# Abstract

# Introduction

# Materials and Methods

## Material

Synthetic DNA dataset

H1N1 Influenza samples

## Model structure

Generative process and model structure

The generative process for RVD2 model is:

Given a data matrix , the model only addresses error/reference positions and does not model individual nucleotide frequencies.

The hierarchical model is

The error read count at position in replicate is modeled by the binomial random variable . is the total counts at position in replicate . The probability of an error at position in replicate is . The error probability has a prior beta distribution with position-specific rate parameter and precision .

The position error rate , has a distribution as prior distribution with parameters and . This is to ensure that the error rate is between 0 and 1. The prior is useful for situations when there is a significant minor allele.

The corresponding probability distribution functions are







Based on this model structure, the complete data likelihood of the hierarchical model is



The complete data log-likelihood is



The log-likelihood is



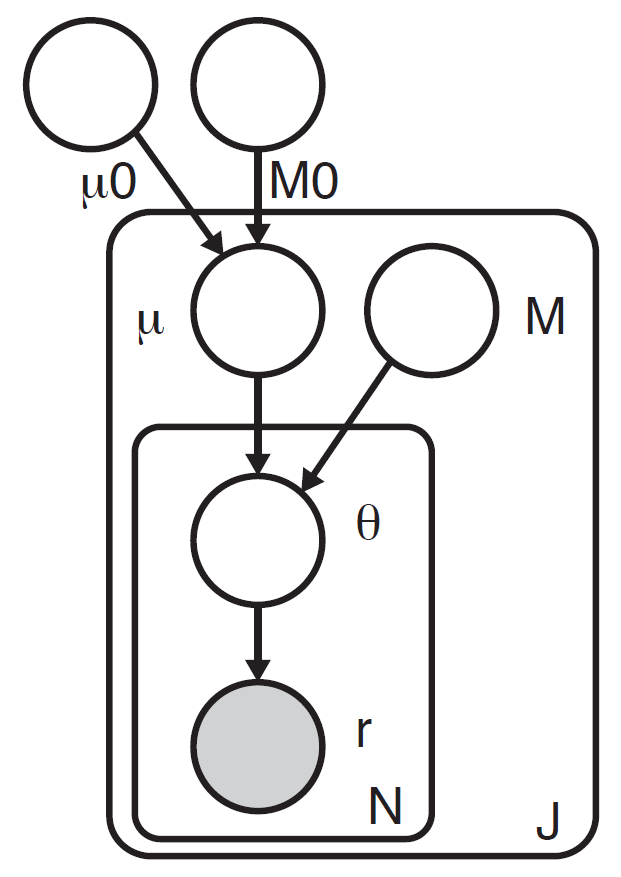


Figure RVD2 Graphical Model

## Inference algorithm and variant calling

### Parameter estimation

Method of moments was adopted for initializing model parameters.

### Metropolis-within-Gibbs

### Bayesian Hypothesis Testing

Variant calling test

# Result

Figure 1 Overall of the process

## Simulation Results

Figure 2 ROC curves for synthetic data

## Comparison Results on Synthetic DNA

Figure 3 ROC comparison with other methods: varScan/Samtools/RVD

## Empirical Results on Clinical Data MS samples

We tested our method on sequence data from clinical H1N1 influenza samples. We expect to see that…

Table: Table of variants

Figure 4

## Performance with Read depth

Samtools way to thin segment data, showing performance as read depth nij decreasing

Figure 5 ROC by read depth