WILLIAM & MARY
CHARTERED 1693

W&M ScholarWorks

Arts & Sciences Articles

Arts and Sciences

4-2017

A Whole-Transcriptome Approach to Evaluating Reference Genes for Quantitative Gene Expression Studies: A Case Study in Mimulus

Kimmy A. Stanton

Patrick P. Edger

Joshua R. Puzey
College of William and Mary, jrpuzey@wm.edu

Taliesin Kinser
College of William and Mary

Philip Cheng

See next page for additional authors

Follow this and additional works at: https://scholarworks.wm.edu/aspubs

Authors

Kimmy A. Stanton, Patrick P. Edger, Joshua R. Puzey, Taliesin Kinser, Philip Cheng, Daniel M. Vernon, Nancy R. Forsthoefel, and Arielle M. Cooley

1 # A Whole Transcriptome Approach to Evaluating Reference

2 # Genes for Quantitative Gene Expression Studies: A Case Study

3 # in *Mimulus*

4

5 **Kimmy A. Stanton, Patrick P. Edger, Joshua R. Puzey, Taliesin Kinser, Philip Cheng,**

6 **Daniel M. Vernon, Nancy R. Forsthoefel, and Arielle M. Cooley**

7

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25

26

27   **ABSTRACT**

28   While qPCR is widely recognized as among the most accurate of methods for

29   quantifying gene expression, it is highly dependent on the use of reliable, stably

30   expressed reference genes. With the increased availability of high-throughput methods

31   for measuring gene expression, whole transcriptome approaches may be increasingly

32   utilized for reference gene selection and validation. In this study, RNA-seq was used to

33   identify a set of novel qPCR reference genes and also to evaluate a panel of traditional

34   "housekeeping" reference genes in two species of the evolutionary model plant genus

35   *Mimulus*. More broadly, the methods proposed in this study can be used to harness the

36   power of transcriptomes to identify appropriate reference genes for qPCR in any study

37   organism, including emerging and non-model systems. We find that RNA-seq

38   accurately estimates gene expression means in comparison to qPCR, and that

39   expression means are robust to moderate environmental and genetic variation.

40   However, measures of expression variability were only in agreement with qPCR for

41   samples obtained from a shared environment. This result, along with transcriptome-

42   wide comparisons, suggests that environmental changes have greater impacts on

43   expression variability than on expression means. We discuss how this issue can be

44   addressed through experimental design, and suggest that the ever-expanding pool of

45   published transcriptomes represents a rich and low-cost resource for developing better

46   reference genes for qPCR.

47

48

49

50

51

52

53

INTRODUCTION

54

55   Quantitative PCR (qPCR) is the premier method for quantifying gene expression

56   because of its simplicity, accuracy, and low cost. However, the quantification accuracy

57   of qPCR is dependent on normalization against reference genes to reduce the impact of

58   technical noise and variation in sample preparation. qPCR data normalization is crucial

59   for the reliable quantification of expression levels, so care must be taken to choose a

60   reliable reference gene that has low variation in expression across diverse sample types

61   (DHEDA *et al.* 2005; GUTIERREZ *et al.* 2008). Traditionally, high expression

62   "housekeeping" genes involved in basic cellular functions were used for qPCR

63   normalization based on the assumption that they would be stably expressed (THELLIN *et*

64   *al.* 1999). Unfortunately, these traditional "housekeeping" reference genes, such as

65   ubiquitin-conjugating enzyme (UBC), polyubiquitin (UBQ), β-actin, α- and β-tubulin,

66   and glyceraldehyde 3-phosphate dehydrogenase (GAPDH), can exhibit surprisingly

67   high expression variance in some species, or among different environmental conditions

68   (SUZUKI *et al.* 2000; BRUNNER *et al.* 2004; DHEDA *et al.* 2004; CZECHOWSKI *et al.* 2005).

69   In efforts to find alternatives to housekeeping genes, high-throughput

70   technologies have been used to survey whole transcriptomes for novel, stably expressed

71   genes. Microarrays have been successfully used for novel reference gene identification

72   in a variety of plants, including *Arabidopsis thaliana*, *Eucalyptus*, and soybean

73   (CZECHOWSKI *et al.* 2005; LIBAULT *et al.* 2008; DE OLIVEIRA *et al.* 2012). However, RNA-seq,

74   a potentially more effective high-throughput method, has rarely been employed. RNA-

75   seq has many advantages over microarrays: it does not require an assembled genome

76   (HAAS AND ZODY 2010; ROBERTSON *et al.* 2010; GRABHERR *et al.* 2011), it has the power to

77   identify novel transcripts and splice variants (TRAPNELL *et al.* 2010), and it is sensitive

78   enough to quantify transcripts with very low expression levels (MARIONI *et al.* 2008). In

79   addition, RNA-seq is fast, relatively inexpensive, and shows minimal variation across

80   technical replicates (MARIONI *et al.* 2008; MORTAZAVI *et al.* 2008; NAGALAKSHMI *et al.*

81 2008; WANG *et al.* 2009). For all of these reasons, RNA-seq is an attractive, whole-

82 transcriptome method for the detection of stably expressed genes and the identification

83 of novel reference genes for qPCR normalization. This approach has rarely been used to

84 evaluate potential qPCR reference genes (but see CHANG *et al.* 2012; YANG *et al.* 2014;

85 ZHUANG *et al.* 2015).

86   A potential pitfall of both the microarray and the RNA-seq approach to reference

87 gene selection is that there are no accepted practices for the analysis of expression

88 variability within whole transcriptomes. Many methods for analyzing expression

89 variability from qPCR data have been developed, including geNorm, BestKeeper, and

90 NormFinder (VANDESOMPELE *et al.* 2002; ANDERSEN *et al.* 2004; PFAFFL *et al.* 2004), but

91 these programs can only analyze the expression data from a handful of genes at a time

92 and thus are not useful for exploring whole transcriptomes. Without an established

93 method for analysis, many diverse methods have been adopted for estimating

94 expression variability within whole transcriptomes, including coefficient of variation

95 (CV) calculations (CZECHOWSKI *et al.* 2005), fold change cut-offs (YANG *et al.* 2014), and

96 p-value cut-offs (LIBAULT *et al.* 2008). However, no comparison of the different methods

97 is currently available; each of the earlier studies included only a single whole-

98 transcriptome measure of expression variability.

99   One system in which a transcriptomic approach to reference gene selection has

100 great potential to advance gene expression studies is the monkeyflower genus *Mimulus*

101 *(*recently split into genera *Mimulus* and *Erythranthe* (BARKER *et al.* 2012)). *Mimulus* has

102 become a widely used model for evolutionary genetic studies because of its phenotypic,

103 ecological, and genetic variation, with centers of species diversity in both North and

104 South America (BEARDSLEY AND OLMSTEAD 2002; WU *et al.* 2008; SOBEL AND STREISFELD

105 2013; TWYFORD *et al.* 2015). *Mimulus* is a powerful system for genetic studies due to the

106 interfertility of diverse species and the availability of genomic resources, including the

107 genome sequence of *M. guttatus*, *M. cardinalis*, *M. lewisii*, and *M. luteus* (HELLSTEN *et al.*

108   2013; YUAN *et al.* 2013; EDGER *et al.* in revision). Yet, despite the utility of *Mimulus* for

109   studying the evolution of genes and gene expression, the only evaluation of qPCR

110   reference genes to date is a non-quantitative study limited to only six housekeeping

111   genes (SCOVILLE *et al.* 2011). A rigorous and quantitative genome-wide analysis of

112   candidate qPCR reference genes is therefore of special utility for advancing

113   evolutionary genetic studies in *Mimulus*.

114   In this study, we systematically and quantitatively evaluate a panel of traditional

115   reference genes and screen whole transcriptomes to identify a set of novel reference

116   genes that can be used for qPCR expression studies in *Mimulus.* We utilize whole

117   transcriptome RNA-seq libraries from two species: *Mimulus guttatus*, a North American

118   diploid, and *Mimulus luteus* var. *luteus*, a Chilean allotetraploid (MUKHERJEE AND

119   VICKERY 1962; VALLEJO‐MARÍN *et al.* 2015). We further develop the toolkit for

120   transcriptome-enabled reference gene selection, by comparing the utility of two distinct

121   methods – the "coefficient of variation (CV) method" and the "fold change cut-off

122   method" – for identifying novel stably expressed genes from RNA-seq data.

123   In these two *Mimulus* species that differ in their ecology, ploidy, and level of

124   resource development, we find that both the CV and fold-change methods identify a

125   similar set of novel reference genes. We propose that these highly stable genes provide

126   a good starting pool of candidate reference genes for qPCR expression studies in

127   *Mimulus*, and report that some traditional reference genes are also satisfactory

128   according to standard quantitative guidelines for qPCR. In addition, we propose a

129   workflow that incorporates either the CV or the fold-change method to screen whole

130   transcriptomes for novel reference genes in other systems. Across environmentally and

131   genetically different plants, we found that gene expression means were relatively

132   similar but expression variability fluctuated dramatically. Based on this finding, we

133   suggest that transcriptomes should either be specific to the samples used for the

134   planned qPCR study or should cover a wide span of biological and environmental

135   diversity, in order for reference genes to be selected with high confidence.

136

137   **MATERIALS AND METHODS**
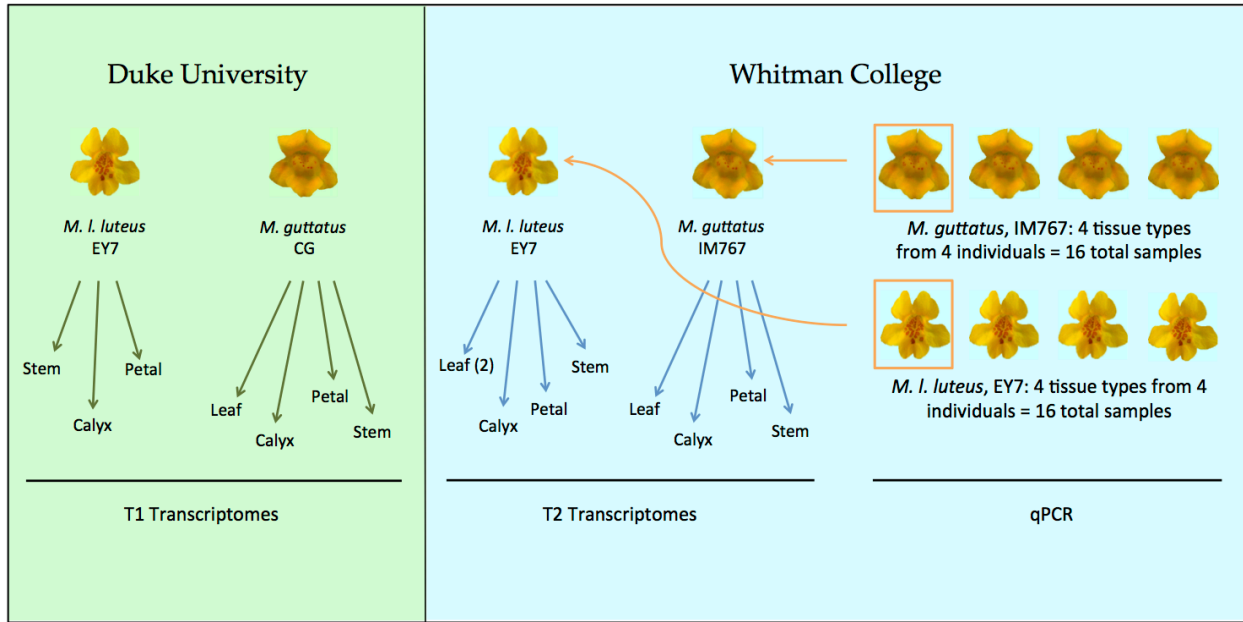
138   **Plant Materials**

139   Two batches of each species were grown in separate greenhouses, providing the RNA

140   samples for both RNA-seq and qPCR (Figure 1). *Mimulus guttatus* genotype CG (wild-collected

141   in Dublane, Scotland) and *Mimulus* luteus var. *luteus* inbred line EY7 (El Yeso, Chile, see Cooley

142   et al. (2008)) were grown at Duke University (NC, USA). RNA, from four tissue types of a single

143   individual of each species grown at Duke University, was sequenced to produce the first set of

144   transcriptomes (T1). *Mimulus guttatus* inbred line IM767 (Iron Mountain, OR, see Willis (1999))

145   and *M. l. luteus* inbred line EY7 (El Yeso, Chile) were grown at Whitman College (WA, USA).

146   RNA from four tissue types of a single individual of each species grown at Whitman College

147   was sequenced to produce the second set of transcriptomes (T2). RNA from four tissue types of

148   four individuals (one of which was the same individual used for the T2 transcriptomes) from

149   each species grown at Whitman College was extracted for use in qPCR.

150   In the Whitman greenhouse, plants were grown with supplemental 14-hour lighting in

151   Miracle-Gro potting soil (N:P:K = 0.21:0.11:0.16, The Scotts Company, Marysville, Ohio, U.S.A.).

152   Plants were maintained on 'self-watering' capillary action flats with once daily top-watering.

153   Greenhouse temperatures, as recorded by a wall sensor, ranged from 16°C to 36°C daily. Plants

154   were fertilized twice weekly with Miracle-Gro Bloom Booster (N:P:K = 15:30:15). In the Duke

155   greenhouse, plants were grown with supplemental 16-hour lighting with twice-daily watering.

156   Greenhouse temperatures ranged from 12°C to 21°C daily. Plants were fertilized with Peter's

157   Professional fertilizer every two weeks, alternating between general purpose (N:P:K = 20:10:20)

158   and low-phosphorus (N:P:K = 15:0:15) formulas, and fertilized with Jack's Classic Blossom

159   Booster (JR Peters INC, PA, USA) (N:P:K = 10:30:20) every week to enhance flowering.

160   Tissue was harvested from young, budding plants, usually between the first and third

161   flower. Four tissue types were collected: young leaf (less than 2.5 cm) near apical and lateral

162   meristems, whole calyx from unemerged buds, petal (with stamen and pistil removed) from

163 unemerged buds, and stem (around 2.5 cm segments) from newer plant growth. Tissue samples

164 were flash frozen in liquid nitrogen and stored at -80°C until the date of RNA extraction.

165

166

Figure 1

Sources of the plant materials that provided RNA for RNA-seq and qPCR. *Mimulus guttatus* genotype CG

was wild-collected in Dublane, Scotland; *M. l. luteus* highly inbred line EY7 was originally collected from

El Yeso, Chile (COOLEY *et al.* 2008); *M. guttatus* highly inbred line IM767 was originally collected from Iron

Mountain, OR (WILLIS 1999).

172

Transcriptome Preparation and Assembly

Total RNA was isolated from four tissue types (stem, leaf, calyx, petal) from both *M.*

guttatus and *M. l. luteus*. At Whitman, the Agilent Plant RNA Isolation Kit (Santa Clara, CA)

was used, and at Duke, the Zymo Research Direct-Zol RNA MiniPrep (Irvine, CA) was used,

following the manufacturer's protocol, with on-column DNase I and elution in nuclease free

water heated to 65°C. RNA concentration and integrity were assessed using a NanoDrop Lite

spectrophotometer (Thermo Fisher Scientific, DE, U.S.A) or Qubit fluorometer (Thermo Fisher

Scientific, DE, U.S.A.).

181   Whole transcriptome, RNA-seq libraries were constructed for four tissue types from

182 each of two biological replicates (T1 and T2) for both *M. guttatus* and *M. l. luteus* (see Figure 1).

183 T1 transcriptomes were prepared using the TruSeq RNA kit (Illumina; San Diego, CA) and then

184 sequenced with single-end 100bp reads using one lane of an Illumina HiSeq-2000 at the

185 University of Missouri DNA core. T2 transcriptomes were prepared using the Kapa Stranded

186 mRNA-Seq kit (Kapa Biosystems; Wilmington, MA) and were sequenced using one lane of an

187 Illumina Hiseq 2500 at the Duke University DNA core.

188   All Illumina reads were quality filtered using NextGENe v2.3.3.1 (SoftGenetics; State

189 College, PA). Adapter sequences and reads with a median quality score of less than 22 were

190 removed; reads were trimmed at positions that had three consecutive bases with a quality score

191 of less than 20; and any trimmed reads with a total length less than 40bp were removed. This

192 resulted in ~87.9% of the reads passing the quality-score filter. Expression levels, in FPKM

193 (fragments per kilobase per million reads), were determined for a total of 25,465 genes in *M.*

194 *guttatus* (diploid) and 46,855 genes in *M. l. luteus* (tetraploid). Quality-filtered reads for each

195 library were aligned to the respective genomes using NextGENe v2.3.3.1. Only uniquely

196 mapped reads were counted, using the following parameters: A. matching Requirement: >40

197 Bases and >99%, B. Allow Ambiguous Mapping: FALSE, and C. Rigorous Alignment: TRUE.

198 This resulted in the alignment of over 74.4 million reads to the diploid *M. guttatus* genome and

199 107.4 million reads to the tetraploid *M. l. luteus* genome.

200   Genome completeness of the allotetraploid *M. l. luteus* in terms of gene content was

201 assessed using BUSCO (SIMÃO *et al.* 2015) with the default setting and a set of universal single-

202 copy orthologs. The vast majority of BUSCO groups, 931 of 956 (97.4 percent), were identified in

203 the *M. l. luteus* genome assembly, and 837 of those had duplicates. The high percentage of

204 duplicate genes in this analysis indicates that homoeologs were not collapsed during the

205 assembly of the genome. This is further supported by comparative genomic analyses of both

206 *Mimulus* genomes (EDGER *et al.* in revision), revealing a 2:1 genome-wide ratio of *M. l. luteus*

207 (tetraploid) *: M. guttatus* (diploid) syntenic blocks.

208

209

Analysis of RNA-seq libraries

Within each species, genes with expression levels lower than 5 FPKM in any of the eight transcriptomes were excluded from any of the further stability analyses. We reasoned that such low-expression genes would make poor qPCR references due to the difficulties in detecting and quantifying their expression. After their removal, a total of 7,225 genes in *M. guttatus* and 10,755 genes in *M. l. luteus* were evaluated. Two methods were used for the analysis of expression stability: simple CV calculations, and exclusion of differentially expressed genes (fold change method (ROBINSON *et al.* 2010)).

For the CV method: Calculations for mean expression (Mean), standard deviation (SD), and the coefficient of variation (CV) were executed in Microsoft Excel or in R (Pumpkin Helmet, v.3.1.2). CV was calculated as SD/Mean. Mean and SD were measured over the four tissue types of both biological replicates (eight samples total) for each species. We adopted a CV cut-off for stable genes of 0.5, which was the cut-off for stable expression across heterogeneous samples advocated by HELLEMANS *et al.* (2007).

For the fold change method: Log fold change was used to evaluate differential expression in pairwise sample comparisons. Genes with a high fold change (greater than 0.4 in *M. guttatus* and 0.3 in *M. l. luteus*) in any pairwise sample comparison were eliminated until a final list of stably expressed genes was obtained (Table S1). The cut-off values used in this study were selected so as to obtain a short list of genes with low variation in expression; the appropriate cut-off value can vary depending on the samples being analyzed and the overall goal of the analysis. The edgeR program (v. 3.12.0) was used to calculate log fold change because the program normalizes expression values by library size for each sample, but any method of fold change calculation can be used. The edgeR program was accessed through Bioconductor and analysis was executed in R.

Gene Annotation

Stably expressed genes were annotated based on the agreement between BLAST results from the NCBI nucleotide database (http://blast.ncbi.nlm.nih.gov/) and from the annotated *Mimulus guttatus* v.2 genome in the Phytozome v.10 database (http://phytozome.jgi.doe.gov/).

239   Traditional reference genes were identified in the RNA-seq datasets in a three-part method.

240   First, known *Arabidopsis thaliana* sequences for traditional reference genes 60s ribosomal protein

241   L8, actin 2/7, actin 11, β-tubulin 2, ubiquitin 5, ubiquitin conjugating enzyme 25, peroxin 4,

242   GAPDH-C1, and EF1-α (see Table S2) were used in a BLAST search against the *M. guttatus* v.2

243   genome in the Phytozome v.10 database in order to identify the appropriate *M. guttatus*

244   homologs. Once a gene match with the correct annotation was identified in Phytozome, a short

245   (approximately 20 bp) sequence from the coding region was then used to identify transcripts

246   from the *M. guttatus* and *M. l. luteus* RNA-seq libraries. The resulting *M. guttatus* and *M. l. luteus*

247   transcripts were used in a BLAST search against the NCBI nucleotide database to ensure that

248   they had been correctly identified.

249

250   **qPCR Genes**

251   Eight genes were selected for validation via qPCR (Table S3). Four traditional reference

252   genes were selected based on both their widespread use in qPCR reference gene literature and

253   on the ease of designing copy specific primers. The four traditional genes chosen were actin 2/7

254   (ACT), GAPDH-C1 (GAP), peroxin 4 (PEX), and ubiquitin conjugating enzyme (UBC). See the

255   above section on Gene Annotation for methods of gene identification within the transcriptome.

256   Four additional genes were chosen based on their apparent stability across T1 tissues in both

257   species, but were later found to be unstably expressed across T2 tissues (see Table S3).

258   However, these genes were retained for analysis in order to compare the qPCR and RNA-seq

259   methods. The four genes chosen were mediator of RNA polymerase 12 (MRP), pectin

260   acetylesterase (PAE), receptor-like kinase (RPK), and FYVE zinc-finger transcription factor

261   (ZNF). The *Mimulus guttatus* GenBank accession numbers for these eight genes, catalogued

262   under *Erythranthe guttata* (BARKER *et al.* 2012), are: ACT = XM_012974510.1, GAP =

263   XM_012999102.1, PEX = XM_013002418.1, UBC = XM_012995233.1, MRP = XM_012984744.1,

264   PAE = XM_012984356.1, RPK = XM_012985914.1, ZNF = XM_013000433.1.

265

266   **qPCR Primer Design**

267   qPCR primers were designed using Primer3

268   (http://biotools.umassmed.edu/bioapps/primer3_www.cgi) with the following criteria: T_m of 60

269   ± 3°C, PCR amplicon length of 130 to 250 bp, primer length of 18 to 25 bp, and GC content of 35

270   – 60%. The Tm criterion was relaxed for UBC to 55 ± 3°C to enable the discovery of suitable

271   primers. Primers were designed to optimally sit as close to the 3' end of the transcript as

272   possible and to span an intron, but these criteria were relaxed in an effort to design primers that

273   are homeolog specific in the allotetraploid *M. l. luteus*. *Mimulus l. luteus* primers were aligned

274   with BLAST against the *M. l. luteus* (Illumina masked v1.1) genome in CoGe

275   (https://genomevolution.org/CoGe/) to ensure homeolog and paralog specificity. *Mimulus*

276   *guttatus* primers were aligned with BLAST against the *M. guttatus* genome (JGI hardmasked

277   vV2) in CoGe to ensure copy specificity. Primers were synthesized by Invitrogen (Life

278   Technologies, U.S.A.). See Table S4 for the full list of primer pairs.

279   To verify primer specificity, PCR products were amplified by Taq DNA polymerase in a

280   Mastercylcer Nexus (Eppendorf, Germany), gel purified using the E.Z.N.A. kit (Omega Biotek,

281   U.S.A.), and Sanger sequenced by Eton Bioscience (U.S.A). Although all primers produced a

282   single band on an agarose gel, the gel extraction step was included to produce cleaner and more

283   concentrated sequencing products. Sequencing confirmed the copy specificity of all primer

284   pairs except for the *M. l. luteus* RPK and PEX primer pairs, which targeted two and three

285   paralogs, respectively.

286

287   **cDNA Synthesis**

288   cDNA was synthesized from 1 µg of total RNA and a mixture of oligo dT and random

289   primers using the Quanta qScript cDNA Synthesis kit (Quanta BioSciences, MD, U.S.A.) and

290   following the manufacturer's protocol. cDNA was stored at 4°C and unused RNA was stored at

291   -80°C.

292   Quality controls for cDNA were two-fold. First, all RNAs and cDNAs were checked for

293   absence of genomic DNA contamination using primers that surround an actin intron (5'-

294   CCCAAGGCTAACAGGGAGAA-3' and 5'- GTGCTGGATTCTGGTGACG-3'). Second, gene

295   expression estimates were obtained from the 3' versus 5' ends of a single gene. A 3'/5' ratio

296   substantially greater or less than 1 may indicate degradation of the mRNA template, or

297   incomplete processivity of the reverse transcription reaction. The MIQE guidelines (BUSTIN *et al.*

298   2009) suggest a range of 0.2 – 5.0 for samples to be used in qPCR. The 3′/5′ ratio of the receptor-

299   like protein kinase cDNA was tested for all tissue types in each individual used in this study,

300   using two primer pairs that amplify in the 5′ region (5′-TGGGCTCGAGTATTTTGCTT-3′ and

301   5′- TGCTTCCTAATCCAAAGATACCA -3′) or the 3′ region (5′-

302   CCTGAGGGTGACAAGACACA -3′ and 5′-ATCAATGGACAAAAGCAGGC -3′) about 1kb

303   away from each other. Some 3′/5′ ratios were found to be > 5 (see Table S5). This could result in

304   an underestimation of expression for genes with primers in the 5′ region of the gene, which

305   includes ACT in both species and the *M. guttatus* ZNF. The 3′/5′ ratios also had a tissue bias,

306   with all stem cDNA samples and some of the calyx cDNA samples having values >5.

307

308   **qPCR Conditions**

309   Comparative qPCR was performed for four biological replicates (all from plants grown

310   at Whitman College, see Plant Materials) and three technical replicates for each tissue type (leaf,

311   stem, petal, and calyx) from each of the two species (*M. guttatus* and *M. l. luteus*). A total of eight

312   genes were selected for qPCR validation (see section qPCR Genes and Table S3) using the

313   primers listed in Table S4. Reactions contained 1X SYBR Green Master Mix (Brilliant III Ultra-

314   Fast SYBR Kit, Agilent Technologies, CA, U.S.A.), 400 nM of primer (except for when

315   amplifying PEX4 from *M. l. luteus*, where 500 nM of primer was used), 1μL of 1:500 diluted ROX

316   dye, and 1μL of cDNA (50 ng/μL) in a final volume of 12.5 μL. PCR reactions were performed

317   in either optical 8-well PCR strips (Agilent Technologies, CA, U.S.A.) or optical 96-well plates

318   (Greiner Bio-One, Belgium) using the Stratagene Mx3000P qPCR system (Agilent Technologies,

319   CA, U.S.A). Samples were amplified for 40 cycles of 10s at 95°C and 20s at the appropriate

320   annealing temperature (see Table S4) after an initial denaturation step at 95°C for 3 minutes. An

321   additional dissociation curve was recorded after cycle 40 by heating from 55°C to 95°C with a

322   ramp speed of 0.01°C per second (Figure S1). Raw qPCR fluorescence data were collected and

323   analyzed by the default settings of the MxPro software v.4.10 (Agilent Technologies, CA,

324   U.S.A.). Cq ("quantification cycle," the cycle in which fluorescence from DNA amplification

325 first exceeds background fluorescence) was determined at a fluorescence threshold of 0.23 for

326 all runs; this fixed threshold was based on the average adaptive threshold of all individual runs.

327 Amplification efficiencies for each primer pair were determined using the Cq values obtained

328 from a ¼ dilution series (1:4, 1:16, 1:64, 1:256, 1:1024) where $E = 10^{(1/-slope)}$. Efficiency for each

329 primer pair was calculated to be between 83% and 102% using the standard curve method

330 (Table S6).

331

332 **Analysis of qPCR Expression Data**

333 Before analysis, the Cq values from qPCR were averaged over the three technical

334 replicates, unless the replicates differed by >1 Cq. In that case, the outlier technical replicate was

335 removed and Cq was averaged over the two remaining technical replicates. These averages

336 were then both calibrator and efficiency normalized using the equation below. GAP amplified

337 from the same sample of *M. l. luteus* young leaf cDNA acted as the inter-plate calibrator.

338 Efficiency values for each gene are listed in Table S6. Relative expression of each gene was

339 calculated as:

340

341 $$\text{Relative Expression} = \text{Efficiency}^{\Delta Cq}, \text{ where } \Delta Cq = Cq^{calibrator} - Cq^{sample}$$

342

343 In order to have a metric of gene stability that could be directly compared to stability

344 estimates from RNA-seq data, the coefficient of variation (CV) was calculated for each gene

345 from the relative qPCR expression data. Calculations for mean expression (Mean), for standard

346 deviation (SD), and for the coefficient of variation (CV= SD/Mean) were executed in Microsoft

347 Excel. SD and Mean were calculated from the relative expression of each of the four tissue

348 types, averaged over the four biological replicates per tissue.

349

350 **Statistical Analyses**

351 All statistical tests were run using R software (Pumpkin Helmet, v.3.1.2). Linear models

352 were fitted to obtain t-test results and Pearson's correlation coefficient.

353

354   **Data Availability**

355   All transcriptomic expression data are provided in Tables S7 and S8. Primer sequences

356   are provided in Table S4. Raw reads from this study are deposited in Dryad

357   (http://dx.doi.org/10.5061/dryad.84655) and are further analyzed in (EDGER *et al.* in revision).

358

359   **RESULTS**

360   **Identification of Novel Reference Genes for *Mimulus***

361   In order to identify potential qPCR reference genes, we compared two simple

362   methods for evaluating variation in expression across tissue types and growing

363   environments: (a) genes with the lowest overall CV across all tissues from both

364   transcriptome sets (T1 and T2; see Figure 1) and (b) exclusion of differentially expressed

365   genes, determined through calculations of fold change, between pairwise comparisons

366   of all tissue samples from both transcriptome sets. We identified 50 genes per species

367   using the CV method and 8 genes per species using the fold change method (Tables S9

368   and S1) that have the potential to be good candidate reference genes for qPCR studies in

369   *Mimulus.*

370   Although CV was not correlated with total expression level (Figure S2), we used

371   a minimum expression cut-off of 5 FPKM in order to exclude genes that are expressed at

372   levels too low to be useful for qPCR normalization. The 50 genes with the lowest CV

373   across both biological replicates of each species are listed in Table S9. Genes on this list

374   have CVs under 0.14 for *M. guttatus* and under 0.12 for *M. l. luteus.* Although a 0.50 CV

375   cut-off has previously been recommended for choosing qPCR reference genes

376   (HELLEMANS *et al.* 2007), we find that a majority of robustly expressed genes fall under

377   this cut-off (Figure 2). In *M. guttatus*, 4,106 genes out of 7,225 had a CV of less than 0.50;

378   in *M. l. luteus,* 6,832 genes out of 10,755 were under this cut-off.

379   For the fold change method, any genes with a log fold change greater than 0.4 in

380   *M. guttatus* or 0.3 in *M. l. luteus*, in any pairwise sample comparison, were excluded.

14

381 Eight *M. guttatus* and eight *M. l. luteus* genes were identified in this manner that had

382 low variation in expression across the four tissue types from two biological replicates

383 (Table S1). The fold change method was consistent with the CV method; five *M. guttatus*

384 genes and one *M. l. luteus* gene identified by the fold change method are also found on

385 the top 50 CV list, and all of the genes identified by the fold change method are listed

386 within the top 200 genes with the lowest CV (Table S10).

387
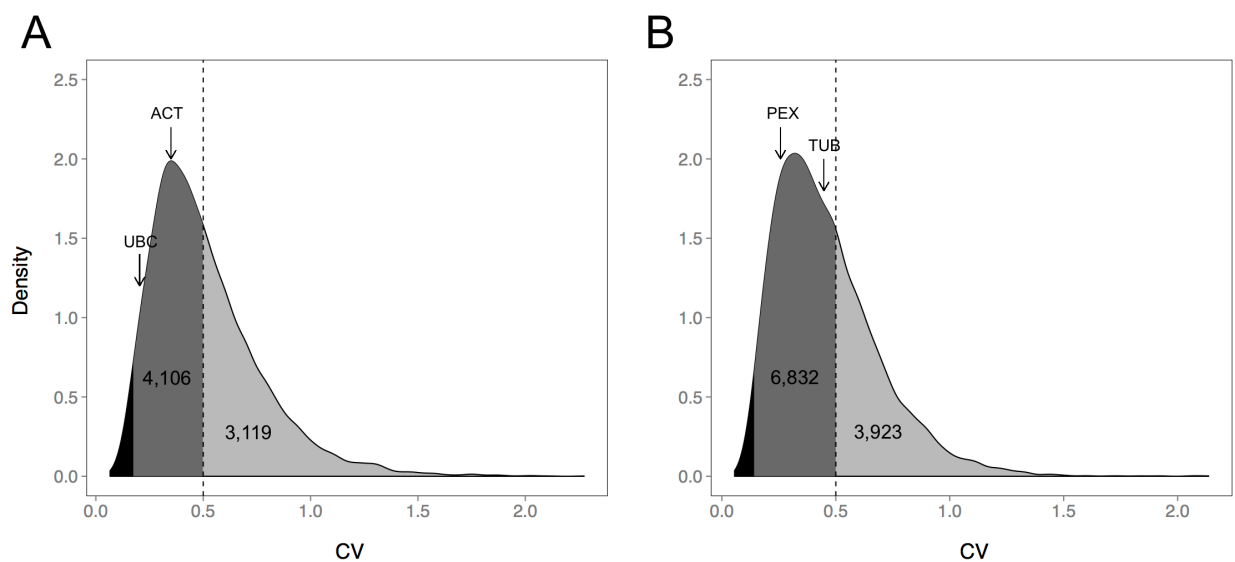
388

389 **Figure 2**

390 Distribution of Coefficient of Variation (CV) for all reliably expressed genes (>5 FPKM in all samples) in

391 (A) *M. guttatus* and (B) *M. l. luteus.* The dashed line marks the 0.50 CV cut-off for stably expressed genes

392 and the arrows point to the two traditional reference genes with the lowest variation in expression for

393 each species (Table S2). The portion of the density curves containing the top 200 genes with the lowest CV

394 are shaded black; all genes selected using the CV and fold change method fall within this region.

395

396 **Traditional Reference Genes in *Mimulus***

397 Since traditional reference genes can be inconsistently expressed in many

398 biological systems (SUZUKI *et al.* 2000; BRUNNER *et al.* 2004; DHEDA *et al.* 2004;

399 CZECHOWSKI *et al.* 2005), we investigated the expression variability of these traditional

400 "housekeeping" reference genes in *Mimulus* using both transcriptomics and qPCR. We

401 chose nine common traditional reference genes to analyze from the RNA-seq datasets:

402 60s ribosomal protein L8 (L8), actin 2/7 (ACT), actin 11 (ACT1), β-tubulin 2 (TUB),

403 ubiquitin 5 (UBQ), ubiquitin conjugating enzyme 25 (UBC), peroxin 4 (PEX), GAPDH-

404 C1 (GAP), and EF1-α (EF1) (see Table S2). We then corroborated the expression

405 variability for four of these nine genes (ACT, PEX, UBC, and GAP) using qPCR (see

406 Table S3).

407   In both *M. guttatus* and *M. l. luteus*, there were thousands of expressed genes

408 with lower CVs than the traditional housekeeping genes (Figure 2, Table S2), and none

409 of the traditional housekeeping genes were amongst the 16 genes identified by the fold-

410 change method. Nevertheless, four traditional genes in *M. guttatus* (GAP, UBC, TUB,

411 and PEX) and four in *M. l. luteus* (L8, GAP, ACT, and UBC) do have CVs lower than 0.5,

412 suggesting that could be useful reference genes for qPCR normalization in these species

413 (Figure 3).

414   The follow-up qPCR validation reported much lower expression variability for

415 the tested subset of traditional genes. This is most likely due to a less variable group of

416 plants being measured for qPCR than were measured for RNA-seq (see Figure 1).

417 Expression variability was even lower when measures from petal tissue are excluded

418 (Figure 3), as expression levels for all four tested genes were substantially higher in

419 petal tissue than in the other three tissue types (Figure S3). This is only the case for the

420 qPCR data and there is no trend in the RNA-seq data when petal is excluded, even

421 though transcriptome T2 was derived from one of the same RNA samples that was used

422 for qPCR. When all tissues were included in the qPCR variability calculations, we found

423 that GAP had the lowest variation in expression in *M. guttatus* and PEX was the least

424 variable in *M. l. luteus*. When petal was excluded, UBC was the least variable traditional

425 reference gene in both species.

426

427
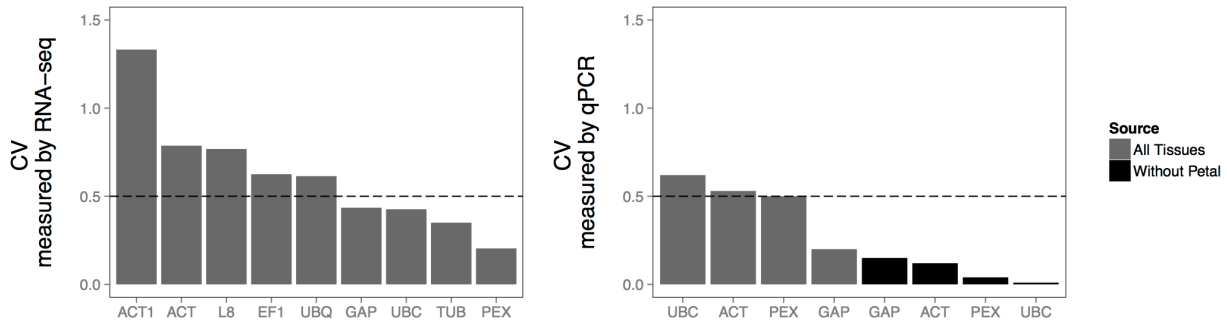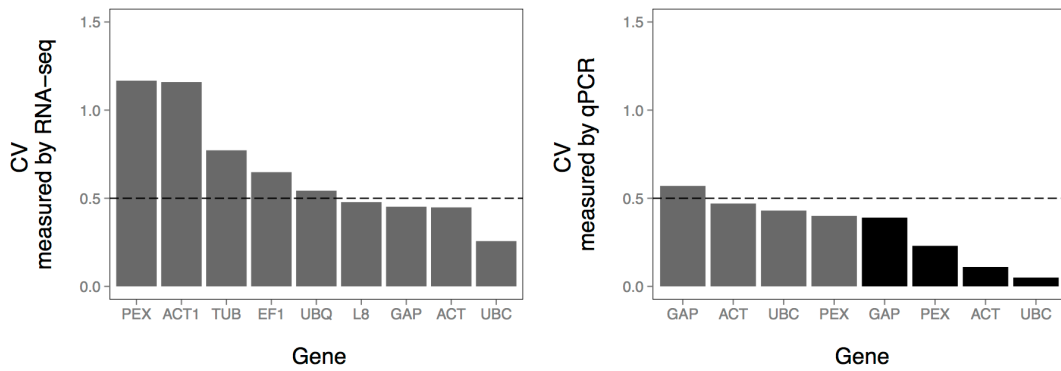
Figure 3

Expression variability estimates for selected traditional reference genes, based on CV. Expression variability in *M. guttatus* (A) and *M. l. luteus* (B) measured via RNA-seq on both T1 and T2 (left column) or via qPCR on T2 samples only (right column). Grey bars show the calculated CV when all tissue types are included and black bars show the calculated CV when petal tissue is excluded. Genes are ordered from most variable to least variable, with a dash line showing a previously suggested cut-off for usable reference genes at 0.50 CV. For the tetraploid *M. l. luteus,* CV reported for the RNA-seq data is the average of both homeologs. The genes tested include 60s ribosomal protein L8 (L8), actin 2/7 (ACT), actin 11 (ACT1), β-tubulin 2 (TUB), ubiquitin 5 (UBQ), ubiquitin conjugating enzyme 25 (UBC), peroxin 4 (PEX), GAPDH-C1 (GAP), and EF1-α (EF1). CV was calculated from FPKM values for genes measured via RNA-seq and from relative expression values, calculated by Efficiency$^{\Delta Cq}$, where $\Delta Cq = Cq_{Calibrator} - Cq_{sample}$, for genes measured via qPCR.

Efficacy of Transcriptomics for Reference Gene Selection

442

443   Although environmental condition was not a purposeful manipulation in our

444 study, the different growth histories of the genetically identical plants used for the two

445 *M. l. luteus* transcriptome sets allowed us to evaluate the robustness of gene expression

446 to moderate environmental variation. This was achieved by comparing both mean

447 expression and expression variability (measured by CV) across the different tissue types

448 between T1 and T2. For comparison, we also evaluate the two *M. guttatus* transcriptome

449 sets, although the plants used in this comparison were genetically as well as

450 environmentally different (see Figure 1).

451   The correlation in CV between T1 and T2 is weaker than the correlation in mean

452 expression for both species, showing a stronger environmental effect on the variance

453 than on the mean (Figure 4). Additionally, CV estimates were more closely correlated

454 between the replicates of *M. l. luteus* than between the replicates of *M. guttatus*, as

455 expected given that the *M. l. luteus* replicates came from the same highly inbred line of

456 plants while the *M. guttatus* replicates came from different lineages.

457

A

r = 0.76 r = 0.59

B

r = 0.75 r = 0.66
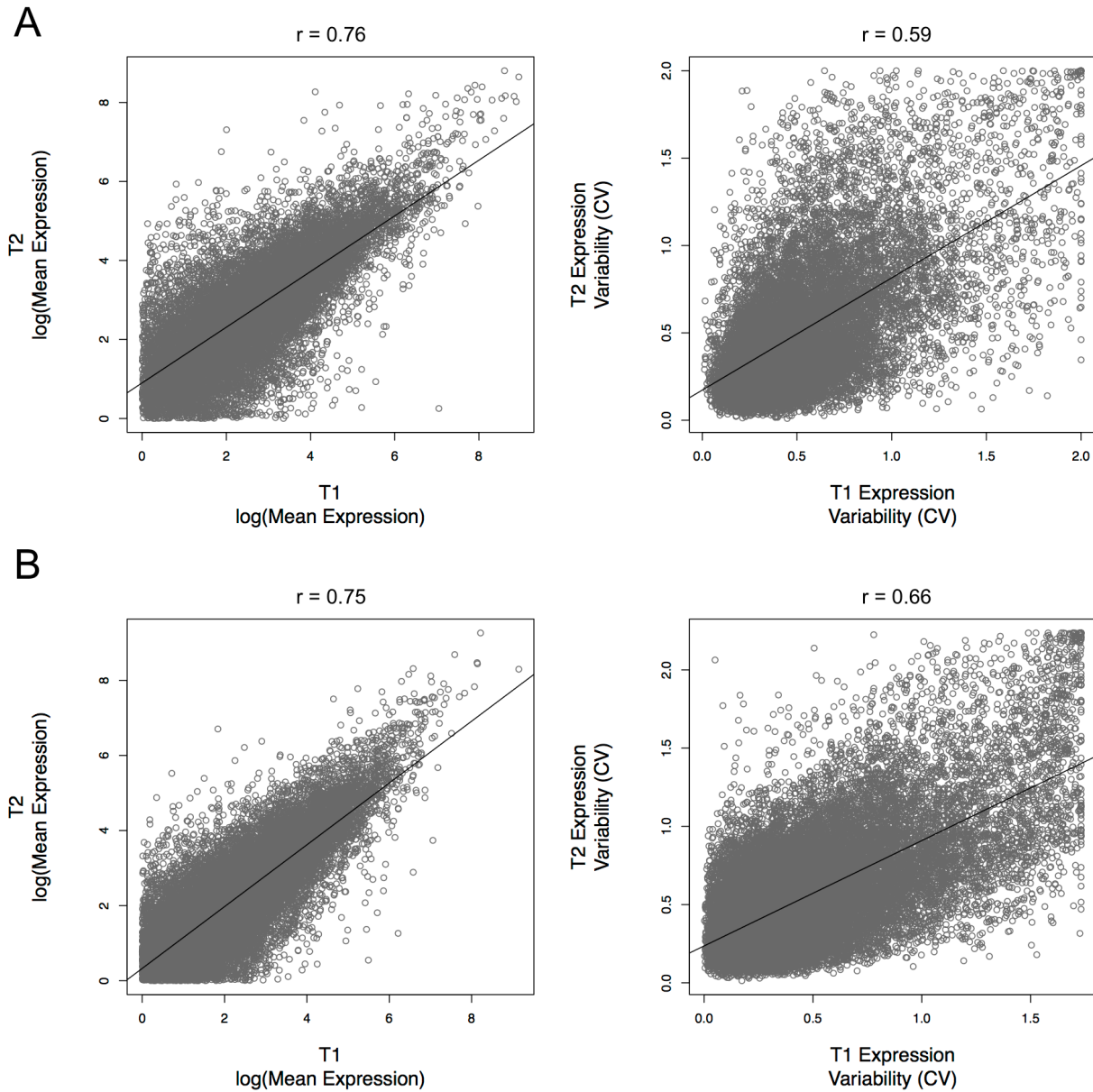
458

459 **Figure 4**

460 Correlation in expression mean (FPKM) and variability (CV), as measured by RNA-seq, between the

461 biological replicates (T1 and T2) of *M. guttatus* (A) and *M. l. luteus* (B). Values for Pearson's correlation

462 coefficient are given above each graph and the line of best fit is shown in black. Genes with mean

463 expression less than 1 FPKM were excluded from the plot.

464

465 The expression data collected via RNA-seq were validated for accuracy using

466 qPCR expression data for a selected group of eight genes including four traditional

19

467   reference genes (Table S3). Mean expression values measured by RNA-seq

468   transcriptomes T1 and T2 were both in agreement with values found via qPCR (Figure

469   5A). In contrast, expression variability estimated by qPCR was significantly correlated

470   with T2 expression variability, but had no significant relationship to T1 expression

471   variability (Figure 5B). This is most likely due to variation in plant lineage and plant

472   growth conditions, as the T2 transcriptomes and the qPCR data derive from genetically

473   identical plants that were grown in the same greenhouse, while the T1 transcriptomes

474   derive from plants grown in a greenhouse at a separate institution. For *M. guttatus*, T1

475   and T2 also differed in the accession used (Fig. 1). This pattern, particularly for the

476   isogenic *M. l. luteus* transcriptomes, suggests that environmental factors may have a

477   greater effect on the "noise" in gene expression than on the expression level itself.

478

479
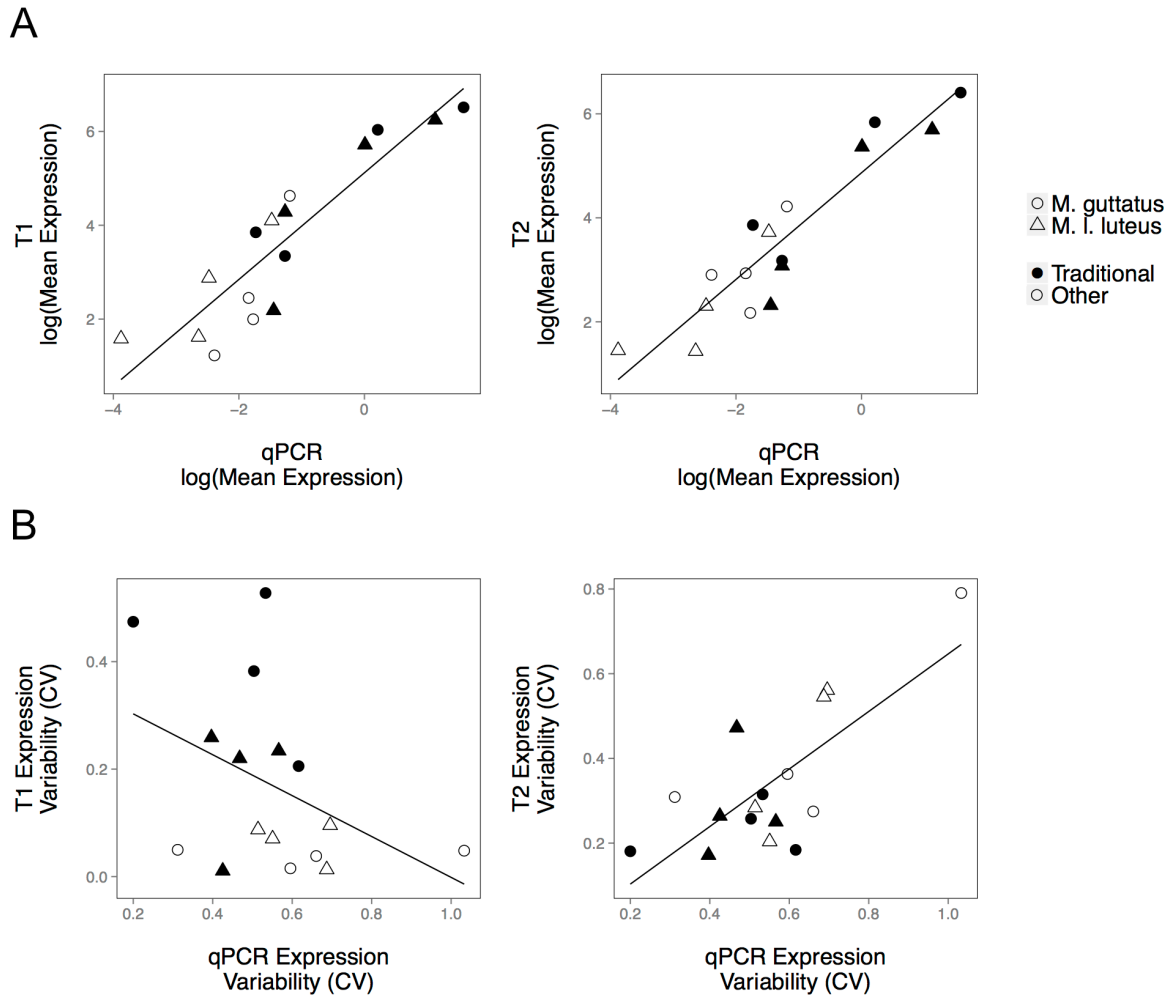
Figure 5

Comparisons of relative gene expression and of expression variability as determined by RNA-seq and by qPCR for a sample of four traditional reference genes (closed symbols) and four additional genes that had initially been found to be stably expressed in transcriptome T1 (open symbols) (Table S3). The T2 RNA-seq transcriptomes and the qPCR data were derived from genetically identical plants grown in the same greenhouse, while the T1 RNA-seq transcriptomes were derived from plants grown in a greenhouse at a separate institution (Fig. 1). (A) There is a strong correlation in relative expression determined by qPCR and RNA-seq, for both T1 (left panel, r = 0.90, $P < 0.001$) and T2 (right panel, r = 0.85, $P < 0.001$). (B) Expression variability (CV) measured via qPCR is correlated with expression variability measured via T2 (right panel, r = 0.74, $P = 0.001$), but is not correlated to expression variability measured via T1 (left panel, r = -0.42, $P = 0.104$). Expression data from both *M. guttatus* and *M. l. luteus* are included together. Relative expression of T1 and T2 is given in FPKM. Relative qPCR expression = Efficiency$^{\Delta Cq}$, where $\Delta Cq = Cq_{Calibrator} - Cq_{sample}$.

21

494

DISCUSSION

Identification of Novel Reference Genes for *Mimulus*

495

496

497   While RNA-seq has the potential to accurately identify genes with low variation

498   in expression, there is still not a universally accepted method for selecting reference

499   genes from RNA-seq data. Most of the programs that are widely used for reference

500   gene selection, such as geNorm, BestKeeper, and NormFinder, were designed

501   specifically for qPCR data and can only process a handful of genes at a time

502   (VANDESOMPELE *et al.* 2002; ANDERSEN *et al.* 2004; PFAFFL *et al.* 2004). We explored two

503   different methods for identifying stably expressed genes from whole transcriptome

504   data: (1) ranking genes based on the coefficient of variation of expression across

505   different samples ("CV method") and (2) excluding unstable genes using a log fold

506   change cut-off value ("fold change method"). We find that both methods identify many

507   stably expressed genes that have the potential to be novel reference genes for qPCR

508   expression studies in *M. guttatus* and *M. l. luteus* (see Tables S9 and S1).

509   Using the CV method, all expressed genes from *M. guttatus* and *M. l. luteus* were

510   ranked based on the variability of their expression across different tissue types and

511   growing conditions and the top 50 genes with the lowest variability were identified

512   (Table S9). Using the fold change method, we identified eight *M. guttatus* and eight *M. l.*

513   *luteus* genes with low variability in expression across four different tissue types and two

514   biological replicates. No traditional reference genes were identified as being among the

515   top 50 most stably expressed genes, by either of our methods. In addition, the novel

516   reference genes we identified had much lower expression variability in our system than

517   any of the most commonly used traditional reference genes (Figure 2), which highlights

518   the utility of the whole transcriptome approach to reference gene selection.

519   The advantage to using either of these methods for reference gene selection is

520   their simplicity in calculation. While the fold change method has the benefit of

521   producing a discrete list of genes with low variation in expression, the CV method has

522   the benefit of quantifying expression variability in a way where genes can be ranked

523   and directly compared. These methods have previously been used in other plant species

524   to select novel reference genes from transcriptomic data (CZECHOWSKI *et al.* 2005;

525   CHANG *et al.* 2012), but we are the first to show that these two methods produce

526   comparable results. All of the genes found on the fold change short-list were among the

527   200 genes with the lowest CV, which corresponds to the top 2-3% most stably expressed

528   transcripts. Ideally, novel reference genes would be selected that score well according to

529   both metrics.

530   The CVs of the novel reference genes we identified are all less than 0.20, whereas

531   a previously suggested cut-off for valid reference genes is a CV of 0.50 (HELLEMANS *et al.*

532   2007). It is important to note that using a 0.50 CV cut-off in our system included a

533   majority of expressed genes (Figure 2), and thus it was not a very discriminating

534   standard for determining expression variability. Since the range of expression can be

535   quite variable, depending on the relatedness of the samples, a single variability cut-off

536   is unlikely to work universally for all experimental designs, and thus a transcriptomic

537   approach appears especially beneficial for selecting stably expressed reference genes.

538

539   **Traditional Reference Genes in *Mimulus***

540   Many studies have pointed to the instability of traditional "housekeeping"

541   reference genes (SUZUKI *et al.* 2000; BRUNNER *et al.* 2004; DHEDA *et al.* 2004; CZECHOWSKI

542   *et al.* 2005). We find that some traditional reference genes in *Mimulus* have the potential

543   to work well for qPCR normalization. Using a whole transcriptome method, we

544   identified four traditional reference genes that have somewhat low variation in

545   expression (CV < 0.50) in *M. guttatus* and *M. l. luteus* (Figure 3). Two genes, UBC and

546   GAP, were even identified as stably expressed in both species and could potentially be

547   good universal reference genes for the *Mimulus* genus. We confirmed our findings for

548   four of these traditional reference genes with qPCR and found that all four (GAP, ACT,

549   UBC, and PEX) could be acceptable as reference genes for both species based on qPCR

550   estimates of expression variability across tissues, although some of the genes were at or

551   slightly above the recommended 0.5 CV cutoff when the relatively-divergent petal

552   tissue samples were included (Figure 3). However, these traditional reference genes

553   were nowhere near the most stably expressed in the transcriptome as a whole (Figure

554   2), which highlights the opportunity to discover dramatically more stable reference

555   genes using a transcriptome-guided approach.

556   Despite the widespread use of *Mimulus* as a model genus for genetics, very few

557   papers have attempted to validate reference genes for use in this genus. SCOVILLE *et al.*

558   (2011) qualitatively ranked the expression variability of six traditional reference genes

559   in *M. guttatus* and found that UBQ and EF1α were the most stably expressed. We

560   quantitatively investigated four of these six traditional reference genes in our own

561   study and found that UBQ and EF1α had higher expression variability than other

562   traditional reference genes and that, in both species, the expression of these two genes

563   was not stable enough for either to be used as a reference gene under our study

564   conditions. SCOVILLE *et al.* (2011) tested different lines of *M. guttatus* and included a

565   wound treatment, which may have resulted in our differing reports of traditional

566   reference gene stabilities. This again highlights the importance of reference gene

567   validation for specific study conditions.

568   Although we found that some traditional reference genes can be used for qPCR

569   normalization, they are not optimal reference genes; the variability in expression of the

570   traditional reference genes is very high when compared to the variability of all robustly

571   expressed genes (Figure 2, Table S2). This indicates that whole transcriptome

572   approaches, such as RNA-seq, have great potential to discover novel reference genes

573   that are stably expressed in the study system of interest. With the current speed and low

574   cost of RNA-seq, as well as the online availability of multi-tissue and/or multi-

24

575 environment RNA-seq data sets, we expect that the whole transcriptome approach will

576 be increasingly useful for reference gene identification and validation.

577

578 **Efficacy of Transcriptomics for Reference Gene Selection**

579   RNA-seq has been shown repeatedly to generate accurate measurements of gene

580 expression (MARIONI *et al.* 2008; MORTAZAVI *et al.* 2008; NAGALAKSHMI *et al.* 2008;

581 NOOKAEW *et al.* 2012). We find similar results in *Mimulus* when comparing the relative

582 expression determined by RNA-seq to the relative expression determined by qPCR for

583 eight selected genes (Figure 5A). We also find that estimates of expression mean are

584 robust to moderate environmental and genetic variation, but that estimates of

585 expression variability across tissue types are only in agreement when the samples were

586 obtained from a shared environment (Figures 4 and 5). These results suggest that

587 environmental changes may have a greater impact on expression variability than on

588 expression means.

589   For the goal of reference gene selection, where expression variability must

590 remain low, this difficulty can be solved in two ways. One approach is to use the same

591 samples for both RNA-seq and the subsequent qPCR analysis, as in Chang *et al.* (2012)

592 and Yang *et al*. (2014). This method would be highly accurate, but would be extremely

593 specific to particular study conditions. A second approach would be to evaluate a large

594 variety of genotypes or growth conditions to discover genes that are maximally stable

595 across genetically and environmentally distinct samples, as was done for *Arabidopsis*

596 *thaliana* in Czechowski *et al*. (2005). This method would allow for the identification of a

597 starting pool of "universally" stable genes.

598

599 **Reference Gene Selection using RNA-seq**

600   We show, using *Mimulus* as a case study, that RNA-seq is a promising tool for

601 selecting genes with low gene expression variance that can be used as novel qPCR

25

602 reference genes. As many research labs regularly use RNA-seq as a first approach to

603 collecting expression data, already completed RNA-seq transcriptomes are a readily

604 available tool that can be used to search for candidate qPCR reference genes in any

605 study system. Although we find that the variance in expression is variable between

606 environmental conditions, we propose that transcriptomes from diverse samples can be

607 pooled in order to identify more universally stable genes. We show that two simple

608 methods for identifying genes with low expression variance, the CV method and the

609 fold change method, both result in comparable evaluations of expression variance.

610 Thus, either of these methods can be used to identify a preliminary set of highly stable

611 candidate reference genes for qPCR experiments.

612

618

619 **LITERATURE CITED**

620 Andersen, C. L., J. L. Jensen and T. F. Orntoft, 2004 Normalization of real-time quantitative
621   reverse transcription-PCR data: A model-based variance estimation approach to
622   identify genes suited for normalization, applied to bladder and colon cancer data
623   sets. Cancer Research 64**:** 5245-5250.
624 Barker, W., G. Nesom, P. M. Beardsley and N. S. Fraga, 2012 A taxonomic conspectus of
625   Phrymaceae: A narrowed circumscription for *Mimulus*, new and resurrected genera,
626   and new names and combinations. Phytoneuron 39**:** 1-60.
627 Beardsley, P. M., and R. G. Olmstead, 2002 Redefining Phrymaceae: The placement of
628   Mimulus, tribe Mimuleae and Phryma. American Journal of Botany 89**:** 1093-1102.
629 Brunner, A. M., I. A. Yakovlev and S. H. Strauss, 2004 Validating internal controls for
630   quantitative plant gene expression studies. BMC Plant Biol 4**:** 14.
631 Bustin, S. A., V. Benes, J. A. Garson, J. Hellemans, J. Huggett *et al.*, 2009 The MIQE guidelines:
632   minimum information for publication of quantitative real-time PCR experiments.
633   Clin Chem 55**:** 611-622.

634 Chang, E. M., S. Q. Shi, J. F. Liu, T. L. Cheng, L. Xue *et al.*, 2012 Selection of Reference Genes
635   for Quantitative Gene Expression Studies in Platycladus orientalis (Cupressaceae)
636   Using Real-Time PCR. Plos One 7.
637 Cooley, A. M., G. Carvallo and J. H. Willis, 2008 Is floral diversification associated with
638   pollinator divergence? Flower shape, flower colour and pollinator preference in
639   chilean Mimulus. Annals of Botany 101**:** 641-650.
640 Czechowski, T., M. Stitt, T. Altmann, M. K. Udvardi and W. R. Scheible, 2005 Genome-wide
641   identification and testing of superior reference genes for transcript normalization in
642   Arabidopsis. Plant Physiol 139**:** 5-17.
643 de Oliveira, L. A., M. C. Breton, F. M. Bastolla, S. D. Camargo, R. Margis *et al.*, 2012 Reference
644   Genes for the Normalization of Gene Expression in Eucalyptus Species. Plant and
645   Cell Physiology 53**:** 405-422.
646 Dheda, K., J. F. Huggett, S. A. Bustin, M. A. Johnson, G. Rook *et al.*, 2004 Validation of
647   housekeeping genes for normalizing RNA expression in real-time PCR.
648   Biotechniques 37**:** 112-119.
649 Dheda, K., J. F. Huggett, J. S. Chang, L. U. Kim, S. A. Bustin *et al.*, 2005 The implications of
650   using an inappropriate reference gene for real-time reverse transcription PCR data
651   normalization. Anal Biochem 344**:** 141-143.
652 Edger, P. P., R. Smith, M. R. McKain, A. M. Cooley, M. Vallejo-Marin *et al.*, in revision
653   Subgenome dominance in an interspecific hybrid, synthetic allopolyploid, and a 140
654   year old naturally established neo-allopolyploid monkeyflower.
655   http://biorxiv.org/content/early/2016/12/19/094797.
656 Grabherr, M. G., B. J. Haas, M. Yassour, J. Z. Levin, D. A. Thompson *et al.*, 2011 Full-length
657   transcriptome assembly from RNA-Seq data without a reference genome. Nat
658   Biotechnol 29**:** 644-652.
659 Gutierrez, L., M. Mauriat, S. Guenin, J. Pelloux, J. F. Lefebvre *et al.*, 2008 The lack of a
660   systematic validation of reference genes: a serious pitfall undervalued in reverse
661   transcription-polymerase chain reaction (RT-PCR) analysis in plants. Plant
662   Biotechnol J 6**:** 609-618.
663 Haas, B. J., and M. C. Zody, 2010 Advancing RNA-Seq analysis. Nat Biotechnol 28**:** 421-423.
664 Hellemans, J., G. Mortier, A. De Paepe, F. Speleman and J. Vandesompele, 2007 qBase
665   relative quantification framework and software for management and automated
666   analysis of real-time quantitative PCR data. Genome Biol 8**:** R19.
667 Hellsten, U., K. M. Wright, J. Jenkins, S. Q. Shu, Y. W. Yuan *et al.*, 2013 Fine-scale variation in
668   meiotic recombination in Mimulus inferred from population shotgun sequencing.
669   Proceedings of the National Academy of Sciences of the United States of America
670   110**:** 19478-19482.
671 Libault, M., S. Thibivilliers, D. D. Bilgin, O. Radwan, M. Benitez *et al.*, 2008 Identification of
672   Four Soybean Reference Genes for Gene Expression Normalization. Plant Genome 1**:**
673   44-54.
674 Marioni, J. C., C. E. Mason, S. M. Mane, M. Stephens and Y. Gilad, 2008 RNA-seq: an
675   assessment of technical reproducibility and comparison with gene expression
676   arrays. Genome Res 18**:** 1509-1517.
677 Mortazavi, A., B. A. Williams, K. McCue, L. Schaeffer and B. Wold, 2008 Mapping and
678   quantifying mammalian transcriptomes by RNA-Seq. Nat Methods 5**:** 621-628.

679 Mukherjee, B. B., and R. K. Vickery, 1962 Chromosome counts in the section *Simiolus* of the
680   genus *Mimulus* (Scrophulariaceae). V. The chromosomal homologies of *M. guttatus*
681   and its allied species and varieties. Madroño 16**:** 141-172.
682 Nagalakshmi, U., Z. Wang, K. Waern, C. Shou, D. Raha *et al.*, 2008 The transcriptional
683   landscape of the yeast genome defined by RNA sequencing. Science 320: 1344-1349.
684 Nookaew, I., M. Papini, N. Pornputtapong, G. Scalcinati, L. Fagerberg *et al.*, 2012 A
685   comprehensive comparison of RNA-Seq-based transcriptome analysis from reads to
686   differential gene expression and cross-comparison with microarrays: a case study in
687   Saccharomyces cerevisiae. Nucleic Acids Research 40**:** 10084-10097.
688 Pfaffl, M. W., A. Tichopad, C. Prgomet and T. P. Neuvians, 2004 Determination of stable
689   housekeeping genes, differentially regulated target genes and sample integrity:
690   BestKeeper - Excel-based tool using pair-wise correlations. Biotechnology Letters
691   26**:** 509-515.
692 Robertson, G., J. Schein, R. Chiu, R. Corbett, M. Field *et al.*, 2010 De novo assembly and
693   analysis of RNA-seq data. Nature Methods 7**:** 909-911.
694 Robinson, M. D., D. J. McCarthy and G. K. Smyth, 2010 edgeR: a Bioconductor package for
695   differential expression analysis of digital gene expression data. Bioinformatics 26**:**
696   139-140.
697 Scoville, A. G., L. L. Barnett, S. Bodbyl-Roels, J. K. Kelly and L. C. Hileman, 2011 Differential
698   regulation of a MYB transcription factor is correlated with transgenerational
699   epigenetic inheritance of trichome density in Mimulus guttatus. New Phytol 191**:**
700   251-263.
701 Simão, F. A., R. M. Waterhouse, P. Ioannidis, E. V. Kriventseva and E. M. Zdobnov, 2015
702   BUSCO: assessing genome assembly and annotation completeness with single-copy
703   orthologs. Bioinformatics**:** btv351.
704 Sobel, J. M., and M. A. Streisfeld, 2013 Flower color as a model system for studies of plant
705   evo-devo. Frontiers in Plant Science 4**:** 1-17.
706 Suzuki, T., P. J. Higgins and D. R. Crawford, 2000 Control selection for RNA quantitation.
707   Biotechniques 29**:** 332-337.
708 Thellin, O., W. Zorzi, B. Lakaye, B. De Borman, B. Coumans *et al.*, 1999 Housekeeping genes
709   as internal standards: use and limits. Journal of Biotechnology 75**:** 291-295.
710 Trapnell, C., B. A. Williams, G. Pertea, A. Mortazavi, G. Kwan *et al.*, 2010 Transcript assembly
711   and quantification by RNA-Seq reveals unannotated transcripts and isoform
712   switching during cell differentiation. Nat Biotechnol 28**:** 511-515.
713 Twyford, A. D., M. A. Streisfeld, D. B. Lowry and J. Friedman, 2015 Genomic studies on the
714   nature of species: adaptation and speciation in Mimulus. Molecular Ecology 24**:**
715   2601-2609.
716 Vallejo‐Marín, M., R. J. Buggs, A. M. Cooley and J. R. Puzey, 2015 Speciation by genome
717   duplication: Repeated origins and genomic composition of the recently formed
718   allopolyploid species *Mimulus peregrinus*. Evolution 69**:** 1487-1500.
719 Vandesompele, J., K. De Preter, F. Pattyn, B. Poppe, N. Van Roy *et al.*, 2002 Accurate
720   normalization of real-time quantitative RT-PCR data by geometric averaging of
721   multiple internal control genes. Genome Biology 3.
722 Wang, Z., M. Gerstein and M. Snyder, 2009 RNA-Seq: a revolutionary tool for
723   transcriptomics. Nature Reviews Genetics 10**:** 57-63.

724 Willis, J. H., 1999 The role of genes of large effect on inbreeding depression in Mimulus
725   guttatus. Evolution 53**:** 1678-1691.
726 Wu, C. A., D. B. Lowry, A. M. Cooley, K. M. Wright, Y. W. Lee *et al.*, 2008 Mimulus is an
727   emerging model system for the integration of ecological and genomic studies.
728   Heredity (Edinb) 100**:** 220-230.
729 Yang, H. L., J. Liu, S. M. Huang, T. T. Guo, L. B. Deng *et al.*, 2014 Selection and evaluation of
730   novel reference genes for quantitative reverse transcription PCR (qRT-PCR) based
731   on genome and transcriptome data in Brassica napus L. Gene 538**:** 113-122.
732 Yuan, Y. W., J. M. Sagawa, R. C. Young, B. J. Christensen and H. D. Bradshaw, 2013 Genetic
733   Dissection of a Major Anthocyanin QTL Contributing to Pollinator-Mediated
734   Reproductive Isolation Between Sister Species of Mimulus. Genetics 194**:** 255-263.
735 Zhuang, H., Y. Fu, W. He, L. Wang and Y. Wei, 2015 Selection of appropriate reference genes
736   for quantitative real-time PCR in Oxytropis ochrocephala Bunge using
737   transcriptome datasets under abiotic stress treatments. Frontiers in plant science 6.
738