Journal of the American Statistical Association

ISSN: 0162-1459 (Print) 1537-274X (Online) Journal homepage: https://www.tandfonline.com/loi/uasa20

Fully Bayesian Analysis of RNA-seq Counts for the Detection of Gene Expression Heterosis

Will Landau, Jarad Niemi & Dan Nettleton

To cite this article: Will Landau, Jarad Niemi & Dan Nettleton (2019) Fully Bayesian Analysis of RNA-seq Counts for the Detection of Gene Expression Heterosis, Journal of the American Statistical Association, 114:526, 610-621, DOI: 10.1080/01621459.2018.1497496

To link to this article: https://doi.org/10.1080/01621459.2018.1497496

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION
2019, VOL. 114, NO. 526, 610–621: Applications and Case Studies
https://doi.org/10.1080/01621459.2018.1497496

Fully Bayesian Analysis of RNA-seq Counts for the Detection of Gene Expression Heterosis

Will Landau, Jarad Niemi, and Dan Nettleton

Department of Statistics, Iowa State University, Ames, IA

ABSTRACT

Heterosis, or hybrid vigor, is the enhancement of the phenotype of hybrid progeny relative to their inbred parents. Heterosis is extensively used in agriculture, and the underlying mechanisms are unclear. To investigate the molecular basis of phenotypic heterosis, researchers search tens of thousands of genes for heterosis with respect to expression in the transcriptome. Difficulty arises in the assessment of heterosis due to composite null hypotheses and nonuniform distributions for *p*-values under these null hypotheses. Thus, we develop a general hierarchical model for count data and a fully Bayesian analysis in which an efficient parallelized Markov chain Monte Carlo algorithm ameliorates the computational burden. We use our method to detect gene expression heterosis in a two-hybrid plant-breeding scenario, both in a real RNA-seq maize dataset and in simulation studies. In the simulation studies, we show our method has well-calibrated posterior probabilities and credible intervals when the model assumed in analysis matches the model used to simulate the data. Although model misspecification can adversely affect calibration, the methodology is still able to accurately rank genes. Finally, we show that hyperparameter posteriors are extremely narrow and an empirical Bayes (eBayes) approach based on posterior means from the fully Bayesian analysis provides virtually equivalent posterior probabilities, credible intervals, and gene rankings relative to the fully Bayesian solution. This evidence of equivalence provides support for the use of eBayes procedures in RNA-seq data analysis if accurate hyperparameter estimates can be obtained. Supplementary materials for this article are available online.

ARTICLE HISTORY
Received May 2017
Revised January 2018

KEYWORDS
CUDA; Empirical Bayes; Graphics processing unit; Hierarchical model; Hybrid vigor; Negative-binomial

1. Introduction

Heterosis, or hybrid vigor, is the biological phenomenon in which hybrid progeny surpasses each of its inbred parents with respect to some characteristic. Ever since Darwin (1876) documented heterosis, the term has usually referred to traits at the phenotypic level, and phenotypic heterosis has long been used to enhance crops and livestock. For example, one well-known maize hybrid described by Hallauer and Miranda (1981) and Hallauer, Carena, and Miranda Filho (2010) has taller, faster-growing stalks with more grain yield than either inbred parent. Similar breeding techniques have used heterosis to improve rice (Yu et al. 1997), alfalfa (Riday and Brummer 2002), tomatoes (Krieger, Lippman, and Zamir 2010), and fish (Wohlfarth 1993). However, the underlying genomic mechanisms of phenotypic heterosis remain unclear (Coors and Pandey 1999; Lippman and Zamir 2007).

Researchers have hypothesized that the enhanced expression of one or more genes in the hybrid relative to both inbred parents, which we call gene expression heterosis, may help account for phenotypic heterosis (Swanson-Wagner et al. 2006; Springer and Stupar 2007). Gene expression heterosis has been measured with a variety of experimental platforms, including microarray and its successor, RNA-sequencing (RNA-seq) (Wang et al. 2006, 2010; Oshlack, Robinson, and Young 2010). Both platforms measure the relative expression levels of genes in organisms across multiple groups or experimental conditions. Relative to microarray, RNA-seq has less noise and higher throughput, among other advantages (Landau and Liu 2013).

However, both microarray and RNA-seq present serious statistical challenges. Since a large number of expressed genes are assayed only a handful of times each, the data analysis is a low-sample-size multiple testing scenario prone to frequent false discoveries. With the additional difficulty of composite null hypotheses for gene expression heterosis detection (Ji, Liu, and Nettleton 2014; Niemi et al. 2015), assessing the false discovery rate (FDR) in the heterosis problem is difficult. In multiple testing scenarios with composite null hypotheses, the distribution of the null *p*-values is typically not uniform (Bayarri and Berger 2000; Robins, Vaart, and Ventura 2000; Sun and McLain 2012; Dickhaus 2013), which violates a key assumption of many ubiquitous FDR control procedures (Benjamini and Hochberg 1995; Storey 2003; Meinhausen and Rice 2006; Dudoit and Laan 2008). Some techniques addressing composite null hypotheses generate null *p*-values that are less likely to violate the uniformity assumption (Bayarri and Berger 2000; Romano and Shaikh 2006; Cabras 2010; Chi 2010; Dickhaus 2013), but they do not entirely remove the assumption itself.

To mitigate these statistical challenges in microarrays, Ji, Liu, and Nettleton (2014) built a normal hierarchical model to borrow information across genes, improve parameter estimation, and provide a data-based Ockham's razor effect to control the false discovery rate. Building on the work of Ji,

CONTACT Jarad Niemi niemi@iastate.edu Department of Statistics, Iowa State University, Ames, IA 50011.
Supplementary materials for this article are available online. Please go to www.tandfonline.com/r/JASA.

Liu, and Nettleton, Niemi et al. (2015) constructed a negative-binomial hierarchical model for use in RNA-seq experiments. Both approaches use an empirical Bayes (eBayes) procedure for parameter estimation and rely on the resulting conditional posterior probabilities of the composite null and alternative hypotheses to identify genes with expression heterosis. They justify the eBayes procedure as an approximation to a fully Bayesian analysis based on asymptotic convergence of the posterior, but provide no supporting evidence for this claim. Nonetheless, eBayes procedures are becoming increasingly popular in statistical genomics (Hardcastle and Kelly 2010; Wu, Wang, and Wu 2012; Leng et al. 2013; Lithio and Nettleton 2015).

Presumably the use of eBayes rather than fully Bayesian procedures is due to the computational difficulties involved in estimating the hundreds of thousands of parameters in these hierarchical models. In our experience, the general-purpose Markov chain Monte Carlo (MCMC) approaches, as implemented in software such as WinBUGS (Lunn et al. 2000), OpenBugs (Lunn et al. 2009), JAGS (Plummer et al. 2003), Stan (Carpenter et al. 2017), and NIMBLE (de Valpine et al. 2017), are computationally intractable for models of this size. Fortunately, new MCMC approaches, based on parallelization on graphics processing units, allow for fully Bayesian analyses of these models in reasonable time frames (Landau and Niemi 2016). In this article, we propose a fully Bayesian analysis of a hierarchical regression model for count data, compare this approach to two best-case-scenario eBayes procedures, and analyze a two-hybrid experiment in maize to identify genes with gene expression heterosis.

In Section 2, we introduce the motivating two-hybrid maize heterosis RNA-seq dataset. Section 3 presents an overdispersed, hierarchical RNA-seq model, useful for the analysis of data from a variety of experimental designs, and describes the fully Bayesian estimation procedure. Section 4 expounds simulation studies based on a two-hybrid plant-breeding scenario to assess our fully Bayesian approach in terms of estimation, inference, and heterosis gene detection, and we compare our method to two best-case-scenario empirical Bayes counterparts. Finally, Section 5 details our analysis of the maize dataset.

2. Two-Hybrid Plant-Breeding Experiment for Heterosis Gene Detection

We focus on the RNA-seq dataset from Paschold et al. (2012), which contains read counts of $G = 39,656$ genes on $N = 16$ biological replicates divided evenly among four genetic varieties. In the underlying experiment, multiple maize seedlings from each variety were germinated according to the procedure by Hoecker et al. (2006). Three and a half days after germination, the primary roots of the seedlings were harvested. Within each variety, four pools of primary roots served as four biological replicates. Following the procedure by Winz and Baldwin (2001), the 16 collections of roots were ground under liquid nitrogen, and the RNA was isolated. Complementary DNA (cDNA) fragments were then synthesized in preparation for sequencing. Next the cDNA from the replicates was divided between two flow cells (i.e., removable compartments for genetic material in the RNA-sequencing platform) as identified in supplementary Table S1. The two flow cells were placed into an Illumina Genome Analyzer II, where the cDNA fragments were read, amplified, and counted. The reads from the sequencing platform were mapped to the B73 reference genome (RefGen_v2) (Schnable et al. 2009), and the preprocessed and amplified read counts for each gene and biological replicate were collected into a data table. The resulting $G \times N$ table of read counts is provided in Table S1.

The varieties in the Paschold et al. data are inbred variety B73, inbred variety Mo17, B73×Mo17 (a first-generation hybrid created by pollinating B73 with Mo17), and Mo17×B73 (a first-generation hybrid created by pollinating Mo17 with B73). This is a special case of a more general plant hybrid scenario where there are two parent varieties and one first-generation hybrid variety for each direction of pollination. For the general scenario, we shall use P1, P2, H12, and H21 for the parents and the first-generation hybrids, respectively. For the Paschold et al. dataset, P1 is B73, P2 is Mo17, H12 is B73×Mo17, and H21 is Mo17×B73.

A major goal is to identify genes that have heterosis with respect to their expression levels: that is, those with significantly higher (in the case of high-parent heterosis) or significantly lower (low-parent heterosis) expression levels in one or both hybrids relative to their parents. For each of the high-parent and low-parent cases, we are interested in heterosis with respect to H12, H21, and the log-scale mean expression level of H12 and H21 together. Table 1 provides the six types of gene expression heterosis parameterized in terms of log-scale mean expression levels μ_{gv} specific to gene g and variety v. (The third column of Table 1 is discussed in Section 4.) "High (low)-parent H" indicates hybrid H has higher (lower) mean expression than both parents while "high (low)-parent mean" indicates that the average of the hybrids is higher (lower) than both parents. Our major objective is to provide a measure of the strength of evidence for each kind of heterosis for each gene which we accomplish using posterior probabilities of heterosis under the model in Section 3.1.

Table 1. Heterosis hypotheses for a two-parent (P1 and P2), two-hybrid (H12 and H21) gene expression experiment represented in terms of the log-scale mean expression μ_{gv} for gene g and variety v and in terms of the parameters $\beta_{g\ell}$ corresponding to columns $\ell = 1, \ldots, L = 5$ of the model matrix X in Equation (1) of Section 4.

Heterosis	With log-scale group means	With $\beta_{g\ell}$ parameters
high-parent H12	$\mu_{g,\mathrm{H12}} > \max(\mu_{g,\mathrm{P1}}, \mu_{g,\mathrm{P2}})$	$2\beta_{g2} + \beta_{g4},\ \ 2\beta_{g3} + \beta_{g4} > 0$
low-parent H12	$\mu_{g,\mathrm{H12}} < \min(\mu_{g,\mathrm{P1}}, \mu_{g,\mathrm{P2}})$	$-2\beta_{g2} - \beta_{g4}, -2\beta_{g3} - \beta_{g4} > 0$
high-parent H21	$\mu_{g,\mathrm{H21}} > \max(\mu_{g,\mathrm{P1}}, \mu_{g,\mathrm{P2}})$	$2\beta_{g2} - \beta_{g4},\ \ 2\beta_{g3} - \beta_{g4} > 0$
low-parent H21	$\mu_{g,\mathrm{H21}} < \min(\mu_{g,\mathrm{P1}}, \mu_{g,\mathrm{P2}})$	$-2\beta_{g2} + \beta_{g4},\ \ 2\beta_{g3} + \beta_{g4} > 0$
high-parent mean	$\mu_{g,\mathrm{H12}} + \mu_{g,\mathrm{H21}} > 2\max(\mu_{g,\mathrm{P1}}, \mu_{g,\mathrm{P2}})$	$\beta_{g2},\ \ \beta_{g3} > 0$
low-parent mean	$\mu_{g,\mathrm{H12}} + \mu_{g,\mathrm{H21}} < 2\min(\mu_{g,\mathrm{P1}}, \mu_{g,\mathrm{P2}})$	$-\beta_{g2}, -\beta_{g3} > 0$

The complement of each of the heterosis hypotheses in Table 1 is a composite null hypothesis. For example, the "no high-parent H12 heterosis" null hypothesis for row 1 of Table 1 can be written as

$$\mu_{g,P1} \leq \mu_{g,H12} \leq \mu_{g,P2} \text{ or } \mu_{g,P2} \leq \mu_{g,H12} \leq \mu_{g,P1}$$

or (equivalently, in terms of the $\beta_{g\ell}$ parameters) as

$$2\beta_{g2} + \beta_{g4} \leq 0 \text{ or } 2\beta_{g3} + \beta_{g4} \leq 0.$$

Currently available software for the analysis of RNA-seq data allows for the testing of point null hypotheses concerning single parameters or linear combinations of parameters but does not (to our knowledge) provide tests for complex composite nulls like those encountered in the search for gene expression heterosis. Even if existing RNA-seq analysis software were modified to provide p-values for tests of composite null hypotheses, the difficulties in multiple testing described in Section 1 would limit their usefulness. Thus, new methodology is needed.

3. Fully Bayesian Methodology

To address the strength of evidence for the various types of heterosis, we build a hierarchical regression model for count data capable of borrowing information across genes and accounting for gene-specific overdispersion (Niemi et al. 2015). We perform a fully Bayesian analysis based on vague proper priors using a slice-sampling-within-Gibbs Markov chain Monte Carlo (MCMC) algorithm that uses general-purpose graphics processing units (GPUs) for efficient computation (Landau and Niemi 2016).

3.1. Hierarchical Model for RNA-seq

Let y_{gn} be the RNA-seq count (i.e., the relative expression level) of gene g ($g = 1, \ldots, G$) in replicate n ($n = 1, \ldots, N$), and let y be the $G \times N$ matrix of the y_{gn}'s. Let X be an $N \times L$ model matrix that connects the N samples (i.e., RNA-seq replicates) to the genotypes, blocking factors, etc., allowing for data from a variety of experimental designs to be analyzed using this methodology. Taking X_n to be the n'th row of X, we let $y_{gn} \overset{\text{ind}}{\sim} \text{Poisson}(\exp(h_n + \varepsilon_{gn} + X_n\beta_g))$. The h_n's are computed from y (as explained in Section 3.2) and are treated as constants that play the role of normalization factors in other RNA-seq models, taking into account sample-specific nuisance effects such as sequencing depth (Anders and Huber 2010; Robinson and Oshlack 2010; Si and Liu 2013). The ε_{gn} parameters account for overdispersion, and we assume $\varepsilon_{gn}|\gamma_g^2 \overset{\text{ind}}{\sim} \text{Normal}(0, \gamma_g^2)$ such that the γ_g^2 parameters are analogous to the gene-specific negative-binomial dispersion parameters widespread in other RNA-seq data analysis methodology (Landau and Liu 2013). We assumed $1/\gamma_g^2 \overset{\text{ind}}{\sim} \text{Gamma}(\nu/2, \nu\tau/2)$, parameterized such that $E[1/\gamma_g^2] = 1/\tau$.

The gene-specific vector-valued parameters β_g account for the effects on gene expression of the experimental variables of interest. Aside from the normalization factor h_n, we interpret $X_n\beta_g$ to be the log-scale mean expression level of gene g in RNA-seq sample n. To borrow information across genes, we assign $\beta_{g\ell} \mid \theta_\ell,\ \sigma_\ell \overset{\text{ind}}{\sim} \text{Normal}(\theta_\ell,\ \sigma_\ell^2)$ for each ℓ.

This model is similar to negative binomial regression models from other RNA-seq data analyses (McCarthy, Chen, and Smyth 2012; Wu, Wang, and Wu 2012), with one difference being that we mix Poisson distributions over log-normal rather than gamma distributions. This choice is made primarily to ease computational implementation by reducing the number of distinct types of full conditionals. Due to the similarity between the gamma and log-normal distributions, we suspect that similar results would be obtained if gamma distributions had been used.

3.2. Inference on Gene-Specific Parameters and Heterosis Probabilities

To perform Bayesian analyses, we assigned independent priors for the hyperparameters. Specifically $\tau \sim \text{Gamma}(a, b)$, $\nu \sim \text{Uniform}(0, d)$, $\theta_\ell \overset{\text{ind}}{\sim} \text{Normal}(0, c_\ell^2)$, and $\sigma_\ell \overset{\text{ind}}{\sim} \text{Uniform}(0, s_\ell)$ for $\ell = 1, \ldots, L$ with the values for Roman letters chosen to provide vague, relatively uninformative priors on these hyperparameters. Before parameter estimation, we calculated the log-scale replicate-specific normalization constants h_n as follows. We first calculated log-scale counts $w_{gn} = \log(y_{gn} + 0.5 \cdot I(y_{gn} = 0))$ where $I(A)$ is 1 if A is true and 0 otherwise, replicate-specific means $\overline{w}_{\cdot n} = \frac{1}{G}\sum_{g=1}^{G} w_{gn}$, and the grand mean $\overline{w}_{\cdot\cdot} = \frac{1}{N}\sum_{n=1}^{N} \overline{w}_{\cdot n}$. Afterward, we set $h_n = \overline{w}_{\cdot n} - \overline{w}_{\cdot\cdot}$ for $n = 1, \ldots, N$. We evaluated alternative ways to calculate normalization constants, and they all resulted in similar inference. As each h_n is calculated by summarizing signals from thousands of genes, the uncertainty associated with each h_n is negligible relative to other sources of variation. Thus, treating normalization factors like h_n as fixed and known is the standard practice in RNA-seq analysis and the strategy we adopt throughout this article.

To estimate the full joint posterior distribution of the parameters, we used the parallelized slice-sampling-within-Gibbs MCMC algorithm described in Landau and Niemi (2016). Without parallel computing, the computational burdens of the MCMC would be heavy, with the elapsed runtime for each dataset stretching over multiple days (see Section 5). However, with the strategy we employed, which uses massively parallel computing that takes advantage of general-purpose graphics processing units (GPUs), we analyzed typical-sized RNA-seq data in just a few hours per dataset. The algorithm accelerates MCMC computation by executing conditionally independent Gibbs steps in parallel and using parallelized reductions to compute the full conditional distributions of the hyperparameters. Efficiency is increased by reducing the data transferred from GPU to CPU, and thus we limited posterior samples to all hyperparameters and a random subset of gene-specific parameters. For each parameter ψ, we also record $\overline{\psi} = \frac{1}{M}\sum_{m=1}^{M} \psi^{(m)}$ and $\overline{\psi^2} = \frac{1}{M}\sum_{m=1}^{M} (\psi^{(m)})^2$, where $\psi_g^{(m)}$ is the mth Monte Carlo sample of ψ_g, and approximate the posterior $p(\psi|y)$ with $N(\overline{\psi}, \overline{\psi^2} - \overline{\psi}^2)$. Finally, we assessed the posterior probabilities of heterosis in Table 1 via their ergodic averages,

for example, according to the first row of Table 1,

$$P(\text{high-parent H12 heterosis for gene } g|y) \approx \frac{1}{M}\sum_{m=1}^{M} I\big(2\beta_{g2}^{(m)} + \beta_{g4}^{(m)} > 0 \text{ and } 2\beta_{g3}^{(m)} + \beta_{g4}^{(m)} > 0\big).$$

For each analysis, we ran four independent Markov chains with overdispersed starting values relative to the full joint posterior distribution of the parameters. For each chain, we used a burn-in period of 10^5 iterations (the first 50 of those iterations without tuning the slice sampler), and then 10^5 true iterations with a thinning interval of 20 so that 5000 samples are retained for a small subset of parameters of interest. We monitored those chains for convergence using Gelman–Rubin potential scale reduction factors $\widehat{R}$ (Gelman and Rubin 1992) which were calculated using $\overline{\psi}$ and $\overline{\psi^2}$ (see Landau and Niemi 2016). Specifically, we monitored $\widehat{R}$ on the $2L + 2$ hyperparameters, the $G \times L$ parameters $\beta_{g\ell}$, and the G hierarchical variance parameters γ_g^2. In our experience, we found $\widehat{R}$ values near one for all but a few of the $\approx 10^5$ gene-specific parameters that varied when rerunning the MCMC. In addition, since we retained Monte Carlo samples of the hyperparameters, we monitored hyperparameter effective sample size, which we generally found to be well above the 10 to 100 effective samples recommended by Gelman et al. (2013).

For computation, we developed and used the R (R Core Team 2016) packages fbseq and fbseqCUDA. In Section 5, we also developed and used fbseqOpenMP, also available on GitHub. The fbseqOpenMP package is a version of fbseqCUDA that replaces CUDA with OpenMP, a less powerful but more accessible parallel computing technology (Dagum and Menon 1998). We also released fbseqStudies, an R package that replicates all the results of this article. The fbseqStudies package is publicly available through the GitHub repository of the same name. Installing fbseqStudies according to the instructions in the package vignette and running the `paper_case()` function reproduces the computation, figures, tables, etc., shown in all the following sections.

4. Studies of Simulated Heterosis Datasets

We assessed coverage of credible intervals (CIs), calibration of posterior probabilities, and the ability of our method to rank genes by constructing simulations with known values of the gene-specific parameters. For CIs, we calculated coverage, that is, the proportion of genes whose true parameter value falls within the interval, across all genes and as a function of the parameter value, and compared this proportion to the intervals' credibility. To assess calibration of posterior probabilities, we constructed kernel-smoothed plots of the true heterosis status of each gene against its estimated posterior probability, and we refer to these figures as calibration curves throughout. For each calibration curve, we calculated the mean absolute vertical distance from the identity line, which we call calibration error. Posterior probabilities provide a ranking of genes of interest for each hypothesis in Table 1. To evaluate these rankings, we constructed receiver operating characteristic (ROC) curves and the areas under these curves (Landau and Liu 2013).

For all the simulation studies in this article, we simulated RNA-seq count datasets under the plant hybrid scenario from Section 2. Each simulated dataset contained count data on $G = 30,000$ genes for $N = 16$ or $N = 32$ total replicates spread evenly over the P1, P2, H12, and H21 varieties. From left to right, the columns in each count data table y corresponded to P1, P2, H12, and then H21, respectively. Within each variety, the all columns for the first block (flow cell) preceded all the columns of the second block. Thus our model matrix, which we also used to analyze the Paschold et al. dataset in Section 5, is compactly represented as

$$X = \left(\begin{bmatrix} 1 & 1 & -1 & 0 \\ 1 & -1 & 1 & 0 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & -1 \end{bmatrix} \otimes J_{(N/4)\times 1} \quad J_{(N/4)\times 1} \otimes \begin{bmatrix} 1 \\ 1 \\ -1 \\ -1 \end{bmatrix} \right), \tag{1}$$

where "⊗" denotes the Kronecker product and $J_{m\times n}$ is the m by n matrix with all entries equal to 1. We chose the first $\ell = 1, \ldots, 4$ columns of the $N \times L$ model matrix ($L = 5$) to strategically model gene expression heterosis. For a maize dataset similar to that of Paschold et al., Lithio and Nettleton (2015) found strong correlations among gene-specific model coefficient parameters, a phenomenon that could potentially violate our model's conditional independence assumptions. To mitigate this effect among columns $\ell = 1, \ldots, 4$, we selected a slightly reparameterized, two-hybrid version of the parameterization used by Ji, Liu, and Nettleton (2014) and Niemi et al. (2015). Column $\ell = 5$ of X is a gene-specific experimental block effect, used in the analysis of the Paschold et al. data (Section 5) to account for the difference between the two flow cells of the sequencing platform in the original experiment. Table 2 provides interpretations for the parameters $\beta_{g\ell}$ in terms of the log-scale group means while Table 1 provides the method to evaluate each heterosis hypothesis using these $\beta_{g\ell}$ parameters.

For Section 4.2, we evaluated coverage, calibration, and ranking for simulations where the assumed model in the analysis matched the model used to generate data. Section 4.3 provides an assessment of robustness under two alternative data-generating scenarios as well as a comparison to an eBayes approach.

4.1. edgeR, a Benchmark Method

Throughout our analyses, to help measure the effectiveness of our fully Bayesian, hierarchical-model-driven scheme, we apply the method by McCarthy, Chen, and Smyth (2012), an alternative approach whose model only borrows information across genes for the overdispersion parameters. McCarthy et al. proposed a negative binomial generalized linear model and implemented the estimation and inference in the `edgeR` package in R. Before estimation, replicate-specific normalization constants are computed with the trimmed mean of M-values (TMM) method by Robinson and Oshlack (2010). Next, gene-specific negative binomial dispersions are obtained by maximizing weighted sums of gene-specific Cox–Reid adjusted profile likelihoods, where the weighting occurs within groups of similar genes to borrow information and improve

Table 2. For the model matrix in Equation (1), interpretations of the parameters $\beta_{g\ell}$ in terms of the group means μ_{gv} (gene g, variety v). Group means and interpretations in the table are given on the natural logarithmic scale. Since β_{g5} cannot be expressed in terms of the group means, only the prose interpretation is given.

$\beta_{g\ell}$	Using group means	Log-scale interpretation
$\ell = 1$	$\frac{\mu_{g,\text{P1}} + \mu_{g,\text{P2}}}{2}$	Parental mean
$\ell = 2$	$\frac{(\mu_{g,\text{H12}} + \mu_{g,\text{H21}})/2 - \mu_{g,\text{P2}}}{2}$	Half difference, hybrid mean versus parent 2
$\ell = 3$	$\frac{(\mu_{g,\text{H12}} + \mu_{g,\text{H21}})/2 - \mu_{g,\text{P1}}}{2}$	Half difference, hybrid mean versus parent 1
$\ell = 4$	$\frac{\mu_{g,\text{H21}} - \mu_{g,\text{H12}}}{2}$	Half the difference between hybrids
$\ell = 5$	–	Flow cell block effect

estimation. Then, gene-specific model coefficient parameters are estimated independently via maximum likelihood. While the method provides a useful baseline for comparing parameter estimates, it does not provide a comparison to posterior heterosis probabilities (Niemi et al. 2015).

4.2. Assessing Performance When the Data-Generation and Analysis Models Agree

For Simulation Study 1, we generated 10 datasets from the model in Section 3.1 using $N = 16$ total replicates per dataset. To generate each dataset, we fixed hyperparameters at $\nu = 3$, $\tau = 0.01$, $\theta_1 = 3, \theta_2 = 0, \theta_3 = -0.007, \theta_4 = -0.005, \theta_5 = 0.008, \sigma_1^2 = 1, \sigma_2^2 = 0.04, \sigma_3^2 = 0.03, \sigma_4^2 = 0.0005$, and $\sigma_5^2 = 0.1$ which are values similar to the posterior modes of the real data shown in Figure 3. Conditioning on those fixed hyperparameter values, we generated the γ_g^2's and $\beta_{g\ell}$'s from their hierarchical distributions under the model. Similarly, we conditioned on those γ_g^2 values to generate the ε_{gn}'s from their hierarchical distributions. Finally, with parameter values in hand and the model matrix X given by Equation (1), we generated RNA-seq counts y_{gn} using the Poisson$(\exp(\varepsilon_{gn} + X_n\beta_g))$ distribution from the model (with $h_n = 0$ for all $n = 1, \ldots, N$ which was also assumed when performing inference).

We used a single node of a computing cluster with a single NVIDIA K20 GPU, two 2.0 GHz 8-Core Intel E5 2650 processors, and 64 GB of memory. The maximum total elapsed runtime per dataset was around 3.2 hr using the GPU-parallelized algorithm. For each dataset, no more than 9 $\widehat{R}$ values were above 1.1 and these all correspond to $\beta_{g\ell}$ parameters. For the hyperparameters, the minimum effective sample size across all simulated datasets was $\sim$ 600 (for σ_4^2). Evidence of lack of convergence was weak overall, though estimation and inference may be poor for the few genes with $\widehat{R} > 1.1$.

With these results, we assessed the accuracy of posterior inference on the hyperparameters. Figure S1 shows the estimated 50% and 95% credible intervals for all 10 datasets, along with the true values used in data generation. There appears to be no apparent overall bias in the location of the intervals and, for each parameter, the number of intervals covering the truth is consistent with the appropriate binomial distribution.

We also assessed posterior inference on the parameters $\beta_{g\ell}$ because they are important for detecting heterosis genes. As described in Section 3.2, we retained full samples of only a few randomly selected $\beta_{g\ell}$ and otherwise approximate posteriors via their normal approximations. Across the simulations, coverage for normal-based 95% CIs ranged from 94.7% to 95.4% for $\ell \neq 4$ and from 92.9% to 96.7% for $\ell = 4$. Figure 1 displays the smoothed coverage proportions plotted against the true parameter values. From the figure, for each ℓ, coverage exceeded desired minimum near the overall mean true parameter value, but dropped for extreme parameter values. The lower row in Figure 1 shows that for $\ell > 1$ the low $\beta_{g\ell}$'s tended to be overestimated and the high $\beta_{g\ell}$'s tended to be underestimated, that is, the CIs shrunk toward the hierarchical mean.

Figure S2 shows the mean squared error (MSE) of the model coefficient estimates for each method, where the mean is computed over all the genes. MSE is significantly lower in our method relative to edgeR, which is a reflection of the benefits of borrowing information.

Finally, Figure S3 shows a receiver operating characteristic (ROC) curve for each kind of heterosis and each dataset. The results, all favorable for our proposed approach, are extremely similar across datasets. With areas under the curves ranging from 0.916 to 0.922 for low-parent heterosis and from 0.930 to 0.936 for high-parent heterosis, our method competently filtered out the heterosis from the null genes. In addition, all the calibration curves in Figure S4 are extremely close to the identity line, so the estimated posterior probabilities of heterosis were extremely accurate and well-calibrated.

4.3. Robust Comparison of Fully Bayes Versus eBayes

In RNA-seq analyses, eBayes is relatively more common than fully Bayes (Hardcastle and Kelly 2010; Wu, Wang, and Wu 2012; Ji, Liu, and Nettleton 2014; Niemi et al. 2015) due to the reduced computational burden even though theoretically, eBayes procedures risk lower quality estimation and posterior inference by ignoring hyperparameter uncertainty. For this study, we considered two eBayes versions of our fully Bayesian approach: the *Oracle* approach fixed hyperparameters at the values used in data generation while the *Means* approach fixed hyperparameters at the posterior means estimated from the fully Bayesian approach. Thus, these methods provided a comparison under the best possible case for eBayes, and we did not address the question of how to obtain eBayes estimates of hyperparameters without running a fully Bayesian analysis.

For a robust comparison of our three methods, we simulated two datasets, one with $N = 16$ total replicates and the other with $N = 32$, under each of the three scenarios below. To

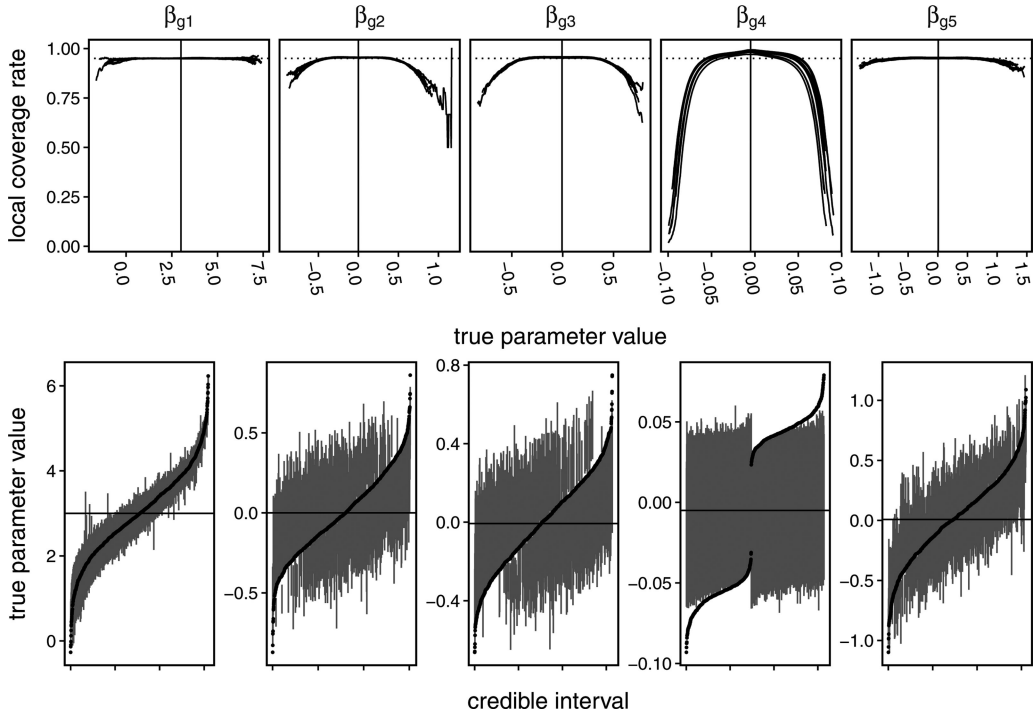

Figure 1. Posterior inference on the $\beta_{g\ell}$ parameters for Simulation Study 1 in Section 4.2. For each $\ell = 1, \ldots, 5$ and each dataset, the top row shows the kernel-smoothed local proportion of $\beta_{g\ell}$ parameters for which 95% CIs cover the true parameter values. The horizontal dashed lines are at 0.95, the desired coverage rate, and the solid black vertical lines indicate the respective true values of the hierarchical means θ_ℓ used to generate the count data. The bottom row shows, for $\ell = 1$ through 5 in the same order, the CIs (dark gray vertical lines) that do not cover the true parameter values (black points). Here, the solid black horizontal lines indicate the true hierarchical mean, θ_ℓ.

generate counts, all scenarios used known values of the parameters $\beta_{g\ell}$, along with known γ_g^2's or negative binomial dispersions, depending on the data-generating mechanism. Thus, parameter estimation and gene detection could be assessed as in the previous simulation study. The following three approaches were used to generate gene-specific parameter values.

Model Datasets were generated exactly as in Simulation Study 1. This was the only scenario where the true hyperparameter values were known, so it was the only scenario where we applied the Oracle eBayes method.

edgeR This scenario used the benchmark method by McCarthy, Chen, and Smyth (2012) explained in Section 4.1. First, we applied the `edgeR` package to the Paschold et al. (2012) data to obtain normalization factors using the default TMM method of Robinson and Oshlack (2010), estimated negative binomial dispersions, and estimated $\beta_{g\ell}$ parameters. Then, using these quantities as truth, we simulated counts using the negative binomial model of McCarthy, Chen, and Smyth.

Simple The β_{g1} and β_{g5} parameter values were generated from normal distributions similar to their counterparts in the Model simulation. For $\ell = 2$, 3, and 4, the $\beta_{g\ell}$'s were drawn from discrete distributions to exaggerate the heterosis effect. We used $P(\beta_{g\ell} = 0) = 0.5$ and $P(\beta_{g\ell} = 1) = P(\beta_{g\ell} = -1) = 0.25$ for $\ell = 2$ and 3, $P(\beta_{g4} = 0) = 0.99$, and $P(\beta_{g4} = 1) = P(\beta_{g4} = -1) = 0.005$. All $\beta_{g\ell}$ parameters were generated independently across $g = 1, \ldots, G$ and $\ell = 1, \ldots L$. With the parameters in hand, count data were generated from a negative binomial model with a single common dispersion for all genes close to value of the dispersions obtained from the Paschold et al. dataset using `edgeR`. With respect to each of the six kinds of heterosis given in Section 2 and Table 1, roughly 6.5% of the simulated genes had some type of heterosis.

The fully Bayesian implementation was exactly the same as the previous simulation study, except that normalization factors are estimated as described in Section 3.2, with essentially the same results in terms of runtime, convergence diagnostics, and effective sample size for hyperparameters. The MCMC step of the eBayes procedure was also performed using the software in the R packages fbseq and fbseqCUDA using an option to skip sampling of hyperparameters. The runtime of this step was similar to the fully Bayesian analysis, for example, up to 2.7 hr for eBayes versus 3.2 hr for fully Bayes for $N = 16$ and up to 4.3 hr versus 4.8 hr for $N = 32$, since the vast majority of time is spent in sampling the gene-specific parameters.

Figure S5 shows the observed rates at which estimated 95% credible intervals cover parameters $\beta_{g\ell}$ for each method under comparison. Coverage was around the nominal 95% for the Model scenario, as well as for the Simple scenario, except for slightly higher-than-nominal coverage of the β_{g4}'s. In the edgeR scenario, coverage was uniformly poor, ranging roughly from 50% to 90%.

Figure 2 provides MSE for the β_{gl} parameters where each mean is taken over all the genes. Here, MSEs are almost the same between the eBayes and fully Bayesian methods, once again showing these methods to be equally matched. Overall, MSE is highest in the edgeR scenario and lowest in the Simple scenario (see the y-axis scales) indicating that parameter estimation is most challenging in the edgeR scenario. MSE for the fully

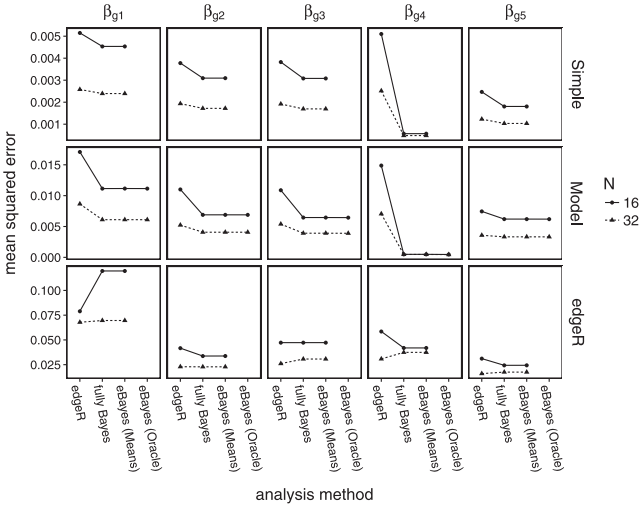

Figure 2. For Simulation Study 2 in Section 4.3, mean squared errors of the estimated model coefficients, where each mean is taken over all the genes. The row labels indicate simulation scenarios, and the lines in each panel correspond to individual simulated datasets.

Bayes and eBayes approaches are smaller or similar to the edgeR results except for β_{g1} in the edgeR simulations with 16 samples. As in Figure 1, the β_{g1}'s have higher variability than the other model coefficients so information borrowed across genes is least useful here. In contrast, MSE dramatically improved relative to edgeR for β_{g4} where true parameters are drawn from distributions tightly concentrated around zero.

Similar overall patterns carry over from parameter estimates to posterior probabilities of heterosis. Figure S11 shows the calibration errors as defined in Section 4, which varied only slightly between the eBayes and fully Bayesian approaches. Calibration error was similar across sample sizes, but increased from the Model scenario to the edgeR scenario and dramatically increased in the Simple scenario.

Figure S9 shows the calibration curves themselves for $N = 16$. (The results for $N = 32$, shown in Figure S10, are similar.) Compared to the Model scenario calibration was worse in the edgeR scenario, where many low probabilities were underestimated and high probabilities were overestimated for some types of high-parent heterosis. For the edgeR scenario, low-parent heterosis probabilities tended to be overestimated overall. Calibration was egregiously poor in the Simple scenario, where posterior probabilities were heavily overestimated.

Figures S6 and S7 provide ROC curves for $N = 16$ and $N = 32$ while Figure S8 provides areas under the ROC curves (AUCs) for all simulations. The AUCs were around 0.85 (0.90) for edgeR, 0.94 (0.96) for Model, and 0.99 (0.99) for Simple with $N = 16$ ($N = 32$). The ROC curves and AUCs were almost identical for the fully Bayes and eBayes methods. Thus, despite a lack of coverage and poor calibration of posterior probabilities, the methodology appears to have provided reasonable rankings of genes even when the model assumed in the analysis disagreed with the data-generating mechanism.

5. Fully Bayesian Analysis of the Paschold Et Al. Dataset

Having assessed our methodology's estimation, inference, and gene detection abilities in the simulation studies in Section 4, we now turn back to the original motivating heterosis dataset in Section 2, where P1 is B73, P2 is Mo17, H12 is B73×Mo17, and H21 is Mo17×B73. The model matrix X and the interpretations of the parameters $\beta_{g\ell}$ are the same as in Section 4.

The dataset contains count data for $G = 39,656$ genes on $N = 16$ biological replicates evenly spread over the four varieties. A large fraction of genes in the reference genome have low expression levels: roughly 21% have mean counts less than 1 and 39% have mean counts less than 10. Still, the mean count is around 255.5, the median is 37, the third quartile is 290, and the maximum is 38,010. Figure S12 shows a kernel density estimate of the log of the counts after incrementing by 1. As the figure suggests, the counts are multimodal, mainly split into low and high count groups.

We applied our fully Bayesian approach to the Paschold et al. dataset using the same number of chains, burn-in length, thinning, number of iterations, hardware, etc., as in Section 4.2. The GPU-accelerated version had a total elapsed runtime of 3.9 hr compared to an OpenMP version with 16 OpenMP threads that would have taken 5 days. Similar to previous convergence diagnostics, $\widehat{R}$ was less than 1.1 for all parameters except three $\beta_{g\ell}$ parameters and one γ_g^2 parameter, and all the hyperparameter effective sample sizes were above 500.

Figure 3 shows posterior distributions for all hyperparameters. The marginal posterior distributions were approximately normal and extremely narrow, so uncertainty in these parameters was small. The marginal posteriors were so concentrated that the prior distributions, which were diffuse and uninformative, would just appear as horizontal lines near zero in the figure.

Figure 4 provides posterior distributions and normal-based approximations to the posterior distributions, as described in Section 3.2, for a random subset of $\beta_{g\ell}$ parameters. For each parameter, the normal approximation closely matched the kernel density estimate, as did the equal-tail 95% CIs computed from each. This finding justifies the computational strategy recommended by Landau and Niemi (2016), which, for the sake of computational tractability, discarded most MCMC parameter samples and retained only the estimated posterior means and mean squares of these parameters.

To assess shrinkage, we compare our hierarchical model approach with the (relatively) nonhierarchical edgeR method by McCarthy, Chen, and Smyth (2012) explained in Section 4.1. Figure 5 compares posterior means of gene-specific parameters from the hierarchical model to the analogous estimates from the nonhierarchical model. Virtually no shrinkage is observed for the β_{g1} since the estimated hierarchical variance, σ_1^2 in Figure 3, is large. In contrast, the rest of the $\beta_{g\ell}$ estimates from the hierarchical model show shrinkage (toward the hierarchical means) relative to the analogous nonhierarchical model estimates with the most severe shrinkage occurring for the β_{g4}'s. The lower-right panel of the figure plots the log of the γ_g^2 parameters versus the log of the edgeR dispersions. The estimates are strongly associated with the identity line, supporting the notion that the γ_g^2's are equivalent to negative binomial dispersions. The odd shape is due to the edgeR approach of borrowing information about overdispersion for genes with similar expression levels and finding that genes with higher expression have lower overdispersion values.

Figure S13 shows the estimated posterior probabilities of each kind of heterosis. Most probabilities are below 0.5, and

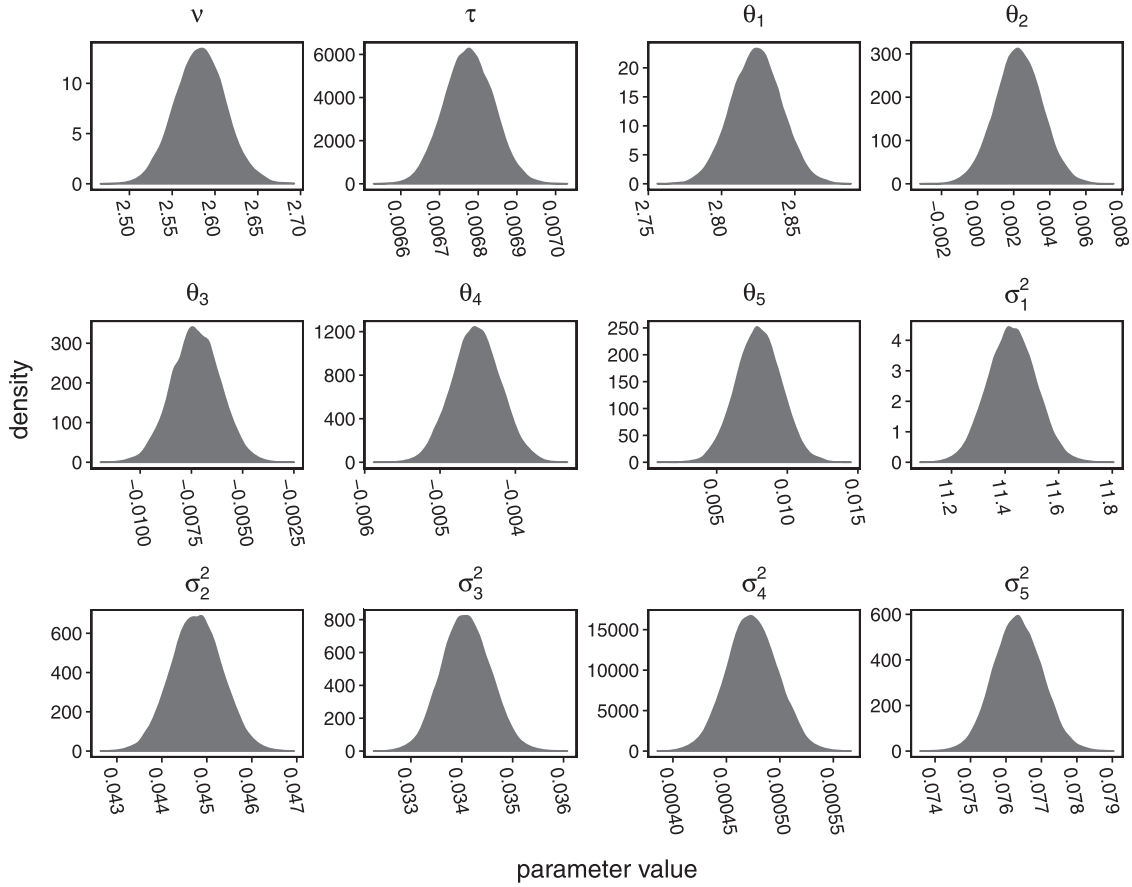

Figure 3. Kernel density estimates from MCMC samples of the hyperparameters from the fully Bayesian analysis of Paschold et al. data.

there is a spike at 0 for each kind of heterosis, so gene-specific heterosis appears uncommon overall. In addition, there is a spike around 0.25 in each histogram, which corresponds to unexpressed and barely expressed genes. The value 0.25 is the estimated predictive probability for a new gene $\tilde{g}$

$$P(\beta_{\tilde{g}2} > 0 \text{ and } \beta_{\tilde{g}3} > 0|y) \approx P(\beta_{\tilde{g}2}/\sigma_2 > 0|y)P(\beta_{\tilde{g}3}/\sigma_3 > 0|y) \approx 0.25$$

since θ_2 and θ_3 are close zero (see Figure 3), β_{g2} and β_{g3} are assumed independent, and the probability that a bivariate, independent, standard normal is in the positive quadrant is 0.25.

We also compare these probabilities to estimated effect size, which we take to be a relative measure of the strength of heterosis in terms of estimated posterior means. For example, consider the heterosis of a gene g with respect to the B73×Mo17 hybrid. From Table 1, heterosis occurs if $2\beta_{g2} + \beta_{g4} > 0$ and $2\beta_{g3} + \beta_{g4} > 0$. For this type of heterosis, we define the effect size of gene g to be the positive part of $\min(2\beta_{g2} + \beta_{g4}, 2\beta_{g3} + \beta_{g4})/\sqrt{\gamma_g^2}$. We define effect size for the other types of heterosis similarly, using the analogous linear combinations of the $\beta_{g\ell}$'s from Table 1. In Figure 6, we plot estimated posterior probabilities against their analogous estimated effect sizes. Each panel shows a so-called "volcano" plot similar to fig. 4 of Niemi et al. (2015). The highest concentrations of genes correspond to low effect sizes and low posterior probabilities with a distinct ridge at low probabilities with an effect size of zero (due to the definition of effect size). Posterior probability and effect size are positively associated and genes of interest for future investigation are genes with high posterior probability and large effect size. Table S1 provides these posterior probabilities and effect sizes enabling scientists to search in maize genome databases for relationships among these genes.

6. Discussion

We presented a fully Bayesian strategy for modeling high-dimensional count data applicable to a variety of experimental designs. The fully Bayesian approach is rare in fields such as RNA-seq data analysis due to the computational challenges, but solidly tractable with new massively parallel computing strategies. We applied our approach to the heterosis problem in RNA-seq data analysis, where no existing methods for RNA-seq analysis are directly applicable. We used simulation studies to assess the fully Bayesian approach and compared it to two best-case-scenario eBayes counterparts. From our simulations, we found that our fully Bayesian method strongly shrunk estimates of important gene-specific parameters toward common means, which generally improved estimation relative to a nonhierarchical model. The fully Bayesian and eBayes methods performed equally well under all metrics considered. This finding supports the use of eBayes methods in RNA-seq analyses when good hyperparameter estimates are available. However, we did not investigate methods of obtaining these hyperparameter estimates. One option is to use moment-matching

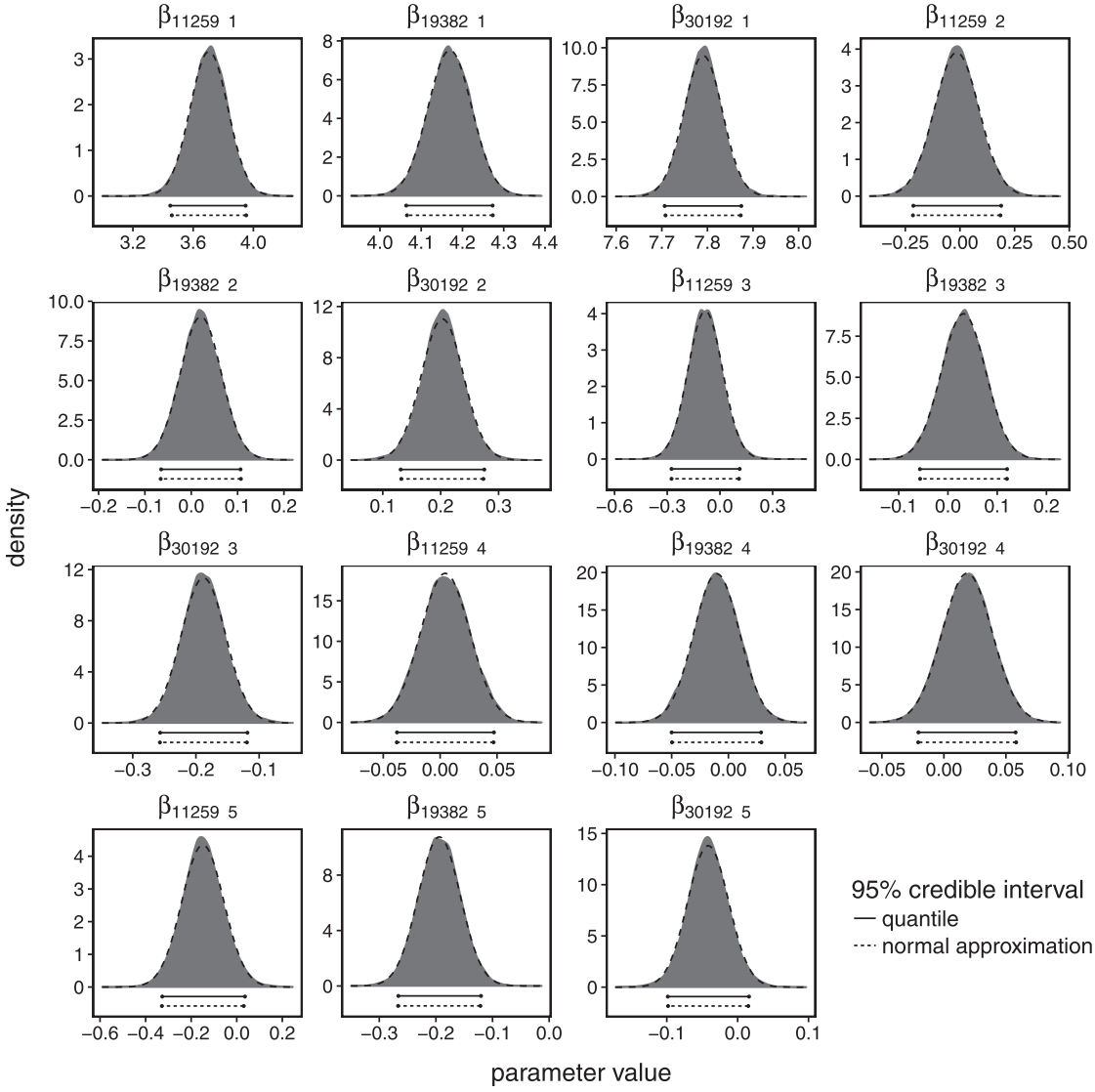

Figure 4. Kernel density estimates (shaded area) and approximate normal densities (dashed lines) of the marginal posterior distributions and 95% equal-tail credible intervals based on MCMC samples (solid) and normal approximation (dashed) of a random subset of $\beta_{g\ell}$ parameters based on the fully Bayesian analysis of the citeauthor-paschold data.

techniques based on independent analyses across genes (Niemi et al. 2015). Another option is the expectation-maximization algorithm which requires nonanalytic integration over the gene-specific parameters in the expectation step. Finally, we analyzed the motivating RNA-seq dataset by Paschold et al. (2012) to evaluate the evidence for each of six types of heterosis and therefore provide guidance on genes that may be involved in the molecular mechanism for heterosis.

Our methodology has important application-level utility for practitioners in multiple-testing scenarios such as genomics. Procedures for controlling the false discovery rate (FDR) usually require the null p-values to have a uniform distribution (Benjamini and Hochberg 1995; Storey 2003; Meinhausen and Rice 2006; Dudoit and Laan 2008), a condition that is typically violated when composite null hypotheses are tested (Bayarri and Berger 2000; Robins, Vaart, and Ventura 2000; Sun and McLain 2012; Dickhaus 2013). We avoid the need for such a tenuous assumption by dispensing with an explicit FDR control procedure altogether, opting instead for a fully Bayesian approach with a hierarchical model that shares information across genes (Muller, Parmigiani, and Rice 2007). FDR control aside, this borrowing of information is associated with improvements in parameter estimation and gene detection (Landau and Liu 2013; Ji, Liu, and Nettleton 2014; Niemi et al. 2015).

Our results suggest some possible improvements to our model for future work particularly with regard to the $\beta_{g\ell} \overset{\text{ind}}{\sim} N(\theta\ell, \sigma_\ell^2)$ assumption. From Figure 1, the gene-specific parameters $\beta_{g\ell}$ were poorly estimated if their true values are extreme for a given index element ℓ. Specifically, low $\beta_{g\ell}$'s tended to be overestimated and high $\beta_{g\ell}$'s tended to be underestimated, that is, the estimates of extreme $\beta_{g\ell}$ parameters were overly shrunk toward their hierarchical means. If we assumed hierarchical distributions with heavier tails, for example, Laplace, Student t, or horseshoe (Carvalho, Polson, and Scott 2009; Thorne 2017), shrinkage should be reduced for extreme $\beta_{g\ell}$'s, and overall estimation, inference, and gene detection could improve. Alternatively, a semiparametric approach, for example, a Dirichlet process mixture (Muller and Mitra 2013;

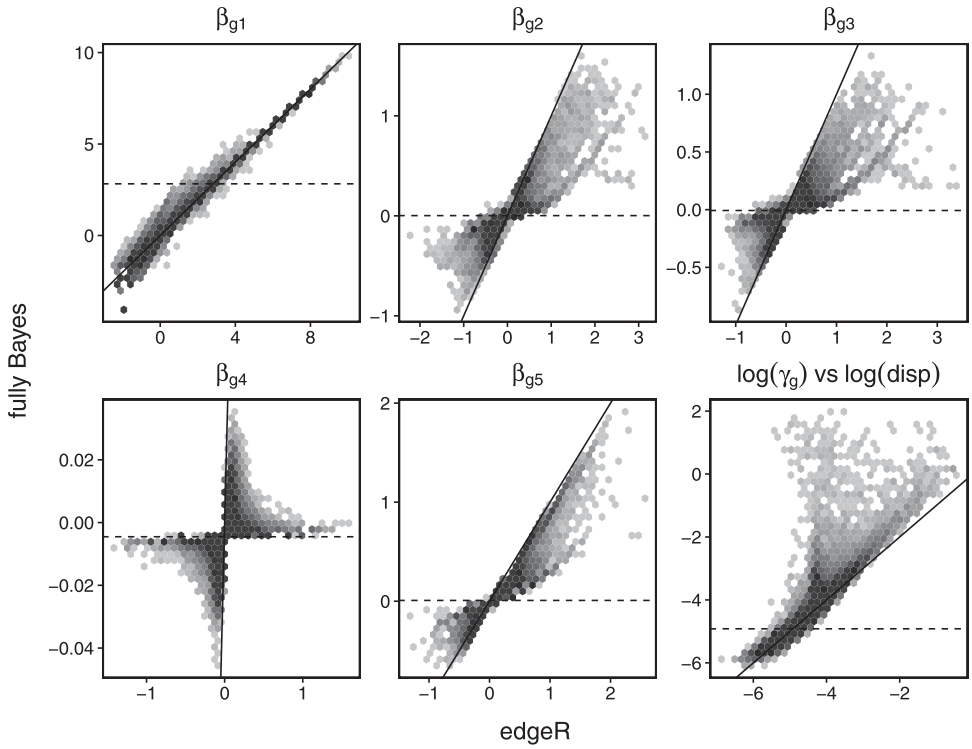

Figure 5. Two-dimensional hexagonal histograms with logarithmic shading of posterior means from the fully Bayesian analysis versus estimates from the edgeR analysis for gene-specific parameters of the Paschold et al. data with the identity line (solid) and hierarchical mean (dashed) under the fully Bayesian analysis. In the lower-right panel, the horizontal axis corresponds to the log-scale gene-specific negative binomial dispersions from edgeR.

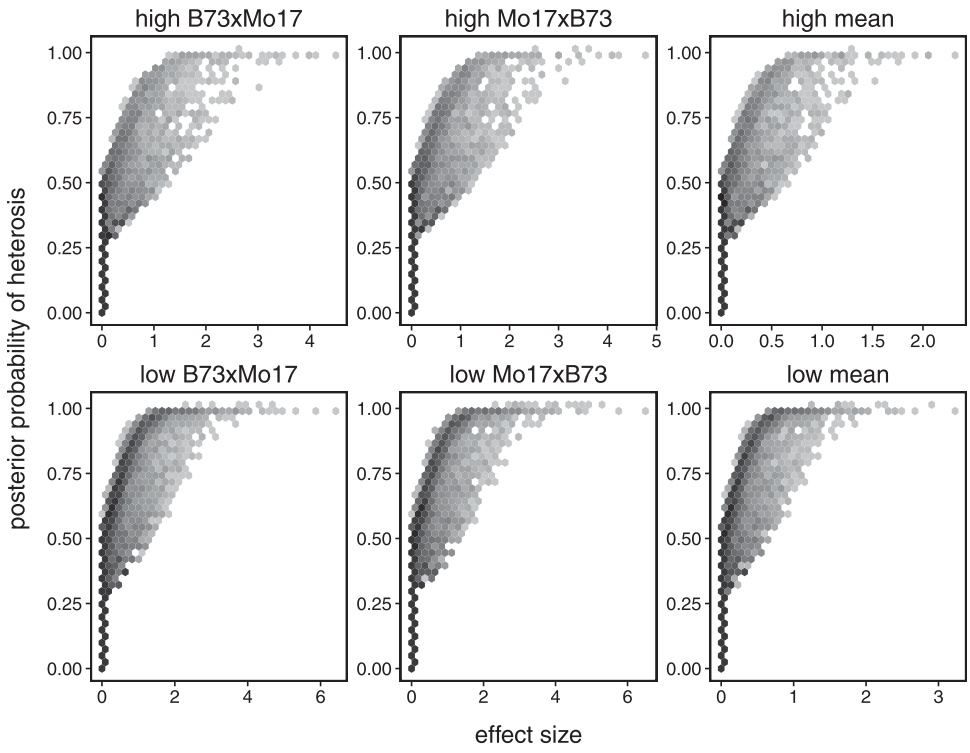

Figure 6. Two-dimensional hexagonal histogram of gene-specific posterior probabilities of heterosis, shaded on a logarithmic scale, against the analogous effect sizes from the fully Bayesian analysis of the Paschold et al. data. Results are shown for high (top row) and low (bottom) heterosis for the B73×Mo17 hybrid (left column) and Mo17×B73 hybrid (middle), and their mean (right).

Liu, Wang, and Liu 2015), could be employed to estimate the distribution of the gene-specific parameters.

Despite the computational gains due to the use of GPUs, computational time is a hinderance for wide adoption of these methods. Luts and Wand (2015) introduced a variational Bayes approximation of posteriors for these types of models. They compared their approach to an MCMC approach based on analyses with 500 genes. As expected, the variational Bayes approach was computationally more efficient but also resulted in a 10% to 25% loss of accuracy based on their metric. It is unclear how these results would translate to a comparison using the 40,000 genes we analyzed here. Because our method allows for a fully Bayesian analysis of a large number of genes, it can be used as a reference for evaluations of variational Bayes approximations or other computationally efficient approaches on datasets of realistic size.

Appendix: Full Conditional Distributions

The model and priors described in Section 3.1 are succinctly represented below.

$$
\begin{aligned}
y_{gn} &\overset{\text{ind}}{\sim} \text{Poisson}\left(\exp\left(h_n + \varepsilon_{gn} + X_n\beta_g\right)\right)\\
\varepsilon_{gn} &\overset{\text{ind}}{\sim} \text{Normal}(0, \gamma_g^2)\\
\frac{1}{\gamma_g^2} &\overset{\text{ind}}{\sim} \text{Gamma}\left(\frac{\nu}{2}, \frac{\nu\tau}{2}\right)\\
\nu &\sim \text{Uniform}(0, d)\\
\tau &\sim \text{Gamma}(a, b)\\
\beta_{g\ell} &\overset{\text{ind}}{\sim} \text{Normal}(\theta_\ell, \sigma_\ell^2)\\
\theta_\ell &\sim \text{Normal}(0, c_\ell^2)\\
\sigma_\ell &\sim \text{Uniform}(0, s_\ell)
\end{aligned}
$$

with $\nu, \tau, \theta_1, \ldots, \theta_L$, and $\sigma_1, \ldots, \sigma_L$ all independent of each other.

Our Gibbs sampler proceeds through the following full-conditional draws for ϵ_{gn}, γ_g^2, ν, τ, $\beta_{g\ell}$, θ_ℓ, and σ_ℓ^2. For convenience, let $p(\psi|\ldots)$ be the full conditional distribution of parameter ψ given all the other parameters and the data. The full conditionals are summarized below.

$$
\begin{aligned}
\theta_\ell|\ldots &\sim \text{Normal}\left(\frac{B}{2A}, \frac{1}{2A}\right)\\
&\quad\times\left(A = \frac{1}{2}\left(\frac{1}{c_\ell^2} + \frac{G}{\sigma_\ell^2}\right),\ B = \frac{1}{\sigma_\ell^2}\sum_{g=1}^{G}\beta_{g\ell}\right)\\
\tau\ |\cdots &\sim \text{Gamma}\left(\text{shape} = a + \frac{G\nu}{2},\ \text{rate} = b + \frac{\nu}{2}\sum_{g=1}^{G}\frac{1}{\gamma_g}\right)\\
\gamma_g|\ldots &\sim \text{Inverse-Gamma}\\
&\quad\times\left(\text{shape} = \frac{N+\nu}{2},\ \text{scale} = \frac{1}{2}\left(\nu\tau + \sum_{n=1}^{N}\varepsilon_{gn}^2\right)\right)\\
\sigma_\ell^2|\ldots &\sim \text{Inverse-Gamma}\\
&\quad\times\left(\text{shape} = \frac{G-1}{2},\ \text{scale} = \frac{1}{2}\sum_{g=1}^{G}(\beta_{g\ell}-\theta_\ell)^2\right)\\
&\quad\times I(\sigma_\ell^2 < s_\ell^2)\\
p(\nu|\ldots) &\propto \exp\left(-G\log\Gamma\left(\frac{\nu}{2}\right) + \frac{G\nu}{2}\log\left(\frac{\nu\tau}{2}\right) - \frac{\nu}{2}\sum_{g=1}^{G}\left[\log\gamma_g^2 + \frac{\tau}{\gamma_g^2}\right]\right)\\
&\quad\times I(0 < \nu < d)
\end{aligned}
$$

$$
\begin{aligned}
p(\varepsilon_{gn}\mid\cdots) &\propto \exp\left(\varepsilon_{gn}y_{gn} - \frac{\varepsilon_{gn}^2}{2\gamma_g^2} - \exp\left(\varepsilon_{gn}\right)\exp\left(h_n + X_n\beta_g\right)\right)\\
p(\beta_{g\ell}\mid\cdots) &\propto \exp\left(\beta_{g\ell}\sum_{n=1}^{N}y_{gn}X_{n\ell} - \frac{(\beta_{g\ell}-\theta_\ell)^2}{2\sigma_\ell^2} - \sum_{x\in S_\ell}\exp\left(x\beta_{g\ell}\right)\right.\\
&\quad\left.\times\sum_{n=1}^{N}I(X_{n\ell}=x)\exp\left[h_n + \varepsilon_{gn} + \sum_{i\neq\ell}X_{ni}\beta_{gi}\right]\right),
\end{aligned}
$$

where S_ℓ is the set of unique nonzero elements of $\{X_{1\ell}, \ldots, X_{N\ell}\}$.

Due to the limited availability of random number generation on GPUs, only the θ_ℓ were sampled directly from their full conditional distributions. The remaining parameters were sampled using a univariate stepping-out slice sampler as given in Neal (2003).

Supplementary Materials

The supplementary figures (Figure S1, Figures S2, etc.) are in `supplement.pdf`. The file `TabelS1.csv` contain the Paschold et al. (2012) data, as well as fully Bayesian posterior estimates of the gene-specific heterosis probabilities, gene-specific parameter means and standard deviations, estimated effect sizes, and gene-specific parameter estimates from the `edgeR` method by McCarthy et al. (2012) from Section 4.1. The `packages` directory contains four R packages including `fbseq` which is the user-interface for the computation, the back-ends `fbseqCUDA` and `fbseqOpenMP` which are suitable for use on computers with and without a CUDA-capable GPU (respectively), and `fbseqStudies` which reproduces the analyses in this article.

Funding

This research was supported by National Institute of General Medical Sciences (NIGMS) of the National Institutes of Health and the joint National Science Foundation/NIGMS Mathematical Biology Program under award number R01GM109458. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or the National Science Foundation.

References

Anders, S., and Huber, W. (2010), "Differential Expression Analysis for Sequence Count Data," *Genome Biology*, 11, R106. [612]

Bayarri, M. J., and Berger, J. O. (2000), "P Values for Composite Null Models," *Journal of the American Statistical Association*, 95, 1127–1142. [610,618]

Benjamini, Y., and Hochberg, Y. (1995), "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing," *Journal of the Royal Statistical Society*, Series B, 57, 289–300. [610,618]

Cabras, S. (2010), "A Note on Multiple Testing for Composite Null Hypotheses," *Journal of Statistical Planning and Inference*, 140, 659–666. [610]

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P. and Riddell, A. (2017), "Stan: A Probabilistic Programming Language," *Journal of Statistical Software*, 76, 1–32. [611]

Carvalho, C., Polson, N., and Scott, J. (2009), "Handling Sparsity via the Horseshoe," in *Proceedings of the 12th International Conference on Artificial Intelligence and Statistics* (Vol. 5), pp. 73–80. [618]

Chi, Z. (2010), "Multiple Hypothesis Testing on Composite Nulls using Constrained p-Values," *Electronic Journal of Statistics*, 4, 271–299. [610]

Coors, J., and Pandey, S. (1999), *The Genetics and Exploitation of Heterosis in Crops*, Madison, WI: American Society of Agronomy, Crop Science Society of America. [610]

Dagum, L., and Menon, R. (1998), "OpenMP: An Industry Standard API for Shared-Memory Programming," *Computational Science & Engineering, IEEE*, 5, 46–55. [613]

Darwin, C. (1876), *The Effects of Cross and Self Fertilisation in the Vegetable Kingdom*, London: John Murray. [610]

de Valpine, P., Turek, D., Paciorek, C. J., Anderson-Bergman, C., Temple Lang, D., and Bodik, R. (2017), "Programming with Models: Writing

Statistical Algorithms for General Model Structures with NIMBLE," *Journal of Computational and Graphical Statistics*, 26, 403–413. [611]

Dickhaus, T. (2013), "Randomized p-values for Multiple Testing of Composite Null Hypotheses," *Journal of Statistical Planning and Inference*, 143, 1968–1979. [610,618]

Dudoit, S., and Laan, M. J. v. d. (2008), *Multiple Testing Procedures with Applications to Genomics*, New York: Springer. [610,618]

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2013), *Bayesian Data Analysis* (3rd ed.), Boca Raton, FL: CRC Press. [613]

Gelman, A., and Rubin, D. B. (1992), "Inference from Iterative Simulation Using Multiple Sequences," *Statistical Science*, 7, 457–472. [613]

Hallauer, A., and Miranda, F. (1981), *Quantitative Genetics in Maize Breeding*, Ames, IA: Iowa State University Press. [610]

Hallauer, A. R., Carena, M. J., and Miranda Filho, J. (2010), *Quantitative Genetics in Maize Breeding* (Vol. 6), New York: Springer. [610]

Hardcastle, T. J., and Kelly, K. A. (2010), "baySeq: Empirical Bayesian Methods for Identifying Differential Expression in Sequence Count Data," *BMC Bioinformatics*, 11, 422. [611,614]

Hoecker, N., Keller, B., Piepho, H., and Hochholdinger, F. (2006), "Manifestation of Heterosis during Early Maize (Zea mays L.) Root Development," *Theoretical and Applied Genetics*, 112, 421–429. [611]

Ji, T., Liu, P., and Nettleton, D. (2014), "Estimation and Testing of Gene Expression Heterosis," *Journal of Agricultural, Biological, and Environmental Statistics*, 19, 319–337. [610,613,614,618]

Krieger, U., Lippman, Z., and Zamir, D. (2010), "The Flowering Gene Single Flower Truss Drives Heterosis for Yield in Tomato," *Nature Genetics*, 42, 459–463. [610]

Landau, W., and Niemi, J. (2016), "A Fully Bayesian Strategy for High-Dimensional Hierarchical Modeling using Massively Parallel Computing," arXiv:1606.06659. [611,612,613,616]

Landau, W. M., and Liu, P. (2013), "Dispersion Estimation and Its Effect on Test Performance in RNA-seq Data Analysis: A Simulation-Based Comparison of Methods," *PLOS ONE*, 8, e81415. [610,612,613,618]

Leng, N., Dawson, J. A., Thomson, J. A., Ruotti, V., Rissman, A. I., Smits, B. M., Haag, J. D., Gould, M. N., Stewart, R. M., and Kendziorski, C. (2013), "EBSeq: An Empirical Bayes Hierarchical Model for Inference in RNA-seq Experiments," *Bioinformatics*, 29, 1035–1043. [611]

Lippman, Z., and Zamir, D. (2007), "Heterosis: Revisiting the Magic," *Trends in Genetics*, 23, 60–66. [610]

Lithio, A., and Nettleton, D. (2015), "Hierarchical Modeling and Differential Expression Analysis for RNA-seq Experiments with Inbred and Hybrid Genotypes," *Journal of Agricultural, Biological, and Environmental Statistics*, 20, 598–613. [611,613]

Liu, F., Wang, C., and Liu, P. (2015), "A Semi-parametric Bayesian Approach for Differential Expression Analysis of RNA-seq Data," *Journal of Agricultural, Biological, and Environmental Statistics*, 20, 555–557. [620]

Lunn, D., Spiegelhalter, D., Thomas, A., and Best, N. (2009), "The BUGS Project: Evolution, Critique and Future Directions," *Statistics in Medicine*, 28, 3049–3067. [611]

Lunn, D. J., Thomas, A., Best, N., and Spiegelhalter, D. (2000), "WinBUGS-A Bayesian Modelling Framework: Concepts, Structure, and Extensibility," *Statistics and Computing*, 10, 325–337. [611]

Luts, J., and Wand, M. P. (2015), "Variational Inference for Count Response Semiparametric Regression," *Bayesian Analysis*, 10, 991–1023. [620]

McCarthy, D., Chen, Y., and Smyth, G. (2012), "Differential Expression Analysis of Multifactor RNA-seq Experiments with Respect to Biological Variation," *Nucleic Acids Research*, 40, 4288–4297. [612,613,616]

Meinhausen, N., and Rice, J. (2006), "Estimating the Proportion of false null Hypotheses among a Large Number of Independently Tested Hypotheses," *Annals of Statistics*, 34, 373–393. [610,618]

Muller, P., and Mitra, R. (2013), "Bayesian Nonparametric Inference - Why and How," *Bayesian Analysis*, 8, 269–302. [618]

Muller, P., Parmigiani, G., and Rice, K. (2007), *Bayesian Statistics* (8th ed.), Oxford, UK: Oxford University Press. [618]

Neal, R. M. (2003), "Slice Sampling," *The Annals of Statistics*, 31, 705–767. [620]

Niemi, J., Mittman, E., Landau, W., and Nettleton, D. (2015), "Empirical Bayes Analysis of RNA-seq Data for Detection of Gene Expression Heterosis," *Journal of Agricultural, Biological, and Environmental Statistics*, 20, 614–628. [610,612,613,614,617,618]

Oshlack, A., Robinson, M. D., and Young, M. D. (2010), "From RNA-seq Reads to Differential Expression Results," *Genome Biology*, 11, 220. [610]

Paschold, A., Jia, Y., Marcon, C., Lund, S., Larson, N. B., Yeh, C.-T., Ossowski, S., Lanz, C., Nettleton, D., Schnable, P. S., and Hochholdinger, F. (2012), "Complementation Contributes to Transcriptome Complexity in Maize (*Zea mays L.*) Hybrids Relative to their Inbred Parents," *Genome Research*, 22, 2445–2454 [611,618]

Plummer, M., et al. (2003), "JAGS: A Program for Analysis of Bayesian Graphical Models using Gibbs Sampling," in *Proceedings of the 3rd International Workshop on Distributed Statistical Computing* (Vol. 124), Technische Universit at Wien, p. 125. [611]

R Core Team (2016), *R: A Language and Environment for Statistical Computing*, Vienna, Austria: R Foundation for Statistical Computing. [613]

Riday, H., and Brummer, E. (2002), "Heterosis of Agronomic Traits in alfalfa," *Crop Science*, 42, 1081–1087. [610]

Robins, J. M., Vaart, A. v. d., and Ventura, V. (2000), "Asymptotic Distribution of P Values in Composite Null Models," *Journal of the American Statistical Association*, 95, 1143–1156. [610,618]

Robinson, M., and Oshlack, A. (2010), "A Scaling Normalization Method for Differential Expression Analysis of RNA-seq Data," *Genome Biology*, 11, R25. [612,613]

Romano, J., and Shaikh, A. (2006), "Stepup Procedures for Control of Generalizations of the Family Wise Error Rate," *Annals of Statistics*, 34, 1850–1873. [610]

Schnable, P., Ware, D., Fulton, R., Stein, J., Wei, F., Pasternak, S., Liang, C., Zhang, J., Fulton, L., and Graves, T.A., and Minx, P. (2009), "The B73 Maize Genome: Complexity, Diversity, and Dynamics," *Science*, 326, 1112–1115. [611]

Si, Y., and Liu, P. (2013), "An Optimal Test with Maximum Average Power While Controlling FDR with Application to RNA-seq Data," *Biometrics*, 69, 594–605. [612]

Springer, N., and Stupar, R. (2007), "Allelic Variation and Heterosis in Maize: How do Two Halves Make more than a Whole?" *Genome Research*, 17, 264–275. [610]

Storey, J. D. (2003), "The Positive False Discovery Rate: A Bayesian Interpretation and the q-Value," *The Annals of Statistics*, 31, 2013–2035. [610,618]

Sun,W.,andMcLain,A.C.(2012), "Multiple Testing of Composite Null Hypotheses in Heteroscedastic Models," *Journal of the American Statistical Association*, 107, 673–687. [610,618]

Swanson-Wagner, R., Jia, Y., DeCook, R., Borsuk, L., Nettleton, D., and Schnable, P. (2006), "All Possible Modes of gene Action are Observed in a Global Comparison of Gene Expression in a maize F1 Hybrid and its Inbred Parents," *Proceedings of the National Academy of Sciences*, 103, 6805–6810. [610]

Thorne, T. (2017), "Approximate Inference of Gene Regulatory Network Models from RNA-Seq Time Series Data," *bioRxiv*, 149674. [618]

Wang, J., Tian, L., Lee, H., Wei, N., Jiang, H., Watson, B., Madlung, A., Osborn, T. C., Doerge, R. W., Comai, L., and Chen, Z. J. (2006), "Genomewide Nonadditive Gene Regulation in Arabidopsis Allotetraploids," *Genetics*, 172, 507–517. [610]

Wang, L., Li, P., and Brutnell, T. P. (2010), "Exploring Plant Transcriptomes using ultra High-Throughput Sequencing," *Briefings in Functional Genomics*, 9, 118–128. [610]

Winz, R., and Baldwin, I. (2001), "Molecular Interactions between the Specialist Herbivore Manduca sexta (Lepidoptera, Sphingidae) and its Natural Host Nicotiana Attenuata. IV. Insect-Induced Ethylene Reduces jasmonate-induced Nicotine Accumulation by Regulating Putrescine N-methyltransferase Transcripts," *Plant Physiology*, 125, 2189–2202. [611]

Wohlfarth, G. (1993), "Heterosis for Growth Rate in Common Carp," *Aquaculture*, 113, 31–46. [610]

Wu, H., Wang, C., and Wu, Z. (2012), "A New Shrinkage Estimator for Dispersion Improves Differential Expression Detection in RNA-seq Data," *Biostatistics*, 1, 1–24. [611,612,614]

Yu, S., Li, J., Xu, C., Tan, Y., Gao, Y., Li, X., Zhang, Q., and Maroof, M. (1997), "Importance of Epistasis as the Genetic Basis of Heterosis in an Elite Rice Hybrid," *Proceedings of the National Academy of Sciences*, 94, 9226–9231. [610]