The effect of reverse transcription enzymes and conditions on high throughput amplicon sequencing of the 16S rRNA

Adam Šťovíček[1], Smadar Cohen-Chalamish[2] and Osnat Gillor[1]

[1] Department of Environmental Hydrology and Microbiology, Ben-Gurion University of the Negev, Beer Sheva, Israel
[2] The Mina and Everard Goodman Faculty of Life Sciences and Advanced Materials and Nanotechnology Institute, Bar-Ilan University, Ramat-Gan, Israel

ABSTRACT

It is assumed that the sequencing of ribosomes better reflects the active microbial community than the sequencing of the ribosomal RNA encoding genes. Yet, many studies exploring microbial communities in various environments, ranging from the human gut to deep oceans, questioned the validity of this paradigm due to the discrepancies between the DNA and RNA based communities. Here, we focus on an often neglected key step in the analysis, the reverse transcription (RT) reaction. Previous studies showed that RT may introduce biases when expressed genes and ribosmal rRNA are quantified, yet its effect on microbial diversity and community composition was never tested. High throughput sequencing of ribosomal RNA is a valuable tool to understand microbial communities as it better describes the active population than DNA analysis. However, the necessary step of RT may introduce biases that have so far been poorly described. In this manuscript, we compare three RT enzymes, commonly used in soil microbiology, in two temperature modes to determine a potential source of bias due to non-standardized RT conditions. In our comparisons, we have observed up to six fold differences in bacterial class abundance. A temperature induced bias can be partially explained by G-C content of the affected bacterial groups, thus pointing toward a need for higher reaction temperatures. However, another source of bias was due to enzyme processivity differences. This bias is potentially hard to overcome and thus mitigating it might require the use of one enzyme for the sake of cross-study comparison.

INTRODUCTION

Massively parallel amplicon sequencing revolutionized our view of microbial world: by sequencing a taxonomic tag such as 16S rRNA encoding gene, it allows taxonomic description of the microbial communities (*Quammen, 2018*). However, the existing approaches introduce caveats: the DNA amplicon sequencing may capture "relic DNA," which is a recalcitrant genetic material from dead cells or naked DNA (*Carini et al., 2016*) in addition, amplicon sequencing carries technical biases due to sample preparation,

DNA extraction methods (*Pan et al., 2010*), amplification reaction (*Pfeiffer et al., 2014*) and analysis (*Pollock et al., 2018*). Moreover, DNA-based microbiome information can describe the total community, but it cannot report which members are metabolically active (*Blazewicz et al., 2013*). In contrast to DNA-based tools, analysis of ribosomes can describe the metabolically active members of a given community. The combination of data generated from rRNA encoding genes and ribosomes led to a wide range of ecological insights, including the response to climatic changes (*Angel et al., 2013*), pH and water availability (*Romanowicz et al., 2016*), and biogeochemical processes (*Freedman et al., 2015*).

Ribosomal analysis studies are based on an assumption that ribosomes are more abundant in active cells compared to dormant ones (*Blazewicz et al., 2013*; *Lennon & Jones, 2011*). However, this assumption may not always be correct. Dormant bacteria may be misclassified as active, when ribosomes are present in cells and spores that are inactive (*Segev et al., 2013*; *Blagodatskaya & Kuzyakov, 2013*). In contrast, active bacteria with low metabolic turnover and low ribosomal count could be labeled as dormant when sequencing depth is insufficient (*Steven et al., 2017*; *Joergensen & Wichern, 2018*). In spite of various biases that introduce discrepancies in the community structure (*Forney, Zhou & Brown, 2004*), ribosomal analysis can capture the biological variability highlighting large differences between samples. However, if more subtle differences are of interest, technical biases could confound biological interpretations (*Lever et al., 2015*; *McCarthy et al., 2015*). This is due to specific challenges introduced RNA-based analysis (*Bustin & Nolan, 2004*, *2017*). Therefore, to confidently compare results across ribosome-based amplicon sequencing studies, we must determine which component of the analysis: RNA extraction, processing or data analysis may influence the outcome and introduce biases.

Prior studies focused on biases in the steps of RNA extraction, amplification and sequencing, but disregard any biases that may occur during RT (*Creer et al., 2016*). At the crucial step of RT, most researches simply "follow the manufacture instructions" (Table 1). However, RT kits typically detail a wide range of temperatures, primer, template and reaction options, which may lead to different results. The reverse transcriptase enzyme requires sequence priming to initiate a reaction. Primers could be poly-A complementary, random or sequence specific. Poly-A priming is limited to eukaryotic mRNA which makes it unsuitable for use with ribosomal taxonomic tags. Opinions vary about the usefulness of random and sequence-specific priming for the analysis of microbiomes: Random priming may produce higher yield of cDNA and improve the detection limit (*Zhang & Byrne, 1999*; *Ståhlberg et al., 2004*), but may decrease the reproducibility and introduce bias (*Bustin & Nolan, 2004*; *Hansen, Brenner & Dudoit, 2010*). Sequence specific primes require fine tuning of the reaction conditions and higher template concentration than random priming (*Ståhlberg, Kubista & Michael, 2004*). Moreover, the results of the RT reaction is determined not only by the type of the RT enzyme used, but also by the reaction conditions (*Curry, McHale & Smith, 2002*; *Ståhlberg et al., 2004*; *Ståhlberg, Kubista & Michael, 2004*; *Bustin & Nolan, 2004*; *Sieber et al., 2010*).

Ideally RT efficiency is near 100%, but in practice it varies dramatically: 90% efficiency was reported for SuperScript III (mutated MMLV RT) (*Ståhlberg et al., 2004*), 20% for

Table 1 Literature overview of RT conditions applied in soil microbiological studies.

Manufacturer	RT enzyme	RT origin	Temperature (°C)		RNA type	Primer type	References
			Suggested	Used			
Promega	MMLV	MMLV	37–42	NA	rRNA	926R	*Carson et al. (2010)*
				NA	rRNA & mRNA	Random hexamers	*Pratscher, Dumont & Conrad (2011)*
	ImProm-II	AMV	37–55	42	rRNA & mRNA	Random hexamers	*Angel et al. (2013)*
				NA	rRNA	Random hexamers	*Ke, Lu & Conrad (2015)*
Qiagen	QuantiTect	Quantiscript	42–50	NA	rRNA	Unique RT Primer Mix	*Barnard, Osborne & Firestone (2015)*
				37	rRNA	Random hexamers	*Placella, Brodie & Firestone (2012)*
	Omniscript	Quantiscript	37	NA	mRNA	Random hexamers	*Paulin et al. (2013)*
				NA	mRNA	invA-R	*García et al. (2010)*
Takara	PrimeScript II	AMV	42–50	NA	mRNA	Random hexamers	*Huang et al. (2016)*
				NA	rRNA	Random hexamers	*Che et al. (2018)*
Roche	Roche reverse transcription kit	AMV	42–60	42 & 50	rRNA	Random hexamers	*Nunes et al. (2018)*
				42 & 50	rRNA	Random hexamers	*Jurburg et al. (2017)*
Thermo Fisher	MMLV	MMLV	37–42	45	rRNA	900R	*Lillis, Doyle & Clipson (2009)*
				NA	rRNA	Random hexamers	*Baldrian et al. (2012)*
	SuperScript-II	MMLV	42–55	NA	rRMA	1492R	*Degelmann et al. (2009)*
				NA	mRNA	Random hexamers	*Nacke et al. (2014)*
	SuperScript-III	MMLV	42–55	NA	rRNA	Random hexamers	*Angel & Conrad (2013)*
				NA	rRNA	27F & LR3	*Romanowicz et al. (2016)*

murine leukemia virus (M-MLV) RT (*Curry, McHale & Smith, 2002*), and as low as 2% for Avian Myeloblastosis Virus (AMV) RT (*Ståhlberg et al., 2004*). Moreover almost two orders of magnitude difference were reported between mutated and wild-type AMV RT (*Ståhlberg et al., 2004*).

To the best of our knowledge no study has yet compared the RT reaction conditions for environmental microbiome profile. We hypothesize that during RT reactions, varying RT enzyme types and temperature conditions will yield different results in microbial diversity and community composition. We further predict that variations in communities will be G-C dependent. To test our prediction, we present a comparative study of commonly used RT enzymes in the field of environmental microbiology as well as a comparison of two different reaction temperatures.

MATERIALS AND METHODS

Study site and sample collection scheme

Soil samples were collected at the central Negev Desert highlands, Israel (Zin Plateau, 30°86′N, 34°80′E) at an established ecological research site. The mean annual precipitation at the sampling site is 90 mm and the mean annual temperature is 30° (LTER data). Samples were collected under the canopy of perennial shrub *Hammada scoparia* in October 2015 at the end of the dry season as previously described (*Baubin et al., 2019*).

Briefly, sampling was conducted in seven random blocks, each providing two technical replicates resulting in 14 samples. Samples were collected from the top five cm of the soil, following the removal of crust and debris. The soil samples were processed within 24 h of collection. Samples were homogenized using two mm sieve and the duplicates from each block were composited. This resulted in seven final replicates.

RNA preparation

Total RNA was extracted from each of the seven samples using a phenol-chlorophorm extraction previously described by _Angel (2012)_. The reaction buffer pH was adjusted to 5. The total RNA was subsequently purified with the MagListo™ Total RNA Extraction Kit (Bioneer, Daejeon, Republic of Korea). Contaminant DNA was removed using a DNAse I from the MasterPure RNA Purification Kit (Epicenter, Madison, WI, USA) with two successive treatments of 30 min according to manufacturer's instructions. The reaction mixture was purified using the MagListo™ Total RNA Extraction Kit (Bioneer, Daejeon, Republic of Korea). The absence of contaminant DNA was verified using total bacterial primers 341F (5′-CCTACGGGAGGCAGCAG-3′) and 515R (5′-TTACCGCGGCTGCTGGCAC-3′) (_Klindworth et al., 2013_) and DreamTaq DNA polymerase (Thermo Scientific, Waltham, MA, USA) under the following conditions: 95 °C for 5 min, followed by 26 cycles of 95 °C for 15 s, 60 °C for 30 s and 72 °C for 30 s for extension followed by 72 °C for 5 min for final extension. If amplification was detected the sample was discarded, re-extracted, purified, and tested. Only DNA-free samples were used in this study.

Reverse transcription reaction conditions

RT kits used in this experiment were chosen to represent the most commonly used enzymes in the field (Fig. 1). All seven samples were reverse transcribed by the same kit to reduce batch effects. Each RT kit used in this study originated from a different source: (I) ImProm-II Reverse Transcription System enzyme (Promega, Madison, WI, USA) originates from AMV RT, (II) SuperScript IV Reverse Transcriptase Kit enzyme (ThermoFisher Scientific, Waltham, MA, USA) originates from MMLV RT and (III) TGIRT originates from the mobile group II introns reverse transcriptase (_Mohr et al., 2013_) TGIRT™-III Enzyme (InGex, St. Louis, MO, USA). Each reaction consisted of 50 ng of total RNA template, measured by Quanti-iT™ RNA Assay Kit (ThermoFisher, Waltham, MA, USA), and random hexamer primers (0.5 µg/reaction). Template and primer mix were heated to 70 °C (ImProm-II) or 65 °C (SuperScript IV). Each reaction was subsequently cooled to 4 °C for 5 min and incubated at 42 °C (ImProm-II), 55 °C (ImProm-II and Superscript) or 57 °C (TGIRT) for 60 min (ImProm-II), 120 min (TGIRT), or 10 min (SuperScript IV). All reactions were terminated and DNA was removed by alkaline lysis using 2 µL of 1M NaOH, incubating for 12 min at 70 °C. After which the reaction was neutralized using 4 µL of 0.5M acetic acid (Table 2).

Illumina sequence preparation

The V3 and V4 regions of the resulting cDNA were amplified using 341F (5′-CCTACGGGAGGCAGCAG-3′) and 806R (5′-GGTCTGGACTACHVGGGTWTCTAAT-3′)
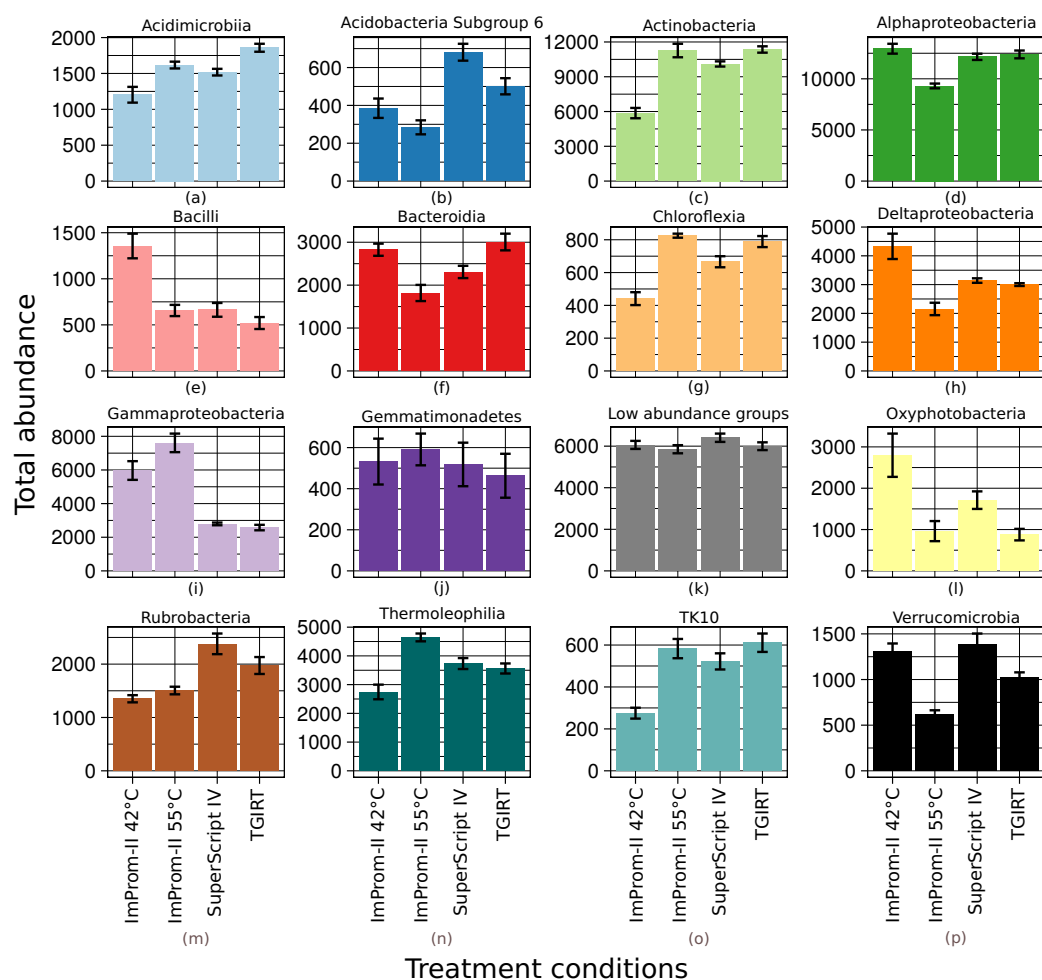
Figure 1 **Relative abundance of main classes across each tested condition (A–P).** Only top 15% of the most abundant classes are represented and the rest is summarized in the "Low abundance groups" category (K). The *x*-axis shows different enzymes and conditions. The *y*-axis shows an average relative abundance. Each category is an average of four samples.   Full-size ☒ DOI: 10.7717/peerj.7608/fig-1

(*Klindworth et al., 2013*) primers. Each reaction was performed in triplicate and consisted of 1 mM bovine serum albumin (Takara, Kusatsu, Japan), 2.5 µL of 10× standard buffer, five µM primers, 0.8 mM dNTPs, 0.4 µL DreamTaq DNA polymerase (Thermo Scientific, Waltham, MA, USA), and 4 µL of template cDNA. The reaction mixtures were subsequently amplified using the following PCR conditions: 95 °C for 30 s, 27 cycles of 95 °C for 15 s, 50 °C for 30 s, 68 °C for 30 s, and 68 °C for 5 min. Resulting amplicon presence was verified using 1.5% agarose gel electrophoresis. Resulting technical triplicates were combined, and the sequencing libraries were constructed using the TruSeq® DNA Sample Preparation Kit (Illumina, San Diego, CA, USA) following the manufacturer's recommendations. The amplicon libraries were sequenced (250 × 2 base pairs, pair-end) on the Illumina MiSeq System platform at the Research Resources Centre at the University of Illinois.

Table 2 A summary of conditions applied to the different reaction conditions.

Manufacturer	RT kit	Primers	Thermo cycling		Reaction mix	
			Temperature (°C)	Time (min)	Reactant	Amount
Promega	Im-Prom II	Random Hexamers (500 ng/reaction)	70	5	DTT	10 µM
			4	5	Tris–HCl	50 mM
			25	5	KCl	75 mM
			42	60	MgCl$_2$	2.5 mM
			70	15	dNTP	0.5 mM
					RNAse inhibitor	0.5/20 µL
Promega	Im-Prom II	Random Hexamers (500 ng/reaction)	70	5	DTT	10 µM
			4	5	Tris–HCl	50 mM
			25	5	KCl	75 mM
			55	60	MgCl$_2$	2.5 mM
			70	15	dNTP	0.5 mM
					RNAse inhibitor	0.5/20 µL
Thermo Fisher	SuperScriptIV	Random Hexamers (2.5 µM)	65	5	DTT	5 µM
			0	1	Tris–HCl	50 mM
			23	10	KCl	50 mM
			55	10	MgCl$_2$	4 mM
			80	10	dNTP	0.5 mM
					RNAse inhibitor	0.5/20 µL
TGIRT	TGIRT-III	Random Hexamers (500 ng/reaction)	65	5	DTT	5 µM
			0	1	Tris–HCl	10 mM
			23	10	EDTA	1 mM
			58	120	MgCl$_2$	4 mM
			80	10	dNTP	0.5 mM
					RNAse inhibitor	0.5/20 µL

Sequence analysis

Resulting paired end sequences were merged using the CASPER program (*Kwon, Lee & Yoon, 2014*), and the resulting merged reads were clustered using the UPARSE pipeline according to the recommended settings (*Edgar, 2013*). The resulting operational taxonomic unit (OTU) representative sequences were taxonomically assigned with the SINA incremental aligner using a lowest common ancestor algorithm (*Pruesse et al., 2007*) and the SILVA database version 132 (*Quast et al., 2013*). All sequences retrieved in this study were uploaded to European Nucleotide Archive (https://www.ebi.ac.uk/ena) submission number PRJEB32237.

Class enrichment plot data preparation

To explore whether G-C content contributed to differences in relative abundances of different taxa, we ran the following analysis: for each reaction condition we calculated pairwise comparisons at the class level: we normalized the proportional

enrichment in each respective reaction conditions following Eq. (1) (Fig. 2; Fig. S2),
where the A and B represent a class at the different conditions.

$$\text{Class}_{\text{normalized}} = \frac{A - B}{A + B} \tag{1}$$

The error bars represent a standard deviation, which have been calculated as a standard
deviation of each category and normalized according to the Eq. (2). The δA and δB
represent the standard deviation of class A and B. Details of deriving this equations are
specified in Equation_S1 and Code_S3.ipynb.

$$\delta\text{Class}_{\text{normalized}} = \frac{2}{(A + B)^2} \sqrt{B^2 \delta A^2 + A^2 \delta B^2} \tag{2}$$

Statistical analysis

All data analysis was performed in R v3.4.3 (*R Core Team, 2018*). The dataset was
sub-sampled (rarified) to an even depth of 9,000 sequences per sample using python
numpy package v1.15.4 (*Van Der Walt, Colbert & Varoquaux, 2011*). The subsampling
removed five samples from the dataset. Additionally, three more samples were removed as
outliers (Code_S1.ipynb). In order to equalize the number of replicates, two random
samples were removed from the TGIRT dataset. This reduced the number of replicates to
four samples per experimental category. The sample diversity was analyzed using the
vegan package v2.5-2 (*Oksanen et al., 2018*). The data was visualized using the R package
ggplot2 package v2.2.1 (*Wickham, 2016*) and python package matplotlib (*Hunter, 2007*).

RESULTS

Sample preparation and diversity analysis

After removing obvious outliers, four samples from each condition were analyzed.
The summary of the analysis and all the code used to produce each figure is included in the
Files S1 and S2. The changes in bacterial diversity among the tested conditions were
expressed using species count, Pielou's evenness index (*Pielou, 1966*) and Shannon
diversity index (*Shannon & Weaver, 1949*). Despite observed trends in the diversity
indices, no statistically significant differences were detected (Fig. S1; Code_S2.ipynb).

Relative abundance plot

A relative abundance of major taxonomic classes is depicted in the Fig. 1. Each column is
an average of four biological replicates. OTUs that were not taxonomically assigned at this
level are summarized as "Unclassified" (≈2%). Various patterns (detailed below) were
detected among the experimental conditions: some patterns could be attributed to
differences in reaction temperature (which ranged from 42 to 55 °C and 57 °C). Other
patterns could be linked to enzyme type. RT reactions with SupeScript IV and TGIRT RT
enzymes yielded no significant differences in class relative abundances. However,
transcription with ImProm-II RT at a similar temperature (≈55 °C) yielded different
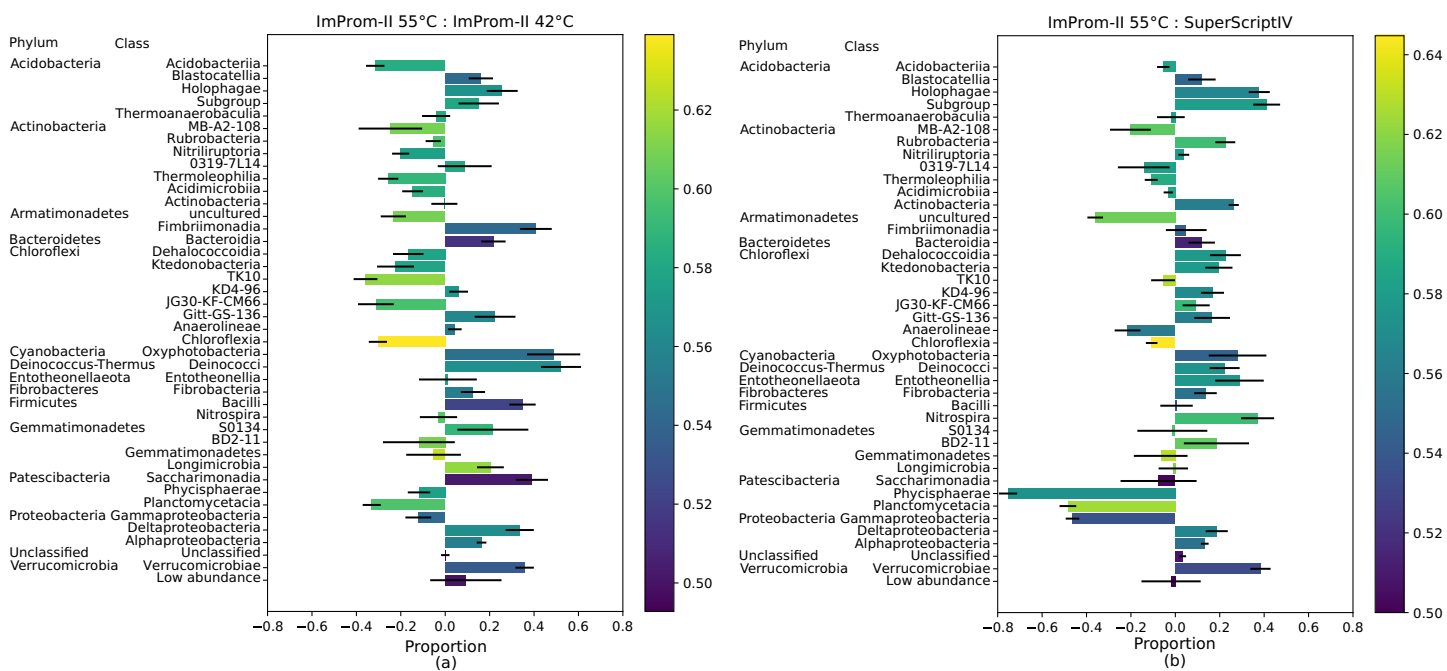abundances: specifically, the abundances of Alphaproteobacteria, Bacteroidia,

Figure 2 A proportional comparison of most abundant classes between the ImProm-II at 42 and 55 °C (A) and the ImProm-II at 55 °C and SuperScriptIV at 55 °C (B). Enrichment is expressed such that a class that is equally proportional in both conditions, has a value of 0. If the class shows in one condition but is absent from another, its value would be equal to 1 or −1 respectively. Furthermore, a weighted average of the GC content of each class is expressed as the bar color. Each value is an average of four biological replicates.   Full-size ◩ DOI: 10.7717/peerj.7608/fig-2

Deltaproteobacteria, Oxyphotobacteria, Rubrobacteria, and Verrucomicrobidae decreased. However, Chloflexia, Gammaproteobacteria and Thermoleophilia abundances increased when their ribosomes were transcribed with ImProm-II RT (Fig. 1; Table S1). When transcription occured at lower temperature (42 °C), relative abundances of Bacilli, Deltaproteobacteria, and Oxyphotobacteria were enriched, while Actinobacteria, Chloroflexia, and Acidimicrobia were depleted under the same conditions (Fig. 1; Table S1).

Class enrichment plot

The Fig. 2 also depicts the weighted average of the G-C content in each class. The proportional comparison is interpreted as follows: A value of zero in the proportional comparison represents the taxonomic class count that is exactly equal between the two compared groups (Fig. 2). A value of 1 or −1 is assigned when a given taxonomic class is only present in one category and absent in another, respectively. In general, there was a tendency toward lower G-C content lower temperature of ImProm-II (Fig. 2A; Figs. S2A and S2B). No statistically significant differences were detected between the profiles resulting from RT of SuprScript-IV or TGIRT (Fig. S2D). The taxa Gemmatimonadetes, Fibrobacteria, and Thermoanaerobaculia were relatively insensitive to the RT conditions. However the majority of the classes were enriched in some conditions. (I) the rate of RT reaction was only sensitive to temperature for the classes Alphaproteobacteria, Gemmatimonadetes, Fibrobacteria, Thermoanaerobaculia, TK10, and Blastocatellia (Fig. 2A). These groups tend to have extreme GC content (both high and low).

Table 3 The linear regression statistics.

Condition 1	Condition 2	Adjusted R^2	t value	p-value	Significance	Note
ImProm-II 42 °C	ImProm-II 55 °C	0.429	5.366	4.89E-06	***	
ImProm-II 42 °C	SuperScript IV	0.1373	2.692	0.0127	*	•
ImProm-II 42 °C	TGIRT	0.2032	3.231	0.00264	**	
ImProm-II 55 °C	SuperScript IV	0.0841	−2.097	0.0431	*	•
ImProm-II 55 °C	TGIRT	0.03624	−1.546	0.131		
SuperScript IV	TGIRT	0.03634	1.548	0.130		

(II) Gammaproteobacteria, Planctomycetacia, and Phycisphaerae are relatively insensitive to the reaction temperature (Fig. 2A), but their abundances vary with different RT enzymes (Fig. 2B; Fig. S2).

Class enrichment statistics

We calculated a linear regression, where the response variable was the relative proportion of each class between two tested categories, and the explanatory variable was the G-C content (the assumptions tests and plots can be found in the File S2). The linear regression assumptions were tested: in case of two condition pairs (ImProm-II at 42 °C & SuperScript-IV as well as ImProm-II at 55 °C & SuperScript-IV), the assumption conditions were not met (File S2). Since we cannot confidently discard the null hypothesis in these cases, we do not consider these two tests significant. Therefore, we are considering only the ImProm-II 42 °C & ImProm-II 55 °C as well as ImProm-II 42 °C & TGIRT as a significant outcome. Differences in the remaining cases cannot be explained by the weighted G-C content alone (Table 3).

We used a GC content as an explanatory variable of a the class enrichment. The rows marked with a • do not fulfill all test assumptions (see Code_S2.ipynb). The statistical significance: *, **, *** at $P < 0.05$, 0.01 and 0.001 respectively.

DISCUSSION

As high throughput sequencing has become increasingly accessible in recent years, researchers urgently call for method standardization to allow for accurate cross-study comparisons (*Pan et al., 2010*; *Blagodatskaya & Kuzyakov, 2013*). With this motivation, researchers developed new platforms that offer protocols and standardized methods for data acquisition from DNA resources, such as the Earth Microbiome Project (http://www.earthmicrobiome.org/) that standardizes DNA amplicon sequencing. Despite the success with standardizing methods and protocols for DNA-based analysis, to this date, there is RNA-based methods have not been standardized, despite the discrepancies repeatedly reported between the RNA- and DNA-based analyses (*Blazewicz et al., 2013*; *Carini et al., 2016*; *Dlott et al., 2015*) and the plethora of methods used in these studies (Table 1). The analysis of rRNA adds specific biases to high throughput sequencing analysis, such as reduced template stability compared to DNA, RT priming bias, and linearity of RT reaction (*Bustin & Nolan, 2004*). These biases need to be either minimized or standardized.

In this study, we focused on one crucial step in the RNA analysis that was previously overlooked: the transcription of RNA to cDNA (Table 1). Several biases connected to the RT reaction have been described for RT-qPCR, such as quantification of expressed genes (*Bustin & Nolan, 2004*, *2017*; *Zhang & Byrne, 1999*) in RNA-Seq, that is, primer related bias of expressed transcripts (*Hansen, Brenner & Dudoit, 2010*). Yet, the role of RT in diversity patterns was not yet investigated in the context of high throughput sequencing of ribosomes. Here, we focus on the role of enzyme and reaction temperature in shaping the diversity and composition of ribosome-based communities. We have compared four RT enzymes commonly used in soil microbiology (Table 1) at two distinct temperature modes (42 °C and 55–57 °C). Then we analyzed temperature and RT enzyme-related effects on the resulting community profiles (Table 2).

Under different reaction conditions, we detected differences in the relative abundance of bacterial classes portrayed by different reaction conditions (Fig. 1). Some observed differences can be attributed to the combined effect of reaction temperature and average template G-C content (Fig. 2; Fig. S2). As expected, this effect is clearest when we applied the same enzyme (ImProm-II) at two reaction temperatures (42 and 55 °C), then the G-C content had the highest prediction power ($t = 5.366$, $p = 4.8 \times 10^{-5}$, Table 3). Likewise, in every comparison of RT enzymes at low and high temperature, G-C content affected the relative abundance of taxonomic classes with statistical significance. Although the RT reactions are commonly performed at 42 °C (Table 1), our results indicate that this reaction temperature is too low to allow successful RT of some soil community taxa, in particular species with higher G-C content.

When transcription was performed with different enzymes under similar reaction temperatures, relative abundances of taxa differed notably (Figs. 1 and 2). Although the RTs of SuperScript-IV and TGIRT originate from different organisms, they yielded similar taxa abundances. Reactions with ImProm-II yielded different profiles. These differences cannot be explained by the G-C content (Table 3), but could be attributed to ribosome properties and the efficiency of RT. The ribosomes extracted from the soil environment were diverse and probably differ in their secondary and tertiary structures (*Yilmaz, Okten & Noguera, 2006*) and post-transcriptional modifications (*Schwartz & Motorin, 2017*). Thus RT enzymes kinetics would differ. The discrepancies reported in this study raise further questions: how would one decide which enzymes or temperatures best reflect the active community composition? It has been suggested that the total community could be used as a reference to accurately deduce the diversity. Furthermore, this study was performed on desert soil samples and the effects in other environments remain to be determined.

We previously demonstrated that the total and active communities in the Negev soil used in this study differ in both abundance profiles and community composition during the dry season, while during the wet season, no differences were detected (*Baubin et al., 2019*). During the dry season, a "phantom taxa" (*Klein et al., 2016*), Deinococcus-Thermus, comprised ≈30% of the total soil community (*Baubin et al., 2019*) but was undetected in the active community of the dry season by any of the methods used here.

These results suggest that the DNA-based total community may differ from the RNA-based community and thus cannot be used as a reliable reference for diversity. Furthermore, our results underline a need to standardize and specify the RT conditions that allow cross-study comparisons. The scale of the effect of RT conditions on the RNA-based community might vary with a studied biome. Dependending on obvious factors such as GC content (discussed above), as well as poorly studied factors such as ribosomal post-transcriptional modifications. Therefore, we recommend verifying each case separately before attempting a cross-study comparison.

CONCLUSION

We have tested commonly utilized RT enzymes at assorted temperatures and observed marked differences in the output community structure. These differences were attributed to RT type and reaction conditions. We suggest that RT reaction conditions may dictate the diversity of a given community and therefore the exact conditions should be detailed in full (i.e., the common notation, "according to manufacturer instructions" does not provide sufficient information (Table 1)). Furthermore, RT should be performed at a sufficiently high temperature to minimize the G-C bias, preferably at 55 °C. Lastly, we suggest that the same RT enzyme should be used across comparable studies, since we detected discrepancies between RT enzymes performing at equivalent conditions (Fig. 2B). Here, we highlight, for the first time, the need for standardisation and careful consideration of RT reaction conditions in studies describing ribosome-based diversity and community composition.

Author Contributions

- Adam Šťovíček conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, approved the final draft.
- Smadar Cohen-Chalamish conceived and designed the experiments, performed the experiments, contributed reagents/materials/analysis tools, authored or reviewed drafts of the paper, approved the final draft.
- Osnat Gillor conceived and designed the experiments, contributed reagents/materials/analysis tools, authored or reviewed drafts of the paper, approved the final draft.

DNA Deposition

The following information was supplied regarding the deposition of DNA sequences:

The sequences retrieved in this study are available at the European Nucleotide Archive: PRJEB32237.

Data Availability

The following information was supplied regarding data availability:

The code used to generate the data is available in the Supplemental Files.

Supplemental Information

Supplemental information for this article can be found online at http://dx.doi.org/10.7717/peerj.7608#supplemental-information.

REFERENCES

Angel R. 2012. Total nucleic acid extraction from soil. *Protocol Exchange* DOI 10.1038/protex.2012.046.

Angel R, Conrad R. 2013. Elucidating the microbial resuscitation cascade in biological soil crusts following a simulated rain event. *Environmental Microbiology* **15**:2799–2815 DOI 10.1111/1462-2920.12140.

Angel R, Pasternak Z, Soares MIM, Conrad R, Gillor O. 2013. Active and total prokaryotic communities in dryland soils. *FEMS Microbiology Ecology* **86(1)**:130–138 DOI 10.1111/1574-6941.12155.

Baldrian P, Kolaiřík M, Štursová M, Kopecký J, Valášková V, Větrovský T, Žifčáková L, Šnajdr J, Rídl J, Vlček Č, Voříšková J. 2012. Active and total microbial communities in forest soil are largely different and highly stratified during decomposition. *ISME Journal* **6(2)**:248–258 DOI 10.1038/ismej.2011.95.

Barnard RL, Osborne CA, Firestone MK. 2015. Changing precipitation pattern alters soil microbial community response to wet-up under a Mediterranean-type climate. *ISME Journal* **9(4)**:946–957 DOI 10.1038/ismej.2014.192.

Baubin C, Farrell AM, Šťovíček A, Ghazaryan L, Giladi I, Gillor O. 2019. Seasonal and spatial variability in total and active bacterial communities from desert soil. *Pedobiologia* **74**:7–14 DOI 10.1016/j.pedobi.2019.02.001.

Blagodatskaya E, Kuzyakov Y. 2013. Active microorganisms in soil: critical review of estimation criteria and approaches. *Soil Biology and Biochemistry* **67**:192–211 DOI 10.1016/j.soilbio.2013.08.024.

Blazewicz SJ, Barnard RL, Daly RA, Firestone MK. 2013. Evaluating rRNA as an indicator of microbial activity in environmental communities: limitations and uses. *ISME Journal* **7(11)**:2061–2068 DOI 10.1038/ismej.2013.102.

Bustin SA, Nolan T. 2004. Pitfalls of quantitative real- time reverse-transcription polymerase chain reaction. *Journal of Biomolecular Techniques* **15(3)**:155–166.

Bustin SA, Nolan T. 2017. Talking the talk, but not walking the walk: RT-qPCR as a paradigm for the lack of reproducibility in molecular research. *European Journal of Clinical Investigation* **47(10)**:756–774 DOI 10.1111/eci.12801.

Carini P, Marsden PJ, Leff JW, Morgan EE, Strickland MS, Fierer N. 2016. Relic DNA is abundant in soil and obscures estimates of soil microbial diversity. *Nature Microbiology* **2**:16242 DOI 10.1038/nmicrobiol.2016.242.

Carson JK, Gonzalez-Quinones V, Murphy DV, Hinz C, Shaw JA, Gleeson DB. 2010. Low pore connectivity increases bacterial diversity in soil. *Applied and Environmental Microbiology* **76(12)**:3936–3942 DOI 10.1128/aem.03085-09.

Che R, Jinling Q, Tahmasbian I, Wang F, Zhou S, Xu Z, Cui X, Che R. 2018. Litter amendment rather than phosphorus can dramatically change inorganic nitrogen pools in a degraded grassland soil by affecting nitrogen-cycling microbes. *Soil Biology and Biochemistry* **120**:145–152 DOI 10.1016/j.soilbio.2018.02.006.

Creer S, Deiner K, Frey S, Porazinska D, Taberlet P, Thomas WK, Potter C, Bik HM. 2016. The ecologist's field guide to sequence-based identification of biodiversity. *Methods in Ecology and Evolution* **7(9)**:1008–1018 DOI 10.1111/2041-210x.12574.

Curry J, McHale C, Smith MT. 2002. Low efficiency of the moloney murine leukemia virus reverse transcriptase during reverse transcription of rare t(8;21) fusion gene transcripts. *BioTechniques* **32(4)**:768–775 DOI 10.2144/02324st02.

Degelmann DM, Kolb S, Dumont M, Murrell JC, Drake HL. 2009. Enterobacteriaceae facilitate the anaerobic degradation of glucose by a forest soil. *FEMS Microbiology Ecology* **68(3)**:312–319 DOI 10.1111/j.1574-6941.2009.00681.x.

Dlott G, Maul JE, Buyer J, Yarwood S. 2015. Microbial rRNA:rDNA gene ratios may be unexpectedly low due to extracellular DNA preservation in soils. *Journal of Microbiological Methods* **115**:112–120 DOI 10.1016/j.mimet.2015.05.027.

Edgar RC. 2013. UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nature Methods* **10(10)**:996–998 DOI 10.1038/nmeth.2604.

Forney LJ, Zhou X, Brown CJ. 2004. Molecular microbial ecology: Land of the one-eyed king. *Current Opinion in Microbiology* **7(3)**:210–220 DOI 10.1016/j.mib.2004.04.015.

Freedman ZB, Romanowicz KJ, Upchurch RA, Zak DR. 2015. Differential responses of total and active soil microbial communities to long-term experimental N deposition. *Soil Biology and Biochemistry* **90**:275–282 DOI 10.1016/j.soilbio.2015.08.014.

García R, Bælum J, Fredslund L, Santorum P, Jacobsen CS. 2010. Influence of temperature and predation on survival of salmonella enterica serovar typhimurium and expression of inva in soil and manure-amended soil. *Applied and Environmental Microbiology* **76(15)**:5025–5031 DOI 10.1128/aem.00628-10.

Hansen KD, Brenner SE, Dudoit S. 2010. Biases in Illumina transcriptome sequencing caused by random hexamer priming. *Nucleic Acids Research* **38(12)**:e131 DOI 10.1093/nar/gkq224.

Huang K, Chen C, Zhang J, Tang Z, Shen Q, Rosen BP, Zhao F-J. 2016. Efficient arsenic methylation and volatilization mediated by a novel bacterium from an arsenic-contaminated paddy soil. *Environmental Science & Technology* **50(12)**:6389–6396 DOI 10.1021/acs.est.6b01974.

Hunter JD. 2007. Matplotlib: A 2D graphics environment. *Computing in Science & Engineering* **9(3)**:90–95 DOI 10.1109/mcse.2007.55.

Joergensen RG, Wichern F. 2018. Alive and kicking: why dormant soil microorganisms matter. *Soil Biology and Biochemistry* **116**:419–430 DOI 10.1016/j.soilbio.2017.10.022.

Jurburg SD, Nunes I, Brejnrod A, Jacquiod S, Priemé A, Sørensen SJ, Van Elsas JD, Salles JF. 2017. Legacy effects on the recovery of soil bacterial communities from extreme temperature perturbation. *Frontiers in Microbiology* **8**:1–13 DOI 10.3389/fmicb.2017.01832.

Ke X, Lu W, Conrad R. 2015. High oxygen concentration increases the abundance and activity of bacterial rather than archaeal nitrifiers in rice field soil. *Microbial Ecology* **70(4)**:961–970 DOI 10.1007/s00248-015-0633-4.

Klein AM, Bohannan BJ, Jaffe DA, Levin DA, Green JL. 2016. Molecular evidence for metabolically active bacteria in the atmosphere. *Frontiers in Microbiology* **7**:1–11 DOI 10.3389/fmicb.2016.00772.

Klindworth A, Pruesse E, Schweer T, Peplies J, Quast C, Horn M, Glöckner FO. 2013. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Research* **41**:e1 DOI 10.1093/nar/gks808.

Kwon S, Lee B, Yoon S. 2014. CASPER: context-aware scheme for paired-end reads from high-throughput amplicon sequencing. *BMC Bioinformatics* **15(Suppl_9)**:S10 DOI 10.1186/1471-2105-15-s9-s10.

Lennon JT, Jones SE. 2011. Microbial seed banks: the ecological and evolutionary implications of dormancy. *Nature Reviews Microbiology* **9(2)**:119–130 DOI 10.1038/nrmicro2504.

Lever MA, Torti A, Eickenbusch P, Michaud AB, Šantl-Temkiv T, Jørgensen BB. 2015. A modular method for the extraction of DNA and RNA, and the separation of DNA pools from diverse environmental sample types. *Frontiers in Microbiology* **6**:476 DOI 10.3389/fmicb.2015.00476.

Lillis L, Doyle E, Clipson N. 2009. Comparison of DNA- and RNA-based bacterial community structures in soil exposed to 2,4-dichlorophenol. *Journal of Applied Microbiology* **107(6)**:1883–1893.

McCarthy A, Chiang E, Schmidt ML, Denef VJ. 2015. RNA preservation agents and nucleic acid extraction method bias perceived bacterial community composition. *PLOS ONE* **10(3)**:e0121659 DOI 10.1371/journal.pone.0121659.

Mohr S, Ghanem E, Smigh W, Sheeter D, Qin Y, King O, Polioudakis D, Iyer VR, Hunicke-Smigh S, Swamy S, Kuersten S, Lambowitz AM. 2013. Thermostable group II intron reverse transcriptase fusion proteins and their use in cDNA synthesis and next-generation RNA sequencing. *RNA* **19(7)**:958–970 DOI 10.1261/rna.039743.113.

Nacke H, Fischer C, Thürmer A, Meinicke P, Daniel R. 2014. Land use type significantly affects microbial gene transcription in soil. *Microbial Ecology* **67(4)**:919–930 DOI 10.1007/s00248-014-0377-6.

Nunes I, Jurburg S, Jacquiod S, Brejnrod A, Falcão Salles J, Priemé A, Sørensen SJ. 2018. Soil bacteria show different tolerance ranges to an unprecedented disturbance. *Biology and Fertility of Soils* **54(2)**:189–202 DOI 10.1007/s00374-017-1255-4.

Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, Minchin PR, O'Hara RB, Simpson GL, Solymos P, Stevens MHH, Szoecs E, Wagner H. 2018. vegan: community ecology package. *Available at* http://vegan.r-forge.r-project.org/.

Pan Y, Bodrossy L, Frenzel P, Hestnes AG, Krause S, Lüke C, Meima-Franke M, Siljanen H, Svenning MM, Bodelier PL. 2010. Impacts of inter- and intralaboratory variations on the

reproducibility of microbial community analyses. *Applied and Environmental Microbiology* **76(22)**:7451–7458 DOI 10.1128/aem.01595-10.

Paulin MM, Nicolaisen MH, Jacobsen CS, Gimsing AL, Sørensen J, Bælum J. 2013. Improving Griffith's protocol for co-extraction of microbial DNA and RNA in adsorptive soils. *Soil Biology and Biochemistry* **63**:37–49 DOI 10.1016/j.soilbio.2013.02.007.

Pfeiffer S, Pastar M, Mitter B, Lippert K, Hackl E, Lojan P, Oswald A, Sessitsch A. 2014. Improved group-specific primers based on the full SILVA 16S rRNA gene reference database. *Environmental Microbiology* **16(8)**:2389–2407 DOI 10.1111/1462-2920.12350.

Pielou EC. 1966. The measurement of diversity in different types of biological collections. *Journal of Theoretical Biology* **13**:131–144 DOI 10.1016/0022-5193(66)90013-0.

Placella SA, Brodie EL, Firestone MK. 2012. Rainfall-induced carbon dioxide pulses result from sequential resuscitation of phylogenetically clustered microbial groups. *Proceedings of the National Academy of Sciences of the United States of America* **109(27)**:10931–10936 DOI 10.1073/pnas.1204306109.

Pollock J, Glendinning L, Wisedchanwet T, Watson M. 2018. The madness of microbiome : attempting to find consensus "best practice" for 16S microbiome studies. *Applied and Environmental Microbiology* **84(7)**:1–12 DOI 10.1128/aem.02627-17.

Pratscher J, Dumont MG, Conrad R. 2011. Ammonia oxidation coupled to CO2 fixation by archaea and bacteria in an agricultural soil. *Proceedings of the National Academy of Sciences of the United States of America* **108(10)**:4170–4175 DOI 10.1073/pnas.1010981108.

Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig W, Peplies J, Glöckner FO. 2007. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Research* **35(21)**:7188–7196 DOI 10.1093/nar/gkm864.

Quammen D. 2018. *The tangled tree: a radical new history of life*. New York: Simon and Schuster.

Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO. 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Research* **41(D1)**:D590–D596 DOI 10.1093/nar/gks1219.

R Core Team. 2018. *R: a language and environment for statistical computing*. Vienna: The R Foundation for Statistical Computing. *Available at* http://www.R-project.org/.

Romanowicz KJ, Freedman ZB, Upchurch RA, Argiroff WA, Zak DR. 2016. Active microorganisms in forest soils differ from the total community yet are shaped by the same environmental factors: The influence of pH and soil moisture. *FEMS Microbiology Ecology* **92(10)**:fiw149 DOI 10.1093/femsec/fiw149.

Schwartz S, Motorin Y. 2017. Next-generation sequencing technologies for detection of modified nucleotides in RNAs. *RNA Biology* **14(9)**:1124–1137 DOI 10.1080/15476286.2016.1251543.

Segev E, Rosenberg A, Mamou G, Sinai L, Ben-Yehuda S. 2013. Molecular kinetics of reviving bacterial spores. *Journal of Bacteriology* **195(9)**:1875–1882 DOI 10.1128/jb.00093-13.

Shannon CE, Weaver W. 1949. *The mathematical theory of information.* Vol. 97. Urbana: The University of Illinois.

Sieber MW, Recknagel P, Glaser F, Witte OW, Bauer M, Claus RA, Frahm C. 2010. Substantial performance discrepancies among commercially available kits for reverse transcription quantitative polymerase chain reaction: a systematic comparative investigator-driven approach. *Analytical Biochemistry* **401(2)**:303–311 DOI 10.1016/j.ab.2010.03.007.

Ståhlberg A, Håkansson J, Xian X, Semb H, Kubista M. 2004. Properties of the reverse transcription reaction in mRNA quantification. *Clinical Chemistry* **50(3)**:509–515 DOI 10.1373/clinchem.2003.026161.

Ståhlberg A, Kubista M, Michael P. 2004. Comparison of reverse transcriptases in gene expression analysis. *Clinical Chemistry* **50(9)**:1675–1678 DOI 10.1373/clinchem.2004.035469.

Steven B, Hesse C, Soghigian J, Gallegos-Graves LV, Dunbar J. 2017. Simulated rRNA/DNA ratios show potential to misclassify active populations as dormant. *Applied and Environmental Microbiology* **83(11)**:1–11 DOI 10.1128/aem.00696-17.

Van Der Walt S, Colbert SC, Varoquaux G. 2011. The NumPy array: a structure for efficient numerical computation. *Computing in Science & Engineering* **13(2)**:22–30 DOI 10.1109/mcse.2011.37.

Wickham H. 2016. *ggplot2: elegant graphics for data analysis*. New York: Springer-Verlag.

Yilmaz SL, Okten HE, Noguera DR. 2006. Making all parts of the 16S rRNA of *Escherichia coli* accessible in situ to single DNA oligonucleotides. *Applied and Environmental Microbiology* **72(1)**:733–744 DOI 10.1128/AEM.72.1.733-744.2006.

Zhang J, Byrne CD. 1999. Differential priming of RNA templates during cDNA synthesis markedly affects both accuracy and reproducibility of quantitative competitive reverse-transcriptase PCR. *Biochemical Journal* **337(2)**:231–241.

Šťovíček et al. (2019), *PeerJ*, DOI 10.7717/peerj.7608

16/16