Illumina short-read sequencing technology (NCBI accession number GCF_001454415.1). To complete the reference genome, the short reads were complemented with long reads obtained after sequencing the genomic DNA of the bacteria using Nanopore sequencing technology. First, high-molecular weight bacterial genomic DNA was extracted using the DNeasy UltraClean Microbial Kit (Qiagen, Hilden, Germany). Next, the DNA sample was carefully prepared using the Rapid Barcoding Sequencing Kit (SQK-RBK004), loaded on a MinION flow cell (FLO-MIN106, R9.4) and sequenced for 24–48 h. The base-calling, demultiplexing and trimming of the raw nanopore reads was performed using Guppy (v3.4.4) and Porechop (v0.2.4) (https://github.com/rrwick/Porechop). The final quality of the sequencing run and the read lengths were evaluated using NanoPlot (v1.28.2) [31]. Afterwards, both short and long-read sequencing datasets were integrated to resolve the US449 genome using de novo hybrid assembly. For this, the Unicycler tool (v0.4.8) was used with default process parameters, followed by an additional polishing step [32,33]. The resulting high-quality genomic sequence of P. aeruginosa US449 was annotated with Prokka (v1.14.6) using default settings [34]. This assembly and annotation was used for further analysis of the transcriptome data. The resolved genome of P. aeruginosa strain US449 was deposited in NCBI GenBank (accession number CP091880). The multi-locus sequence type of US449 was identified using the mlst typing tool (https://github.com/mcginnislab/mlst).

2.3. Infection conditions and total RNA extraction

An overnight culture of P. aeruginosa US449 was inoculated in 20 mL fresh LB medium and grown to the early exponential phase (OD600 = 0.3). At this moment, a 4.5 mL culture sample was collected to serve as an uninfected control sample (t = 0 min). The remaining culture was infected with phage LU77 with a multiplicity of infection (MOI) of 50 to ensure a synchronous infection. The culture was incubated at 37 °C and 4.5 mL samples for RNA isolation were collected at different timepoints during infection that represent the early (5 min), middle (10 min) and late (20 min) infection stage, as determined from previous growth experiments [25,35]. All collected samples were immediately mixed with 0.5 mL of stop mix solution (95% v/v ethanol, 5% v/v phenol) and snap-frozen in liquid nitrogen to preserve the transcriptional state of the cells. Prior to total RNA extraction, high infection rates (>95% cells infected) were confirmed by comparing the number of colony-forming units (CFU/mL) in the uninfected control and the 5 min post-infection sample after overnight incubation.

Next, the samples were thawed on ice and centrifuged (20 min, 4,000 g, 4 °C). Total RNA was isolated from the cell pellet by lysozyme treatment, followed by hot phenol extraction and subsequent ethanol precipitation. The samples were treated with DNase I and cleaned using ethanol precipitation and subsequent RNA Clean & Concentrator-5 spin-column purification (Zymo Research). Successful removal of genomic DNA was confirmed by PCR using a host-specific primer pair (Supplementary Figure S1). The purity and concentration of the RNA samples were measured by the spectrophotometric SimpliNano device (Biochrom US, Inc.) and a Qubit 4 fluorometer using an RNA HS assay kit (ThermoFisher Scientific), respectively. Evaluation of RNA integrity was performed using an Agilent 2100 Bioanalyzer system in combination with the RNA 6000 Pico Kit, and only samples with RNA integrity numbers (RIN) greater than 9 were used for downstream processing and sequencing.

2.4. In vitro transcription of RNA spike-in

The HiScribe T7 High Yield RNA Synthesis Kit (New England Biolabs) was used for the in vitro synthesis of RNA transcripts from
2.5. ONT-cappable-seq library preparation and sequencing

2.5.1. Enrichment of primary transcripts

The enrichment of primary transcripts was performed by an adapted protocol based on the SMRT- Cappable-seq method [9,15]. Total RNA (5 μg) was supplemented with 1 ng of RNA spike-in in a total volume of 30 μL, incubated for 5 min at 65 °C, and placed on ice. The RNA was capped by adding 0.5 mM of 3′-desethylbiotin-GTP (3′-DTB-GTP) (New England Biolabs), 50 units of Vaccinia Capping Enzyme (New England Biolabs), 5 μL of 10X VCE Buffer (New England Biolabs) and 0.5 units of yeast pyrophosphatase (New England Biolabs), followed by incubation for 40 min at 42 °C. The capped RNA was cleaned by spin-column purification using the Zymo RNA Clean & Concentrator-5 kit, the quality and quantity of the purified transcripts was assessed using a SimpliNano spectrophotometer. The RNA products were run on a RNA 6000 Pico chip using an Agilent 2100 Bioanalyzer device to evaluate transcript integrity and length (1.8 kb).

2.5.2. Reverse transcription and selection for full-length transcripts

Reverse transcription and PCR amplification was performed according to the cDNA-PCR Barcoding protocol (SQK-PCS109 with SQK-PBB004; Oxford Nanopore Technologies), with an additional selection for full-length primary transcripts in the enriched samples. Briefly, 9 μL of enriched RNA (enriched sample) or 1 μL of non-enriched RNA (control) was mixed with oligo(dT) VN primers and 10 mM dNTPs in a total volume of 11 μL, incubated for 5 min at 65 °C, and snap-cooled on ice. Next, first strand synthesis was carried out as described in the protocol with 200 units of Maxima H Minus Reverse Transcriptase (Thermo Fisher Scientific), followed by incubation for 90 min at 42 °C and inactivation for 5 min at 85 °C. The enriched samples were then treated with 50 units of RNase H (New England Biolabs) for 30 min at 37 °C. All samples were purified using Ampure XP beads (Beckman Coulter) and subsequently eluted in Low TE buffer (1 mM Tris-HCl pH7.5; 0.1 mM EDTA). Before proceeding to second-strand synthesis, the enriched samples were subjected to a second round of streptavidin-based enrichment to select full-length cDNA molecules. For this, the cDNA/RNA duplexes were incubated again with an equal amount of prewashed Hydrophilic Streptavidin Magnetic Beads and incubated as stated before. The beads were washed three times with Washing Buffer and resuspended in 20 μL of low TE buffer. Afterwards, second-strand cDNA synthesis and PCR Barcoding was carried out on all samples using the manufacturers’ guidelines. For PCR barcoding, 16 cycles of amplification were performed, each with an extension time of 15 min to ensure full-length amplification.

2.5.3. Library preparation and nanopore sequencing on the PromethION platform

The cDNA reaction products were treated with 20 units of Exonuclease 1 (New England Biolabs) and subsequently purified and concentrated using AMPure XP beads. The quantity of the amplified cDNA barcoded samples was measured using the Qubit 1X HS dsDNA Assay kit (Thermo Fisher Scientific). Equimolar amounts were pooled to a total of 100 fmol, in a 23 μL sample volume. Finally, the cDNA library was mixed with nanopore sequencing adapters, loaded on a PromethION flow cell (R9.4.1) and run on a PromethION 24 device with live base-calling and demultiplexing enabled. After 48 h, the flow cell was refuelled and reloaded with the amplified cDNA library, and the sequencing run was continued for an additional 48 h until all pores were exhausted.

2.6. Data analysis

2.6.1. Raw read processing and QC

All fastq files that exceeded the default phred-like quality score threshold after base calling (≥7) were used as an input for downstream data analysis. Overall sequencing performance, throughput and raw read quality were evaluated by NanoComp (v1.11.2) [31]. Next, the raw reads were processed by Pychopper (v2.5.0) (https://github.com/nanoporetech/pychopper) to identify, rescue and correctly orient full-length cDNA reads based on the presence and direction of the reverse transcription (VNP) and strand-switching (SSP) primer sequences in the cDNA. Cutadapt (v2.7) [36] was used to remove 3′ polyA stretches and sequence remnants on the 5′ end that match the primers used during reverse transcription, as described previously [16].

2.6.2. Read mapping to the reference genomes

The trimmed reads were mapped to the reference genomes of Pseudomonas phage LUZ7 (NC_013691.1) and P. aeruginosa strain US449 using minimap2 (v2.17) with settings recommended for nanopore cDNA reads (–ax -map-ont -k14) [37]. To avoid spurious alignments, reads with more than 10 clipped bases at their 5′ and 3′ end were discarded from the SAM file using samclip (v0.4.0) (https://github.com/tseemann/samclip). SAMtools (v1.9) was used to convert the alignment format, generate sorted and indexed BAM files and evaluate the mapping metrics of each sample [38]. Subsequently, the mapped reads were assigned to the genomic features of P. aeruginosa or LUZ7 using featureCounts (v2.0.1) in long-read mode (–l –O) [39]. Alignments were visualized in the Integrative Genomics Viewer (IGV) software [40].

2.6.3. Identification of phase TSSs

The identification of transcriptional boundaries was carried out using custom scripts inspired by previously published work [16], with modifications tailored to our ONT-cappable-seq approach (https://github.com/LoCT-KUleuven/ONT-cappable-seq). For TSS detection, the alignment files were used to generate strand-specific bed files that report the number of reads starting at each
position in the LUZ7 genome using bedtools genomewalk (-3 -bga) [41]. Next, a peak calling algorithm (https://github.com/NICHD-BSPC/termseq-peaks) was applied to determine genomic positions of local maxima from 5’ read ends [16,42]. Nearby peaks within a distance of 5 base pairs (bp) on the same strand were clustered, retaining only the peak position with the highest number of 5’ read ends. Peak positions where less than five reads began were neglected.

In a second step, the read count per million mapped reads (RPM) was calculated for all remaining peak positions in the individual samples. To assess whether the accumulation of 5’ read ends arises from primary transcripts, an enrichment ratio was determined for each peak position i by dividing the RPM in the enriched sample by the RPM in the corresponding control sample with maximum 1 bp positional difference (i ± 1):

$$\text{enrichment ratio}(i) = \frac{\text{RPM}_{\text{enriched sample (i)}}}{\text{RPM}_{\text{control sample}}}$$

In case the enrichment ratio surpassed the threshold value, the peak position was annotated as a TSS. To accommodate for differences in enrichment levels across the samples, this threshold (TSS) was set by assessing the enrichment ratio of the TSS of the T7 promoter of the RNA spike-in in the samples taken 5 min (TSS = 40.9), 10 min (TSS = 16.7) and 20 min (TSS = 24) post-infection. The −50 to +5 region of the annotated TSS was analysed using MEME [43] and late promoter regions (-100 to +1) were uploaded in SAPPHIRE (v2) [44] for Pseudomonas sy7 promoter sequence prediction.

2.6.4. Identification of phage TTSs

LUZ7 TTSs were pinpointed in a similar manner as the TSSs detection. In a first step, strand-specific bed files of the 3’ read ends were generated from the alignments of the enriched samples (bedtools genomewalk −3 −bga) [41]. Next, genomic positions with an accumulation of 3’ read ends were identified using the peak calling algorithm as described previously [16,42]. However, as ONT-cappable-seq does not allow direct differentiation between random termination events and true TTSs, more stringent filtering steps were implemented for 3’ end detection as opposed to TSS identification. Peak positions located on the same strand within 25 bp were clustered together. Here, the site with the largest number of reads was selected as a representative. Candidate TTSs with <20 reads ending at that position were filtered out.

In a second step, the strength of the candidate terminators was estimated using custom scripts inspired by previously described methods [14,45]. Briefly, all reads that start upstream of the TTS were extracted and used to calculate the coverage drop before and after the terminator, averaged over a window of 20 bp. Only terminators that showed a read reduction of at least 20% were annotated as putative TTSs and used for further analysis. Intrinsic transcriptional terminators were predicted using TransTermHP [46] and ARNold [47] after uploading the −50 to +50 region of the ONT-cappable-seq predicted TTSs. RNAfold (v2.4.13) was used to predict and calculate the secondary structure and minimum free energy of the −50 to +1 region of the TTS [48].

2.6.5. Determination of transcriptional units

The transcription units (TUs) were delineated based on adjacent TSSs and TTSs defined in this study. As a validation, we identified reads that encompass the candidate TUs in full-length. In case no clear TSS-TTS pair could be identified, the most distant 3’ end was used as the putative TU boundary. The unique combination of phage genes encoded in the transcription units, defined as the transcriptional context [15], was determined by finding overlaps between the ONT-cappable-seq reads and the genomic features of LUZ7 using the bedtools intersect tool [41]. Only the genes encoded on the same strand that are least 90% covered by the transcription unit were included in the transcriptional context (-F 0.9 -s). Overlapping transcriptional units that were transcribed from the same strand and had at least one common gene in their transcriptional context were annotated in an operon structure [49].

2.7. Primer extension assays

Primer extension assays were carried out using the Superscript IV First-Strand Synthesis System (Thermo Fisher Scientific), according to the manufacturer’s instructions. For this, specific primers (Supplementary Table S1) were radiolabelled with T4 Polynucleotide kinase (NEB) and [γ-32P]ATP (6,000 Ci/mmol), purified on a Micro-Bio Spin P-6 Gel Column (Bio-Rad) and subsequently annealed to 5 μg of RNA extract. The RNA-primer samples were shortly incubated at 65 °C, supplemented with the SuperScript IV Reverse Transcriptase reaction mix and incubated at 55 °C for 10 min. The reaction product was mixed with formamide loading dye, boiled for 1 min and loaded on an 8% polyacrylamide-7 M urea gel.

In parallel, for each TSS candidate, the surrounding DNA region was amplified from the LUZ7 genome using the primer for primer extension, paired with a primer located upstream the TSS of interest (Supplementary Table S1). The resulting PCR fragment was used as a template for Sanger sequencing to infer the nucleotide sequence of the transcription initiation region. The Sanger sequencing reaction was carried out using the USB Thermo Sequene Cycle Sequencing Kit (Affymetrix), following the guidelines for radiolabelled primer cycle sequencing. Briefly, after mixing the radiolabelled primers, the DNA template, and a dNTP mix containing a dideoxy variant of one of the nucleotides (ddATP, ddTTP, ddCTP or ddGTP), four sequencing reactions were performed for 45 cycles (95 °C, 30 s; 55 °C, 30 s and 72 °C, 1 min). The samples were supplemented with loading dye, boiled for 1 min and loaded on an 8% polyacrylamide-7 M urea gel alongside their respective primer extension product. The gel was fixed, dried and exposed on a phosphor screen before visualising with a Typhoon phosphorimager (GE Healthcare).

In parallel, for each TSS candidate, the surrounding DNA region was amplified from the LUZ7 genome using the primer for primer extension, paired with a primer located upstream the TSS of interest (Supplementary Table S1). The resulting PCR fragment was used as a template for Sanger sequencing to infer the nucleotide sequence of the transcription initiation region. The Sanger sequencing reaction was carried out using the USB Thermo Sequenase Cycle Sequencing Kit (Affymetrix), following the guidelines for radiolabelled primer cycle sequencing. Briefly, after mixing the radiolabelled primers, the DNA template, and a dNTP mix containing a dideoxy variant of one of the nucleotides (ddATP, ddTTP, ddCTP or ddGTP), four sequencing reactions were performed for 45 cycles (95 °C, 30 s; 55 °C, 30 s and 72 °C, 1 min). The samples were supplemented with loading dye, boiled for 1 min and loaded on an 8% polyacrylamide-7 M urea gel alongside their respective primer extension product. The gel was fixed, dried and exposed on a phosphor screen before visualising with a Typhoon phosphorimager (GE Healthcare).

2.8. In vivo promoter activity assay

The functionality of selected LUZ7 late promoters was confirmed experimentally in vivo using our recently established SEVAtile-based expression systems. For this, adaptors (generated by overlapping primer pairs) encoding putative promoters were cloned upstream of a standardized ribosomal binding site (BCD2) and a green fluorescent reporter protein (msfGFP) gene in a pBCGDes vector using SEVAtile assembly [50]. In addition, a vector lacking a promoter sequence (pBCGDes BCD2-msfGFP) and a vector with a
constitutive promoter (pBGDes P_{M7}-BCD2-msfGFP) were created to serve as a negative and positive control, respectively. As a separate negative control, we constructed three vectors with a random LUZ7 sequence as a decoy promoter site (pBGDes P_{decoy}1-3-BCD2-msfGFP). Vectors, primers and insert sequences used are listed in Supplementary Table S1. The genetic constructs were transfected into both E. coli PIR2 and P. aeruginosa PA01 host cells. In case of P. aeruginosa PA01, the resulting vectors were electroporated together with a pTNS2 helper plasmid according to the method by Choi et al. (2005) to ensure site-specific genomic integration [51]. To quantify the in vivo expression levels of msfGFP in the transformed P. aeruginosa PA01 and E. coli PIR2 cells, fluorescent expression assays were performed using four biological replicates. The bacteria were inoculated in M9 minimal medium (1 × M9 salts (BD Biosciences), 0.5% casein amino acids (LabM, Neogen), 2 mM MgSO_4, 0.1 mM CaCl_2 (Sigma Aldrich), 0.2% citrate (Sigma Aldrich)) supplemented with 50 μg/μl kanamycin (E. coli PIR2) or 30 μg/ml gentamicin (P. aeruginosa PA01) and grown overnight. The next day, the cultures were diluted in fresh M9 medium with the appropriate antibiotic in a Corning® 96 Well Black Polystyrene Microplate with Clear Flat Bottom and incubated with shaking. After 3.5 h, msfGFP and OD600 levels were measured on the CLARIOstar® Plus Microplate Reader (BMG Labtech, Ortenberg, Germany), as described previously [50]. The msfGFP fluorescence units were normalized for the associated OD600 values and converted to absolute units using an independent calibrate, 5(6)-carboxyfluorescein (5(6)-FAM) (Sigma Aldrich) [50,52].

2.9. In vivo terminator validation

A subset of the LUZ7 terminators was validated in vivo by a terminator trap system, as described by Lammens et al. (2021). Briefly, pBGDes vectors bearing the terminator regions of interest between an msfGFP reporter upstream and an mcherry reporter downstream were generated using SEVatile assembly [50] (Supplementary Table S1). Expression of both reporters is driven by the P_{M7} constitutive promoter and standardized translation initiation elements BCD1 and BCD2 [53,54]. Hosts carrying constructs without terminator (pBGDes P_{M7}-BCD2-msfGFP-BCD1-mCherry) and the wildtype T7 terminator from coliphage T7 (pBGDes P_{M7}-BCD2-msfGFP-T7(wt)-BCD1-mCherry) were used as a negative and positive control, respectively. Next, fluorescent expression assays were performed to quantify the transcriptional termination activity of the terminators of interest. Similar to the promoter activity assay described above, four biological replicates of P. aeruginosa PA01 carrying the constructs were grown in M9 minimal medium with antibiotic and subsequently assayed for fluorescence measurements. After 3–4 h incubation in fresh M9 medium the next day, the msfGFP, mCherry and OD600 levels were measured on the CLARIOnstar device. The fluorescence intensity of msfGFP and mCherry was determined for each sample and normalized for the corresponding OD600 value. The termination activity was subsequently calculated according to previously published methods [50,55]:

\[
\text{termination activity} = \frac{\text{msfGFP}_{\text{normalized}}/\text{mCherry}_{\text{normalized}}}{\text{msfGFP}_{\text{normalized}}/\text{mCherry}_{\text{normalized}}_{\text{control}}}
\]

3. Results and discussion

3.1. Profiling the full-length primary phage transcriptome using ONT-cappable-seq

In this study, we combine the Cappable-seq enrichment strategy and the nanopore sequencing platform to isolate and sequence the full-length primary transcriptome of Pseudomonas phage LUZ7 at different infection stages. The key steps of the ONT-cappable-seq workflow are depicted in Fig. 1a. The RNA transcripts are subjected to an enzymatic capping reaction that specifically labels the 5′ triphosphate group hallmarking the beginning of primary transcripts with a desthiobiotin tag [9]. Next, the transcripts are polyadenylated to preserve their 3′ ends and make them compatible with nanopore PCR-cDNA sequencing library preparation. To selectively enrich primary transcripts, the desthiobiotinylated and polyadenylated transcripts are captured on streptavidin beads and separated from their processed counterparts [9]. Due to low RNA amounts (only 1–5% of the total prokaryotic RNA pool represent primary transcripts), we have opted for a PCR-cDNA sequencing approach instead of direct cDNA or native RNA sequencing, which require higher input amounts [16]. After the first enrichment procedure, the RNA is reverse transcribed and the second round of enrichment is performed to select for full-length transcripts [15]. The cDNA is barcoded and PCR-amplified, followed by adapter addition to feed and translocate the cDNA fragments through the nanopore. In parallel, an unenriched control library is similarly prepared from the same RNA sample by excluding streptavidin enrichment steps from the workflow. Depending on the number of samples and desired read depth, the final pooled library can be sequenced on a MinION or a PromethION device.

To obtain temporal transcriptional landscapes of LUZ7 throughout infection of its P. aeruginosa US449 host, uninfected (0 min), early (5 min), middle (10 min) and late (20 min) infection stage samples were taken and, together with their corresponding controls, multiplexed and loaded on a single PromethION flow cell. The sequencing experiment ran for 48 h, generating a total of 43.3 million reads (18.2 Gb) that passed the default quality score threshold after base calling (Supplementary Figure S2). The full-length reads were trimmed, oriented and mapped against the reference genomes of the phage and the host for further downstream analyses (Fig. 1b). The number of reads that could be mapped to both genomes ranged between 69.9% and 78.4% in the individual samples (Supplementary Figure S3a).

The enrichment of primary transcripts is a critical step in the ONT-cappable-seq method and can be evaluated by analysing the fraction of reads that originate from processed RNA species, such as ribosomal RNA (rRNA) and transfer RNA (tRNA) [15,56]. We compared the distribution of reads on the genomic features of LUZ7 and P. aeruginosa US449 between the enriched and the control libraries of each dataset. We observed that more than 97.5% of the mapped reads in the control libraries correspond to rRNA and tRNA. In contrast, the number of reads from rRNA and tRNA was significantly reduced (28.6%–60.4%) in the enriched libraries of each timepoint, indicating successful selection against processed transcripts (Fig. 2). After sequencing the enriched samples, 96–99% of the annotated genes are covered by at least one read in all datasets, whereas only 65–80% of the genes are covered in the control samples. This shows that the enrichment procedure and the reduced number of processed RNA species enables deeper sequencing of the prokaryotic transcriptome.

To further evaluate whether the ONT-cappable-seq procedure provides a comprehensive picture of the time-resolved full-length transcriptional landscape during phage infection, we analysed the length of the mapped reads in individual samples (Supplementary Figure S3b). As expected, the aligned read length distribution was impacted significantly by the enrichment procedure with N50 read lengths varying between 358 and 372 bp for the enriched samples and 1,511–1,541 bp for the controls. The length difference is more pronounced on the bacterial side and can be largely attributed to rRNA depletion, as the mean reads lengths and N50 values of phage-derived transcripts are similar between enriched and control libraries across different samples.
Despite modest average read lengths, the top five longest mapped phage reads in each sample span 3.3 kb–8.3 kb, indicating that the procedure allows end-to-end sequencing of long RNA molecules. The advantage of this long-read RNA sequencing method to recover the boundaries of individual phage transcripts is illustrated when comparing the Integrative Genomics Viewer (IGV) data tracks of our technique to standard RNA-seq experiments performed previously on another N4-like phage, LIT1 [28] (Fig. 1c).
3.2. Delineation of the transcriptional features of *Pseudomonas* phage LUZ7

ONT-cappable-seq enables full-length profiling of primary transcripts, providing information on their 5' and 3' extremities simultaneously. Using this approach, we generated a comprehensive inventory of phage LUZ7 transcripts, unveiling transcription initiation and termination events, complex operon structures and gene expression levels across different infection stages, all in a single experiment (Fig. 3).

3.2.1. Identification of LUZ7 transcription start sites

Candidate transcription start sites (TSSs) are identified by determining the genomic positions where a significant number of reads begin. Afterwards, comparison of the relative strength of each TSS candidate between the enriched library and the control reveals the positions of original 5' transcript ends, enabling TSS determination at high resolution. TSS identification by our method was first verified using an *in vitro* transcribed RNA spike-in, which was added...
to RNA from phage-infected cells prior to library preparation (Supplementary Figure S4). This demonstrated accurate TSS annotation of the spike-in transcript initiated from a T7 RNA polymerase promoter. Next, the boundaries of LUZ7 transcripts were delineated. Comparative transcriptomic analysis, followed by manual curation, yielded a total of 46 unique TSSs across the LUZ7 genome, which were classified according to their genomic context and expression level as described previously [8] (Supplementary Table S2). Primary TSSs are defined as TSSs that reside within ±500 bp upstream of annotated genes and show a higher expression level than additional TSSs for the same gene (secondary TSSs). Intragenic TSSs on the same strand of an encoded gene, antisense TSSs located inside or within 100 bp distance of an oppositely oriented gene, and TSSs with no genes nearby are designated as internal, antisense, and orphan TSSs, respectively. In this manner, we observed 25 primary and three secondary phage TSSs, and found a considerable overlap between the five different TSS categories (Fig. 4a).

Multiple sampling points throughout the LUZ7 infection allowed us to distinguish between early, middle, and late phage promoters, from which three different RNAPs successively initiate transcription during infection. Transcription of the LUZ7 genome begins after the phage co-injects its viral DNA along with a large vRNAP that drives transcription from early phage promoters. After analysing the viral primary transcriptome during the early infection stage, nine TSSs and their cognate early promoter sequences were identified, including two promoters that were previously predicted after inspection of the LUZ7 genome [25]. The majority of early promoters contain a typical hairpin structure similar to that of coliphage N4 early promoters (Fig. 4b), with a ≥4 bp stem and a conserved three nucleotide loop [58]. Motif analysis revealed that the preferred TSS nucleotide of the identified early promoters is guanosine, located 13–14 bp downstream of the hairpin centre. Notably, the few early promoters that lack the distinctive hairpin structure all reside within the coding region of genes that are preceded by a hairpin promoter, yielding truncated transcripts. These non-canonical early promoters may therefore depend on a prior initiation from their canonical hairpin counterpart, though this requires further investigation.

Interestingly, two of the strongest early TSSs are organized in tandem, 33 nucleotides apart, and drive the expression of early phage protein Drc, making it the most highly expressed viral gene during infection, accounting for up to 64% of phage transcripts. The

---

**Fig. 4. TSS identification in LUZ7.** A. Venn diagram showing viral TSS classifications. B. Motif analysis of the early promoters reveals a conserved hairpin structure similar to that found in N4. C. Motif analysis of LUZ7 middle promoters reveals a GC-rich hairpin structure.
Drc protein was recently shown to be a key transcriptional regulator during LUZ7 infection, activating the progression from the early to the middle transcription stage by recruiting the LUZ7 RNAP to middle promoters [26]. In addition, in vitro this protein can fully cover large stretches of ssDNA when present at high concentrations and since its transcripts are so abundant this is likely also the case in vivo during phage infection.

Middle promoters have thus far been impossible to identify in LUZ7 and related phages such as LIF1, despite the availability of middle stage transcription initiation sites in the prototypical phage N4 and RNA-seq data on phage LIF1 [28, 59]. Our ONT-capable-seq data on RNA from the mid-infection stage did enable us to identify four viral TSSs associated with middle stage infection. Collectively, these middle promoters show a GC-rich hairpin structure that resembles that of the early hairpin promoters, with a stem of ≥6 bp and a 3–4 nucleotide (nt) loop positioned ~15 bp upstream of the identified TSS (Fig. 4c). Similarly to the early TSSs, guanosine is the preferred nucleotide for transcription initiation from middle promoters. The architecture of middle promoters is significantly different from the RNAPII transcription initiation sites that have been documented for coliphage N4, suggesting that LUZ7 and N4 have distinct transcriptional initiation mechanisms on their middle promoters, as in fact was proposed previously [26]. In N4, RNAPII prefers to start transcription from thymidine and recognizes promoters with two consensus motifs in a sequence-specific manner, whereas transcription initiation by LUZ7’s RNAPII appears to be directed by the recognition of hairpin structures [24]. These results indicate that ONT-capable-seq can help infer the requirements of RNAPII promoter recognition in N4-like phages, which has proven to be very challenging in the past. However, given how extensive the middle genomic region is, it cannot be excluded that the limited set of identified LUZ7 middle promoters is an underestimation of the true number of TSSs associated with the middle transcription stage.

The third and final transcription stage of LUZ7 is driven by the host RNAP, which is recruited to late promoters on the phage genome. We discovered 33 late TSSs scattered across the LUZ7 genome, of which more than 80% are located on the Crick strand. This strand bias is largely in agreement with the orientation of the structural and lysis gene cassettes, which is opposite to that of the early and middle genes. Notably, all genes on the Crick strand are strictly transcribed during the late stage, even when these genes are found inserted within early gene cassettes (e.g., genes orf34 and orf35). In N4, late promoters show remote resemblance to the σ70 promoter sequences of the host. To assess whether this is also true for LUZ7, nucleotides –100 to +1 of the identified late TSSs were analysed using the Pseudomonas promoter prediction tool, SAPPHIRE.CNN [44]. Indeed, the vast majority of late promoters displayed significant similarity to the Pseudomonas σ70 promoter consensus sequence.

3.2.2. Identification of phage transcription termination sites

In addition to TSSs, the annotation of 3′ transcript ends is essential to obtain a complete picture of an organism’s transcription. The longer reads from nanopore-based sequencing allow the detection of the 3′ ends. However, the determination of true TTs is often blurred by the activity of 3′ to 5′ exonucleases that obfuscate the true 3′ boundaries of RNA molecules. As some of the major bacterial ribonucleases show a substrate preference for transcripts with 5′ monophosphate termini, primary transcripts with intact 5′ triphosphate groups are less susceptible to RNA processing and are more likely to contain their original termination ends [60, 61]. Consequently, the enrichment of primary transcripts performed in this study increases the likelihood that the observed termination signals represent genuine TTs, but it cannot be excluded that a number of 3′ ends originate from post-transcriptional processing.

In bacteria and phages, the termination of transcription largely occurs via two main mechanisms, intrinsic termination and factor-dependent termination. In general, transcription termination motifs were located on the phage genome by detecting local maxima of 3′ read ends in our enriched datasets, revealing a total of 61 termination sites, among which 14 are predicted to be intrinsic, factor-independent terminators (Supplementary Table S3). The vast majority of identified TTs (65.6%) end transcription of RNA molecules transcribed from the Watson strand, whereas 21 of the termination signals are encoded on the Crick strand. Almost half of the termination sites (47.5%) are located in intragenic regions, directly downstream of the preceding gene on the same strand. The rest of the phage TTs reside either in intergenic regions (32.8%) or in antisense orientation relative to annotated genes (19.7%). The length of the 3′ untranslated regions (3′UTR) varies extensively, with more than 30% of the 3′UTRs exceeding 100 nt in length (Supplementary figure S5a). Analysis of regions upstream (~50 to +1) of TTs indicated that most phage terminators are prone to form stable secondary structures (Fig. 6a). For each terminator, the termination efficiency (TE) was determined by calculating the level of readthrough across the termination site, and a wide range of terminator strengths was observed (Supplementary figure S5b). On average, the termination efficiency is weakly correlated with the folding energy of RNA upstream of the TTS sequence (R = -0.252, p < 0.05) (Supplementary Figure S5c). Interestingly, ten terminators showed altered readthrough levels (~30%) throughout the infection. This suggests that these dynamic transcriptional terminators may help tweak LUZ7 gene expression levels throughout the infection (Fig. 6b,c). For example, the TE of terminator T8 is <5% at the early infection stage, allowing almost all transcripts to fully span the early transcribed orf15, whereas
10 min post-infection the TE of T8 increases and almost 50% of the transcripts end prematurely within the orf15 gene body. Similarly, the TE of terminator T42 decreases from 88% in the middle stage of infection to 57% at the late infection stage, enabling more transcripts to read through to gene orf65, which codes for an important regulator in the final transcriptional stage of LUZ7. Modulation of transcription termination is a powerful strategy to regulate gene expression programs and has also been investigated in detail in a handful of phages, including coliphages \( \kappa \) and HK022 and Thermus thermophilus phage P23-45 [62–65]. The phage-encoded antiterminators manipulate the host RNAP and enable it to bypass terminators at the appropriate transcription stage during infection. In LUZ7, the varying terminator readthrough levels throughout infection might point to similar or novel antitermination strategies that successively modify the different RNAPs during phage transcription in response to regulatory stimuli, though this warrants further investigation.

In addition to unidirectional transcriptional termination events, we discovered four overlapping bidirectional terminator regions comprised of pairs of convergent TTSs located within 4–40 nt from each other (Fig. 6d). All but one of these bidirectional terminators reside between genes that are oriented in a head-to-head manner. Comparison of the readthrough levels in both directions suggests that termination tends to be more efficient in the orientation of the gene that is predominantly expressed (Supplementary figure S5d).

### 3.2.2.1. Experimental validation of LUZ7 terminator regions

The potential of ONT-cappable-seq to uncover TTSs and terminators was supported by validating the termination activity of a subset of the LUZ7 terminators, including T1, T47 and bidirectional terminator pair T60 and T61 in \( P. \ aeruginosa \). To this end, we constructed a terminator trap using the SEVAtile assembly method, in which the phage terminators are inserted between two fluorescent...
Fig. 6. Characteristics of LUZ7 TTSs and terminators. A. Distribution of the minimum free energy (kcal/mol) of the −50 to +1 sequences of identified phage TTS (dark blue) compared to sequences of the same length randomly selected across the LUZ7 genome (light blue). B. Bar plot showing the termination efficiency of a set of phage terminators detected at the early (dark blue), middle (light blue) and late (orange) infection stages. C. Comparison of ONT-cappable-seq data tracks of early and middle infection for phage terminators downstream of LUZ7 orf15. The fraction of transcripts that span the entire length of the gene is significantly decreased at middle infection stage. D. ONT-cappable-seq data track of the late infection stage showing an example of a bidirectional terminator region located between a convergent gene pair. E. In vivo validation of LUZ7 terminators in P. aeruginosa using a terminator trap assay [50,55]. The terminator of interest is trapped between an msfgfp and an mcherry reporter gene, which are transcribed from a constitutive promoter (Pem7). BCD1 and BCD2 are used as ribosomal binding sites to initiate translation of the reporters. The control represents a terminator trap vector without a terminator sequence. The termination activity was quantified by comparing the ratio of the normalized msfGFP and mCherry expression levels with the ratio of normalized msfGFP and mCherry levels in the control (Methods). The bidirectional LUZ7 terminator pair is connected with a bracket. Data represent mean values for four replicates and standard deviation is indicated by error bars. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

reporter genes, msfgfp and mcherry, whose expression is driven by a constitutive promoter [50]. To quantify termination activity, the intensity of msfgfp and mcherry expression was measured for all samples, and compared to control constructs containing either no terminator sequence or the native T7 terminator of phage T7 between the two genes, as described earlier [50,55]. Relative to the control, all phage terminators show reduced levels of downstream mcherry expression and, therefore, significant termination activity (Student’s t-test, p < 0.05), confirming their functionality in vivo (Fig. 6e). These results further support the assumption that most TTSs annotated in this work represent genuine terminators.

3.2.3. Analysis of the operon organization

Taking advantage of the potential of ONT-cappable-seq to uncover both ends of transcripts, the operon architecture of LUZ7 can also be elucidated. In general, the boundaries of transcription units (TUs) were annotated based on adjacent TSSs and TTSs defined in this study. Next, the candidate TUs were validated by identifying ONT-cappable-seq reads spanning the full-length of a TU, leading to a total of 86 unique TUs (Supplementary Table S4). The length of the TUs varied extensively and it was found that LUZ7 TUs cover, on average, 1.9 genes (Supplementary Figure S5a,b). Analysis of the transcriptional context of TUs, defined by the unique combination of encoded genes [15], revealed significant overlap between individual TUs, as more than 60% of the genes in the annotated TUs are transcribed in more than one context (Fig. 7a). This suggests that the alternative usage of TUs could serve as an additional regulatory layer to fine-tune the expression of individual genes [66].

In the majority of cases, we observed that the gene content of transcripts altered as a result of sequential readthrough across different TTSs, as exemplified by the stepwise transcription termination pattern of LUZ7 genes coding for gp64-gp69 (Fig. 7b). This is consistent with the finding that more than 60% of the identified phage termination sites allow substantial downstream transcript extension (average TE <50%). Previous studies have already observed that modulation of the degree of readthrough at termination sites gives rise to numerous complex operon structures in bacteria [15]. In addition, the use of inefficient terminators as a transcriptional strategy to properly balance gene expression levels has also been demonstrated for model coliphage T7 [67].

Based on the proposed definition of a complex operon as a gene cluster of which the transcription is orchestrated by various overlapping TUs that have at least one gene located on the same strand in common [49], the LUZ7 genome can be roughly organized in 14 complex operons, which are illustrated in summarizing Fig. 3 (Supplementary Table S5). In general, genes that are included in the same operon structure are likely to be functionally related. This observation also appears to apply for LUZ7. For example, the first three operons residing in the leftmost genomic region comprise most of the early phage ORFans involved in the host hijacking process, whereas operons 7–8 and operon 13 encompass structural genes involved in the formation of the tail and the virion capsid, respectively [25,28]. It should be noted that not all annotated genes are included in identified operons. For example, a gene of the tail fiber protein (orf56, 3.2 kb) and of the virion-associated RNAP gene (orf703, 13.2 kb), two of the largest genes encoded by the phage, are not found in detected operons. This could indicate that unusually long transcripts pose more of a challenge to be fully captured by ONT-cappable-seq, making it more difficult to detect transcriptional unit variants that extend beyond these gene borders.

3.2.4. Identification of other regulatory features

In addition to cataloguing regulatory elements of LUZ7, ONT-cappable-seq allowed us to reveal peculiar transcriptional patterns of the phage. At the initial stages of infection, the genes encoding for hypothetical protein gp01 and the Drc protein (gp14) are extensively transcribed, reaching a combined total of 90% of early phage transcripts. Both gp01 and Drc have homologues in a number of N4-like Pseudomonas phages, including phage LIT1 [26,28]. Former analysis of the LIT1 transcriptome five minutes after P. aeruginosa PA01 infection found comparable expression levels for the corresponding genes [28], corroborating that the encoded proteins play an important and seemingly conserved role in the infection strategy of N4-like phages targeting Pseudomonas. Whereas the crucial involvement of Drc in the transcriptional regulation of LUZ7 has already been demonstrated [26], a function for gp01 has not yet been identified.

Consistent with the transcriptional landscape of LIT1, we observe that the LUZ7 genome is almost exclusively transcribed from the Watson strand during early and middle timepoints [28]. Interestingly, the transcripts captured mid-infection collectively cover the entire phage genome, including the rightmost region that encodes 29 small ORFans (gp87–gp115) [25]. While the function of these genes is unknown, the temporal transcriptional information gained in this study could provide clues on the role of this unique gene cluster during the phage cycle. Given that the middle genes of LUZ7 are primarily involved in phage genome replication, these genes are likely to have a related function. In addition, ONT-cappable-seq data show that middle transcripts extend throughout genomic regions that have late structural gene cassettes encoded on the opposite strand. The same phenomenon was observed in LIT1, where it was suggested that the antisense transcripts are responsible for silencing premature expression of the late genes [28]. During the final infection stages, the dominant transcription direction switches in both phases. Presumably, the surge in RNA molecules transcribed from the Crick strand overpowers the antisense transcripts to enable the timely expression of structural and lysis-associated genes [28].
Our data also revealed interesting transcriptional activity in the intergenic space between LUZ7 genes orf69 and orf70. This region is extensively transcribed during the late infection stage, giving rise to numerous RNA species of varying lengths (Fig. 8). Most of the reads that align with this region are <250 nt long and lack any apparent protein-coding potential, suggesting the presence of putative small non-coding RNAs (sRNA) in the LUZ7 transcriptome. Closer examination of this condensed transcriptional pattern roughly distinguishes three abundant sRNA candidates, designated as sRNA1, sRNA2, and sRNA3, that partially overlap with each other. Two of these RNA species share the same 5' extremity but use an alternative termination site, resulting in a short and extended version of the transcript. By contrast, the shortest sRNA candidate, sRNA3, seemingly lacks an annotated TSS, suggesting that this fragment is derived from a primary precursor through processing events. Functional sRNA biogenesis through RNase activity has already been demonstrated in various bacteria [68–70], and was recently also suggested to occur in Pseudomonas phage phiKZ [71]. In addition, the biological relevance of small non-coding RNAs in the regulation of phage and host development is increasingly being recognized in several other Pseudomonas phages [5,72]. However, in LUZ7, the regulatory scope of the putative sRNA species and the exact mechanisms that underlie their biogenesis remain to be elucidated.

4. Conclusions and perspectives

The plethora of high-throughput microbial transcriptomic approaches that have emerged in the past decades has led to major advances in our understanding of the different layers of gene expression in bacteria. However, in contrast to bacterial RNA biology, the diverse transcription strategies of phages still remain poorly understood. We here took advantage of the Cappable-seq enrichment strategy and the nanopore sequencing platform to develop ONT-cappable-seq and generated a comprehensive transcriptional map of an N4-like Pseudomonas phage, LUZ7. We showed that ONT-cappable-seq allows full-length transcriptional profiling and enables the simultaneous identification of both 3' and 5' transcriptional boundaries of individual phage transcripts. Using this approach, we pinpointed key phage-encoded transcriptional features at the early, middle, and late infection stages, unveiled complex operon organizations and revealed the presence of novel potential regulatory elements, such as sRNAs. We found many parallels between the transcriptional progression schemes of LUZ7 and coliphage N4, however, the middle promoter motifs identified in this study, together with the Drc-mediated recruitment of RNAPII [26], hint to different transcriptional initiation mechanisms from middle promoters of the two viruses. Furthermore, our analysis revealed, extensive and time-dependent readthrough at phage terminator regions, leading to numerous extended transcription unit variants that incorporate additional genes. Consistent with previous observations in bacteria [15,73],
our results suggest that weak termination sites have an important regulatory role in the transcriptional strategy of phages, allowing them to adjust their dense transcriptional patterns in response to different infection stages. The systematic discovery of dynamic transcription terminators by ONT-cappable-seq could help identify and dissect novel regulatory mechanisms to modulate the strength of terminators in phages and their bacterial hosts.

While the underlying mechanisms through which the annotated transcriptional features orchestrate gene regulation in LU7 warrant further investigation, this work already demonstrates the breadth of knowledge that can be gained from ONT-cappable-seq over classical RNA-seq approaches, which generally fail to capture primary RNA species and delineate transcriptional boundaries. Although this study primarily focused on viral transcripts, similar analyses can be performed on the bacterial side to shed further light on the regulatory features that govern the complex interaction between phages and their host.

In recent years, several full-length RNA-seq protocols have emerged to comprehensively characterize prokaryotic transcripts [14–16]. Notably, the SMRT-Cappable-seq protocol includes size-selection steps to eliminate all cDNA fragments below 1 kb from the total library [15]. This approach would be less suitable to study the LU7 transcriptome, as illustrated by the staggering amount of transcripts encoding small proteins such as gp01 and Drc, and the unique intergenic transcription hotspot of small RNA species revealed in this work. Nevertheless, future efforts to mitigate the challenges in end-to-end sequencing of particularly long RNA molecules could further enhance the annotation of transcriptional boundaries and operon organizations by ONT-cappable-seq and may increase the number of identified TSSs and TTSs in this study.

Taken together, this work introduced ONT-cappable-seq as a novel method to profile prokaryotic transcripts in full-length and yielded a first global dataset of TSSs, TTSs and operons in a phage infection model. In particular, we found that phages, whose genomes are densely encoded with small genes and that rely on complex transcriptional regulation systems, are a promising sandbox for the application of ONT-cappable-seq. In phage biology, our current understanding of transcriptome architecture and temporal gene regulation is mainly restricted to a handful of model phages. We envision that the pipeline provided in this study can serve as a gateway to move beyond these model phages and chart the transcriptional landscapes of diverse and uncharacterized phages in unprecedented detail. In addition to the PromethION sequencing platform used in this study, our technique is also compatible with the portable MinION device, offering a highly accessible and cost-effective approach to obtain a bird’s eye view of the transcriptional scheme of phages. In the future, the global elucidation of viral transcriptional regulatory sequences can help unravel the complexity of phage-host interactions, which can ultimately drive phage-based applications in medicine and biotechnology forward.

5. Data availability
The resolved genome of P. aeruginosa strain US449 was deposited in NCBI GenBank (accession number CP091880). Raw RNA sequencing files are deposited under GEO accession number GSE196845. All scripts and codes used in this study are made available on Github (https://github.com/LoGT-KULeuven/ONT-cappble-seq). Any additional information is accessible from the authors upon request.

Declaration of Competing Interest
The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements
We thank Genomics Core Leuven and its staff for providing the PromethION flow cell and assistance with sequencing.

Funding
This work was supported by the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme [819880]. MB is funded by a grant from the Special Research Fund [BOF/21/092]. EML is supported by the Fonds voor Wetenschappelijk Onderzoek Vlaanderen (FWO) [grant No. G096519N].

Appendix A. Supplementary data
Supplementary data to this article can be found online at https://doi.org/10.1016/j.csbj.2022.05.034.

References
[1] Hör J, Gorski SA, Vogel J. Bacterial RNA Biology on a Genome Scale. Mol Cell 2018;70(5):785–99.
[2] Parmar KM, Gajiwad SL, Dhakephalkar PK, Kothari R, Singh RP. Intriguing interaction of bacteriophage-host association: An understanding in the era of omics. Front Microbiol 2017;8(APR).
[3] Guell M, Yus E, Lluch-Senar M, Serrano L. Bacterial transcriptomics: What is beyond the RNA horiz-one? Nat Rev Microbiol 2011;9(9):658–69.
[4] Brandão A, Pires DP, Coppons L, Voet M, Lavigne R, Azeredo J. Differential transcription profiling of the phage LUZ19 infection process in different growth media. RNA Biol 2021;18(11):1778–90.
[5] Chevallereau A, Blasdel BG, De Smet J, Monot M, Zimmermann M, Kodageeva M, et al. Next-Generation “omics” Approaches Reveal a Massive Alteration of Host RNA Metabolism during Bacteriophage Infection of Pseudomonas aeruginosa. PLoS Genet 2016;12(7):1–20.
[6] Lood C, Danis-Wlodarczyk K, Blasdel BG, Bin JH, Vandenheuvel D, Briers Y, et al. Integrative omics analysis of Pseudomonas aeruginosa virus PA5oct highlights the molecular complexity of jumbo phages. Environ Microbiol 2020;22(6):2165–81.
[7] Leskenin K, Blasdel BG, Lavigne R, Skurnik M. RNA-sequencing reveals the progression of Phage-Host interactions between eQTL-37 and Yersinia enterocolitica. Viruses 2016;8(4).
[8] Sharma CM, Hoffmann S, Darfeuille F, Reignier J, Findesil S, Sittka A, et al. The primary transcriptome of the major human pathogen Helicobacter pylori. Nature 2010;464(7286):250–5.
[9] Etrvoller L, Buswell J, Yigit E, Schildkraut I. A novel enrichment strategy reveals unprecedented number of novel transcription start sites at single base resolution in a model prokaryote and the gut microbiome. BMC Genomics 2016;17(1):1–14.
[10] Wicke L, Ponath F, Coppons L, Gerovac M, Lavigne R, Vogel J. Introducing differential RNA-seq mapping to track the early infection phase for Pseudomonas phage φKZ. RNA Biol 2021;18(8):1099–110.
[11] Sharma CM, Vogel J. Differential RNA-seq: The approach behind and the biological insight gained. Curr Opin Microbiol 2014;19(1):97–105.
[12] Stark R, Grzelak M, Hadfield J. RNA sequencing: the teenage years. Nat Rev Genet 2019;20(11):631–56.
[13] Byrne A, Cole C, Volden R, Vollmers C. Realizing the potential of full-length transcriptome sequencing. Philos Trans R Soc B Biol Sci 2019;374(1786).
[14] Ju X, Li D, Liu S. Full-length RNA profiling reveals pervasive bidirectional transcription terminators in bacteria. Nat Microbiol 2019;4(11):1907–18.
[15] Yan B, Boitano M, Clark TA, Ettwiller L. SMRT-Cappable-seq reveals complex operon variants in bacteria. Nat Commun 2018;9(11).
[16] Grüninger F, Ferreira-Cerca S, Grohmann D. Nanopore sequencing of RNA and cDNA molecules in Escherichia coli. RNA. 2021. 1:rna.078937.121.
Pust M-M, Davenport CF, Wiehlmann L, Burkhard T. Direct RNA nanopore sequencing of *Pseudomonas aeruginosa* clone C transcriptomes. *J Bacteriol* 2020;182(7):2391–5.

Smith AM, Jain M, Mulroney L, Gararde DL, Akeston M. Reading canonical and modified nucleotides in 16S ribosomal RNA using nanopore direct RNA sequencing. *PLoS ONE* 2017;12(5):1–15.

Pust M-M, Davenport CF, Wiehlmann L, Burkhard T. Direct RNA nanopore sequencing of *Pseudomonas aeruginosa* clone C transcriptomes. *J Bacteriol* 2020;182(7):2391–5.

Smith AM, Jain M, Mulroney L, Gararde DL, Akeston M. Reading canonical and modified nucleotides in 16S ribosomal RNA using nanopore direct RNA sequencing. *PLoS ONE* 2017;12(5):1–15.

Pust M-M, Davenport CF, Wiehlmann L, Burkhard T. Direct RNA nanopore sequencing of *Pseudomonas aeruginosa* clone C transcriptomes. *J Bacteriol* 2020;182(7):2391–5.

Smith AM, Jain M, Mulroney L, Gararde DL, Akeston M. Reading canonical and modified nucleotides in 16S ribosomal RNA using nanopore direct RNA sequencing. *PLoS ONE* 2017;12(5):1–15.

Pust M-M, Davenport CF, Wiehlmann L, Burkhard T. Direct RNA nanopore sequencing of *Pseudomonas aeruginosa* clone C transcriptomes. *J Bacteriol* 2020;182(7):2391–5.

Smith AM, Jain M, Mulroney L, Gararde DL, Akeston M. Reading canonical and modified nucleotides in 16S ribosomal RNA using nanopore direct RNA sequencing. *PLoS ONE* 2017;12(5):1–15.