1. Inference and Hypothesis Testing

Metropolis-within-Gibbs sampling is evolved for inference. Algorithm 1 shows the inference process and the detail are also illustrated.

Algorithm 1 Inference process for Metropolis-within-Gibbs

```
Initialize θ, μ, M_j, μ_0, M_0
repeat
    for each location j do
        Samples from p(μ_j|θ_ij, μ_0, M_0)                ▷ Sample μ_j
        Set μ_j to the sample median for the samples
        Samples from p(M_j|μ, σ, θ_ji, μ_j)               ▷ Sample M_j
        for each replicate i do
            Sample from p(θ_ij|r_ij, n_ij, μ_j, M)        ▷ Sample θ_ji
        end for
    end for
until sample size sufficient
```

1.1. **Initialization.** [This paragrah COPIED]The initial values for the model parameters and latent variables is obtained by a method-of-moments (MoM) procedure. MoM works by setting the population moment equal to the sample moment. A system of equations is formed such that the number of moment equations is equal to the number of unknown parameters and the equations are solved simultaneously to give the parameter estimates. We simply start with the data matrices r and n and work up the hierarchy of the graphical model solving for the parameters of each conditional distribution in turn.

The initial parameter estimates and derivations are provided in Appendix **??**. Below is the MoM estimate for replicate-level parameters $\tilde{\theta}_{ji} = \frac{r_{ji}}{n_{ji}}$. The estimates for the position-level parameters are $\tilde{\mu}_j = \frac{1}{N}\sum_{i=1}^{N}\theta_{ji}$ and $\tilde{M}_j = \frac{\tilde{\mu}_j(1-\tilde{\mu}_j)}{\frac{1}{N}\sum_{i=1}^{N}\theta_{ji}^2} - 1$. The estimates for the genome-level parameters are $\tilde{\mu}_0 = \frac{1}{J}\sum_{j=1}^{J}\mu_j$ and $\tilde{M}_0 = \frac{\tilde{\mu}_0(1-\tilde{\mu}_0)}{\frac{1}{J}\sum_{j=1}^{J}\mu_j^2} - 1$.

1.2. **Sampling from** $p(\theta_{ij}|r_{ij}, n_{ij}, \mu_j, M)$**.** Because the Bayesian conjugacy between the prior $p(\theta_{ji}|\mu_j, M_j) \sim \text{Beta}(\mu_j, M_j)$ and the likelihood $p(r_{ji}|n_{ji}, \theta_{ji}) \sim \text{Binomial}(\theta_{ji}, n_{ji})$, we draw the samples from the posterior distribution $p(\theta_{ji}|r_{ji}, n_{ji}, \mu_j, M_j)$ The posterior distribution is

$$p(\theta_{ji}|r_{ji}, n_{ji}, \mu_j, M_j) \sim \text{Beta}\left(\frac{r_{ji} + M_j\mu_j}{n_{ji} + M_j}, n_{ji} + M_j\right). \tag{1}$$

1.3. **Sampling from** $p\left(\mu_j|\theta_{ji}, M_j, \mu_0, M_0\right)$**.** Based on Markov blanket, the posterior distribution over μ_j is

$$p(\mu_j|\theta_{ji}, M_j, \mu_0, M_0) \propto p(\mu_j|\mu_0, M_0)p(\theta_{ji}|\mu_j, M_j). \tag{2}$$

[This paragrah COPIED]Since the prior, $p(\mu_j|\mu_0, M_0)$, is not conjugate to the likelihood, $p(\theta_{ji}|\mu_j, M_j)$, we sample from the posterior distribution using the Metropolis-Hastings algorithm. By experience when $\mu_j^{(p)} \in (10^{-3}, 1-10^{-3})$, the proposal distribution variance for all the Metropolis-Hastings steps within a Gibbs iteration is set to $\sigma_j = 0.1 \cdot \mu_j^{(p)}$; otherwise, we set $\sigma_j = 10^{-4}$ if $\mu_j^{(p)} < 10^{-3}$ and $\sigma_j = 10^{-1} - 10^{-4}$ if $\mu_j^{(p)} > 1 - 10^{-3}$. We have found that the algorithm performance improves when we take the median of five or more M-H samples as a single Gibbs step for each position.

We resampled from the proposal if the sample is outside of the support of the posterior distribution. We throw away a burn-in period - 20% of the samples, and thin the chain by a factor 2 to reduce autocorrelation among samples, resulting in a sample with size 1600 from the posterior distribution.

1.4. **Sampling from** $p\left(M_j|\mu, \sigma, \theta_{ji}, \mu_j\right)$**.** Since Jeffreys prior is from the Fisher information and in RVD3 model $\theta_{ji} \sim Beta\left(\mu_j, M_j\right)$,

$$I\left(M_j\right) = E_{M_j}\left[-\frac{\delta^2 \log p\left(\theta_j|\mu_j, M_j\right)}{\delta M_j^2}\right] \tag{3}$$

We calculated the equations Appendix **??** and obtained the Jeffreys' prior for M_j:

$$[-\left(\Psi_1(M_j) - \Psi_1(\mu_j M_j)\mu_j^2 - \Psi_1((1-\mu_j)M_j)(1-\mu_j)^2\right)]^{\frac{1}{2}} \tag{4}$$

For log-normal prior, the posterior distribution over M_j given its Markov blanket is

$$p(M_j|\mu, \sigma, \theta_{ji}, \mu_j) \propto p(\theta_{ji}|\mu_j, M_j)p(M_j|\mu, \sigma) \tag{5}$$

We have $\theta_{ji} \sim$ Beta(μ_j, M_j), and $M_j \sim$ log-normal(μ, σ). Instead of computing the posterior distribution directly, Metropolis-Hastings algorithm was taken to sample from the posterior distribution.

1.5. **Posterior Density Test.** Posterior distributions of μ_j for the control and case are achieved - $\tilde{\mu}_j^{\text{case}}$ and $\tilde{\mu}_j^{\text{control}}$, by Metropolis-within-Gibbs. So we called a variant when $\tilde{\mu}_j^{\text{case}} > \tilde{\mu}_j^{\text{control}}$ with $1-\alpha$ confidence,

$$\Pr(\tilde{\mu}_j^{\text{case}} - \tilde{\mu}_j^{\text{control}} \geq \tau) > 1 - \alpha, \tag{6}$$

where τ is a detection threshold and $1-\alpha$ is the confidence level. We set $\tau = 0$ in our experiment [XXX].

1.6. χ^2 **test for non-uniform base distribution.** [This part is COPIED]

An abundance of non-reference bases at a position called by the posterior density test may be due to a true mutation or due to a random sequencing error; we would like to differentiate these two scenarios. We assume non-reference read counts caused by a non-biological mechanism results in a uniform distribution over three non-reference bases. In contrast, the distribution of counts among three non-reference bases caused by biological mutation would not be uniform.

We use a χ^2 goodness-of-fit test on a multinomial distribution over the non-reference bases to distinguish these two possible scenarios. The null hypothesis is $H_0 : p = (p_1, p_2, p_3)$ where $p_1 = p_2 = p_3 = 1/3$. Cressie and Read (1984) identified a power-divergence family of statistics, indexed by λ, that includes as special cases Pearson's $\chi^2(\lambda = 1)$ statistic, the log likelihood ratio statistic ($\lambda = 0$), the Freeman-Tukey statistic ($\lambda = -1/2$), and the Neyman modified statistic $X^2(\lambda = -2)$. The test statistic is

$$2nI^{\lambda} = \frac{2}{\lambda(\lambda+1)} \sum_{k=1}^{3} r_{ji}^{(k)} \left[\left(\frac{r_{ji}^{(k)}}{E_{ji}^{(k)}} \right)^{\lambda} - 1 \right]; \lambda \in R, \tag{7}$$

where $r_{ji}^{(k)}$ is the observed frequency for non-reference base k at position j in replicate i and $E_{ji}^{(k)}$ is the corresponding expected frequency under the null hypothesis. **?** recommended $\lambda = 2/3$ when no knowledge of the alternative distribution is available and we choose that value.

We control for multiple hypothesis testing in two ways. We use Fisher's combined probability test (**?**) to combine the p-values for N replicates into a single p-value at position j,

$$X_j^2 = -2 \sum_{i=1}^{N} \ln(p_{ji}). \tag{8}$$

Equation (8) gives a test statistic that follows a χ^2 distribution with $2N$ degrees of freedom when the null hypothesis is true. Finally, we use the Bejamini-Hochberg method to control the family-wise error rate (FWER) over positions that have been called by the Bayesian hypothesis test (6) (**??**).