Assessing and reducing sources of gene expression variability in *Staphylococcus epidermidis* biofilms

Cármen Sousa, Angela França, and Nuno Cerca

CEB–Centre of Biological Engineering, LIBRO–Laboratory of Research in Biofilms Rosário Oliveira, University of Minho, Braga, Portugal

**BioTechniques** 57:295-301 (December 2014) doi 10.2144/000114238

**Keywords:** biofilms; gene expression variability; biofilm pool; RNA extraction; reverse transcriptase; qPCR

Gene expression quantification can be a useful tool in studying the properties of bacterial biofilms. Unfortunately, techniques such as RNA extraction, cDNA synthesis, and quantitative PCR (qPCR) can introduce variability into mRNA transcript measurements, obscuring biologically relevant results. Here we sought to identify the steps that impair accurate gene expression quantification from *Staphylococcus epidermidis* biofilm samples. We devised an experimental setup that could be used to determine the contribution of each experimental step to the variability of mRNA transcript measurement. Among factors tested, biofilm growth contributed the most bias to gene expression quantification. Additional experiments demonstrated that pooling biofilms together reduced this variability, resulting in more accurate gene expression analysis results. We therefore recommend pooling in order to reduce the variability associated with gene expression quantification from biofilm samples.

**Material and methods**

**Experimental workflow**

To analyze the source of variability associated with quantitative gene expression analysis of biofilms, we devised an experimental setup that allowed us to determine the contributions of biofilm growth, RNA extraction, cDNA synthesis, and qPCR run to variability. The experimental workflow is illustrated in Figure 1A. In brief, to determine the overall variability of the gene expression quantification process, RNA was isolated from four independent biofilms, and the expression of five different genes was then quantified. To exclude the contribution of the biological variability, RNA was isolated from four independent biofilms, and the expression of five different genes was then quantified. To exclude the contribution of RNA extraction, one RNA sample randomly selected from the previous four RNA extractions performed was used to synthesize four different cDNA samples. Lastly, to rule out the contribution of reverse transcriptase reaction, one cDNA sample randomly selected from the four previously synthesized was used.

**METHOD SUMMARY**

Here we identified biofilm growth as the major factor influencing the variability often observed in gene expression quantification assays. We showed that biofilm pooling resulted in more accurate, precise, and meaningful gene expression data analysis.
to perform four independent qPCR runs.

For the biofilm pooling strategy (Figure 1B), 4 independent pools of 10 or 20 biofilms were thoroughly mixed by gentle vortexing, and then each pool was divided in 2 equivalent samples. Each of these samples was used for RNA extraction, cDNA synthesis, and qPCR. Nevertheless, for the determination of the coefficient of variation (CV), we only used the values from four independent pools. The other samples were used as internal controls for the RNA extraction experiment, being the experiments validated when the variability between pairs was similar. It is important to stress that the same reagent brands, the same operator, and the same equipment were strictly maintained throughout the experiments because these factors are also known to introduce variability in gene expression quantification assays (12).

Bacterial and growth conditions

For quantification of the variability associated with gene expression assays, the *S. epidermidis* strains RP62A (ATCC 35984) (13), clinical isolate 9142 (14), and commensal isolate JI6 (15) were used. Other strains included in the study are listed in Table 1. A single colony of each strain was inoculated into 1 mL of tryptic soy broth (TSB) (Liofilchem, Roseto degli Abruzzi, Italy) and incubated overnight at 37°C with shaking at 120 rpm. Thereafter, a bacterial suspension with an optical density at a wavelength of 640 nm ($A_{640}$) of 0.250 ± 0.05 was prepared in fresh TSB, and a 15 µL aliquot of this suspension was inoculated into 1 mL TSB supplemented with 1% (v/v) glucose (Fisher Scientific, Waltham, MA) to induce biofilm formation in 24-well plates (Orange Scientific, Braine-l’Alleud, Belgium). The plates were incubated for 24 h at 37°C with shaking at 120 rpm. Before any analysis, spent medium was removed, and biofilms were washed twice and suspended in 2 mL 0.9% NaCl (AnalaR Normapur, Radnor, PA).

*S. epidermidis* isolates characterization

*S. epidermidis* isolates used for this study were selected from a collection of clinical and commensal isolates that were characterized in terms of their ability to form biofilms and for the presence of the genes of interest. In brief, biofilm formation capability was determined by $A_{640}$ as described elsewhere (16). The presence of the genes of interest was determined by PCR using DreamTaq PCR Master Mix.

Table 1. Biofilm formation ability and presence of the genes of interest in different clinical and commensal *Staphylococcus epidermidis* isolates.

<table>
<thead>
<tr>
<th>Strain</th>
<th>Genes of interest</th>
<th>Biofilm formation ($OD_{640}$)</th>
<th>Collection</th>
</tr>
</thead>
<tbody>
<tr>
<td>RP62A</td>
<td>aap, lrgB, fmtC, icaA, pgi</td>
<td>1.269 ± 0.017</td>
<td>Intravascular catheter-associated sepsis (13)</td>
</tr>
<tr>
<td>9142</td>
<td>aap, lrgB, fmtC, icaA, pgi</td>
<td>0.823 ± 0.353</td>
<td>Blood clinical isolate (14)</td>
</tr>
<tr>
<td>1457</td>
<td>aap, lrgB, fmtC, icaA, pgi</td>
<td>0.810 ± 0.243</td>
<td>Central venous catheter-associated infection (14)</td>
</tr>
<tr>
<td>M129</td>
<td>aap, lrgB, fmtC, icaA</td>
<td>0.410 ± 0.117</td>
<td>Dialysis-associated peritonitis (15)</td>
</tr>
<tr>
<td>IE186</td>
<td>aap, lrgB, fmtC, icaA</td>
<td>0.729 ± 0.218</td>
<td>Infective endocarditis (15)</td>
</tr>
<tr>
<td>FJ6</td>
<td>aap, lrgB, fmtC, icaA, pgi</td>
<td>0.384 ± 0.203</td>
<td>Skin of healthy volunteers (15)</td>
</tr>
<tr>
<td>JI6</td>
<td>aap, lrgB, fmtC, icaA, pgi</td>
<td>0.573 ± 0.119</td>
<td>Skin of healthy volunteers (15)</td>
</tr>
<tr>
<td>LE7</td>
<td>lrgB, fmtC, icaA, pgi</td>
<td>0.853 ± 0.335</td>
<td>Skin of healthy volunteers (15)</td>
</tr>
</tbody>
</table>
(Thermo Scientific, Waltham, MA). The oligonucleotide sequences of the primers used in this study are listed in Table 2, and the thermal cycler conditions were the same used for qPCR, which are described below.

RNA extraction

RNA extraction was performed as optimized previously (10). In brief, this protocol uses both chemical (phenol; AppliChem, Darmstadt, Germany) and mechanical (glass beads; Sigma-Aldrich, St. Louis, MO) lysis together with column systems for RNA isolation (E.Z.N.A Total RNA kit I, Omega Bio-Tek, Norcross, GA). Genomic DNA was digested with DNase I (Thermo Scientific) following the manufacturer’s instructions, and RNA concentration and purity were determined using a NanoDrop 1000 spectrophotometer (Thermo Scientific). The absorbance ratios A_{260}/A_{280} and A_{260}/A_{230} were used as indicators, respectively, of protein contamination and polysaccharide, phenol, or chaotropic salt contamination (17). To determine RNA integrity, 23S and 16S rRNA banding patterns were evaluated in nondenaturing gel electrophoresis. Electrophoresis was carried out at 80 V for 60 min in a 1.5% agarose gel. The gel was stained with Midori Green Advanced DNA stain (Nippon Genetics Europe, Dueren, Germany) and visualized using a ChemiDoc XRS (Bio-Rad, Hercules, CA). Because RNA integrity significantly influences the quantification of gene expression (5,6), only samples

<table>
<thead>
<tr>
<th>Target gene</th>
<th>Primer sequence (5’ to 3’)</th>
<th>Melting temperature (°C)</th>
<th>Amplicon (bp)</th>
<th>Priming efficiency (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>16S rRNA</td>
<td>FW GGGCTACACGCTGCTCAAA</td>
<td>59.79</td>
<td>176</td>
<td>93</td>
</tr>
<tr>
<td></td>
<td>RV GTACAAGACCCACGAACTGTA</td>
<td>59.85</td>
<td></td>
<td></td>
</tr>
<tr>
<td>aap</td>
<td>FW GACACCGCGCTGCTGAC</td>
<td>59.22</td>
<td>190</td>
<td>95</td>
</tr>
<tr>
<td></td>
<td>RV GCGATGCTGCTGATAGCTCA</td>
<td>59.98</td>
<td></td>
<td></td>
</tr>
<tr>
<td>icaA</td>
<td>FW TGCATCTGCGAGGGATCA</td>
<td>60.20</td>
<td>134</td>
<td>90</td>
</tr>
<tr>
<td></td>
<td>RV TAACTGCGGCTATTTGATT</td>
<td>59.99</td>
<td></td>
<td></td>
</tr>
<tr>
<td>lrgB</td>
<td>FW ATATCGACAGGCAGGAGATAT</td>
<td>59.87</td>
<td>165</td>
<td>90</td>
</tr>
<tr>
<td></td>
<td>RV ATTTGTGCTGCTGACGTT</td>
<td>59.61</td>
<td></td>
<td></td>
</tr>
<tr>
<td>pgi</td>
<td>FW TACCTGACAGGACCCACGAC</td>
<td>54.05</td>
<td>170</td>
<td>85</td>
</tr>
<tr>
<td></td>
<td>RV CATCGGATGACCCACGTC</td>
<td>53.95</td>
<td></td>
<td></td>
</tr>
<tr>
<td>fmtC</td>
<td>FW CGCCCTCATCATAGCGTTG</td>
<td>60.19</td>
<td>182</td>
<td>97</td>
</tr>
<tr>
<td></td>
<td>RV CCAATTGATACCCCCAAC</td>
<td>60.03</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
presenting comparable RNA quality, as determined by the presence of sharp bands and a 2:1 intensity ratio between 23S and 16S rRNA, were used in this study.

Complementary DNA synthesis
Complementary DNA (cDNA) was synthesized as described elsewhere (10). In brief, 0.5 μg of total RNA was converted into cDNA in the presence of the enzyme RevertAid H minus reverse transcriptase (RT) (Thermo Scientific) in an MJ Mini Gradient Thermal cycler (Bio-Rad). Random primers (NZYTech, Lisbon, Portugal) were used as the priming strategy. To determine genomic DNA carry-over and contamination of the reagents used, control reactions lacking RT (no-RT control) and the template were prepared.

Quantitative PCR run
For mRNA quantification, a previously optimized qPCR reaction was performed (10). The primers (Metabion, Steinkirchen, Germany) used were designed using Primer3 software (18) (Table 2). The experiment was performed in a CFX96 Thermal cycler (Bio-Rad) with the following cycling parameters: 10 min at 94°C followed by 40 repeats of 5 s at 94°C, 10 s at 58°C, and finally 15 s at 72°C using iQ SYBR Green supermix (Bio-Rad). Reaction efficiency was determined by the dilution method (19) and using a temperature gradient from 50° to 65°C. At 58°C, all set of primers used had the best and most similar efficiencies. RNA samples were considered free from significant genomic DNA contamination if the quantification cycle (Cq) difference between the specific signal and the respective no-RT control was greater than 10. Neither unanticipated products nor primer dimers were detectable by melting curve analysis. The quantification of the specific transcripts for each gene under study was determined using the delta Cq method (EΔCq), a variation of the Livak method (20), where ∆Cq = Cq (reference gene) - Cq (target gene) and E is the experimentally determined reaction efficiency.

Statistical analysis
The CV was determined as a measure of variability. Mean and standard deviation were determined using Microsoft Office Excel 2007.

Results and discussion
Gene expression quantification using qPCR requires the collection of biological material; RNA extraction; and cDNA synthesis, amplification, and quantification. Each of these steps can introduce variability in the quantification process, resulting in inaccurate gene expression quantification (7–10,21). We conceived a simple experimental design (Figure 1A) that allowed us to determine the relative contributions of RNA extraction, cDNA synthesis, and qPCR, as well as the impact of biofilm growth, on gene expression quantification variability. Because gene-associated particulars can impact mRNA transcript measurements in different ways, we analyzed the expression of five genes with distinct functions, plus the 16S RNA control. The genes included (i) aap (22), which is involved in protein-associated biofilm formation and accumulation; (ii) icaA, which is involved in both polysaccharide-mediated biofilm accumulation (23) and immune evasion (15); (iii) pgd, which plays a role in glucose metabolism (24); (iv) fmtC (also known as mprF), which is involved in L-lysine modification of phosphatidylglycerol and immune evasion (25); and (v) lrgB, which has been linked with programmed cell death (26). We quantified the expression of these selected genes in four independently grown biofilms and calculated the experimental variability. As depicted in Figure 2, the overall CV of our experimental setup was 61% ± 26%, with the highest disparity in pgd gene (87%) expression and the lowest in lrgB gene (27%) expression.

To exclude the influence of biological variability and to assess the variability introduced by the RNA isolation procedure, we performed four parallel RNA extractions from the same biofilm samples. Interestingly, our experimental setup revealed that RNA extraction and subsequent steps accounted for a CV of only 23% ± 10% (Figure 2). As before, gene-to-gene variation was observed, with the highest variation associated with the pgd gene (37%) and the lowest with the icaA gene (15%). Next, we examined intrinsic variability associated with cDNA synthesis process by performing four different synthesis reactions from the same randomly selected RNA sample. The overall CV observed was 26%, with the highest disparity in pgd gene (37%) and the lowest with the icaA gene (15%).
24% ± 5%, and again, the highest variation was associated with the pgI gene (30%) and the lowest with the icaA gene (17%) (Figure 2).

To rule out variability inherently associated with cDNA synthesis and determine the variability introduced during qPCR, we randomly selected one of the previously synthesized cDNAs and performed four independent qPCR amplifications. The CV determined for this experiment was 15% ± 4%, with the pgI gene again showing the greatest variation (22%) and the lrgB gene having the least variation (11%) (Figure 2). We repeated the experiment with different randomly selected RNAs and cDNAs obtaining similar results (data not shown).

It is important to emphasize that the variation introduced by each step is strictly dependent on the kit used, as different commercially available kits will exhibit differences in reproducibility (7,9) and consequently, the CVs determined here are merely representative. Nevertheless, our findings clearly indicate that among the variables studied, biofilm growth made the greatest contribution to the variability detected in gene expression quantification. Given this finding, we wanted to devise a strategy to decrease such variability.

Considering biofilms are very heterogeneous samples (11), we attempted to pool several biofilms (Figure 1B) to determine whether such a pooling strategy could decrease the variability. We pooled together either 10 or 20 biofilms that were grown in 24-well plates and then performed 2 RNA extractions from each pool. To assess the gene expression variability in the biofilm pools, we selected the gene with the highest variation detected in our initial experiments (pgi) and a gene with lower variability (aap), being the last directly involved in biofilm formation. As can be seen in Figure 3, biofilm pooling reduced the high variability associated with S. epidermidis RP62A biofilm growth (an average of 3.5-fold for the pool of 10 biofilms and 3.8-fold for the pool of 20 biofilms). Interestingly, the values obtained from the pool of 20 biofilms were as low as the variability introduced by qPCR amplification, suggesting that this pooling strategy was able to eliminate the variability introduced by the biofilm itself. When 10 biofilms were pooled, the variability detected was slightly higher, but nevertheless, there was a drastic reduction in the gene expression variability observed under our experimental conditions.

To verify that these findings were strain independent, the experiment was repeated using different S. epidermidis isolates. For these experiments, we characterized a collection of clinical and commensal isolates for the ability to form biofilm and the presence of the genes of interest (Table 1), selecting the clinical isolate 9142 and the commensal isolate J16 for the validation of our strategy. Importantly, the same trend was observed for both clinical (average of 2.6-fold reduction) and commensal isolates (average of 2.1-fold reduction), further suggesting this biofilm pooling strategy is gene and strain independent (Figure 4).

In conclusion, our data show that biofilm growth, among the studied variables, is the major factor influencing the variability of biofilm gene expression quantification assays. Application of a corrective strategy, pooling multiple biofilms, led to a meaningful decrease in variance and more accurate and feasible gene expression analysis. Although our experimental design was validated for qPCR gene quantification, this strategy should also prove valid for biofilm transcriptomic analysis using RNA sequencing and microarrays.

Author contributions

N.C. designed the experiments. C.S. and A.F. carried out the laboratory experiments. C.S., A.F., and N.C. analyzed the data, interpreted the results, discussed analyses, interpretation, and presentation, and wrote the paper. All authors have contributed to, seen, and approved the manuscript.

Acknowledgments

We would like to thank Kimberly K. Jefferson at Virginia Commonwealth University for reviewing the manuscript. This work was co-funded by Fundação para a Ciência e a Tecnologia (FCT) and COMPETE grants PTDC/BIA-MIC/113450/2009 and FCOMP-01-0124-FEDER-014309, FCT Strategic Project PEst-OE/EBB/LA0023/2013, and FCT project RECI/BBB-EBI/0179/2012 (FCOMP-01-0124-FEDER-027462), and by QREN, FEDER, ON2 project. N.C. is an Investigador FCT.

Competing interests

The authors declare no competing interests.

References

epidermidis biofilms: a comparison between commercial kits and a customized protocol. PLoS ONE 7:e37480.


Received 27 August 2014; accepted 8 October 2014.

Address correspondence to Nuno Cerca, CEB–Centre of Biological Engineering, University of Minho, Braga, Portugal. E-mail: nuncerca@ceb.uminho.pt

To purchase reprints of this article, contact: biotechniques@fosterprinting.com