*Article*

# Metatranscriptomic Analysis of Bacterial Communities on Laundered Textiles: A Pilot Case Study

Susanne Jacksch [1,2] , Christoph König [1] , Dominik Kaiser [1], Mirko Weide [2], Stefan Ratering [3] , Sylvia Schnell [3] and Markus Egert [1,*]

[1] Faculty of Medical and Life Sciences, Institute of Precision Medicine, Microbiology and Hygiene Group, Furtwangen University, 78054 Villingen-Schwenningen, Germany; Susanne.Jacksch@hs-furtwangen.de (S.J.); koenig.christoph93@gmail.com (C.K.); dominik.kaiser87@gmail.com (D.K.)

[2] International Research & Development–Laundry & Home Care, Henkel AG & Co. KGaA, 40191 Düsseldorf, Germany; mirko.weide@henkel.com

[3] Research Centre for BioSystems, Land Use, and Nutrition (IFZ), Institute of Applied Microbiology, Justus-Liebig-University Giessen, 35392 Giessen, Germany; stefan.ratering@umwelt.uni-giessen.de (S.R.); sylvia.schnell@umwelt.uni-giessen.de (S.S.)

[*] Correspondence: Markus.Egert@hs-furtwangen.de; Tel.: +49-7720-3074554; Fax: +49-7720-3074207

**Abstract:** Microbially contaminated washing machines and mild laundering conditions facilitate the survival and growth of microorganisms on laundry, promoting undesired side effects such as malodor formation. Clearly, a deeper understanding of the functionality and hygienic relevance of the laundry microbiota necessitates the analysis of the microbial gene expression on textiles after washing, which—to the best of our knowledge—has not been performed before. In this pilot case study, we used single-end RNA sequencing to generate de novo transcriptomes of the bacterial communities remaining on polyester and cotton fabrics washed in a domestic washing machine in mild conditions and subsequently incubated under moist conditions for 72 h. Two common de novo transcriptome assemblers were used. The final assemblies included 22,321 Trinity isoforms and 12,600 Spades isoforms. A large part of these isoforms could be assigned to the SwissProt database, and was further categorized into "molecular function", "biological process" and "cellular component" using Gene Ontology (GO) terms. In addition, differential gene expression was used to show the difference in the pairwise comparison of the two tissue types. When comparing the assemblies generated with the two assemblers, the annotation results were relatively similar. However, there were clear differences between the de novo assemblies regarding differential gene expression.

**Keywords:** RNA sequencing; metatranscriptome; laundry hygiene; cotton; polyester

## 1. Introduction

A multitude of microorganisms live in modern washing machines. The common routes of contamination are worn clothing, tap water and air [1,2]. Promoted by the warm, humid and nutrient-rich environment, microorganisms such as bacteria and fungi can settle and multiply inside the machine [2–4]. The negative effects of such a contamination are unattractive staining, malodor and biofilm formation [2,4]. In particular, the formation of resistant biofilms might pose a risk for susceptible persons, as biofilms might represent a reservoir for (potentially) pathogenic microorganisms that re-contaminate the laundry during washing [5,6].

The microbial contamination of washing machines and laundry is further promoted by largely sustainability-driven adaptations to the washing process that are common today, such as reduced water consumption, low washing temperatures and the increased use of bleach-free liquid detergents [4].

Using the molecular approach of 16S rRNA gene pyrosequencing, we recently showed that the relatively most abundant sequence types in domestic washing machines were

closely related to potentially pathogenic bacteria, such as *Brevundimonas vesicularis* or *Pseudomonas aeruginosa* inside the detergent drawer, and *Moraxella osloensis* or *Acinetobacter parvus* inside the door seal [7]. While this and other structural studies have looked at the microbial community composition of washing machines and laundry items [3,8–14], studies on the metabolic activities of the laundry microbiota are often limited to distinct functionalities, such as the formation and prevention of malodor [2,15,16]. Malodor is often associated with a lack of hygiene, and can negatively affect the life cycle of a textile [17].

In contrast to metagenomics studies, metatranscriptome studies unravel the totality of the genes that are expressed in a complex microbial community [18]. The next-generation sequencing of RNA (RNASeq) can determine the metabolic potential at the time of sampling by quantifying almost all of the transcripts from the present cells, and can thus help to obtain a profound insight into the expression profiles of an entire microbiota in a single experiment, characterizing the functionality of a microbial community [19,20].

Transcriptome studies are computationally challenging and usually require several bioinformatics tools [21,22]. The major steps in a typical metatranscriptome analysis include quality trimming and the removal of contaminating sequencing reads, the reconstruction of the individual transcripts, the annotation of these transcripts and genes, and the quantification of their expression [21,23–25]. In order to reconstruct the transcriptome, de novo assemblers based on de Bruijn graphs, such as Spades (rna mode) [26], Trinity [27], Velvet/Oases [28,29], or SOAPdenovo-trans [30], focus on the relationship between substrings of a fixed length k (*k*-mers). They can be used if no reference genome is available [22,31,32]. Each of these assemblers can produce useful assemblies, but when comparing different assembler software, a considerable degree of variability becomes evident [33,34].

In this pilot case study, we aimed to analyze the expression profiles of the post-wash bacteriota on two common fabric tissue types washed in a domestic washing machine, by generating de novo assembled transcriptomes using two different assemblers. To the best of our knowledge, our study represents the first study using metatranscriptomics for the field of laundry hygiene.

## 2. Materials and Methods

### 2.1. Sample Preparation

Two common fabric types, cotton and polyester, were used for the washing experiments. Textile samples were cut from locally purchased, new, white, cotton and polyester T-shirts with an area of $8 \times 15$ cm (120 cm$^2$), and were subsequently sterilized by autoclaving. In each washing experiment (Figure 1), the cut textile samples ($2 \times 120$ cm$^2$ per fabric type) were washed in a private household washing machine (an approximately five years old EcoActive W1900 appliance (Miele, Gütersloh, Germany) with approximately 4 Kg of ballast laundry consisting of worn cotton T-shirts and jean pants. A mild and short washing program for synthetics was used at 30 °C for 59 min, with a final 600 rpm spinning cycle using 30 mL of a commercial, bleach-free liquid detergent (Spee AktivGel, Henkel, Düsseldorf, Germany). The detergent was under-dosed to aid microbial survival on the washed textiles. Because many components of commercial laundry detergents are antimicrobial, we used a bleach-free liquid detergent and under-dosed it, assuming that this might increase the amount of active bacteria on the laundry after washing [35,36].

After washing, and virtually simulating "forgetting" the laundry in the washing drum, the textile samples were placed separately in a closed 38 l plastic box (Rotho Kunststoff, Würenlingen, Switzerland) and incubated at room temperature for 72 h, together with 4 pieces of washed ballast laundry (two cotton and two polyester T-shirts, with no contact with the test fabrics). This pre-incubation step was necessary to obtain sufficient RNA for the downstream analysis. Afterwards, the textile samples were stored at −80 °C until further processing.
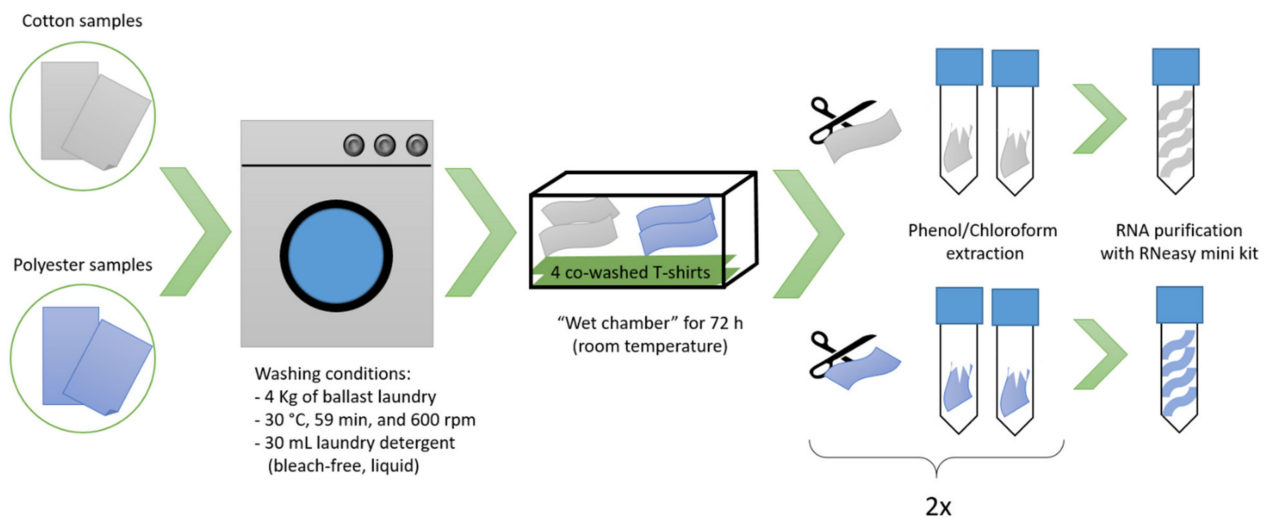
**Figure 1.** Schematic overview of a single washing experiment. Two sterile cotton (grey) and polyester (blue) textile samples (120 cm$^2$ each) were washed in a standard household washing machine under mild conditions. Subsequently, the textile samples were incubated in a "wet chamber" for 72 h at room temperature. Each lobe was cut into smaller pieces under sterile conditions and distributed over two reaction tubes. After the phenol/chloroform extraction, the RNA from the two cotton and polyester lobes, respectively, were combined into a single RNA extract using the RNeasy Mini Kit. This experiment was replicated once, finally yielding two independent cotton and polyester RNA extracts each (n = 2).

*2.2. RNA Extraction and Sequencing*

The total RNA was isolated using a modified phenol–chloroform extraction method from Zoetendal et al. [37]. For the cell disruption, each textile sample (120 cm$^2$) was cut under sterile conditions into pieces of approximately 1.5 cm$^2$, and was evenly distributed into two sterile extraction tubes containing 15 sterile glass spheres (Ø 3 mm, Sigma-Aldrich, St. Louis, MO, USA) and approximately 4.3 g of a sterile ceramic silica extraction powder (Ø 0.1 mm, BioSpec Products, Bartlesville, OK, USA). Subsequently, 500 µL 10% SDS (Ambion, Carlsbad, CA, USA) and 9 mL phenol (Carl Roth, Karlsruhe, Germany) were added to each extraction tube and treated for 3 × 45 s with a FastPrep24 instrument (MP Biomedicals, Eschwege, Germany) at 5.5 m/s. Then, the extraction tubes were centrifuged for 15 min at $3220\times g$ and 4 °C. After the centrifugation, the upper aqueous phase of an extraction tube was transferred to two Phase Lock Gel (PLG) Heavy Tubes (5Prime, Hilden, Germany). Then, 250 µL acid phenol and 250 µL chloroform (Sigma-Aldrich, Taufkirchen, Germany) were added to each PLG tube and gently mixed. Then, the tubes were centrifuged at $13,500\times g$ for 5 min to separate the phases. The aqueous phase was transferred into new PLG tubes, and the procedure was repeated. Afterwards, the aqueous phase was transferred to a new PLG tube, mixed with 500 µL chloroform and centrifuged again at $13,500\times g$ for 5 min. The supernatant was transferred to a new 2 mL tube, and the RNA was purified using an RNeasy Mini Kit (Qiagen, Venlo, The Netherlands) according to the manufacturer's protocol. The RNA extracts belonging to the same textile sample were pooled during their application to the RNeasy column by the repeated transfer of 700 µL to the column, each followed by centrifugation. In order to exclude any contamination with DNA, a DNA digestion was performed two times for 20 min at room temperature. The final elution of the total RNA was performed with 30 µL TE buffer (Sigma-Aldrich, Taufkirchen, Germany). Two independent washing experiments were conducted (n = 2), yielding two RNA extracts from cotton and two from polyester samples, respectively.

The obtained RNA extracts were stored at −80 °C until the library preparation. For this, the RNA samples were reverse transcribed using the ScriptSeq Complete Kit for Bacteria (Epicentre, Madison, WI, USA) according to the manufacturer's protocol. The quality of the cDNA library was determined using an Agilent 2100 Bioanalyzer (Agilent, Waldbronn, Germany). Finally, the samples were sequenced on an Illumina MiSeq platform using a MiSeq Reagent Kit v2 (both Illumina, Munich, Germany).

### 2.3. Sequence Data Analysis

The raw sequences of each single-end library were subjected to a quality control prior to their assembly. The remaining adapters and reads with a Phred quality score < 20 were removed using Trim-galore (version 0.6.6, [38]). Furthermore, human contaminations were excluded using Bowtie2 (version 2.4.2, [39]) and the human GRCh38.p13 (Release 35, [40]). In addition, ribosomal RNA was also removed from the dataset using sortmeRNA (version 4.2.0, [41]) and its provided databases. In the next step, reads with a length of less than 50 bp were discarded, truncated at a length of 280 bp and filtered out with an average Phred quality score of 20 using Trimmomatic (version 0.36, [42]). The final quality of the preprocessed reads was visualized using FastQC (version 0.11.9, [43]) and MultiQC (version 1.9 [44]). Prior to their assembly, the preprocessed reads were error-corrected using Rcorrector (version 1.0.4, [45]), as error correction is considered the best practice for transcriptome assembly [46]. In order to assemble the reads de novo to a transcriptome, Trinity software (version 2.8.5, [27]) with default parameters and Spades software (version 3.14.1, [26,47]) with the -rna-flag and different *k*-mers (13, 15, 19, 21, 25, 29, 31, 43, 55, 67, 79, 91, 103, 115, 127) were used.

The assembly statistics were calculated by the TransRate software (version 1.0.3, [48]). As mentioned in other studies, non-redundant transcripts were removed with the CD-HIT package (version 4.8.1, [49]) with an identity threshold of 95% and a word size of 10 [50–52]. In order to reassess the quality of the clustered transcriptome, assembly statistics were generated using Quast (version 5.0.2, [53]) and rnaQuast (version 2.0.1, [54]). In order to further determine the quality of the assembly, the read representation was calculated by aligning the input reads against the transcriptome using Bowtie2.

The transcriptome completeness was evaluated using the Benchmarking Universal Single-Copy Orthologs (BUSCO) tool (version 3.0.2, [55]) against the bacteria_odb10, archaea_odb10 (both, creation date: 2020-03-06), fungi_odb10 and eukaryota_odb10 (both, creation date: 2020-09-10) databases to quantify the percentage of single-copy orthologues. Full-length transcripts were calculated using BLAST (version 2.2.31, [56,57]) against the SwissProt/UniProtKB database (version 2020_05, [58]) with the parameters max_target_seqs 1, -evalue $1.0 \times 10^{-20}$. The BLAST-results were analyzed using the "analyze_blastPlus_topHit_coverage.pl" script from the Trinity software. As recommended by the Trinity website, ExN50 statistics were calculated using RSEM software (version 1.3.3, [59]) with Bowtie2 and the "contig_ExN50_statistic.pl" script from the Trinity software.

In order to perform the annotation of the transcripts, open reading frames (ORFs) within the assemblies were determined using TransDecoder (version 5.5.0, [60]). The transcript ORFs with less than 150 bp were excluded from the dataset, and the BLAST (BLASTx and BLASTp; E-value: $1.0 \times 10^{-3}$) analysis against the SwissProt/UniProtKB database and the HMMER search (version 3.3.1, [61,62]) against the PFAM database (version 33.1 (May 2020, 18259 entries [63,64]) were performed. Finally, the annotation results were loaded into the Trinotate classification tool (version 3.2.1, [65,66]) to determine the functionality by means of Gene Ontology (GO) [67]. Non-supervised Orthologous Groups (eggNOG) [68] were used to visualize the gene expression profiles grouped according to gene genealogy. Modified trinotateR [69] functions were used to evaluate the Trinotate output with R.

For the taxonomic annotation, the different assemblies were aligned against the NCBI nucleotide database (from May 2021) using BLAST (BLASTx, E-value: $1.0 \times 10^{-3}$).

For differential expression, the transcript abundance was calculated using the "align_and_estimate_abundance.pl" and the "abundance_estimates_to_matrix.pl" scripts from the Trinity bundle using RSEM software with Bowtie2. The differential expression analysis was carried out using R (version 3.5.3, [70]), RStudio (version 1.1.463, [71]) and edgeR (version 3.24.3, [72]). The edgeR package uses negative binomial models to detect dispersion, and later determines the differential expression with the exact test, which is analogous to the Fisher exact test [72]. An overview of the bioinformatics pipeline used here is shown in Figure S1.

## 3. Results and Discussion

### 3.1. Reads and de novo Transcriptome Assembly

In total, 8.3 million raw sequences with sequence lengths between 35 and 300 bp were obtained from the four samples after sequencing on the Illumina MiSeq instrument. After the various quality filtering steps, 6.8 million reads remained, with lengths of 50–280 bp. This corresponds to a loss of ~ 18%. A detailed listing of the numbers of sequences after the different quality filtering steps is given in Table S1.

In order to generate a de novo assembly, two different assemblers were used, i.e., Trinity and Spades. Both assemblers showed constant and good assemblies in a comparison of multiple assemblers with datasets of several different species, and were therefore selected [73]. The de novo assembly from the pre-processed reads generated 24,386 isoforms with lengths ranging from 201 to 64,155 bp using Trinity, and 13,147 isoforms of lengths between 365 and 112,899 bp using Spades.

The determined N50 value was 2192 bp for the Trinity assembly, and 2641 bp for the Spades assembly. The N50 value quantifies the average length of a contig, which comprises 50% of the sequence within the total assembly [23].

### 3.2. Evaluation of the Different de novo Transcriptome Assemblies

In order to investigate the quality of the assemblies, different measures were carried out (Table S2). First, the individual isoforms were assembled into clusters with 95% similarity, reducing the number of non-redundant isoforms to 22,321 for Trinity and 12,600 for Spades, respectively. For further quality control, the pre-processed input reads were aligned against the different clustered transcriptome assemblies to determine the read representation. In general, 80% of the reads mapping back to the transcriptome is considered an indication of a good assembly [74]. Our four sample reads aligned approximately 80–89% with each of the transcriptome assemblies.

In order to obtain an impression of the completeness of the generated assemblies, we applied BUSCO as a reference-based pipeline. This pipeline indicated that a large proportion of the single-copy genes were found in the bacterial domain. More precisely, of the 124 BUSCO groups examined, 69% were recovered from the Trinity assembly and 66% were recovered from the Spades assembly. Nonetheless, a large number of BUSCOs were missing or were too fragmented to be considered, more precisely 29% from the Trinity and 32% Spades assemblies. Although not complete, the values for the bacterial database showed good coverage of the transcripts to the known single copy orthologues, indicating an almost complete expected gene content [75]. The other sets of BUSCOs (archaea, eukaryotea, and fungi) revealed a completeness of less than 10%.

As suggested on the Trinity website, the number of full-length transcripts and the ExN50 values were determined for the further evaluation of the assembly. In order to determine the number of full-length transcripts, a BLAST analysis was first performed against the SwissProt database. To do so, a relatively low e-value ($1.0 \times 10^{-20}$) was used to store only the single best matching proteins and to discard hits for very short sequences, as these usually do not deliver a BLAST hit [76,77]. Under these conditions, 32% of the Trinity generated isoforms and 42% of the Spades generated isoforms could be assigned to SwissProt proteins, respectively. Nevertheless, out of these, 42% (Trinity) and 51% (Spades) of the near-full-length transcripts (> 70%) could be recovered with these assemblies, respectively. However, both assemblies seem to contain a high proportion of fragmented or incorrectly assembled transcripts [78].

The ExN50 value indicates the N50 value by using only the most highly expressed transcripts [79]. Both assemblers peak at low percentages, indicating a tendency to detect highly expressed isoforms (Figure S2). Therefore, under the experimental conditions used here, neither assembler could adequately detect low-expressed transcripts [80]. However, increasing the sequencing depth in follow-up experiments might enhance the ExN50 value to obtain a more complete representation of the transcriptome.

All in all, the above-mentioned indices show that the assemblies are of good quality, but the method clearly needs further improvement, e.g., by increasing the sequencing depth or by using multiple *k*-mers lengths to account for variable transcript expression [81,82].

### 3.3. Transcript Annotation

The BLAST analysis showed that bacterial sequences represented the majority (approximately 99%) of all of the sequences within the two assemblies. Eukaryotic and viral sequences were minorly abundant, probably because the ScriptSeq Complete Kit for Bacteria was used for the library preparation.

Based on the transcript counts, several bacterial genera known to be typical for washing machines and laundered textiles [7,10,13] were detected, such as *Acinetobacter* (48.5%, 51.7%), *Aeromonas* (26.1%, 21.6%), *Rhizobium* (6.0%, 6.5%), *Agrobacterium* (2.9%, 2.4%), *Moraxella* (1.8%, 2.1%) and *Pseudomonas* (0.4%, 0.4%) (the brackets show the relative abundances based on the Spades and Trinity assemblies, respectively, averaged over all of the samples). However, we also detected genera which were, to the best of our knowledge, previously not reported as being typical for washing machines or laundered textiles, such as *Sphingorhabdus* (9.9%, not detected), *Anderseniella* (2.1%, 12.1%), *Epilithonimonas* (0.9%, 1.0%), *Haematobacter* (0.5%, 0.6%) and *Escherichia* (0.04%, 0.3%). Figure S3 shows the relative abundances of the 11 most relatively abundant genera for both experiments, based on Trinity and Spades assemblies, respectively.

The identification of the functional classes was achieved by following the GO term classification using SwissProt gene symbols. These GO terms are divided in three categories, i.e., "biological process", "cellular component" and "molecular functions", which describe the attributes of a gene product [67,83]. In each of the assemblies, we detected a wide range of GO terms from all three functional categories, suggesting the active and diverse microbial gene expression of the investigated textiles (Figure 2).
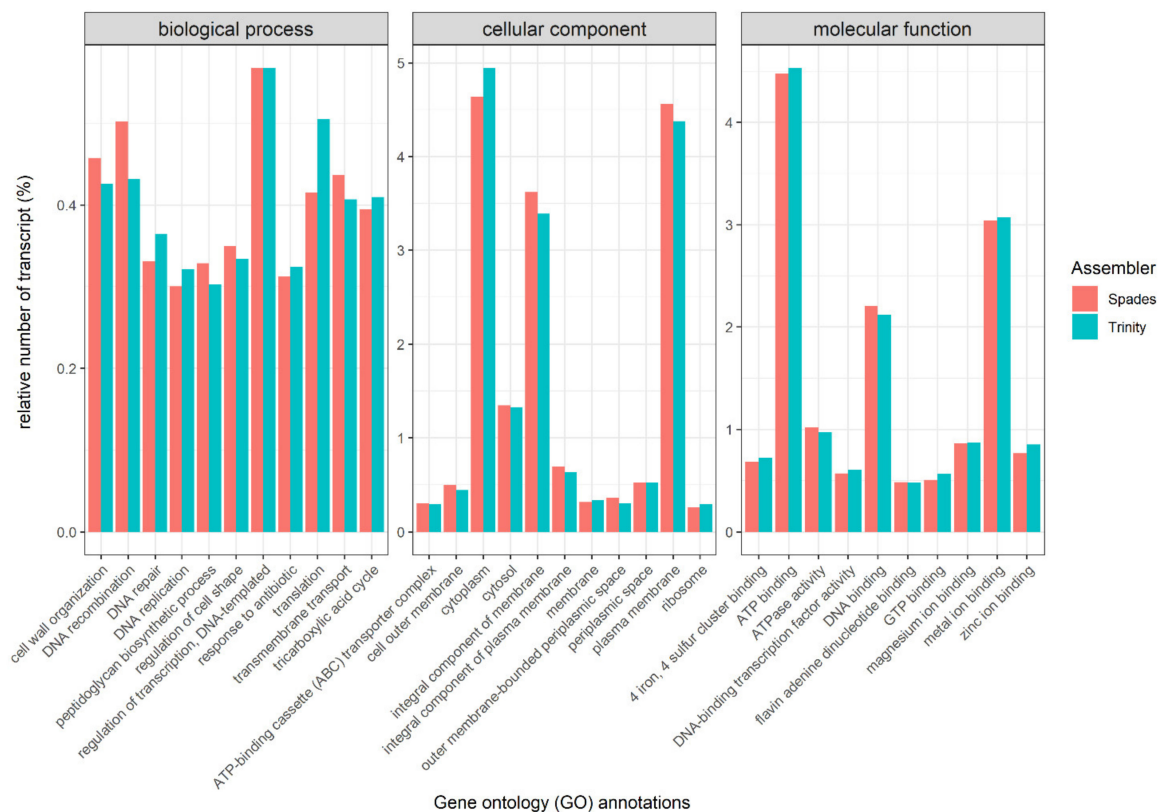


**Figure 2.** Bar chart of the assigned Gene Ontology (GO) terms of the annotated genes assembled with Spades (**red**) or Trinity (**blue**). The GO terms are categorized into biological process, cellular components, and molecular functions, respectively. For each category, the 10 GO terms with the highest relative number of transcripts are displayed in alphabetical order.

In total, 3881 unique GO terms were identified with the Spades assembly, whereas 4390 unique GO terms were determined with the Trinity assembly. Both assemblies shared 3686 unique GO terms. The most frequently assigned GO terms allocated to "cellular components" were related to cytoplasm (Spades: 4.6%, Trinity: 4.9%), plasma membranes (Spades: 4.6%, Trinity: 4.4%), and the integral components of the membrane (Spades: 3.6%, Trinity: 3.4%). On the other hand, most of the "molecular functions" were associated with binding: ATP binding (Spades: 4.5%, Trinity: 4.5%), metal ion binding (Spades: 3.0%, Trinity: 3.1%) and DNA binding (Spades: 2.2%, Trinity: 2.1%). Only a small proportion of the transcripts could be assigned to "biological processes". Here, the regulation of transcription (DNA-templated) (Spades: 0.6%, Trinity: 0.6%), DNA recombination (Spades: 0.5%, Trinity: 0.4%), and cell wall organization (Spades: 0.5%, Trinity: 0.4%) predominated. In summary, these data suggest that a metabolically active bacterial community was present on the investigated fabric patches.

Notably, despite using different assembly software, there was a high level of similarity between the assigned GO terms, which suggests the accuracy of the different assemblies and the assigned annotations [76]. Furthermore, previous research has shown that GO terms have high consistency across multiple species despite the intrinsic differences between the different assembly strategies and annotation pipelines, suggesting the usage of GO for comparisons with other studies [77].

### 3.4. Differential Expression

In order to investigate whether the tissue type had an influence on the microbial gene expression, a differential expression analysis was performed using the edgeR-R-package. After filtering and normalization using the built-in-functions of edgeR, both assemblies clearly showed a separation between the fabric types, indicating differences in their expression profiles (Figure 3). In addition, Figure 3 shows marked differences between the replicate washing experiments, strongly requiring follow-up-studies with larger sample sizes and more standardized conditions. The differences between the two experiments might originate from the different ballast laundry used. While the vertical separation of the samples seems to be textile-dependent, the horizontal separation seems to experiment-dependent.
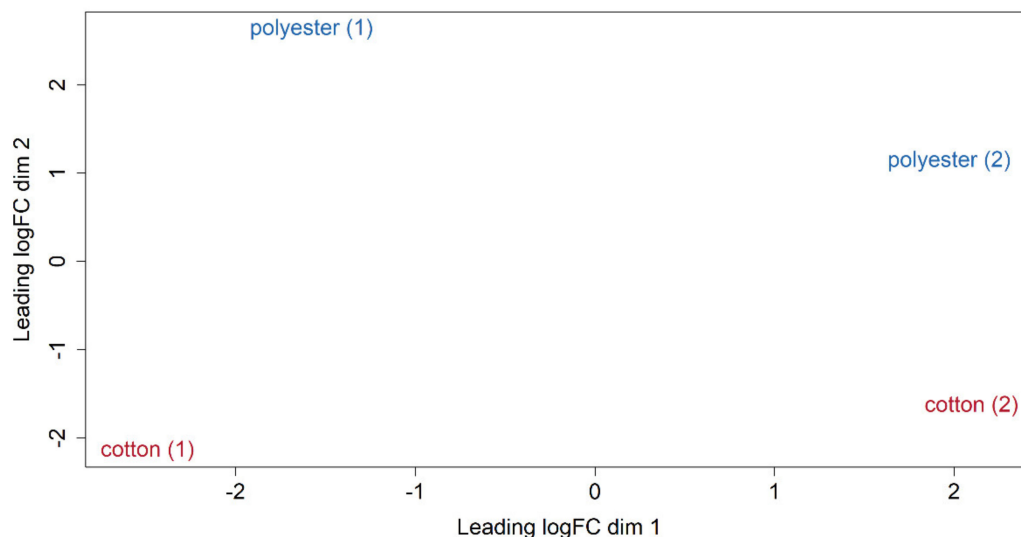


**Figure 3.** Multi-dimensional scaling (MDS) plot based on RNAseq expression profiles from different tissue type samples. The distances between the samples in the plot were calculated based on leading log2 fold changes (logFC) between each cotton and polyester sample, which were defined as the average log2-fold change for the 500 most differential genes between each pair of profiles [84]. The numbers in parentheses define the respective washing experiments. Here, the MDS plot for the Trinity assembly is shown. The very similar MDS plot for the Spades assembly can be found in Figure S4.

In contrast to the annotation, the differential expression analysis showed differences between the two assemblers. Overviews of the different gene expression profiles from the different assemblers, categorized into higher functional groups, are shown in Figures S5 and S6.

Performing the differential expression, we identified 8146 genes with a log2 fold change (4447 upregulated and 3699 downregulated) for the Trinity assembly and 4037 (2248 upregulated and 1789 downregulated) for the Spades assembly when comparing the cotton and polyester samples. Out of these, 16 genes were statistically significantly differentially expressed (adjusted *p*-value < 0.01) in the case of the Trinity assembly (Table 1). Of these, nine genes were identified as being up-regulated in the cotton samples, and seven genes were identified as being down-regulated in the cotton samples. For the Spades assembly, we identified 4037 differentially expressed genes (2248 upregulated and 1789 downregulated). However, for only one gene, the expression was significantly different. It was up-regulated in the cotton samples (Table 1).

Figures S5 and S6 suggest that the number of genes with significant differences in expression between the cotton and polyester might be higher than the 17 genes detected here. In order to prove this, more standardized studies with a bigger sample size are needed. Interestingly, two of the differentially expressed genes were affiliated with *Moraxella* (Table 1), a genus which is well known for laundry malodor production [15,85].

The genes, which were significantly up- or down-regulated, were mainly enzymes that are predominantly involved in metabolic pathways, as well as the transport of substances across the cell membrane (Table 1). The identified genes are common genes that are found in different microorganisms and have a variety of cellular activities, such as the "AAA (ATPases Associated with diverse cellular Activities) family ATPases" or the "LysR family transcriptional transporter", which is involved in virulence, metabolism, quorum sensing and motility [86,87]. In addition, our data slightly suggest that carbohydrates might have acted as substrates on the fabric samples, and might have led to differences in gene expression, as indicated by the different expression of the genes for "Sucrose-6-phosphate hydrolase" and "PTS system sucrose-specific EIIBC component". The phosphoenolpyruvate-dependent phosphotransferase system (PTS) is found in various microorganisms, such as *Escherichia coli*, *Streptococcus mutans* and *Bacillus subtilis*, and—in addition to the transport and phosphorylation of carbohydrates—it is also involved in the movement towards carbon sources [88,89]. It is tempting to speculate that the chemical differences between cotton (made of cellulose fibers) and polyester (made of polyethylene terephthalate fibers) are responsible for these differences. Unlike natural fibers, synthetic fibers are less susceptible to bacterial degradation [90].

The observed differences in the number of differentially expressed genes may also be due to the different ways in which the used assemblers work. As there are a number of different tools and parameters that can be used to reconstruct transcripts, it is difficult to determine a single robust method [91–94]. For future studies, merging the different assemblies created with different programs and parameters might lead to a more reliable representation of the post-wash laundry transcriptome [91].

Clearly, in this pilot case study, we only worked with small sample sizes (n = 2) for each fabric type and poorly standardized washing conditions (different ballast laundry). Small sample sizes increase the variances in gene expression, resulting in lower confidence and increasing *p*-values [95]. Therefore, the number of biological replicates should be increased in future experiments, and even technical replicates should be taken into account to reduce the technical noise [78,95].

**Table 1.** Overview of the significantly differentially expressed genes. Shown here are the log fold changes (logFC), log counts per million (logCPM) and the *p*-value after the false discovery rate (FDR) correction, as calculated by edgeR for the pairwise comparison between cotton and polyester. If available, the SwissProt entry or BLAST accession number (March, 2021), as well as the organism determined by means of BLAST against the nucleotide database (May, 2021), are shown, as well as the determined up- or down-regulation of the respective gene for the cotton samples in comparison to the polyester samples. The functional orthologous groups were obtained using the EggNOG database [68]. If a protein has multiple domains, more than one functional orthologous groups is possible.

| Assembly | Gene | Uniprot Entry/ Blast Accession # | Percent Identity | E-Value | Ontology | Name | Genus | log FC | log CPM | FDR | Regulation |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Trinity | TRINITY_DN12211_c0_g1 | WP_082183191.1 | 100.00 | $2.0 \times 10^{-08}$ | Energy production and conversion | FAD-binding oxidoreductase | *Rhizobium* | −9.871 | 5.431 | 0.003 | down |
| | TRINITY_DN12794_c0_g1 | WP_126090172.1 | 73.08 | $2.5 \times 10^{-02}$ | Transcription | LysR family transcriptional regulator | none available | −9.988 | 5.545 | 0.003 | down |
| | TRINITY_DN16017_c0_g1 | KIV68812.1 | 100.00 | $2.0 \times 10^{-36}$ | Carbohydrate transport and metabolism | Sucrose−6−phosphate hydrolase | *Rhizobium* | −9.852 | 5.412 | 0.003 | down |
| | TRINITY_DN16123_c0_g1 | IHFA_CHRVO | 41.54 | $1.3 \times 10^{-10}$ | Transcription | Integration host factor subunit alpha | none available | −9.635 | 5.202 | 0.008 | down |
| | TRINITY_DN19317_c0_g1 | GLO22_ECOLI | 51.90 | $4.3 \times 10^{-15}$ | Inorganic ion transport and metabolism | Hydroxyacylglutathione hydrolase GloC | *Rhizobium* | −9.756 | 5.319 | 0.010 | down |
| | TRINITY_DN6555_c0_g1 | WP_164056586.1 | 100.00 | $2.0 \times 10^{-17}$ | Replication, recombination and repair | AAA family ATPase | *Rhizobium* | −9.705 | 5.270 | 0.005 | down |
| | TRINITY_DN8425_c0_g1 | WP_142779495.1 | 100.00 | $1.0 \times 10^{-18}$ | Cell cycle control, cell division, chromosome partitioning | ParA family protein | *Rhizobium* | −9.602 | 5.170 | 0.008 | down |
| | TRINITY_DN12673_c0_g1 | WP_042878669.1 | 97.53 | $2.0 \times 10^{-25}$ | Amino acid transport and metabolism, Carbohydrate transport and metabolism | DMT family transporter | *Aeromonas* | 9.611 | 5.184 | 0.008 | up |
| | TRINITY_DN14275_c0_g1 | WP_174060752.1 | 100.00 | $5.0 \times 10^{-23}$ | Amino acid transport and metabolism, Inorganic ion transport and metabolism | ABC transporter permease | *Rhizobium* | 9.590 | 5.163 | 0.010 | up |

**Table 1.** *Cont.*

| Assembly | Gene | Uniprot Entry/ Blast Accession # | Percent Identity | E-Value | Ontology | Name | Genus | log FC | log CPM | FDR | Regulation |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | TRINITY_DN14310_c0_g1 | WP_124801776.1 | 100.00 | $6.0 \times 10^{-11}$ | Inorganic ion transport and metabolism | cation:proton antiporter | *Epilithonimonas* | 9.577 | 5.150 | 0.008 | up |
| | TRINITY_DN16876_c0_g1 | YDDG_ECOLI | 66.67 | $6.9 \times 10^{-08}$ | Amino acid transport and metabolism, Carbohydrate transport and metabolism | Aromatic amino acid exporter YddG | *Acinetobacter* | 9.751 | 5.318 | 0.003 | up |
| | TRINITY_DN16965_c0_g1 | STY97430.1 | 95.16 | $1.0 \times 10^{-33}$ | Coenzyme transport and metabolism | Dihydroneopterin aldolase | *Moraxella* | 9.920 | 5.481 | 0.003 | up |
| | TRINITY_DN19448_c0_g1 | WP_204155761.1 | 82.05 | $4.0 \times 10^{-15}$ | Cell motility, Intracellular trafficking, secretion, and vesicular transport | prepilin-type N-terminal cleavage/methylation domain-containing protein | *Moraxella* | 9.832 | 5.396 | 0.003 | up |
| | TRINITY_DN19763_c0_g1 | Y2604_PSEAE | 69.09 | $6.5 \times 10^{-21}$ | Cell wall/ membrane/ envelope biogenesis | Uncharacterized protein PA2604 | *Pseudomonas* | 9.985 | 5.544 | 0.004 | up |
| | TRINITY_DN24104_c0_g1 | PTSBC_SALTM | 79.22 | $7.3 \times 10^{-33}$ | Carbohydrate transport and metabolism | PTS system sucrose-specific EIIBC component | *Aeromonas* | 10.140 | 5.695 | 0.005 | up |
| | TRINITY_DN9443_c0_g1 | WP_074855682.1 | 100.00 | $4.0 \times 10^{-18}$ | Inorganic ion transport and metabolism | ArsC family reductase | *Pseudomonas* | 9.803 | 5.368 | 0.003 | up |
| **Spades** | NODE_11013_length_ 508_cov_2.086614_g10506 | YOXD_BACSU | 46.39 | $1.7 \times 10^{-35}$ | Function unknown | Uncharacterized oxidoreductase YoxD | *Epilithonimonas* | 9.511 | 5.043 | 0.000 | up |

## 4. Conclusions

Our study delivered the first laundry metatranscriptome, and it suggests a differential gene expression of the post-wash bacteriota on two commonly used types of fabrics. Our data provide an initial overview and characterization of the bacterial laundry transcriptome, as well as a comparison between the two de novo transcriptome assemblers used, i.e., Trinity and Spades. Clearly, the approach needs further optimization, such as a higher sequencing depth and further biological and technical replicates, ideally in combination with DNA shotgun sequencing, in order to identify the microorganisms of which the metabolic activity shapes the microbial community on laundry. Nevertheless, the assemblies created here represent a solid basis for further metatranscriptomic studies.

## References

1. Novak Babič, M.; Gostinčar, C.; Gunde-Cimerman, N. Microorganisms populating the water-related indoor biome. *Appl. Microbiol. Biotechnol.* **2020**, *104*, 6443–6462. [CrossRef]
2. Munk, S.; Johansen, C.; Stahnke, L.H.; Adler-Nissen, J. Microbial survival and odor in laundry. *J. Surfact. Deterg.* **2001**, *4*, 385–394. [CrossRef]
3. Nix, I.D.; Frontzek, A.; Bockmühl, D.P. Characterization of Microbial Communities in Household Washing Machines. *Tenside. Surfactants Deterg.* **2015**, *52*, 432–440. [CrossRef]
4. Egert, M. The BE-Microbiome-Communities with Relevance for Laundry and Home Care. *SOFW J.* **2017**, *143*, 44–48.
5. Bloomfield, S.F.; Exner, M.; Goroncy-Bermes, P.; Hartemann, P.; Heeg, P.; Ilschner, C.; Krämer, I.; Merkens, W.; Oltmanns, P.; Rotter, M.; et al. Lesser-known or hidden reservoirs of infection and implications for adequate prevention strategies: Where to look and what to look for. *GMS Hyg. Infect. Control* **2015**, *10*, Doc04. [CrossRef]
6. Gibson, L.L.; Rose, J.B.; Haas, C.N. Use of quantitative microbial risk assessment for evaluation of the benefits of laundry sanitation. *Am. J. Infect. Control.* **1999**, *27*, S34–S39. [CrossRef]
7. Jacksch, S.; Kaiser, D.; Weis, S.; Weide, M.; Ratering, S.; Schnell, S.; Egert, M. Influence of Sampling Site and other Environmental Factors on the Bacterial Community Composition of Domestic Washing Machines. *Microorganisms* **2020**, *8*, 30. [CrossRef]

8. Babič, M.N.; Zalar, P.; Ženko, B.; Schroers, H.-J.; Džeroski, S.; Gunde-Cimerman, N. *Candida* and *Fusarium* species known as opportunistic human pathogens from customer-accessible parts of residential washing machines. *Fungal Biol.* **2015**, *119*, 95–113. [CrossRef] [PubMed]

9. Boonstra, M.B.; Spijkerman, D.C.M.; Voor In 't Holt, A.F.; van der Laan, R.J.; Bode, L.G.M.; van Vianen, W.; Klaassen, C.H.W.; Vos, M.C.; Severin, J.A. An outbreak of ST307 extended-spectrum beta-lactamase (ESBL)-producing *Klebsiella pneumoniae* in a rehabilitation center: An unusual source and route of transmission. *Infect. Control Hosp. Epidemiol.* **2020**, *41*, 31–36. [CrossRef]

10. Callewaert, C.; van Nevel, S.; Kerckhof, F.M.; Granitsiotis, M.S.; Boon, N. Bacterial Exchange in Household Washing Machines. *Front. Microbiol.* **2015**, *6*, 1381. [CrossRef]

11. Gattlen, J.; Amberg, C.; Zinn, M.; Mauclaire, L. Biofilms isolated from washing machines from three continents and their tolerance to a standard detergent. *Biofouling* **2010**, *26*, 873–882. [CrossRef]

12. Honisch, M.; Stamminger, R.; Bockmühl, D.P. Impact of wash cycle time, temperature and detergent formulation on the hygiene effectiveness of domestic laundering. *J. Appl. Microbiol.* **2014**, *117*, 1787–1797. [CrossRef]

13. Jacksch, S.; Zohra, H.; Weide, M.; Schnell, S.; Egert, M. Cultivation-Based Quantification and Identification of Bacteria at Two Hygienic Key Sides of Domestic Washing Machines. *Microorganisms* **2021**, *9*, 905. [CrossRef] [PubMed]

14. Schmithausen, R.M.; Sib, E.; Exner, M.; Hack, S.; Rösing, C.; Ciorba, P.; Bierbaum, G.; Savin, M.; Bloomfield, S.F.; Kaase, M.; et al. The Washing Machine as a Reservoir for Transmission of Extended-Spectrum-Beta-Lactamase (CTX-M-15)-Producing *Klebsiella oxytoca* ST201 to Newborns. *Appl. Environ.* **2019**, *85*, e01435-19. [CrossRef] [PubMed]

15. Kubota, H.; Mitani, A.; Niwano, Y.; Takeuchi, K.; Tanaka, A.; Yamaguchi, N.; Kawamura, Y.; Hitomi, J. *Moraxella* species are primarily responsible for generating malodor in laundry. *Appl. Environ. Microbiol.* **2012**, *78*, 3317–3324. [CrossRef]

16. Stapleton, K.; Hill, K.; Day, K.; Perry, J.D.; Dean, J.R. The potential impact of washing machines on laundry malodour generation. *Lett. Appl. Microbiol.* **2013**, *56*, 299–306. [CrossRef]

17. Van Herreweghen, F.; Amberg, C.; Marques, R.; Callewaert, C. Biological and Chemical Processes that Lead to Textile Malodour Development. *Microorganisms* **2020**, *8*, 1709. [CrossRef] [PubMed]

18. Bashiardes, S.; Zilberman-Schapira, G.; Elinav, E. Use of Metatranscriptomics in Microbiome Research. *Bioinform. Biol. Insights* **2016**, *10*, 19–25. [CrossRef]

19. Kukurba, K.R.; Montgomery, S.B. RNA Sequencing and Analysis. *Cold Spring Harb. Protoc.* **2015**, *2015*, 951–969. [CrossRef]

20. Cardoso-Silva, C.B.; Costa, E.A.; Mancini, M.C.; Balsalobre, T.W.; Canesin, L.E.; Pinto, L.R.; Carneiro, M.S.; Garcia, A.A.; de Souza, A.P.; Vicentini, R. De novo assembly and transcriptome analysis of contrasting sugarcane varieties. *PLoS ONE* **2014**, *9*, e88462. [CrossRef]

21. Bharti, R.; Grimm, D.G. Current challenges and best-practice protocols for microbiome analysis. *Brief. Bioinform.* **2021**, *22*, 178–193. [CrossRef]

22. Ghurye, J.S.; Cepeda-Espinoza, V.; Pop, M. Metagenomic Assembly: Overview, Challenges and Applications. *Yale J. Biol. Med.* **2016**, *89*, 353–362.

23. Wajid, B.; Serpedin, E. Do it yourself guide to genome assembly. *Brief. Funct. Genom.* **2016**, *15*, 1–9. [CrossRef]

24. Koch, C.M.; Chiu, S.F.; Akbarpour, M.; Bharat, A.; Ridge, K.M.; Bartom, E.T.; Winter, D.R. A Beginner's Guide to Analysis of RNA Sequencing Data. *Am. J. Respir. Cell Mol. Biol.* **2018**, *59*, 145–157. [CrossRef]

25. Oshlack, A.; Robinson, M.D.; Young, M.D. From RNA-seq reads to differential expression results. *Genome Biol.* **2010**, *11*, 220. [CrossRef]

26. Bushmanova, E.; Antipov, D.; Lapidus, A.; Prjibelski, A.D. rnaSPAdes: A de novo transcriptome assembler and its application to RNA-Seq data. *Gigascience* **2019**, *8*. [CrossRef] [PubMed]

27. Grabherr, M.G.; Haas, B.J.; Yassour, M.; Levin, J.Z.; Thompson, D.A.; Amit, I.; Adiconis, X.; Fan, L.; Raychowdhury, R.; Zeng, Q.; et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **2011**, *29*, 644–652. [CrossRef] [PubMed]

28. Zerbino, D.R.; Birney, E. Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.* **2008**, *18*, 821–829. [CrossRef]

29. Schulz, M.H.; Zerbino, D.R.; Vingron, M.; Birney, E. Oases: Robust de novo RNA-seq assembly across the dynamic range of expression levels. *Bioinformatics* **2012**, *28*, 1086–1092. [CrossRef] [PubMed]

30. Xie, Y.; Wu, G.; Tang, J.; Luo, R.; Patterson, J.; Liu, S.; Huang, W.; He, G.; Gu, S.; Li, S.; et al. SOAPdenovo-Trans: De novo transcriptome assembly with short RNA-Seq reads. *Bioinformatics* **2014**, *30*, 1660–1666. [CrossRef]

31. Liu, J.; Yu, T.; Mu, Z.; Li, G. TransLiG: A de novo transcriptome assembler that uses line graph iteration. *Genome Biol.* **2019**, *20*, 81. [CrossRef]

32. Chen, Q.; Lan, C.; Zhao, L.; Wang, J.; Chen, B.; Chen, Y.-P.P. Recent advances in sequence assembly: Principles and applications. *Brief. Funct. Genom.* **2017**, *16*, 361–378. [CrossRef] [PubMed]

33. Earl, D.; Bradnam, K.; St John, J.; Darling, A.; Lin, D.; Fass, J.; Yu, H.O.K.; Buffalo, V.; Zerbino, D.R.; Diekhans, M.; et al. Assemblathon 1: A competitive assessment of de novo short read assembly methods. *Genome Res.* **2011**, *21*, 2224–2241. [CrossRef]

34. Bradnam, K.R.; Fass, J.N.; Alexandrov, A.; Baranay, P.; Bechner, M.; Birol, I.; Boisvert, S.; Chapman, J.A.; Chapuis, G.; Chikhi, R.; et al. Assemblathon 2: Evaluating de novo methods of genome assembly in three vertebrate species. *Gigascience* **2013**, *2*, 10. [CrossRef] [PubMed]

35.　Falk, N.A. Surfactants as Antimicrobials: A Brief Overview of Microbial Interfacial Chemistry and Surfactant Antimicrobial Activity. *J Surfactants Deterg.* **2019**, *22*, 1119–1127. [CrossRef] [PubMed]

36.　Schages, J.; Stamminger, R.; Bockmühl, D.P. A New Method to Evaluate the Antimicrobial Efficacy of Domestic Laundry Detergents. *J Surfactants Deterg.* **2020**, *23*, 629–639. [CrossRef]

37.　Zoetendal, E.G.; Booijink, C.C.; Klaassens, E.S.; Heilig, H.G.; Kleerebezem, M.; Smidt, H.; de Vos, W.M. Isolation of RNA from bacterial samples of the human gastrointestinal tract. *Nat. Protoc.* **2006**, *1*, 954–959. [CrossRef]

38.　Krueger, F. *Trim Galore*. Babraham Institute. 2012. Available online: http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/ (accessed on 28 April 2021).

39.　Langmead, B.; Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **2012**, *9*, 357–359. [CrossRef]

40.　Frankish, A.; Diekhans, M.; Ferreira, A.-M.; Johnson, R.; Jungreis, I.; Loveland, J.; Mudge, J.M.; Sisu, C.; Wright, J.; Armstrong, J.; et al. GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Res.* **2019**, *47*, D766–D773. [CrossRef]

41.　Kopylova, E.; Noé, L.; Touzet, H. SortMeRNA: Fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics* **2012**, *28*, 3211–3217. [CrossRef]

42.　Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **2014**, *30*, 2114–2120. [CrossRef] [PubMed]

43.　Andrews, S. *FastQC*. Babraham Institute. 2010. Available online: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/ (accessed on 28 April 2021).

44.　Ewels, P.; Magnusson, M.; Lundin, S.; Käller, M. MultiQC: Summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **2016**, *32*, 3047–3048. [CrossRef] [PubMed]

45.　Song, L.; Florea, L. Rcorrector: Efficient and accurate error correction for Illumina RNA-seq reads. *Gigascience* **2015**, *4*, 48. [CrossRef]

46.　Macmanes, M.D.; Eisen, M.B. Improving transcriptome assembly through error correction of high-throughput sequence reads. *PeerJ* **2013**, *1*, e113. [CrossRef] [PubMed]

47.　Bankevich, A.; Nurk, S.; Antipov, D.; Gurevich, A.A.; Dvorkin, M.; Kulikov, A.S.; Lesin, V.M.; Nikolenko, S.I.; Pham, S.; Prjibelski, A.D.; et al. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **2012**, *19*, 455–477. [CrossRef] [PubMed]

48.　Smith-Unna, R.; Boursnell, C.; Patro, R.; Hibberd, J.M.; Kelly, S. TransRate: Reference-free quality assessment of de novo transcriptome assemblies. *Genome Res.* **2016**, *26*, 1134–1144. [CrossRef]

49.　Li, W.; Godzik, A. Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **2006**, *22*, 1658–1659. [CrossRef]

50.　Moreno-Santillán, D.D.; Machain-Williams, C.; Hernández-Montes, G.; Ortega, J. De Novo Transcriptome Assembly and Functional Annotation in Five Species of Bats. *Sci. Rep.* **2019**, *9*, 6222. [CrossRef]

51.　Ranjan, A.; Ichihashi, Y.; Farhi, M.; Zumstein, K.; Townsley, B.; David-Schwartz, R.; Sinha, N.R. De novo assembly and characterization of the transcriptome of the parasitic weed dodder identifies genes associated with plant parasitism. *Plant Physiol.* **2014**, *166*, 1186–1199. [CrossRef]

52.　Carradec, Q.; Pelletier, E.; Da Silva, C.; Alberti, A.; Seeleuthner, Y.; Blanc-Mathieu, R.; Lima-Mendez, G.; Rocha, F.; Tirichine, L.; Labadie, K.; et al. A global ocean atlas of eukaryotic genes. *Nat. Commun.* **2018**, *9*, 373. [CrossRef]

53.　Gurevich, A.; Saveliev, V.; Vyahhi, N.; Tesler, G. QUAST: Quality assessment tool for genome assemblies. *Bioinformatics* **2013**, *29*, 1072–1075. [CrossRef]

54.　Bushmanova, E.; Antipov, D.; Lapidus, A.; Suvorov, V.; Prjibelski, A.D. rnaQUAST: A quality assessment tool for de novo transcriptome assemblies. *Bioinformatics* **2016**, *32*, 2210–2212. [CrossRef]

55.　Seppey, M.; Manni, M.; Zdobnov, E.M. BUSCO: Assessing Genome Assembly and Annotation Completeness. *Methods Mol. Biol.* **2019**, *1962*, 227–245. [CrossRef]

56.　Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W.; Lipman, D.J. Basic local alignment search tool. *J. Mol. Biol.* **1990**, *215*, 403–410. [CrossRef]

57.　Blast. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **2018**, *46*, D8–D13. [CrossRef]

58.　UniProt Consortium. UniProt: A worldwide hub of protein knowledge. *Nucleic Acids Res.* **2019**, *47*, D506–D515. [CrossRef]

59.　Li, B.; Dewey, C.N. RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinform.* **2011**, *12*, 323. [CrossRef] [PubMed]

60.　Haas, B. TransDecoder. Available online: https://github.com/TransDecoder/TransDecoder.wiki.git (accessed on 28 April 2021).

61.　Finn, R.D.; Clements, J.; Eddy, S.R. HMMER web server: Interactive sequence similarity searching. *Nucleic Acids Res.* **2011**, *39*, W29–W37. [CrossRef]

62.　Eddy, S.R. *HMMER*. HMMER. 2020. Available online: http://hmmer.org/ (accessed on 28 April 2021).

63.　El-Gebali, S.; Mistry, J.; Bateman, A.; Eddy, S.R.; Luciani, A.; Potter, S.C.; Qureshi, M.; Richardson, L.J.; Salazar, G.A.; Smart, A.; et al. The Pfam protein families database in 2019. *Nucleic Acids Res.* **2019**, *47*, D427–D432. [CrossRef] [PubMed]

64.　Finn, R.D.; Bateman, A.; Clements, J.; Coggill, P.; Eberhardt, R.Y.; Eddy, S.R.; Heger, A.; Hetherington, K.; Holm, L.; Mistry, J.; et al. Pfam: The protein families database. *Nucleic Acids Res.* **2014**, *42*, D222–D230. [CrossRef] [PubMed]

65. Bryant, D.M.; Johnson, K.; DiTommaso, T.; Tickle, T.; Couger, M.B.; Payzin-Dogru, D.; Lee, T.J.; Leigh, N.D.; Kuo, T.-H.; Davis, F.G.; et al. A Tissue-Mapped Axolotl De Novo Transcriptome Enables Identification of Limb Regeneration Factors. *Cell Rep.* **2017**, *18*, 762–776. [CrossRef]
66. Trinotate. Available online: https://github.com/Trinotate/Trinotate.github.io/wiki (accessed on 28 April 2020).
67. Ashburner, M.; Ball, C.A.; Blake, J.A.; Botstein, D.; Butler, H.; Cherry, J.M.; Davis, A.P.; Dolinski, K.; Dwight, S.S.; Eppig, J.T.; et al. Gene ontology: Tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* **2000**, *25*, 25–29. [CrossRef]
68. Huerta-Cepas, J.; Szklarczyk, D.; Heller, D.; Hernández-Plaza, A.; Forslund, S.K.; Cook, H.; Mende, D.R.; Letunic, I.; Rattei, T.; Jensen, L.J.; et al. eggNOG 5.0: A hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res.* **2019**, *47*, D309–D314. [CrossRef]
69. Stubben, C. *trinotateR*. 2016. Available online: https://github.com/cstubben/trinotateR (accessed on 28 April 2021).
70. The R Project for Statistical Computing. 2020; Available online: https://www.R-project.org/ (accessed on 28 April 2021).
71. RStudio Team. *R Studio: Integrated Development for R*; RStudio, Inc.: Boston, MA, USA, 2016; Available online: http://www.rstudio.com/ (accessed on 28 April 2021).
72. Robinson, M.D.; McCarthy, D.J.; Smyth, G.K. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **2010**, *26*, 139–140. [CrossRef] [PubMed]
73. Hölzer, M.; Marz, M. De novo transcriptome assembly: A comprehensive cross-species comparison of short-read RNA-Seq assemblers. *Gigascience* **2019**, *8*, giz039. [CrossRef] [PubMed]
74. Haas, B.J.; Papanicolaou, A.; Yassour, M.; Grabherr, M.; Blood, P.D.; Bowden, J.; Couger, M.B.; Eccles, D.; Li, B.; Lieber, M.; et al. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* **2013**, *8*, 1494–1512. [CrossRef] [PubMed]
75. Simão, F.A.; Waterhouse, R.M.; Ioannidis, P.; Kriventseva, E.V.; Zdobnov, E.M. BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **2015**, *31*, 3210–3212. [CrossRef]
76. Carruthers, M.; Yurchenko, A.A.; Augley, J.J.; Adams, C.E.; Herzyk, P.; Elmer, K.R. De novo transcriptome assembly, annotation and comparison of four ecological and evolutionary model salmonid fish species. *BMC Genom.* **2018**, *19*, 32. [CrossRef]
77. Riesgo, A.; Andrade, S.C.S.; Sharma, P.P.; Novo, M.; Pérez-Porro, A.R.; Vahtera, V.; González, V.L.; Kawauchi, G.Y.; Giribet, G. Comparative description of ten transcriptomes of newly sequenced invertebrates and efficiency estimation of genomic sampling in non-model taxa. *Front. Zool.* **2012**, *9*, 33. [CrossRef]
78. Conesa, A.; Madrigal, P.; Tarazona, S.; Gomez-Cabrero, D.; Cervera, A.; McPherson, A.; Szcześniak, M.W.; Gaffney, D.J.; Elo, L.L.; Zhang, X.; et al. A survey of best practices for RNA-seq data analysis. *Genome Biol.* **2016**, *17*, 13. [CrossRef]
79. Geniza, M.; Jaiswal, P. Tools for building de novo transcriptome assembly. *Curr. Plant Biol.* **2017**, *11–12*, 41–45. [CrossRef]
80. Sahraeian, S.; Mohiyuddin, M.; Sebra, R.; Tilgner, H.; Afshar, P.T.; Au, K.F.; Bani Asadi, N.; Gerstein, M.B.; Wong, W.H.; Snyder, M.P.; et al. Gaining comprehensive biological insight into the transcriptome by performing a broad-spectrum RNA-seq analysis. *Nat. Commun.* **2017**, *8*, 59. [CrossRef] [PubMed]
81. Robertson, G.; Schein, J.; Chiu, R.; Corbett, R.; Field, M.; Jackman, S.D.; Mungall, K.; Lee, S.; Okada, H.M.; Qian, J.Q.; et al. De novo assembly and analysis of RNA-seq data. *Nat. Methods* **2010**, *7*, 909–912. [CrossRef] [PubMed]
82. Zhao, Q.Y.; Wang, Y.; Kong, Y.M.; Luo, D.; Li, X.; Hao, P. Optimizing de novo transcriptome assembly from short-read RNA-Seq data: A comparative study. *BMC Bioinform.* **2011**, *12* (Suppl. 14), S2. [CrossRef]
83. Harris, M.A.; Clark, J.; Ireland, A.; Lomax, J.; Ashburner, M.; Foulger, R.; Eilbeck, K.; Lewis, S.; Marshall, B.; Mungall, C.; et al. The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res.* **2004**, *32*, D258–D261. [CrossRef]
84. Pal, B.; Chen, Y.; Vaillant, F.; Capaldo, B.D.; Joyce, R.; Song, X.; Bryant, V.L.; Penington, J.S.; Di Stefano, L.; Tubau Ribera, N.; et al. A single-cell RNA expression atlas of normal, preoplastic and tumorigenic states in the human breast. *EMBO J.* **2021**, *40*, e107333. [CrossRef]
85. Denawaka, C.J.; Fowlis, I.A.; Dean, J.R. Source, impact and removal of malodour from soiled clothing. *J. Chromatogr. A* **2016**, *1438*, 216–225. [CrossRef] [PubMed]
86. Frickey, T.; Lupas, A.N. Phylogenetic analysis of AAA proteins. *J. Struct. Biol.* **2004**, *146*, 2–10. [CrossRef] [PubMed]
87. Maddocks, S.E.; Oyston, P.C.F. Structure and function of the LysR-type transcriptional regulator (LTTR) family proteins. *Microbiology* **2008**, *154*, 3609–3623. [CrossRef]
88. Postma, P.W.; Lengeler, J.W.; Jacobson, G.R. Phosphoenolpyruvate:carbohydrate phosphotransferase systems of bacteria. *Microbiol. Rev.* **1993**, *57*, 543–594. [CrossRef]
89. St Martin, E.J.; Wittenberger, C.L. Regulation and function of sucrose 6-phosphate hydrolase in *Streptococcus mutans*. *Infect. Immun.* **1979**, *26*, 487–491. [CrossRef] [PubMed]
90. Callewaert, C.; de Maeseneire, E.; Kerckhof, F.-M.; Verliefde, A.; van de Wiele, T.; Boon, N. Microbial odor profile of polyester and cotton clothes after a fitness session. *Appl. Environ. Microbiol.* **2014**, *80*, 6611–6619. [CrossRef] [PubMed]
91. Kumar, S.; Blaxter, M.L. Comparing de novo assemblers for 454 transcriptome data. *BMC Genom.* **2010**, *11*, 571. [CrossRef]
92. Moreton, J.; Dunham, S.P.; Emes, R.D. A consensus approach to vertebrate de novo transcriptome assembly from RNA-seq data: Assembly of the duck (*Anas platyrhynchos*) transcriptome. *Front. Genet.* **2014**, *5*, 190. [CrossRef]

93. Scholz, M.; Lo, C.-C.; Chain, P.S.G. Improved assemblies using a source-agnostic pipeline for MetaGenomic Assembly by Merging (MeGAMerge) of contigs. *Sci. Rep.* **2014**, *4*, 6480. [CrossRef]
94. Wences, A.H.; Schatz, M.C. Metassembler: Merging and optimizing de novo genome assemblies. *Genome Biol.* **2015**, *16*, 207. [CrossRef]
95. Martin, J.A.; Wang, Z. Next-generation transcriptome assembly. *Nat. Rev. Genet.* **2011**, *12*, 671–682. [CrossRef] [PubMed]