MOLECULAR ECOLOGY
RESOURCES
Molecular Ecology Resources (2011) doi: 10.1111/j.1755-0998.2011.03058.x

Primer design and transcript quantification of a highly multiplexed RT-PCR for a nonmodel avian species

SUSAN L. BALENGER, CHRISTOPHER J. W. McCLURE and GEOFFREY E. HILL
331 Funchess Hall, Department of Biological Sciences, Auburn University, Auburn, AL 36849, USA

Abstract

Multiplexed qRT-PCR assays are currently lacking for nearly all species without genome or transcriptome resources. Here, we present a strategy for primer design of highly multiplexed qRT-PCR assays, evaluate Beckman Coulter's Quant Tool gene expression quantification software and provide details of our assay for the North American songbird *Carpodacus mexicanus* (house finch), for which only small sections of genome sequence are available. We combined Beckman Coulter's eXpress Designer module for creating custom multiplex primers with the free, online program Amplify 3 to design and evaluate primers computationally before testing them empirically. We also generated a standard curve for each gene included in the final multiplex. We compared models using cubic and quadratic polynomial estimators that did and did not force the intercept through zero. Ultimately, we used the sequences available for 316 clones differentially expressed in cDNA macroarray and microarray comparisons, and from these sequences, we were able to generate a set of transcript-specific primers for use with the GeXP analyser for 20 house finch genes.

Keywords: *Carpodacus mexicanus*, gene expression, house finch, standard curve estimator

Received 4 March 2011; revision received 21 June 2011; accepted 27 June 2011

Introduction

The expanding use of large-scale, high-throughput genome and transcriptome data sets generated from microarrays and pyrosequencing projects has renewed interest in examining variation in functionally relevant and adaptive markers in the context of individual and population fitness (Vasemägi & Primmer 2005; Hoffmann & Willi 2008; Shiu & Borevitz 2008; Ungerer *et al.* 2008; Piertney & Webster 2010). Studies of gene expression expand our ability to identify and understand the function of genes, but traditional platforms for assaying transcript abundance are not always ideal for studies of natural populations. Specifically, microarrays and real-time PCR platforms are useful for transcriptome-scale and single-gene studies respectively, but the cost and complexity of microarrays limit their use by many researchers, while extensive lists of candidate genes often cannot feasibly be examined using real-time PCR. The emergence of RNA-seq pyrosequencing-based platforms has undercut many of these difficulties, but bioinformatic limitations still remain. Employing multiplexed RT-PCR platforms to simultaneously measure the expression of multiple genes can significantly reduce time and cost as well as improve measures of intrasample variation. Furthermore, studies employing RNA-seq and/or microarray technologies to investigate expression changes in many thousands of genes could benefit from the development of highly multiplexed gene expression reactions by allowing for the statistical validation of these experiments across several dozen genes. Currently, multiplexed RT-PCR assays have been primarily implemented in biomedical research, including neuropathological disease diagnosis, cancer biomarker signatures and viral infection identification (Therianos *et al.* 2004; Chen *et al.* 2007; Nagel *et al.* 2009; and Rai *et al.* 2009). The lack of ecological and evolutionary studies utilizing these assays is likely due in large part to the difficulty involved in designing primers that will not amplify off-target transcripts, which is especially difficult when working with a species lacking genome or transcriptome sequences. Furthermore, each gene in a multiplex typically requires its own distinct fluorescent probe so that it can be distinguished from other genes (Brisson *et al.* 2004). The need for individual labelling typically limits multiplexed platforms to between four and six genes. The cost of multiple probes can increase the cost of development of the assay to the extent that it makes multiplexing impractical.

Beckman Coulter offers its GeXP gene expression platform for examining up to 35 genes in a single multiplexed reaction. Using the Beckman CEQ8000 DNA

Correspondence: Susan L. Balenger, Fax: 334 844 1645;
E-mail: balensl@auburn.edu

series gel capillary electrophoresis sequencer, genes are differentiated based on a unique amplicon size rather than the fluorescence wavelength of the probe. Thus, only a single dye is used for every gene of interest by virtue of a complementary universal sequence tag attached to each forward primer.

Recently, Rai *et al.* (2009) demonstrated an analytical validation of the GeXP method using a standard curve to assess inter- and intra-assay precision along a range of concentrations. More recently, Beckman Coulter released a normalization macro it calls 'Quant Tool', which is available for free download to CEQ8000 users. This tool incorporates sample expression data into the standard curve values, which allows for the relative quantification of transcripts of each gene included in the multiplex reaction according to its own standard curve. Quant Tool calculates the best-fit line for each gene using a third-order polynomial function with a y-intercept equal to zero.

This study (i) presents our primer design strategy, which reduces the time and cost of primer design and optimization, (ii) evaluates Beckman Coulter's Quant Tool gene expression quantification software as well as other possible polynomial estimators and (iii) provides details of the multiplexed qRT-PCR assay we designed according to these protocols for the North American songbird *Carpodacus mexicanus* (house finch), which has only a tiny amount of genome sequence currently available. We chose candidate genes for the *C. mexicanus* assay based on their association with an adaptive response to infection with the bacterial pathogen, *Mycoplasma gallisepticum* (Bonneaud *et al.* 2011), hereafter referred to as MG.

Materials and methods

Transcript selection and primer design

We initially selected 30 house finch genes of interest as well as three housekeeping genes to attempt to include in a single, multiplexed reverse transcription–PCR based on differential expression in a microarray study (Bonneaud *et al.* 2011) and interesting GO (gene ontology) functions (http://www.geneontology.org/). According to Beckman Coulter standard protocols, we also included a pair of primers designed to amplify an external RNA transcript spiked into each reaction.

We generated 'first-pass' multiplex primers using the Primer Design module of the GeXP eXpress Profiler software (Beckman Coulter, Fullerton, CA, USA). Each forward and reverse primer included a 5′ end containing a universal priming sequence and a 3′ end containing a transcript-specific sequence. Primer pairs were designed to yield RT-PCR products at least four nucleotides apart in length within a range of 100–400 nt, as well as having similar GC content and similar melting temperatures. As a result of necessary attenuation of reverse primer concentrations, direct comparisons can be made between samples but not between genes; thus, we did not evaluate differences in amplification efficiencies between fragments.

We imported primer and available gene sequence information from this 'first-pass' multiplex into Amplify 3 (http://engels.genetics.wisc.edu/amplify/), which is a free online PCR simulation program for Mac OSX. To adapt Amplify 3 for RT-PCR primers, we created two sequence files in alternate orientations for each gene. We prepared two separate lists of primers: one containing all of the reverse primers and another containing both reverse and forward primers. We then simultaneously compared the entire set of primers to be multiplexed against each individual gene sequence, comparing only reverse primers against the 'RNA' sequence and both reverse and forward primers against its reverse complement. Only fragments predicted to overlap with the size range of interest (i.e. 100–400 nt) were used in comparisons. We evaluated fragment sizes as well as the quality estimates given by Amplify 3, termed 'primability' and 'stability'. We then considered the entire pool of fragments and excluded target primers from the multiplex if they amplified any undesigned fragment that overlapped the size (±2 nucleotides) of a designed fragment. Amplification was expected when each primer in the pair had a primability value ≥80 and a stability value ≥50.

We needed to design new primers for Amplify 3 testing for genes with primers that were excluded in the process described above. We subjected the genes to the Primer Design module of the eXpress Designer software. We used the Primer Design module to redesign primers for individual genes using default parameters and tested these newly designed primers along with the rest of the primers in the multiplex using Amplify 3 as described above. We repeated this process until we identified primer pairs for each gene that met our criteria from a bioinformatic standpoint.

Animals and RNA

We collected total RNA from spleen tissue of male house finches that were and were not experimentally infected with MG. Specific details of capture, housing, care and experimental manipulations of house finches are described in Bonneaud *et al.* (2011). Briefly, wild-caught birds originating from two distinct populations were inoculated with 10 μL of SP4 media to each eye or with 10 μL of a stock culture containing approximately 1×10^4 to 1×10^6 colour-changing units/mL of an MG field isolate collected in Auburn, AL, January 2007 (BUA #243). Control birds were euthanized 14 days after sham

inoculation, while infected birds were euthanized either 3 or 14 days postexposure (IACUC protocol #2007–1197). We immediately removed the spleens from all euthanized birds and stored them in RNAlater at 4 °C for 24 h before being placing them at −80 °C for future characterization of gene expression.

We extracted total RNA from approximately 15 mg of each individual's homogenized spleen tissue using Qiagen RNeasy miniprep spin columns, followed by the digestion of genomic DNA according to the manufacturers' protocols (Turbo DNase, Ambion). We determined the quantity of purified total RNA using a Nanodrop spectrophotometer and determined RNA integrity on an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). All RNA extracts were stored at −80 °C until further processing.

Primer pool testing and multiplex optimization

We followed Beckman Coulter protocols for testing primer combinations and attenuating primer concentrations at the bench (http://www.beckmancoulter.com/literature/Bioresearch/A29143AC.pdf). Briefly, we tested all of the following possible combinations of primers with each sample pool in duplicate: (i) single reverse primer–single forward primer, (ii) single reverse primer–multiplex of all forward primers, (iii) multiplex of all reverse primers–single forward primer and (iv) multiplex of all reverse primers–multiplex of all forward primers. We then compared these results to those predicted by Amplify 3. Whenever (i), (ii) or (iii) produced undesigned peaks (UDPs) of significant size (>2000 rpu) within 2 bp of any expected fragment (see Tables 1 and 2), we either removed the problematic primers from the multiplex or, whenever possible, extended the primer by inserting a 1- to 2-nt spacer between its gene-specific and universal tag portions to shift the amplified size away from the UDP.

Multiplexed qRT-PCR conditions

We generated a standard curve for all genes in the multiplex using a twofold series of dilutions (250.0, 125, 62.5, 31.3, 15.6, 7.8, 3.9, 2.0 and 1.0 ng) of a reference pool of total RNA. Each concentration was run in quadruplicate. The house finch reference pool comprised equal parts of RNA from each of two control and four treatment groups of birds (see Table S1 for details). We used Beckman Coulter protocols for each reverse transcription reaction using GeXP Start kit reagents unless otherwise noted: 4.0 μL of RT buffer (5×), 2.0 μL of a pool of attenuated reverse primers (10×), 1.0 μL of reverse transcriptase, 3.0 μL of DNase-free water, 5.0 μL of 0.625 ng/μL

Table 1 Oligonucleotide primer sequences and expected PCR product sizes for house finch multiplexed RT-PCR

Gene symbol	GenBank accession no.	Forward primer w/o universal tag (5′–3′)*	Reverse primer w/o universal tag (5′–3′)†	Product size	Reverse primer dilution‡
MAP	GW346167	TCTGGCCCAAAACTTCCATA	CCACATTCCCTTCTTCTTCTG	139	1:64
TXN	GW346164	GCAGCCTGGTTGAATTTGAG	AAAGGGCTTGATCATTTTGC	145	None
Ig4A	GW346137	CTCGTAGTGCAGTGTTACCGATGT	GGCCGAAATTTTGCAATCTA	149	1:64
DSTN	GW346131	CAGATGCTTCTGAGGCTTTTA	AGACTGGGGCTGAAAATACCA	152	1:256
LCP	DR782758	GGCATAGACTCTTGCTCCGA	TCAACCAGGGTCCATCAAGT	158	1:256
PSAP	DR782822	GCTCTTGATGTGAAGGGTCC	CTCCTTCCCAAAGGTCTCCT	181	1:256
RHOA	GW346157	CGCCAAGCTCAGAATTAACC	CTCAGGAGATTGGCAGAAGC	188	1:128
SEC61	GW346160	TGGGATATGGATCAACCTGA	ATGCAGTTTGTGGAACCCAG	206	1:8
TCR	GW346163	AAACTGGCAACACACTCGAA	ACCCTGACCACCTGACTCTG	213	1:128
HSP90	DR782747	TGGAGACCTTTGCCTTTCAG	CAGCTCTTTCCCAGAATCCA	218	1:128
ICK	GW346139	AAACCATGCCAATGTGGTAAA	CCTGTGAAAGAACCCATGCT	240	None
TIF	DR782722	TTGGTTCAGCTCCCAATCTC	AAACAGGGATGCTGTTACGC	248	None
NADH4	DR782776	GCTGTGGGTTCGTTCGTAGT	AGGAGCAATCATAACCAGCG	255	1:512
UBC7	GW346169	CTGAGAGGTGGGATGCAGAT	ACAAGGTGCAGGGTGGATT	258	1:1024
NABP	GW346152	ATAGCTTCAGACAGGGCGAA	TCCCAGCTAGCACTTAGGGA	263	1:8
PTMS	DR782728	ATCCTCGTGATGTCTGTGCC	CTGGCCCTCGTGAATTTTT	268	1:32
MHCIi	DR782864	TGCTCCTTCAGCTCCTGATT	GGTGTTGTTGGAGGTCGAGT	275	1:256
ARP2/3	bankit1234533	TGGCACAATTGACTTTCCAG	CGTCTTCACTTCGCTGTCAT	311	1:16
IgJ	GW346136	AACCTCAGACTCGTGCCATC	TGAAGGTTGTGCAGAGGTCA	318	1:16
KanR				325	
CAL2	bankit1324554	GGAGTTCAAGGCCAAGGAG	ATGGACAAGGAGGACAGGAA	341	3×

*Forward universal primer sequence: AGGTGACACTATAGAATA.

†Reverse universal primer sequence: GTACGACTCACTATAGGGA.

‡Each reverse primer was separately diluted in 10 μM Tris–HCl from a 100-μM stock solution prior to being pooled.

Table 2 Comparison of standard curve line estimators generated for each gene in the house finch multiplexed qRT-PCR assay

Gene symbol	Quant Tool line equation (r^2)	Preferred Line equation according to AIC (r^2)
MAP	$y = -0.076x^3 + 2.220x^2 + 1.688x$ (0.9991)	$y = (2.2 \times 10^{-6})x^3 - 0.001x^2 + 0.142x$ (0.9726)
TXN	$y = -1.035x^3 + 14.200x^2 - 12.939x$ (0.9969)	$y = (1.5 \times 10^{-6})x^3 - 0.001x^2 + 0.078x + 0.484$ (0.9627)
Ig4A	$y = -0.017x^3 + 1.910x^2 + 14.113x$ (0.9989)	$y = (4.7 \times 10^{-7})x^3 - 0.000x^2 + 0.064x$ (0.9665)
DSTN	$y = -0.020x^3 + 0.962x^2 + 0.527x$ (0.9988)	$y = (3.6 \times 10^{-6})x^3 - 0.002x^2 + 0.230x$ (0.9802)
LCP	$y = -0.023x^3 + 0.879x^2 + 2.958x$ (0.9991)	$y = (2.5 \times 10^{-6})x^3 - 0.001x^2 + 0.178x$ (0.9829)
PSAP	$y = -0.035x^3 + 1.457x^2 - 3.990x$ (0.9988)	$y = (3.0 \times 10^{-6})x3 - 0.001x2 + 0.2179x + 2.641$ (0.9465)
RHOA	$y = -0.141x^3 + 3.010x^2 + 3.559x$ (0.9993)	$y = (1.7 \times 10^{-6})x^3 - 0.001x^2 + 0.109x$ (0.9800)
SEC61	$y = -0.011x^3 + 0.596x^2 + 0.637x$ (0.9993)	$y = (4.5 \times 10^{-6})x^3 - 0.002x^2 + 0.283x$ (0.9848)
TCR	$y = -0.054x^3 + 1.626x^2 + 8.889x$ (0.9997)	$y = (8.0 \times 10^{-7})x^3 - 0.000x^2 + 0.089x$ (0.9874)
HSP90	$y = -0.051x^3 + 1.428x^2 + 4.989x$ (0.9999)	$y = (1.7 \times 10^{-6})x^3 - 0.001x^2 + 0.127x$ (0.9934)
ICK	$y = -0.010x^3 + 0.576x^2 + 6.146x$ (0.9998)	$y = (1.1 \times 10^{-6})x^3 - 0.001x^2 + 0.136x$ (0.9937)
TIF	$y = -0.017x^3 + 0.711x^2 + 6.277x$ (0.9998)	$y = (1.2 \times 10^{-6})x^3 - 0.001x^2 + 0.130x$ (0.9918)
NADH4	$y = -0.034x^3 + 1.373x^2 + 0.375x$ (0.9992)	$y = (3.2 \times 10^{-6})x^3 - 0.001x^2 + 0.199x$ (0.9904)
UBC7	$y = -0.117x^3 + 3.644x^2 + 6.308x$ (0.9990)	$y = (1.1 \times 10^{-6})x^3 - 0.001x^2 + 0.085x$ (0.9897)
NABP	$y = -0.123x^3 + 3.021x^2 - 0.491x$ (0.9982)	$y = (2.6 \times 10^{-6})x^3 - 0.001x^2 + 0.146x$ (0.9910)
PTMS	$y = -0.027x^3 + 1.267x^2 + 1.052x$ (0.9995)	$y = (2.9 \times 10^{-6})x^3 - 0.001x^2 + 0.190x$ (0.9908)
MHCIi	$y = -0.333x^3 + 5.546x^2 + 3.664x$ (0.9987)	$y = (1.3 \times 10^{-6})x^3 - 0.001x^2 + 0.084x$ (0.9953)
ARP2/3	$y = -0.027x^3 + 1.189x^2 + 3.326x$ (0.9999)	$y = (2.0 \times 10^{-6})x^3 - 0.001x^2 + 0.154x$ (0.9913)
IgJ	$y = -0.021x^3 + 1.011x^2 - 1.064x$ (0.9991)	$y = (3.9 \times 10^{-6})x^3 - 0.002x^2 + 0.236x + 0.806$ (0.9763)
CAL2	$y = -1.467x^3 + 15.756x^2 + 10.446x$ (0.9994)	$y = (5.7 \times 10^{-7})x^3 - 0.001x^2 + 0.044x$ (0.9830)

external spike-in control RNA (KAN^R) and 5.0 μL of 5 ng/μL sample mRNA. The concentration of each reverse primer varied from 0.01 to 5 μM to adjust the signal of each gene to within the dynamic range of the CEQ8000 fluorescence detector (see Table 1 for reverse primer dilutions). Thermal reaction parameters for the RT reaction were 48 °C for 1 min, 42 °C for 60 min and 95 °C for 5 min. Each PCR consisted of 4.0 μL of $MgCl_2$ (ABgene, Rockford, IL, USA), 2.0 μL of a 10× pool of forward primers all at 2 μM concentration, 0.7 μL of Taq polymerase (ABgene) and 9.3 μL of cDNA from the RT reaction. Cycling parameters were 95 °C for 10 min, followed by 35 cycles for 94 °C for 30 s, 55 °C for 30 s and 50 °C for 1 min.

Quantification models

We generated a standard curve for each gene using the Quant Tool estimator, which produces a cubic polynomial estimator with a *y*-intercept forced through zero and uses the mean value measured at each concentration in generating the coefficients. We also examined cubic polynomial equations, which included each individual replicate in the generation of the coefficients. The cubic lines included all replicates that were and were not forced through a zero intercept. We also generated a quadratic equation to examine the appropriateness of the cubic polynomial as the correct shape across the range of the instrument.

We were unable to rank the Quant Tool line estimator in comparison with our best-fit lines owing to unequal sample sizes, so we limited our analysis to the three best-fit lines generated from replicates for each gene. We ranked and compared the three models separately for each gene using Akaike's Information Criterion (AIC, Akaike 1974) corrected for small sample size (Burnham & Anderson 2002). Model ranking was carried out using R. We then used a chi-square test to determine whether one model consistently received the most model weight.

Validation of multiplexed qRT-PCRs

Bonneaud *et al.* (2011) give details of the house finch cDNA microarray construction and hybridization. Cross-platform validation was performed using both a Spearman rank correlation to test whether the magnitude of differential expression in the microarray was consistent with that in the multiplexed qRT-PCR and a one-sampling *t*-test framework to examine whether the direction of expression in the microarray and in the qRT-PCR was consistent. *P* values represent one-tailed estimates.

Results

Utilizing only the partial sequences available for 30 genes identified as differentially expressed in a cDNA microarray comparison, we were able to generate a set of transcript-specific primers for use with the GeXP analyser for 20 house finch genes (Table 1; Fig. 1). The final multiplex contained 22 peaks: 18 genes of interest, two housekeeping genes, one external spike-in control and 1 UDP.

We had to exclude between seven and nine replicates (mean = 7.2, SD = 0.6) from the standard curve of each

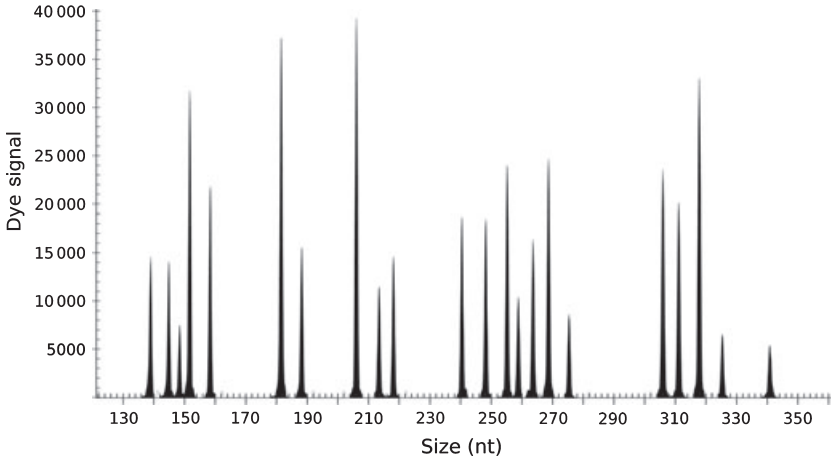

Fig. 1 Electropherogram of fluorescence-labelled RT-PCR products synthesized with primers described in Table 1.

gene owing to limits in the detection capabilities of the CEQ8000. Therefore, although each concentration of each gene should have included four replicates, the number of replicates was most commonly three and ranged from one to four. Table 2 gives equations for best-fit lines and r^2 values generated by the Quant Tool for each gene in the multiplex as well as those for the best model according to AIC ranks. In general, the model that included the cubic polynomial of the known concentration with the intercept forced through the origin received the most model weight ($\chi^2 = 24.72$, d.f. = 2, $P < 0.0001$). However, the cubic polynomial model without the intercept forced through the origin was the best model for three genes (Table S2), suggesting that the intercept may best be chosen quantitatively for some genes.

The house finch multiplexed qRT-PCR assay was validated using two different analytical approaches to compare it to a cDNA microarray study (data also given in Bonneaud *et al.* 2011). A correlational approach showed the magnitude of differential expression in the microarray was consistent with that in the qRT-PCR (Spearman rank correlation; Rs = 0.40, $N = 44$, $P = 0.002$). A one-sampling t-test framework showed a similar direction of expression in the microarray and in the qRT-PCR: control birds from two distinct populations ($T10 = 2.26$, $P = 0.024$) and infected birds from two distinct populations ($T15 = 3.33$, $P = 0.002$).

Discussion

Multiplexed qRT-PCR assays are currently lacking for nearly all species with little available sequence because of the difficulty of designing primers that will work appropriately in a multiplex for such species. We present a protocol for primer design and analysis of multiplexed quantitative RT-PCR assays that is useful for a species with little available sequence. Using a machine purchased by our laboratory for conventional sequencing and fragment size analysis of microsatellite regions of DNA, we were able to use the existing system to pursue studies of gene expression. Other systems already in place in many single PI laboratories and core sequencing facilities, most notably ABI sequencers, could, in theory, be adapted for similar purposes, although the optimization involved would be much greater without the active pursuance of this use by ABI. Fortunately, for CEQ users, Beckman Coulter has developed products and protocols specifically for gene expression studies. However, even when utilizing Beckman Coulter protocols and bioinformatic testing like that described here, design optimization of such a highly multiplexed reaction will probably require at least one to two months of both computational and benchtop work. When a researcher is interested in the expression of many genes but requires information from only a very small number of samples, a technology like RNA-seq will likely be more efficient. If, however, measuring expression of at least 20 specific genes is useful to your study and, for instance, you are interested in natural populations so that your total sample size is generally >40 individuals, then the method presented here will ultimately save time, money and resources over generating the same data set from singlet qRT-PCRs.

By combining Beckman Coulter's eXpress Profiler primer design software with Amplify 3, we were able to successfully select primers for 18 of 30 genes of interest and two of three housekeeping genes for inclusion in a single house finch multiplexed qRT-PCR gene expression assay. In particular, we used Amplify 3 to test primers designed by eXpress Profiler, and results were consistent *in vitro* and bioinformatically.

Beckman Coulter's eXpress Profiler suite of applications includes an option for designing primers for use in custom multiplexes (eXpress Designer). These programs provide what Beckman Coulter refers to as a 'first-pass'

multiplex. In this module, users select which gene sequences or accession numbers to include in the multiplex and can also define the minimum desired difference in size between amplified fragments and the range of fragment sizes desired. This latter feature is limited by the maximum size standard used in the reaction, which is 400 bp. The eXpress Designer will immediately return one pair of primers for each gene selected for inclusion according to default parameters for melting temperature and primer length. Primer pairs from this 'first pass', however, are designed by the software without regard for (i) sequences of the other genes of interest, (ii) primers designed for use in the same multiplex or (iii) additional genome sequence information that may be available on public databases.

For researchers studying species with little genome sequence information publicly available, this type of search can be frustrating and potentially unproductive. On the other hand, when a genome is available, results of searches can yield overwhelming amounts of information. Not performing a bioinformatic check, however, is also not advisable because of the cost required to empirically test and optimize each new primer pair to be added to a multiplex. Software such as Amplify 3 for Mac OSX provides a free and efficient method for evaluating multiple primers simultaneously against DNA sequences in a PCR-simulated reaction. Amplify 3 software currently only simulates PCRs, but it is easily adapted for use in testing RT-PCR primers.

Amplify 3 allowed us to do two things that would otherwise not have been possible: (i) to compare the entire set of pooled primers simultaneously against a gene sequence and (ii) to compare the primer plex against other sequences contained in our cDNA library, which are currently not available through NCBI. The ability to compare a multiplexed primer pool against each gene of interest dramatically increased our ability to detect instances of nonpaired primers amplifying undesigned fragments while limiting the scope of the search to a manageable set of comparisons.

In the case of the house finch, we began with extremely limited sequence information [i.e. 316 clones ranging in length from 91 to 1011 bp, which were sequenced from a SSH cDNA library as part of macroarray (Wang *et al.* 2006) and microarray studies (Bonneaud *et al.*2011)], and thus, our ability to redesign primers was limited for many genes of interest. Of the genes included in the final multiplex, clone lengths ranged from 216 to 806 bp. Thus, while we began with some sequence information pertaining to each gene of interest, approximately one-third of genes were ultimately excluded either as a result of predicted UDPs or after empirical testing at the bench. Having incomplete gene sequences for our candidate genes limited the ability of Amplify 3 to detect potential UDPs as well as our ability to redesign primers. Many of our sequences were so short that redesigning primers would have shifted the fragment size below the size standard included in the reactions. In such situations, it would be useful to have an extended size standard available. Of course, many of the transcripts expressed in our samples were unknown and so could not be tested with Amplify 3 and are likely responsible for any UDPs we eventually detected [including one at 306 bp that remained in the final multiplex (Fig. 1)]. By combining bioinformatic and simulated reactions with laboratory generated data, however, we were still able to create primers for a highly multiplexed gene expression assay in a comparatively short period of time and at a reduced cost.

One of the advantages of using multiplexed RT-PCRs rather than microarrays is the comparatively tiny amount of RNA required for this type of assay. Generation of a standard curve, however, requires microgram quantities of RNA rather than the nanogram quantities required for running samples. The ability to generate expression data for up to 35 genes at once, typically requiring only 25 ng of total RNA for each reaction, lends itself to the types of investigations into individual variation that are of interest to ecologists and evolutionary biologists. However, if one is interested in establishing a quantitative scale using a standard curve, it is essential to consider the amount of sample that is required for generation of such a curve when planning experiments.

Finally, we examined of the suitability of fitting a cubic polynomial function to the standard curve of each gene and consistently found it performed better than did a simpler quadratic equation. This function represents the changing response of the instrument over its dynamic range and appears to accurately predict the behaviour of the CEQ8000. That the shape is a third-order polynomial, however, means that data from the lower and upper concentration extremes could influence the lines such that very small errors or very large errors could have unduly large and small effects, respectively. One way to reduce this effect on sample data is to follow Beckman Coulter's protocol and guidelines for reverse primer concentration attenuation, which ensures that, for a given concentration, each gene in the standard curve is producing a comparably sized peak (i.e. within approximately 30 000 units) and increases the likelihood that expression of your samples will fall within the more linear portion of the curve. In addition, when estimating transcript abundance from a standard curve, we recommend considering whether or not to force the curve through zero, as is the default option for Beckman's Quant Tool software. Given that the line function of the standard curve is a third-order polynomial, forcing the fit through zero can have large effects on the values at the lower end of the range. In our study of gene expression in house finch

spleen, we have used a simple estimator using R (http://www.r-project.org) that allows the user to determine whether to force the y-intercept through zero and that also allows for each replicate of the curve to be included as a data point in the fit of the predicted line. Including replicates in the fit of the line increases the degrees of freedom of the model and weights each concentration along the line according to the number of replicates. This is important because, especially at the high and low ends of the dynamic range of the instrument, one or more of the replicates may not be interpretable by the CEQ8000. When this occurs, some concentrations along the curve will have more replicates than others, and thus, some points can be estimated with greater certainty than others. Unequal variance is a violation of the assumptions of linear regression and should thus be avoided. Further, using mean values in regression artificially reduces the variation in the data and inflates r^2-values. In many cases, it may be possible to simply rerun the standard curve until each point is represented in quadruplicate. When samples are limited in quantity, as was the case of in our house finch study, multiple runs may not be an option. The Quant Tool estimator completely ignores this possibility, and we argue that the r^2-values produced by the Quant Tool estimator are invalid. Thus, our estimator appears to provide a better representation of the data generated by the GeXP system although analyses of additional experimental systems are necessary.

Acknowledgements

We thank Vaclav Majer, Frank F. Bartol, Camille Bonneaud, Laura Estep, John Beaty, Kevin Fielman and Scott Santos for technical advice and Eva Simyani, Emily Brennan and Lauren Drake for assistance in the laboratory. This work was funded by a Biogrant from the Office for Vice President for Research at Auburn University and by funds from the Center for Environmental and Cellular Signal Transduction (Auburn University) and NSF grant IOS-0923600 (GEH). SLB was supported by a grant from the Alabama Experimental Program to Stimulate Competitive Research.

References

Akaike H (1974) A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, **19**, 716–723.

Bonneaud C, Balenger SL, Russell AF, Zhang J, Hill GE, Edwards SV (2011) Rapidly evolving disease resistance manifest through functional changes in gene expression in a wild bird. *Proceedings of the National Academy of Sciences of the United States of America*, **108**, 7866–7871.

Brisson M, Hall S, Hamby RK, Park R, Srere HK (2004) Optimization of single and multiplex real-time PCR. In: *A-Z of Quantitative PCR* (ed. Bustin SA), pp. 619–642. International University Line, La Jolla, California.

Burnham KP, Anderson DR (2002) *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*, pp. 488. Springer-Verlag, New York.

Chen QR, Vansant G, Oades K *et al*. (2007) Diagnosis of the small round blue cell tumors using multiplex polymerase chain reaction. *Journal of Molecular Diagnostics*, **9**, 80–88.

Hoffmann AA, Willi Y (2008) Detecting genetic responses to environmental change. *Nature Reviews Genetics*, **9**, 421–432.

Nagel MA, Gilden D, Shade T, Gao B, Cohrs RJ (2009) Rapid and sensitive detection of 68 unique varicella zoster virus gene transcripts in five multiplex reverse-transcription-polymerase chain reactions. *Journal of Virological Methods*, **157**, 62–68.

Piertney SB, Webster LMI (2010) Characterising functionally important and ecologically meaningful genetic diversity using a candidate gene approach. *Genetica*, **138**, 419–432.

Rai AJ, Kamath RM, Gerald W, Fleisher M (2009) Analytical validation of the GeXP analyzer and design of workflow for cancer-biomarker discovery using multiplexed gene-expression profiling. *Analytical & Bioanalytical Chemistry*, **393**, 1505–1511.

Shiu SH, Borevitz JO (2008) The next generation of microarray research: applications in evolutionary and ecological genomics. *Heredity*, **100**, 141–149.

Therianos S, Zhu M, Pyun E, Coleman PD (2004) Single-channel quantitative multiplex reverse transcriptase-polymerase chain reaction for large numbers of gene products differentiates nondemented from neuropathological Alzheimer's disease. *American Journal of Pathology*, **164**, 795–806.

Ungerer MC, Johnson LC, Herman MA (2008) Ecological genomics: understanding gene and genome function in the natural environment. *Heredity*, **100**, 178–183.

Vasemägi A, Primmer CR (2005) Challenges for identifying functionally important genetic variation: the promise of combining complementary research strategies. *Molecular Ecology*, **14**, 3623–3642.

Wang Z, Farmer KL, Hill GE, Edwards SV (2006) A cDNA macroarray approach to parasite-induced gene expression in a songbird host: genetic response of House Finches to experimental infection by *Mycoplasma gallisepticum*. *Molecular Ecology*, **15**, 1263–1273.

Data Accessibility

DNA sequences: Provided in Table 1.

Supporting Information

Additional supporting information may be found in the online version of this article.

Table S1 Contents of each reference pool of total RNA used to generate standard curves.

Table S2 Rank comparisons of three best-fit lines for each gene in multiplex.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.