Transcriptome and translatome profiles of *Streptomyces* species in different growth phases

Woori Kim1,6, Soonkyu Hwang1,6, Namil Lee1,6, Yongjae Lee1, Suhyung Cho1, Bernhard Palsson2,3,4 & Byung-Kwan Cho1,4,5 ✉

*Streptomyces* are efficient producers of various bioactive compounds, which are mostly synthesized by their secondary metabolite biosynthetic gene clusters (smBGCs). The smBGCs are tightly controlled by complex regulatory systems at transcriptional and translational levels to effectively utilize precursors that are supplied by primary metabolism. Thus, dynamic changes in gene expression in response to cellular status at both the transcriptional and translational levels should be elucidated to directly reflect protein levels, rapid downstream responses, and cellular energy costs. In this study, RNA-Seq and ribosome profiling were performed for five industrially important *Streptomyces* species at different growth phases, for the deep sequencing of total mRNA, and only those mRNA fragments that are protected by translating ribosomes, respectively. Herein, 12.0 to 763.8 million raw reads were sufficiently obtained with high quality of more than 80% for the Phred score Q30 and high reproducibility. These data provide a comprehensive understanding of the transcriptional and translational landscape across the *Streptomyces* species and contribute to facilitating the rational engineering of secondary metabolite production.

**Background & Summary**

*Streptomyces*, which comprise the largest genus of Actinobacteria, are huge natural reservoir of secondary metabolites, including antibiotics, immuno-suppressants, and other medicinal compounds1–6. Recent advancements in high-throughput sequencing have led to the development of the genome mining approach, which implicates that the genome of each *Streptomyces* species has more than 30 secondary metabolite biosynthetic gene clusters (smBGCs) with potential to produce various unexplored secondary metabolites2. These secondary metabolites are synthesized by a series of enzymatic reactions, which depend on the supply of precursor molecules from primary metabolism, such as acetyl-coenzyme A and amino acids7. After active growth terminates, an overall metabolic transition occurs, which leads to the activation of secondary metabolite production8,9; this metabolic transition from primary to secondary metabolism is governed by multi-layered regulatory mechanisms at transcriptional, translational, and post-translational levels10,11. Thus, understanding the complex regulatory systems of the metabolic transition is important to enhance secondary metabolite production. The overall metabolic transition encompasses diverse genome-wide gene expression changes, which are regulated by signaling cascades from the pleiotropic regulators to pathway-specific regulators8,10,12,13. To understand the underlying molecular mechanisms of metabolic transitions, transcriptional changes that occur between growth phases have been studied14,15. For example, the time-series transcriptome analysis of *Streptomyces coelicolor* demonstrated that coherent genes that are involved in specific metabolism and their regulatory genes exhibit similar expression patterns during metabolic transitions; this suggests that primary metabolism-related genes are functionally connected to the smBGCs through regulatory gene expression. Based on this suggestion, putative regulatory genes and their interconnected networks could be identified by screening genes that have similar expression patterns15.

1Department of Biological Sciences and Ki for the BioCentury, Korea Advanced Institute of Science and Technology, Daejeon, 34141, Republic of Korea. 2Department of Bioengineering, University of California San Diego, La Jolla, CA, 92093, USA. 3Department of Pediatrics, University of California San Diego, La Jolla, CA, 92093, USA. 4Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Lyngby, 2800, Denmark. 5Intelligent Synthetic Biology Center, Daejeon, 34141, Republic of Korea. 6These authors contributed equally: Woori Kim, Soonkyu Hwang, Namil Lee. ✉e-mail: bcho@kaist.ac.kr
Bacteria can fine-tune gene expression both at the transcriptional and translational levels. For example, Escherichia coli proteome analysis revealed that only approximately half of protein abundance is determined by transcriptional regulation, which indicates the existence of various post-transcriptional regulation. In this regard, deciphering translational dynamics is important for understanding post-transcriptional regulations that are closely related to cellular protein levels. Recently, ribosome profiling has been used to measure translational levels by deep sequencing of the ribosome-protected mRNA fragments (RPFs) at the position of the translating ribosome. Several ribosome profiling studies in Streptomyces have been reported by our research group for S. coelicolor, S. clavuligerus, and S. lividans, which revealed translational buffering of secondary metabolism-related genes at a later growth phase and that translational abundance is more consistently maintained than transcript abundance. Translational regulations are advantageous for the tight control of secondary metabolite biosynthesis, as translation requires the highest energy costs among all cellular reactions. Moreover, the expression of smBGC-associated genes can be more rapidly regulated at the translational level than at the transcriptional level in response to dynamic environmental changes. Given the dynamic relationship between transcription and translation, as exhibited by translational buffering, integrative analysis at both levels should unravel complex regulations in Streptomyces. However, transcriptomic and translatomic data have covered only a small portion of approximately 350 reported Streptomyces genomes, which have not been systematically validated at the multi-species level.

In this study, we provide RNA-Seq and ribosome profiling data of five Streptomyces species at four different growth phases, followed by validation of the read quality. The species were S. avermitilis MA-4680, S. clavuligerus ATCC27064, S. lividans TK24, S. venezuelae ATCC15439, and S. tsukubaensis NRRL 18488, which are industrial strains that produce antifungal avermectin, β-lactamase inhibitor clavulanic acid, and immunosuppressant FK506, respectively. S. lividans and S. venezuelae were characterized by fast growth and ease of genetic manipulation, and have been employed as heterologous expression hosts. An overview of the preparation of transcriptomic and translatomic data is illustrated in Fig. 1. A total of 12 to 83.5 million raw reads for RNA-Seq and 113 to 763.8 million raw reads for ribosome profiling were obtained. Although the RNA-Seq and ribosome profiling data of two species (S. clavuligerus and S. lividans) among the five species were already reported in previous studies by our research group, this study provided a uniformly processed and mapped dataset of all five species. This facilitates the efficiency of the comparative transcriptome and translome analysis at multi-time points between multi-species. Further, understanding the transcriptional and translational regulatory mechanisms and developing regulatory synthetic parts, such as promoters, ribosome-binding sequences, 5′ untranslated regions, and terminators from the dataset allows rational genome engineering for efficient secondary metabolite production by Streptomyces.

Methods

Strains and cell growth. Streptomyces strains were inoculated from their 20% glycerol stock of spores into 50 mL of R5—liquid medium with 8 g of glass beads (3 ± 0.3 mm diameter) in a 250 mL baffled flask, grown at 30 °C, and pre-cultured at 250 rpm. The R5—liquid medium consists of 103 g/L sucrose, 0.25 g/L K2SO4, 10.12 g/L MgCl2·6H2O, 10 g/L−1 glucose, 0.1 g/L−1 casamino acids, 5 g/L−1 yeast extract, 5.73 g/L−1 TES (pH 7.2), 0.08 mg/L−1 ZnCl2, 0.4 mg/L−1 FeCl3·6H2O, 0.02 mg/L−1 CuCl2·2H2O, 0.02 mg/L−1 MnCl2·4H2O, 0.02 mg/L−1 Na2MoO4·10H2O, 0.02 mg/L−1 (NH4)6Mo7O24·4H2O, and 0.28 g/L−1 NaOH. The growth mycelium was inoculated to fresh R5—medium with an initial optical density of 0.05 at 600 nm for the main culture as biological duplicates and grown under the previously mentioned conditions. The cells were sampled at four different time points based on the growth profile of each strain, as follows: early-exponential (E), transition (T), late-exponential (L), and stationary (S) phases. The E, T, L, and S time points were 13, 17, 19.5, and 33.5 h for S. clavuligerus and grown under the previously mentioned conditions. The cultures were then incubated for 5 min at 4 °C for 10 min at 16,000 × g. The supernatant was collected and stored at −80 °C. Following the preparation of lysates from four growth phases as biological duplicates, the lysates were mixed with a solution of phenol:chloroform:isoamyl alcohol (25:24:1, v/v), and the mixtures were separated by centrifugation. DNA in the extracted RNA samples was removed by treatment with 2 μL RNase Inhibitor (Thermo Scientific, Waltham, MA, USA). Lastly, the DNase I-treated RNA samples were purified using phenol:chloroform:isoamyl alcohol (25:24:1, v/v) and ethanol precipitation. To eliminate rRNAs in the recovered RNA samples, the Ribo-Zero rRNA Removal Kit for Bacteria (Epicentre, Madison, WI, USA) was used according to the manufacturer’s instructions. The quality of rRNA-depleted RNA samples was checked using 2% agarose gel electrophoresis. The suitable RNA samples were then used to construct RNA sequencing libraries using the TruSeq Stranded mRNA Library Prep Kit (Illumina, San Diego, CA, USA). The size distributions of the final libraries were checked using the Agilent 2200 TapeStation System (Agilent, Santa Clara, CA, USA). The constructed libraries were sequenced on the HiSeq. 2500 platform using either a 100-bp (S. lividans, S. avermitilis, S. clavuligerus, and S. venezuelae) or 50-bp (S. tsukubaensis) single-end read recipe (Fig. 1a).

RNA-Seq library preparation and high-throughput sequencing. The overview of the library construction of RNA-Seq is illustrated in Fig. 1a. The harvested cells were washed with polysome buffer (20 mM Tris–HCl, pH 7.5; 140 mM NaCl and 5 mM MgCl2), and then resuspended with 500 μL lysis buffer (0.3 M sodium acetate, pH 5.2; 10 mM ethylenediaminetetraacetic acid and 1% Triton X-100). The resuspended cells were frozen with liquid nitrogen and ground using a mortar and pestle. The ground mycelium was thawed and centrifuged at 4 °C for 10 min at 16,000 × g. The supernatant was collected and stored at −80 °C. Following the preparation of lysates from four growth phases as biological duplicates, the lysates were mixed with a solution of phenol:chloroform:isoamyl alcohol (25:24:1, v/v), and the mixtures were separated by centrifugation. DNA in the extracted RNA samples was removed by treatment with 2 μL DNase I (NEB, Ipswich, MA, USA), 5 μL 10 × DNase I buffer, and 1 μL SUPERase-In RNase Inhibitor (Thermo Scientific, Waltham, MA, USA). Lastly, the DNase I-treated RNA samples were purified using phenol:chloroform:isoamyl alcohol (25:24:1, v/v) and ethanol precipitation. To eliminate rRNAs in the recovered RNA samples, the Ribo-Zero rRNA Removal Kit for Bacteria (Epicentre, Madison, WI, USA) was used according to the manufacturer’s instructions. The quality of rRNA-depleted RNA samples was checked using 2% agarose gel electrophoresis. The suitable RNA samples were then used to construct RNA sequencing libraries using the TruSeq Stranded mRNA Library Prep Kit (Illumina, San Diego, CA, USA). The size distributions of the final libraries were checked using the Agilent 2200 TapeStation System (Agilent, Santa Clara, CA, USA). The constructed libraries were sequenced on the HiSeq. 2500 platform using either a 100-bp (S. lividans, S. avermitilis, S. clavuligerus, and S. venezuelae) or 50-bp (S. tsukubaensis) single-end read recipe (Fig. 1a).
Data processing of RNA-Seq reads. Raw FASTQ files were processed using the CLC Genomics Workbench (CLC Bio, Aarhus, Denmark). Raw reads were trimmed by their overall quality (score: 0.05; maximum ambiguous nucleotides: 2) and length (minimum length: 15 nucleotides). The filtered reads were mapped
to each reference genome sequence with the default parameters (mismatch cost: 2; insertion cost: 2; deletion cost: 3; length fraction: 0.9; similarity fraction: 0.9; and ignore non-specific matches). The accession number of each reference genome is as follows: S. avermitilis MA-4680 (NC_003155), S. clavuligerus ATCC27064 (chromosome NZ_CP027858, plasmid NZ_CP027859), S. lividans TK24 (NZ_CP009124), S. venezuelae ATCC15439 (NZ_CP013129), and S. tsukubaensis NRRL18488 (chromosome CP027063, plasmid CP027071, and CP027072). The statistics pertaining to quality trimming and reference mapping are summarized in Table 1. The number of uniquely mapped reads to each gene were counted using the RNA-Seq analysis tool in the CLC Genomics Workbench and the read counts were normalized using the DESeq2 package in R33.

### Ribosome profiling library preparation and high-throughput sequencing

An overview on the library construction of ribosome profiling is illustrated in Fig. 1a. The mycelium that was treated by thiostrepton was collected by centrifugation at 4°C for 10 min at 3,000 × g, and the cell pellet was washed with 2 mL of polysome buffer that was composed of 20 mM Tris–HCl (pH 7.5), 140 mM NaCl, and 5 mM MgCl2 with 20 μM thiostrepton. The washed pellet was re-suspended in 1 mL of lysis buffer composed of 950 μL of polysome buffer

---

### Table 1. Overall statistics of RNA-Seq data.

| Species          | Growth phase | Number of raw reads | Average length (bp) | Number of trimmed reads | Percentage trimmed | Trimmed reads length (bp) | Number of randomly mapped reads | Number of uniquely mapped reads | Percentage of uniquely mapped reads (%) | Raw read FASTQ accession |
|------------------|--------------|---------------------|---------------------|------------------------|--------------------|--------------------------|-------------------------------|-----------------------------------|----------------------------------------|--------------------------|
| S. avermitilis   | E1           | 35,222,700          | 101                 | 15,222,324             | 100.00             | 100.9                    | 14,743,001                    | 14,475,094                        | 95.09                                  | SRP158023                |
| MA-4680          | E2           | 35,540,304          | 101                 | 15,339,540             | 100.00             | 100.8                    | 14,842,965                    | 14,232,752                        | 91.59                                  |                         |
| S. clavuligerus  | T1           | 18,962,695          | 101                 | 18,961,958             | 100.00             | 100.8                    | 17,584,931                    | 16,820,053                        | 88.70                                  |                         |
| ATCC 27064       | T2           | 18,054,983          | 101                 | 18,054,302             | 100.00             | 100.9                    | 16,337,750                    | 14,948,455                        | 82.80                                  |                         |
| S. livesdans     | L1           | 13,904,005          | 101                 | 13,903,462             | 100.00             | 100.9                    | 12,858,182                    | 12,238,050                        | 88.02                                  |                         |
| TK24             | L2           | 16,814,305          | 101                 | 16,813,651             | 100.00             | 100.8                    | 15,778,544                    | 15,212,127                        | 90.47                                  |                         |
| S. tsukubaensis  | S1           | 16,662,552          | 101                 | 16,661,924             | 100.00             | 100.8                    | 15,627,234                    | 14,761,494                        | 88.59                                  |                         |
| NRRL18488        | S2           | 16,278,766          | 101                 | 16,278,123             | 100.00             | 100.9                    | 14,643,706                    | 12,519,514                        | 76.91                                  |                         |

---

**www.nature.com/scientificdata/**
and 50 μL of 20% Triton X-100 with 20 μM thioistrepton. The resuspended cells were dripped into a mortar filled with liquid nitrogen and then grounded with a pestle. The cell debris was removed by centrifugation at 4°C for 5 min at 3,000 × g. The supernatant was further clarified and collected by centrifugation at 4°C for 10 min at 16,000 × g. To digest RNA in the lysate (containing 50 μg RNA), the S. avermitilis and S. tsukubaensis samples were treated with 750 U of RNase I (Invitrogen, Waltham, MA, USA) at 37°C for 45 min, and the remaining strains were treated with 400 U of Micrococcal Nuclease (MNase) (NEB), 20 μL of 10× MNase buffer, and 2 μL of 100× Bovine Serum Albumin (BSA) (NEB) at 37°C for 2 h. The samples were then loaded onto Illustra MicroSpin S-400 HR Columns (GE Healthcare Life Sciences, Marlborough, MA, USA) that were previously washed three times with 500 μL of washing buffer, which was composed of 50 mM Tris-HCl (pH 8.0), 250 mM NaCl, 50 mM MgCl2, 25 mM EGTA, and 1% Triton X-100. The column was centrifuged at 4°C for 2 min at 400 × g, and the flow-through was further purified by a phenol-chloroform-isoamyl alcohol extraction and ethanol precipitation. rRNA was depleted with the Ribo-Zero rRNA Removal Kit (Epigence) according to the manufacturer's instructions. The ribosome-protected RNA fragments (RPF) of between 26 and 34 bp were separated by electrophoresis for 65 min at 200 V using 15% polyacrylamide TBE-urea gel (Invitrogen), and eluted in 400 μL of RNA gel extraction buffer, which was composed of 300 mM sodium acetate pH 5.5, 1 mM EDTA, and 0.25% (w/v) SDS. The samples were frozen for 30 min at −80°C and then incubated at 37°C for 4 h. The eluted RNAs were isolated by ethanol precipitation and purified once again with the RNasey MinElute Column (Qiagen, Hilden, Germany) using the manufacturer's protocol. The enriched RPFs were then denatured for 90 s at 80°C and incubated for 1 h at 37°C with 5 μL of 10× T4 Polynucleotide Kinase (PNK) buffer (NEB), 20 μL of SUPERase-In RNase Inhibitor, and 10 μL of T4 PNK (NEB) to dephosphorylate the 3’ end. The dephosphorylated RNAs were purified using the RNasey MinElute Column (Qiagen). The sequencing library was constructed from the end-repaired RPFs using the NEBNext Multiplex Small RNA Library Prep Set for Illumina (NEB) according to the manufacturer's instructions. The final library of approximately 150–160 bp was size-selected by gel electrophoresis for 90 min at 100 V using a 2% agarose gel that was dyed with SYBR Gold Nucleic Acid Gel Stain (Bio-Rad, Hercules, CA, USA). The concentration of the final library was measured using a Qubit 2.0 Fluorometer (Invitrogen) and the Qubit dsDNA HS Kit. The size distribution was assessed using the Agilent 2200 TapeStation System (Agilent). The constructed library was sequenced on the Illumina HiSeq. 2500 platform using the 50-bp single-end read recipe (Fig. 1b).

Data processing of ribosome profiling reads. The libraries of seven samples of S. avermitilis—except for the E1 sample—and six samples of S. venezuelae—except for the T2 and S2 samples—were prepared and sequenced twice to increase the output and merged for further data processing (Table 2). The sequencing results were de-multiplexed and processed by CLC Genomics Workbench (CLC Bio). A total of 113,065,267 to 763,831,282 raw reads were generated for each replicate and were exported in the FASTQ format for the data upload. The reads were then mapped to the PhiX control sequences (NCBI Genbank accession number: NC_001422) to eliminate the PhiX control reads with the following parameters: mismatch cost: 2; insertion cost: 3; deletion cost: 3; length fraction: 0.9; similarity fraction: 0.9; and non-specific matches were randomly mapped. A total of 112,376,633 to 661,109,040 reads were unmapped. As these reads were sequenced from the 5’ end of the enriched RPF to 50 bp downstream, which is longer than the size-selected RPF (26 to 34 bp), the 5’ end sequences of the 3’ adapter sequence of the NEBNext Multiplex Small RNA Library Prep Set for Illumina (NEB) were also included. To remove the adapter sequences from the reads prior to mapping, the sequences were trimmed by the following parameters: action: remove adapter; strand: minus; mismatch cost: 2; gap cost: 3; internal match minimum score: 3; and end match minimum score: 3. Ultimately, the removed adapter sequence was 5’−ATACGAGATNNNNNNGGGTAGGGGATTGAGCTACGACGTGTGTCCTTCCGATTT−3’, in which the NNNNNN sequences were CACTGT, ATTTGC, TACAAG, and TTTCA for index 5, 6, 12, and 19, respectively. The reads were additionally trimmed based on their overall quality (score: 0.05, maximum ambiguous nucleotides: 2) and length (>15 bp). The trimming steps yielded 90.68 to 98.47% of the PhiX control unmapped reads. A total of 84,947,464 to 590,644,871 reads with an average read length of 25.8 to 33.7 bp were mapped with random mapping of non-specific matches, while 1,833,155 to 103,819,037 reads with an average read length of 25.8 to 33.1 bp were mapped with ignored mapping of non-specific matches (mismatch cost: 2; insertion cost: 3; deletion cost: 3; length fraction: 0.9; similarity cost: 0.9). The overall statistics of the data processing are summarized in Table 2. The mapped information was exported in a BAM file format, and the number of mapped reads at each genomic position was counted as the read count. Normalized read value and principal component analysis (PCA) plots were generated using the DESeq. 2 package in R34.

Data Records
Raw read FASTQ files, trimmed read FASTQ files, mapped read BAM files, and the gene expression text files of all samples were uploaded to the public databases (Tables 1 and 2). Raw read FASTQ files of RNA-Seq and ribosome profiling of three species (S. avermitilis, S. clavuligerus, S. tsukubaensis) were deposited at the National Center for Biotechnology Information Sequence Read Archive (NCBI SRA)35–37. Raw read FASTQ files of RNA-Seq and ribosome profiling of S. lividans were deposited at the European Nucleotide Archive (ENA)38. Raw read FASTQ files of RNA-Seq of S. venezuelae were deposited at the ENA39. Raw read FASTQ files of ribosome profiling of S. venezuelae were deposited at the NCBI SRA40–47. Trimpread FASTQ files and mapped read BAM files of the raw read FASTQ files in the NCBI SRA (RNA-Seq and ribosome profiling data of S. avermitilis, S. clavuligerus, S. tsukubaensis, and ribosome profiling data of S. venezuelae) were deposited at the ENA with a new accession48. Trimpread FASTQ files and mapped read BAM files of the raw read FASTQ files in the ENA (RNA-Seq and ribosome profiling data of S. lividans, and RNA-Seq data of S. venezuelae) were also deposited at the ENA with the same accession as each corresponding raw read FASTQ file38,39. The gene expression profile as raw read counts of
RNA-Seq and ribosome profiling data of *S. avermitilis*, *S. clavuligerus*, *S. tsukubaensis*, and *S. lividans*, are available in a text file format in the Gene Expression Omnibus (GEO) database. Also, the gene expression profiles of all datasets (RNA-Seq and ribosome profiling of the five species), including raw read counts, DESeq 2 normalized values, fold change values between growth phases, and p-values for the fold changes, are available in a text file format in the Figshare. The raw read FASTQ data of *S. clavuligerus* in NCBI SRA was published in the previous study. Also, the raw read FASTQ data of *S. lividans* in ENA was published in the previous study. Note that the ribosome profiling data of *Streptomyces griseus* was uploaded under the same accession with those of *S. venezuelae*, but they are not described in this study.

### Technical Validation

**RNA-Seq read quality validation.** A total of 40 RNA-Seq runs that were applied to five species at four time points as duplicates yielded on average 16,430,039 reads (*S. avermitilis*), 13,961,155 reads (*S. clavuligerus*), 15,686,025 reads (*S. lividans*), 12,899,493 reads (*S. venezuelae*), and 51,144,650 reads (*S. tsukubaensis*). After trimming the sequencing reads by quality score and nucleotide length, more than 99.8% of the sequencing reads.
remained, which indicated high-sequencing quality. The remaining reads were used as input to generate sequencing QC reports in the CLC Genomics Workbench to validate the quality of the reads. At first, the overall read lengths were extremely long, corresponding to the sequencing read recipe (Fig. 2a, Table 1). For the four species that were sequenced with the 100-bp read recipe, the percentage of read lengths that were over 100bp was more than 97.9%, and for S. tsukubaensis, which was sequenced with the 50-bp read recipe, the percentage of read lengths that were over 50 bp was more than 93.8%. Further, more than 98.6% (S. avermitilis), 98.9% (S. clavuligerus), 98.9% (S. lividans), 98.9% (S. venezuelae), and 96.4% (S. tsukubaensis) of the total reads exhibited an average Phred score of greater than 30, which indicates 99.9% base call accuracy (Fig. 2b). In addition, the quality of each base of the obtained reads was examined. The overall base positions of the sequencing reads were highly covered, and even the lowest average values of the coverage were 97.5% (S. avermitilis), 97.6% (S. clavuligerus), 97.8% (S. lividans), 98.3% (S. venezuelae), and 95.4% (S. tsukubaensis) at the last position, respectively (Fig. 2c). Moreover, the median values of the Phred scores per base position of the reads were consistently high across reads, with 40 scores in four species and 38 scores in S. tsukubaensis (Fig. 2d). From these quality validation results, we validated the quality of all obtained RNA sequencing reads for subsequent analysis.

Assessment of transcriptome data. The qualified reads were mapped to each reference genome with a uniquely mapped percentage that ranges from 76.91% to 95.09% (S. avermitilis), 52.19% to 90.88% (S. clavuligerus), 60.15% to 87.08% (S. lividans), 67.35% to 86.23% (S. venezuelae), and 76.33% to 96.26% (S. tsukubaensis) (Table 1). The number of uniquely mapped reads at each gene was counted and normalized using the DESeq.2 package in R33 to reduce variation between samples. Using the normalized values, principal component analysis (PCA) was performed, which validated the high reproducibility of the sequencing data (Fig. 2e). Using the distribution of log (DESeq normalized value + 1) broadly ranged from 0 to 20 in the different growth phase samples (Fig. 2f).

Ribosome profiling read quality validation. A total of 40 ribosome profiling reads were obtained from five Streptomyces species at four time points as duplicates. Unlike the RNA-Seq data, the trimmed reads were considered as raw sequences of the enriched RPF sequences, as additional PhiX control and adapter sequences that were involved in the ribosome profiling steps must be removed (Fig. 1a,b). Since the RPF fragments were selected by size, ranging from 26 to 34 bp, the 3′ end of the total 50 bp sequencing read contained non-RPF sequences, such as the 3′ adapter sequences, which do not represent the quality of the RPF reads. Thus, the QC reports on the trimmed reads were exported from CLC Genomics Workbench (Qiagen) to assess the quality of the RPF reads. The read length distribution exhibited a broad range from 20 to 40 bp, with one or two enriched peaks (Fig. 3a). The enriched peak sizes were comparable to the monosome-protected sizes, and they varied for different species, while they were more conserved for different growth phase samples of the same species. The differences in RNA degradation efficiency of RNase I or MNorth across species may be the primary reason for the observed size differences46. Further, the read quality that was measured by the average Phred scores was generally high in all samples of the five species; the quality of more than 94% of the reads was higher than Q20, and more than 80% were higher than Q30 (Fig. 3b). Both per-sequence and per-base analyses of the read quality were observed. As the read size ranged mostly between 25 and 35 bp after adapter trimming, the base number coverage at each position of the 50 bp read dramatically decreased at the 3′ end (Fig. 3c). In terms of species, most of them exhibited the highest decline at 28 to 30 bp, which was consistent with the read length distribution (Fig. 3a). Given the base coverage, the median Phred score per base was demonstrated to be from 1 to 35 bp (Fig. 3d). The overall median of the quality score was approximately Q38, while the median score at the 5′ end was slightly lower than that of the middle section, and the score at the 3′ end of select species showed dramatic reductions. The low quality at the 3′ end may be due to some portions of identical long reads, which were somehow enriched during the size selection step of library construction, which stimulates wrong base calling. For the S. clavuligerus E2 sample, enriched peaks of less than 20 bp in length for approximately 10% of the total reads were unexpectedly found, along with decreased coverage at 15 bp, but these peaks did not seem to affect the overall read quality (Figs. 3a,c,d). Overall, most of the reads were shorter than 35 bp, and the read quality of all samples was high and suitable for downstream analyses.

Assessment of translatome data. To examine the additional quality of the reads for the translational abundance of each gene, the trimmed reads were mapped to their corresponding genome. Based on the mapping parameter, some reads would be non-specifically aligned to more than one genomic position due to highly repetitive genomic regions, including rRNA genes. Approximately, 43.1 to 590.6 million reads (75.3 to 96.8% of the trimmed reads) were mapped when the non-specifically matched reads were randomly assigned to one of the mapped positions, while 1.8 to 103.9 million reads (1.3 to 54.9% of the reads) were uniquely mapped when the non-specifically matched reads were excluded (Table 2). These results suggest that the non-specifically matched reads were generally more than half of the total mapped reads. Further, the proportion of these reads varied in different samples even within the same species, which is because the RNA was enriched during the monosome recovery step, and the efficiency of RNA removal differed across samples96. S. clavuligerus showed the highest uniquely mapped read number and ratio among five species, with an average of 82.4 million reads (47% of the trimmed reads). The S. clavuligerus E2 sample showed a relatively lower mapped number (61.3 million reads, 30.6% of the trimmed reads) compared to other S. clavuligerus samples. S. venezuelae showed 34.2 to 69 million mapped reads (average 13.9% of the trimmed reads), respectively. The two early samples of S. lividans showed 97.2 and 81.1 million mapped reads (46 and 43.8% of the trimmed reads), respectively, while other samples showed low numbers (9.7 to 24.8 million mapped reads, 5.5 to 15.4% of the trimmed reads). S. tsukubaensis showed various ranges for the mapping read number; L2 and S2 samples showed 56 and 103 million mapped reads (35 and 53.9% of the trimmed reads), respectively, while other samples showed a lower number of mapped
reads (1.9 to 8.8 million mapped reads, 1.5 to 6.9% of the trimmed reads). *S. avermitilis* showed the lowest number of mapped reads and ratio among the five species, with 2 to 13.1 million mapped reads (1.3 to 5.9% of the trimmed reads). Although the minimum mapped read number among the samples was 1.8 million, the numbers

---

**Fig. 2** Read quality analysis of RNA-Seq samples of five *Streptomyces* species at four growth phases. The replicate of each growth phase is represented as “1” or “2” after the growth phase. (a) Read length distribution of trimmed reads. (b) Distribution of average Phred scores of the trimmed reads. (c) The number of sequences that cover individual base positions normalized to the total number of sequences at each base position. (d) The distribution of the median Phred quality scores that were observed at each base position. (e) PCA plot of RNA-Seq mapped reads of each gene. (f) Violin and box plot of the log2 normalized expression values.
obtained are, based on several bacterial transcriptome studies, considered sufficient for analysis of the whole translational profile and differential expression levels of genes, as 1 to 5 million reads are suggested for high statistical significance. Among the uniquely mapped reads, some reads were mapped within RNA genes, rather
than protein-coding genes, which mostly corresponded to tRNA genes. These reads may be the fragments of tRNA and rRNA that were bound to the ribosome and then enriched during monosome recovery\(^2\). Therefore, further validation was performed using only the mapped reads of the protein-coding genes. A total of 1.1 to 19.5 million reads were mapped to protein-coding genes, which were 4.3 to 81.0% of the uniquely mapped genes, which indicates a high ratio of tRNA gene-mapped reads (Table 2). To validate the mapped read quality, the reproducibility of the mapped read number among biological replicates was investigated by PCA. All replicates were found to exhibit high reproducibility (Fig. 3e). The mapped read quality for quantitative analysis, such as the differential expressibility of the mapped read number among biological replicates was investigated by PCA. All replicates were found to exhibit high reproducibility (Fig. 3f). In conclusion, the mapped reads were confirmed to exhibit high quality in terms of sequencing depth, reproducibility, and translational abundance.

### Code availability

Versions and parameters of all the bioinformatic tools that were used in this work are described in the “Methods” section.

**Received:** 3 January 2020; **Accepted:** 2 April 2020; **Published:** 8 May 2020

**References**

1. Flardh, K. & Rutten, M. J. *Streptomyces* morphogenetics: dissecting differentiation in a filamentous bacterium. *Nat Rev Microbiol* 7, 36–49, https://doi.org/10.1038/nrmicro1968 (2009).

2. Hwang, K. S., Kim, H. U., Charusanti, P., Palsson, B. O. & Lee, S. Y. Systems biology and biotechnology of *Streptomyces* species for the production of secondary metabolites. *Biotechnol. Adv.* 32, 255–269, https://doi.org/10.1016/j.biotechadv.2013.10.008 (2014).

3. Procopio, R. E., Silva, I. R., Martins, M. K., Azevedo, J. L. & Araujo, J. M. Antibiotics produced by *Streptomyces*. *Braz. J. Infect. Dis* 16, 466–471, https://doi.org/10.1016/j.bjid.2012.08.014 (2012).

4. Lee, N. et al. Synthetic biology tools for novel secondary metabolite discovery in *Streptomyces*. *J Microbiol Biotechnol* 29, 667–686, https://doi.org/10.1014/jmb.1904.04015 (2019).

5. Wörthen, D. B. *Streptomyces* in nature and medicine: The antibiotic makers. *Journal of the History of Medicine and Allied Sciences* 63, 273–274, https://doi.org/10.1093/jhmas/jrt016 (2008).

6. Demain, A. L. Importance of microbial natural products and the need to revitalize their discovery. *J Ind Microbiol Biotechnol* 41, 185–201, https://doi.org/10.1007/s10295-013-1325-z (2014).

7. Hodgson, D. A. Primary metabolism and its control in *Streptomyces*: A most unusual group of bacteria. *Adv Microb Physiol* 42, 47–238, https://doi.org/10.1016/s0065-2911(00)42003-5 (2000).

8. Alam, M. T. et al. Metabolic modeling and analysis of the metabolic switch in *Streptomyces coelicolor*. *BMC Genomics* 11, 202 (2010).

9. Rokem, J. S., Lantz, A. E. & Nielsen, J. Systems biology of antibiotic production by microorganisms. *Nat Prod Rep* 24, 1262–1287, https://doi.org/10.1039/b617765b (2007).

10. Bibb, M. The regulation of antibiotic production in *Streptomyces coelicolor* A3(2). *Microbiology* 142, 1335–1344, https://doi.org/10.1099/mic.0.053887-142.6-1335 (1996).

11. Jeong, Y. et al. The dynamic transcriptional and translational landscape of the model antibiotic producer *Streptomyces coelicolor* A3(2). *Nat Commum* 7, 11605, https://doi.org/10.1038/ncomms11605 (2016).

12. Wentzel, A. et al. Optimized submerged batch fermentation strategy for systems scale studies of metabolic switching in *Streptomyces coelicolor* A3(2). *BMC Syst Biol* 6, 59, https://doi.org/10.1186/1752-0509-6-59 (2012).

13. Nieselt, K. et al. The dynamic architecture of the metabolic switch in *Streptomyces coelicolor*. *BMC Genomics* 11, 10, https://doi.org/10.1186/1471-2164-11-10 (2010).

14. Huang, J. Q., Li, C. J., Pan, K. H. & Cohen, S. N. Global analysis of growth phase responsive gene expression and regulation of antibiotic biosynthetic pathways in *Streptomyces coelicolor* using DNA microarrays. *Gene* 15, 3183–3192 (2001).

15. Chen, L. et al. Transcriptomics analyses reveal global roles of the regulator AveI in *Streptomyces avermitilis*. *FEBS Microbiol Lett* 298, 199–207 (2009).

16. Berghoff, R. A. et al. Integrative “omics”-approach discovers dynamic and regulatory features of bacterial stress responses. *Plos Genet* 9, e1003576, https://doi.org/10.1371/journal.pgen.1003576 (2013).

17. Waters, L. S. & Storz, G. Regulatory RNAs in bacteria. *Cell* 136, 615–628, https://doi.org/10.1016/j.cell.2009.01.043 (2009).

18. Lu, P., Vogel, C., Wang, R., Yao, X. & Marckote, E. M. Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation. *Nat Biotechnol* 25, 117–124, https://doi.org/10.1038/nbt1270 (2007).

19. Brar, G. A. & Weissman, J. S. Ribosome profiling reveals what, when, where and how of protein synthesis. *Nat Rev Mol Cell Biol* 16, 651–664, https://doi.org/10.1038/nrmicro4069 (2015).

20. Ingolia, N. T., Ghaemmaghami, S., Newman, J. R. & Weissman, J. S. Genome-wide analysis in *vivo* of translation with nucleotide resolution using ribosome profiling. *Science* 324, 218–223, https://doi.org/10.1126/science.1168978 (2009).

21. Hwang, S. et al. Primary transcriptome and translrome analysis determines transcriptional and translational regulatory elements encoded in the *Streptomyces clavuligerus* genome. *Nucleic Acids Res* 47, 6114–6129, https://doi.org/10.1093/nar/gkz471 (2019).

22. Lee, Y. et al. The transcription unit architecture of *Streptomyces lividans* TK24. *Frontiers in Microbiology* 10, https://doi.org/10.3389/fmicb.2019.02074 (2019).

23. Jones, G. H. RNA degradation and the regulation of antibiotic synthesis in *Streptomyces*. *Future Microbiol* 5, 419–429, https://doi.org/10.21273/fmth.10.14 (2010).

24. Paradkar, A. Clavulanic acid production by *Streptomyces clavuligerus*: biogenesis, regulation and strain improvement. *J Antibiot (Tokyo)* 66, 411–420, https://doi.org/10.1038/ja.2013.26 (2013).

25. Barreiro, C. et al. Draft genome of *Streptomyces tsukubaensis* NRRL 18488, the producer of the clinically important immunosuppressant tacrolimus (FK506). *J Bacteriol* 194, 3756–3757, https://doi.org/10.1128/JB.00692-12 (2012).

26. Hotson, I. K. The avermectins: A new family of antiparasitic agents. *J Am Vet Assoc* 53, 87–90 (1982).

27. Nepal, K. K. & Wang, G. *Streptomyces*: surrogate hosts for the genetic manipulation of biosynthetic gene clusters and production of natural products. *Biotechnol Adv* 37, 1–20, https://doi.org/10.1016/j.biotechadv.2018.10.003 (2019).

28. Myronovskiy, M. & Luzhetskyy, A. Heterologous production of small molecules in the optimized *Streptomyces* hosts. *Nat Prod Rep* 36, 1281–1294, https://doi.org/10.1039/c9np00023b (2019).

29. Jung, W. S. et al. Heterologous expression of tylosin polyketide synthase and production of a hybrid bioactive macrolide in *Streptomyces venezuelae*. *Appl Microbiol Biotechnol* 72, 763–769, https://doi.org/10.1007/s00253-006-0318-5 (2006).
30. Kim, E. J., Yang, I. & Yoon, Y. J. Developing Streptomyces venezuelae as a cell factory for the production of small molecules used in drug discovery. *Archives of Pharmacal Research* **38**, 1606–1616, https://doi.org/10.1007/s12272-015-0638-z (2015).

31. Vecchione, I. J., Alexander, B. Jr. & Sello, J. K. Two distinct major facilitator superfamily drug efflux pumps mediate chloramphenicol resistance in *Streptomyces coelicolor*. *Antimicrob Agents Chemotherapy* **53**, 4673–4677, https://doi.org/10.1128/AAC.00853-09 (2009).

32. Garcia-Dominguez, M., Martin, J. F., Mahoro, B., Demain, A. L. & Lisas, P. Efficient plasmid transformation of the 3-lactam producer *Streptomyces clavuligerus*. *Appl Environ Microbiol* **53**, 1376–1381 (1987).

33. Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol* **11**, R106, https://doi.org/10.1186/gb-2010-11-10-r106 (2010).

34. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq. 2. *Biostatistics* (2014).

35. Calviello, L. & Ohler, U. Beyond read-counts: Ribo-seq data analysis to understand the functions of the transcriptome.

36. Haas, B. J., Chin, M., Nusbaum, C., Birren, B. W. & Livny, J. How deep is deep enough for RNA-Seq profiling of bacterial transcriptomes?

37. Vecchione, J. J., Alexander, B. Jr. & Sello, J. K. Two distinct major facilitator superfamily drug efflux pumps mediate chloramphenicol resistance in *Streptomyces coelicolor*. *Antimicrob Agents Chemotherapy* **53**, 4673–4677, https://doi.org/10.1128/AAC.00853-09 (2009).

38. Garcia-Dominguez, M., Martin, J. F., Mahoro, B., Demain, A. L. & Lisas, P. Efficient plasmid transformation of the 3-lactam producer *Streptomyces clavuligerus*. *Appl Environ Microbiol* **53**, 1376–1381 (1987).

39. Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol* **11**, R106, https://doi.org/10.1186/gb-2010-11-10-r106 (2010).

40. Haas, B. J., Chin, M., Nusbaum, C., Birren, B. W. & Livny, J. How deep is deep enough for RNA-Seq profiling of bacterial transcriptomes?

41. Calviello, L. & Ohler, U. Beyond read-counts: Ribo-seq data analysis to understand the functions of the transcriptome. *Trends Genet* **33**, 728–744, https://doi.org/10.1016/j.tig.2017.08.003 (2017).

42. Diamant, A. & Tuller, T. Estimation of ribosome profiling performance and reproducibility at various levels of resolution. *Biol Direct* **11**, 24, https://doi.org/10.1186/s13082-016-0127-4 (2016).

43. Haas, B. J., Chm, M., Nusbaum, C., Birren, B. W. & Livny, J. How deep is deep enough for RNA-Seq profiling of bacterial transcriptomes? *BMC Genomics* **13**, 734, https://doi.org/10.1186/1471-2164-13-734 (2012).

44. Rey, F. E. et al. Dissecting the in vivo metabolic potential of two human gut acetogens. *J Biol Chem* **285**, 22082–22090, https://doi.org/10.1074/jbc.M110.117713 (2010).

45. Lee, N. & et al. Ribosome pausing at the AT-rich codons regulates the protein expression of secondary metabolite gene clusters in the *Streptomyces tsukubaensis* NRRL 18488. *Gene Expression Omnibus*, https://identifiers.org/GEO:GSE97637 (2019).

46. Kim, W. et al. Transcriptome and translatome profiles of Streptomyces species in different growth phases. *European Nucleotide Archive*, https://identifiers.org/embl:PRJEB34219 (2019).

47. NCBI Sequence Read Archive, https://identifiers.org/ncbi/insdc.sra:SRX6932518 (2020).

48. Lee, N. & et al. Ribosome pausing at the AT-rich codons regulates the protein expression of secondary metabolite gene clusters in the *Streptomyces tsukubaensis* NRRL 18488. *Gene Expression Omnibus*, https://identifiers.org/GEO:GSE97637 (2019).

49. Kim, W. et al. Transcriptome and translatome profiles of Streptomyces species in different growth phases. *Figgshare* https://doi.org/10.6084/m9.figshare.c.4867830 (2020).

50. Hwang, S. et al. Primary transcriptome and translatome analysis determines transcriptional and translational regulatory elements encoded in the *Streptomyces clavuligerus* genome. *Gene Expression Omnibus*, https://identifiers.org/GEO:GSE128216 (2019).

51. Hwang, S. et al. Ribosome profiling of *Streptomyces griseus* NRBC13350 and *Streptomyces venezuelae* ATCC15439. *Gene Expression Omnibus*, https://identifiers.org/GEO:GSE138278 (2020).

52. Lee, N. et al. Ribosome pausing at the AT-rich codons regulates the protein expression of secondary metabolite gene clusters in the *Streptomyces tsukubaensis* NRRL 18488. *Gene Expression Omnibus*, https://identifiers.org/GEO:GSE97637 (2019).

53. Kim, W. et al. Transcriptome and translatome profiles of Streptomyces species in different growth phases. *Figgshare* https://doi.org/10.6084/m9.figshare.c.4867830 (2020).

54. Calviello, L. & Ohler, U. Beyond read-counts: Ribo-seq data analysis to understand the functions of the transcriptome. *Trends Genet* **33**, 728–744, https://doi.org/10.1016/j.tig.2017.08.003 (2017).

55. Diamant, A. & Tuller, T. Estimation of ribosome profiling performance and reproducibility at various levels of resolution. *Biol Direct* **11**, 24, https://doi.org/10.1186/s13082-016-0127-4 (2016).

56. Haas, B. J., Chm, M., Nusbaum, C., Birren, B. W. & Livny, J. How deep is deep enough for RNA-Seq profiling of bacterial transcriptomes? *BMC Genomics* **13**, 734, https://doi.org/10.1186/1471-2164-13-734 (2012).

57. Rey, F. E. et al. Dissecting the in vivo metabolic potential of two human gut acetogens. *J Biol Chem* **285**, 22082–22090, https://doi.org/10.1074/jbc.M110.117713 (2010).

58. Westernmann, A. J., Gorski, S. A. & Vogel, J. Dual RNA-seq of pathogen and host. *Nat Rev Microbiol* **10**, 618–630, https://doi.org/10.1038/nrmicro2852 (2012).

59. McClure, R. et al. Computational analysis of bacterial RNA-Seq data. *Nucleic Acids Res* **41**, e140, https://doi.org/10.1093/nar/gkt444 (2013).

60. Latif, H. et al. A streamlined ribosome profiling protocol for the characterization of microorganisms. *Biotechniques* **58**, 329–332, https://doi.org/10.2144/000114302 (2015).

Acknowledgements

This work was supported by a grant from the Novo Nordisk Foundation (NNF10CC1016517) and Bio & Medical Technology Development Program (2018M3A9F3079664 to B.-K.C.) through the National Research Foundation of Korea (NRF) funded by the Ministry of Science and ICT (MSIT).

Author contributions

B.-K.C. conceived and supervised the study. W.K., S.H., N.L. and B.-K.C. designed the experiments. W.K., S.H., N.L. and Y.L. performed the experiments. W.K., S.H., N.L., Y.L., S.C., B.P. and B.-K.C. analyzed the data. W.K., S.H., N.L., B.P. and B.-K.C. wrote the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to B.-K.C.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.
