Choosing reference genes for RT-qPCR for *Fusarium graminearum* plant infection (*In Planta*) and *In Vitro* growth studies based on transcriptomic data

Xiaorong Lin 1, 2, Hongchen Li 1, Zonghua Wang 1, 2, Stefan Olsson 1

1 Fujian University Key Laboratory for Plant-Microbe Interaction, Fujian Agriculture and Forestry University, Fuzhou, Fujian Province, China
2 Institute of Oceanography, Minjiang University, Fuzhou, Fujian Province, China

Corresponding Authors: Zonghua Wang, Stefan Olsson
Email address: wangzh@fafu.edu.cn, stefan@olssonstefan.com

**Background.** Choosing reference genes for RT-qPCR for the study of transcriptomic responses of target genes is often done using “standard” reference genes (housekeeping genes) selected before the genomic era. Now, published transcriptome data can be used to aid in this selection to avoid the selection of a reference gene that varies and obscure results.

**Methods.** We use transcriptome data for the model pathogen fungus *Fusarium graminearum* to select housekeeping genes for *In Vitro* and *In Planta* conditions. Transcriptome data was downloaded from a publicly available database. We selected a database where transcriptome chip data from many experiments using the same chip has been deposited divided the downloaded data into *In Vitro* and *In Planta* conditions based on the information about the experiments.

**Results.** We ranked the genes with the least variation (relative difference between maximum and minimum expression) across each dataset. Genes previously shown to perform well as reference genes for *In Vitro* conditions in a similar analysis as ours also performed well for *In Vitro* conditions in our dataset but worked less well for *In Planta* conditions. We found 5 reference genes that performed well under both *In Planta* conditions and *In Vitro* conditions.

**Discussion.** Even if these 5 reference genes performed well, for other (new) target conditions we recommend making a transcriptome analysis to select well performing reference genes for RT-qPCR if possible. Alternatively, select 2 of the 5 genes that we show here performed well under both *In Planta* and *In Vitro* conditions.
Choosing reference genes for RT-qPCR for *Fusarium graminearum* plant infection (*In Planta*) and *In Vitro* growth studies based on transcriptomic data.

Xiaorong Lin¹², Hongchen Li¹, Zonghua Wang¹², Stefan Olsson¹

¹ Fujian University Key Laboratory for Plant-Microbe Interaction, Fujian Agriculture and Forestry University, Fuzhou, 350002, China.
² Institute of Oceanography, Minjiang University, Fuzhou, 350108, China.

Corresponding Authors:
Stefan Olsson¹
Fujian Agriculture and Forestry University, Fuzhou 350002, China
Email address: stefan@olssonstefan.com
Zonghua Wang¹
Fujian Agriculture and Forestry University, Fuzhou 350002, China
Email address: wangzh@fafu.edu.cn

Abstract

**Background.** Choosing reference genes for RT-qPCR for the study of transcriptomic responses of target genes is often done using “standard” reference genes (housekeeping genes) selected before the genomic era. Now, published transcriptome data can be used to aid in this selection to avoid the selection of a reference gene that varies and obscure results.

**Methods.** We use transcriptome data for the model pathogen fungus *Fusarium graminearum* to select housekeeping genes for *In Vitro* and *In Planta* conditions. Transcriptome data was downloaded from a publicly available database. We selected a database where transcriptome chip data from many experiments using the same chip has been deposited divided the downloaded data into *In Vitro* and *In Planta* conditions based on the information about the experiments.

**Results.** We ranked the genes with the least variation (relative difference between maximum and minimum expression) across each dataset. Genes previously shown to perform well as reference genes for *In Vitro* conditions in a similar analysis as ours also performed well for *In Vitro* conditions in our dataset but worked less well for *In Planta* conditions. We found 5 reference genes that performed well under both *In Planta* conditions and *In Vitro* conditions.

**Discussion.** Even if these 5 reference genes performed well, for other (new) target conditions we recommend making a transcriptome analysis to select well performing reference genes for RT-
qPCR if possible. Alternatively, select 2 of the 5 genes that we show here performed well under both In Planta and In Vitro conditions.

Introduction
To investigate a set of target genes expression during a set of target conditions there is a need for RT-qPCR reference genes (housekeeping genes) that are stably expressed under all investigated target conditions (Czechowski, 2005; Eisenberg & Levanon, 2013; Stanton et al., 2017; Carmona et al., 2017; Gao et al., 2018). Traditional reference genes selected before the genomic era has been shown to be far from stable under many conditions, and choosing them can obscure the results (Eisenberg & Levanon, 2013; Stanton et al., 2017). Methods for selecting good reference genes have been devised although these often relies on qPCR data which produces a recursive problem since the same method is used to evaluate if the reference genes are working well as references (Carmona et al., 2017). An alternative strategy using publicly deposited RNAseq data in SRA datafiles have been devised (Carmona et al., 2017) to select stably expressed reference genes under different conditions. In a relative recent study aiming to find stable and reliable reference genes for Fusarium graminearum (a model organism in plant pathology) under In Vitro conditions it was found that GzUBH (FGSG_01231) and EF1A (FGSG_08811) showed the best performance (Kim & Yun, 2011).

As far as we know a similar analysis has not been done for Fusarium graminearum for In Planta conditions and is thus the purpose of this study. It is also of interest to select housekeeping genes with a set of different relative expression so as to match these with the expression of the target genes under target conditions. Affymetrix chips for transcriptomic analysis were made for F. graminearum relatively early and many experiments by different labs were carried out with the same chip (FusariumPLEX http://www.plexdb.org/modules/PD_general/pathogens_list.php). This data is publicly available and is thus an alternative to RNAseq data for this fungus for evaluating stable expressed reference genes for In Vitro and In Planta conditions.

Materials & Methods
Transcriptome data download and preparation
Transcriptome data for Fusarium graminearum (Gibberella zeae (Schwein.) Petch), strain Ph1 was downloaded from PLEX DB. We used the datasets FusariumPLEX (http://www.plexdb.org/modules/PD_general/pathogens_list.php). The probe id:s were translated to gene ID following the BROAD protein annotation (fusarium_graminearum_ph-1_3_proteins.fasta.gz downloaded from ftp.broadinstitute.org, path: /distribution/annotation/fungi/fusarium/genomes/fusarium_graminearum_ph-1). The translation of the probe-set to the BROAD database FGSG-codes is found in the Supplemental Data S1. The chip experiment data (Supplemental Table S1 Showing a listing experiments) were split in two sets, In Planta related data and In Vitro related data and two Excel sheet matrixes with the
different transcriptome expression data were produced, *In Planta* expression (Supplemental Data S2) and *In Vitro* expression (Supplemental Data S3).

**Finding genes with the least varying gene expression in the two datasets and handling of the results.**

A simple method was used to find genes with least variation of relative expression in the two datasets. Since gene expression of a particular gene can be normally distributed or have many other distributions, we chose to look for genes with least relative difference between maximum and minimum expression. Thus, we calculated for the set of treatments maximum expression minus minimum gene expression divided by average gene expression ((Max-Min)/Average) and ranked the genes with the lowest values as the least varying genes. Plots illustrating the results were prepared in Excel or in the statistics freeware PAST (https://folk.uio.no/ohammer/past/).

**Results**

The transcriptomic responses of genes were evaluated for the *In Planta* and *In Vitro* datasets (Supplemental Data S2 and S3). The *In Planta* datasets contains 64 full transcriptome datasets and the *In Vitro* contains 98 datasets from different experiments. We used a simple method for detecting which genes showed least relative difference between maximum and minimum expression in respective datasets (see methods). We then ranked the genes for their suitability as reference genes for qPCR and compared the values for the found genes with the values found for commonly used reference genes for *F. graminearum*, gamma-actin (FGSG_07335) (Brown et al., 2011), GAPDH (FGSG_06257) (Kim & Yun, 2011; Harris et al., 2016), EF1A (FGSG_08811) (Kim & Yun, 2011; Harris et al., 2016) and B-Tubulin (FGSG_09530) (Kim & Yun, 2011; Harris et al., 2016). We plotted the average expression level versus relative difference between maximum and minimum expression for all genes for both the *In Planta* and the *In Vitro* data (Supplemental Figures 1 and 2 and Supplemental Data S2 and S3). To highlight the genes of most interest Fig. 1 show the average expression level versus relative difference between maximum and minimum expression for the least varying genes at different levels of average expression for the *In Planta* experiments. A similar plot for the *In Vitro* experiments is shown in Fig. 2. In both plots we have inserted or marked commonly used reference genes. The 4 common reference genes performed well for the *In Vitro* dataset (Fig. 2) but were not very good as reference genes for *In Planta* conditions (Fig. 1). To find genes that performed relatively well as housekeeping genes under both *In Planta* and *In Vitro* conditions we ranked the genes for their suitability as reference genes under the two conditions and plotted the found ranks for the genes (Fig. 3). As can be seen in Fig. 3 only five genes were performing well as reference genes under both *In Planta* and *In Vitro* conditions. To further illustrate the difference between *In Planta* and *In Vitro* conditions we investigated how much overlap it was among the 100 top ranked genes under both conditions and found that only 20 genes were found on both lists (Fig. 4 and Supplemental Table S2). Table 1 lists the 10 best reference genes for each condition as well
as their rank under the other condition and a list of the 5 genes that performed well under both conditions (Se also Figure 3). Table 1 also shows that the genes with least varying gene expression for the In Planta conditions were more likely to perform well also under the In Vitro conditions than the reverse, and most interestingly EF1A (FGSG_08811) that varied least in expression in the In Vitro dataset did not perform well In Planta.

Discussion

As often found, the transcription of genes involved in transcription, translation, and protein quality control are often among the ones that are most stably related to general transcription level and thus perform well as reference genes (Eisenberg & Levanon, 2013; Carmona et al., 2017; Gao et al., 2018). The two reference genes previously found to be most reliable for F. graminearum In Vitro conditions (Kim & Yun, 2011) also performed very well in our In Vitro dataset, thus giving support to our approach. In our In Vitro datasets and EF1A was also a top performing reference gene with little variation between treatments.

However, EF1A was not a top performing gene under In Planta conditions (Table 1) illustrating that although many conditions were used in the In Vitro dataset, In Vitro growth only shows a limited repertoire of gene expression variation. Interestingly, we found that genes performing well as reference genes under In Planta conditions more frequently performed well also under In Vitro conditions than the reverse. This could indicate that In Vitro conditions are more likely to expose only a limited repertoire of gene expression regulations. Thus, In Vivo the pathogen appears to display more variable gene expression for most genes than In Vitro (see conceptual model in Fig. 5) and that most In Vitro patterns of expression are found also In Vivo.

Conclusions

Our results can be summarized in the following conclusions concerning selection of reference genes for F. graminearum RT-qPCR-studies.

i. Use published transcriptome data to find reference genes if this data is available for the target conditions and use simple techniques similar to what is used here or techniques for more advanced and automatic analysis (Carmona et al., 2017).

ii. If no published suitable transcriptome data is available for your conditions of interest, generate such data and analyze it as in (i).

iii. If no published data is available for the conditions of interest and b. is not possible due to time and money constrains, choose at least two reference genes from the five that are here shown to perform well both In Planta and In Vitro. Then choose genes involved in different processes. Hopefully, analysis of gene expression of target genes in relation to both these reference genes will give similar conclusions.

Acknowledgements
This work was supported by Fujian Agriculture and Forestry University and The 100 Talent Program of Fujian Province.

References


Figure legends

Figure 1. Plot showing genes with low relative difference between maximum and minimum expression for different average expression levels in the *In Planta* dataset. Red dots indicate the four standard reference genes (Table 1). X-axis=Log2 for average expression of each gene for the *In Planta* data. Y-axis= Log scale for Log2 variation values as Log2 maximum relative difference in expression times 100.

Figure 2. Plot showing genes with low relative difference between maximum and minimum expression for different average expression levels in the *In Vitro* dataset. Red dots indicate the four standard reference genes (Table 1). X-axis=Log2 for average expression of each gene for the *In Vitro* data. Y-axis= Log scale for Log2 variation values as Log2 maximum relative difference in expression times 100.

Figure 3. Plot showing rank for the relative difference between maximum and minimum expression for the best performing reference genes in the *In Planta* dataset versus the rank in the *In Vitro* dataset. The 5 genes in the lower left corner are the genes that performs best in both datasets (see also Table 1).

Figure 4. Illustrates how much overlap (red) it is among the 100 best reference genes in the *In Planta* (green) dataset and The *In Vitro* dataset (blue).

Figure 5. Thought model of number of conditions *F. graminearum* can grow under in relation to the two datasets. Black= Number of conditions *F. graminearum* can grow under. Green=Number of conditions *F. graminearum* is exposed to in the *In Planta* dataset. Red=Number of conditions *F. graminearum* is exposed to in the *In Vitro* dataset.
Table 1 (on next page)

Genes with least varying gene expression for 3 different set of conditions. *In Planta, In Vitro, In Planta AND In Vitro.*
Table 1. Genes with least varying gene expression for 3 different set of conditions. *In Planta, In Vitro, In Planta AND In Vitro*

<table>
<thead>
<tr>
<th>Id</th>
<th>Rank <em>In Vitro</em></th>
<th>Rank <em>In Planta</em></th>
<th>Annotation</th>
</tr>
</thead>
<tbody>
<tr>
<td>FGSG_08811</td>
<td>1</td>
<td>207</td>
<td>elongation factor-1</td>
</tr>
<tr>
<td>FGSG_04289</td>
<td>2</td>
<td>5</td>
<td>probable histone H4</td>
</tr>
<tr>
<td>FGSG_10733</td>
<td>3</td>
<td>16</td>
<td>probable ribosomal protein S28</td>
</tr>
<tr>
<td>FGSG_11627</td>
<td>4</td>
<td>629</td>
<td>probable HTA2 - histone H2A.2</td>
</tr>
<tr>
<td>FGSG_08743</td>
<td>5</td>
<td>89</td>
<td>conserved hypothetical protein</td>
</tr>
<tr>
<td>FGSG_02523</td>
<td>6</td>
<td>4804</td>
<td>probable IgE-dependent histamine-releasing factor</td>
</tr>
<tr>
<td>FGSG_01425</td>
<td>7</td>
<td>2530</td>
<td>probable H+-transporting ATPase</td>
</tr>
<tr>
<td>FGSG_10089</td>
<td>8</td>
<td>1026</td>
<td>related to sporulation-specific gene SP52</td>
</tr>
<tr>
<td>FGSG_01504</td>
<td>9</td>
<td>913</td>
<td>probable ribosomal protein L31.e.A cytosolic</td>
</tr>
<tr>
<td>FGSG_10235</td>
<td>10</td>
<td>4650</td>
<td>related to rasp f 7 allergen</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Id</th>
<th>Rank <em>In Planta</em></th>
<th>Rank <em>In Vitro</em></th>
<th>Annotation</th>
</tr>
</thead>
<tbody>
<tr>
<td>FGSG_09733</td>
<td>1</td>
<td>19</td>
<td>related to 20S proteasome maturation factor</td>
</tr>
<tr>
<td>FGSG_00656</td>
<td>2</td>
<td>62</td>
<td>related to F1F0-ATP synthase subunit G</td>
</tr>
<tr>
<td>FGSG_06921</td>
<td>3</td>
<td>23</td>
<td>probable RPL39 - 60S large subunit ribosomal protein L39.e</td>
</tr>
<tr>
<td>FGSG_10001</td>
<td>4</td>
<td>214</td>
<td>probable 13 kD U4/U6.U5 snRNP associate protein Snu13</td>
</tr>
<tr>
<td>FGSG_04289</td>
<td>5</td>
<td>2</td>
<td>probable histone H4</td>
</tr>
<tr>
<td>FGSG_02461</td>
<td>6</td>
<td>142</td>
<td>probable MDH1 - malate dehydrogenase precursor mitochondrial</td>
</tr>
<tr>
<td>FGSG_09667</td>
<td>7</td>
<td>17</td>
<td>ubiquinol cytochrome c reductase 8.5 kDa subunit</td>
</tr>
<tr>
<td>FGSG_09545</td>
<td>8</td>
<td>75</td>
<td>conserved hypothetical protein</td>
</tr>
<tr>
<td>FGSG_01897</td>
<td>9</td>
<td>1793</td>
<td>related to microsomal glutathione S-transferase 3</td>
</tr>
<tr>
<td>FGSG_06021</td>
<td>10</td>
<td>13</td>
<td>ADP/ATP carrier</td>
</tr>
</tbody>
</table>

Top 5 among top 20 found within top 100 of both *in Planta* and *in Vitro*:

<table>
<thead>
<tr>
<th>Rank</th>
<th>Rank <em>In Vitro</em> × rank <em>In Planta</em></th>
<th>Annotation</th>
</tr>
</thead>
<tbody>
<tr>
<td>FGSG_04289</td>
<td>1 × 10</td>
<td>probable histone H4</td>
</tr>
<tr>
<td>FGSG_09733</td>
<td>2 × 19</td>
<td>related to 20S proteasome maturation factor</td>
</tr>
<tr>
<td>FGSG_10733</td>
<td>3 × 48</td>
<td>probable ribosomal protein S28</td>
</tr>
<tr>
<td>FGSG_06921</td>
<td>4 × 69</td>
<td>probable RPL39 - 60S large subunit ribosomal protein L39.e</td>
</tr>
<tr>
<td>FGSG_06021</td>
<td>5 × 130</td>
<td>ADP/ATP carrier</td>
</tr>
</tbody>
</table>
Figure 1 (on next page)

Plot showing genes with low relative difference between maximum and minimum expression for different average expression levels in the *In Planta* dataset.

Red dots indicate the four standard reference genes (Table 1). X-axis=Log2 for Average expression of each gene for the *In Planta* data. Y-axis= Log scale for Log2 variation values as Log2 maximum relative difference in expression times 100.
Figure 1. Plot showing genes with low relative difference between maximum and minimum expression for different average expression levels in the In Planta dataset. Red dots indicate the four standard reference genes (Table 1). X-axis=Log2 for Average expression of each gene for the In Planta data. Y-axis=Log scale for Log2 variation values as Log2 maximum relative difference in expression times 100.
Plot showing genes with low relative difference between maximum and minimum expression for different average expression levels in the *In Vitro* dataset.

Red dots indicate the four standard reference genes (Table 1). X-axis=Log2 for Average expression of each gene for the *In Vitro* data. Y-axis= Log scale for Log2 variation values as Log2 maximum relative difference in expression times 100.
Figure 2. Plot showing genes with low relative difference between maximum and minimum expression for different average expression levels in the *In Vitro* dataset. Red dots indicate the four standard reference genes (Table 1). X-axis=Log2 for Average expression of each gene for the *In Vitro* data. Y-axis= Log scale for Log2 variation values as Log2 maximum relative difference in expression times 100
Figure 3 (on next page)

Plot showing rank for the relative difference between maximum and minimum expression for the best performing reference genes in the *In Planta* dataset versus the rank in the *In Vitro* dataset.

The 5 genes in the lower left corner are the genes that performs best in both datasets (see also Table1).
Figure 3. Plot showing rank for the relative difference between maximum and minimum expression for the best performing reference genes in the *In Planta* dataset versus the rank in the *In Vitro* dataset. The 5 genes in the lower left corner are the genes that perform best in both datasets (see also Table1).
**Figure 4** (on next page)

Illustrates how much overlap (red) it is among the 100 best reference genes in the *In Planta* (green) dataset and The *In Vitro* dataset (blue).
Figure 4. Illustrates how much overlap (red) it is among the 100 best reference genes in the *In Planta* (green) dataset and The *In Vitro* dataset (blue).
Thought model of number of conditions *F. graminearum* can grow under in relation to the two datasets.

Black = Number of conditions *F. graminearum* can grow under. Green = Number of conditions *F. graminearum* is exposed to in the *In Planta* dataset. Red = Number of conditions *F. graminearum* is exposed to in the *In Vitro* dataset.
Figure 5. Thought model of number of conditions *F. graminearum* can grow under in relation to the two datasets. Black= Number of conditions *F. graminearum* can grow under. Green= Number of conditions *F. graminearum* is exposed to in the *In Planta* dataset. Red=Number of conditions *F. graminearum* is exposed to in the *In Vitro* dataset.