Population Genetics Homework - Week 6

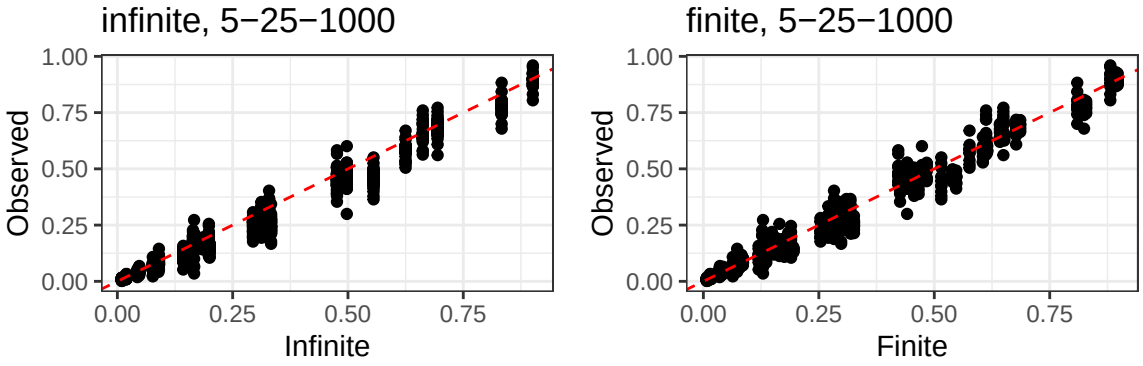
Christian Polania

12/17/2021

Intro

I ran 7 sets of simulations to explore the variability generated by, and the usefulness of, the infinite and finite island approximations. Starting with the original parameters to be used as a middle ground, I then ran an "over" and "under" set of simulations using a low and high value for n_loci, n_sample, and n_gen. I'll show the low, middle and high value simulations for each, with two plots and two model summaries (one plot and summary each for infinite and finite). Then, I'll discuss.

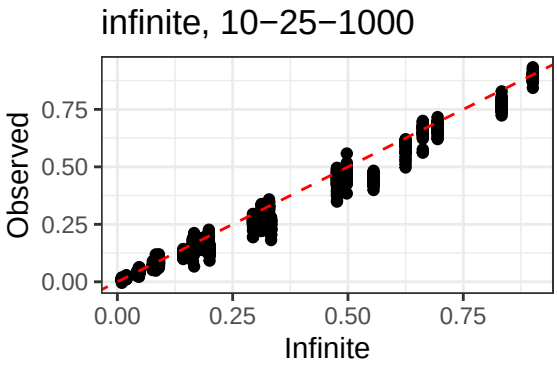
Effect of number of number of loci (2, 10, 20)

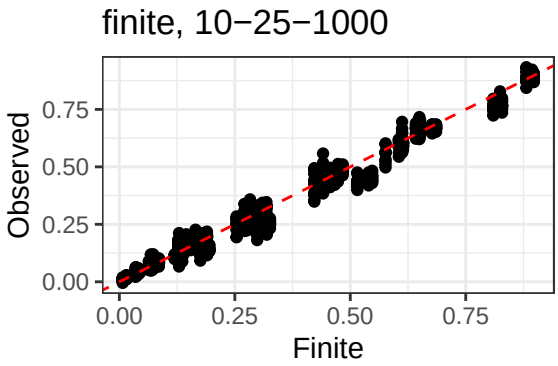



```
##
## Call:
## lm(formula = Observed ~ Infinite, data = df)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.168861 -0.013447  0.005154  0.014868  0.134695
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.007805   0.001724  -4.528  6.7e-06 ***
## Infinite     0.957284   0.004604 207.907  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.03627 on 958 degrees of freedom
## Multiple R-squared:  0.9783, Adjusted R-squared:  0.9783
## F-statistic: 4.323e+04 on 1 and 958 DF,  p-value: < 2.2e-16

##
```

```
## Call:
## lm(formula = Observed ~ Finite, data = df)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.140771 -0.012523  0.000508  0.014906  0.168592
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.002148   0.001720   1.249    0.212
## Finite      0.976126   0.004782 204.121   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.03692 on 958 degrees of freedom
## Multiple R-squared:  0.9775, Adjusted R-squared:  0.9775
## F-statistic: 4.167e+04 on 1 and 958 DF,  p-value: < 2.2e-16
```

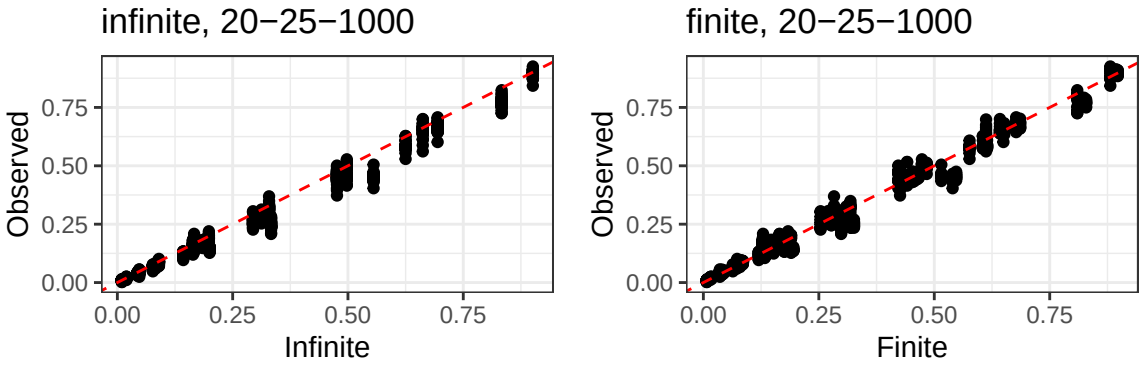





```
##
## Call:
## lm(formula = Observed ~ Infinite, data = df)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.127636 -0.008931  0.005794  0.014997  0.093354
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.006875   0.001458  -4.716 2.77e-06 ***
## Infinite     0.947874   0.003895 243.380  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.03068 on 958 degrees of freedom
## Multiple R-squared:  0.9841, Adjusted R-squared:  0.9841
## F-statistic: 5.923e+04 on 1 and 958 DF,  p-value: < 2.2e-16

##
## Call:
## lm(formula = Observed ~ Finite, data = df)
```

```
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.109908 -0.008509  0.000702  0.014680  0.128603
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.002849   0.001419   2.009   0.0448 *
## Finite      0.967035   0.003944 245.177   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.03045 on 958 degrees of freedom
## Multiple R-squared:  0.9843, Adjusted R-squared:  0.9843
## F-statistic: 6.011e+04 on 1 and 958 DF,  p-value: < 2.2e-16
```

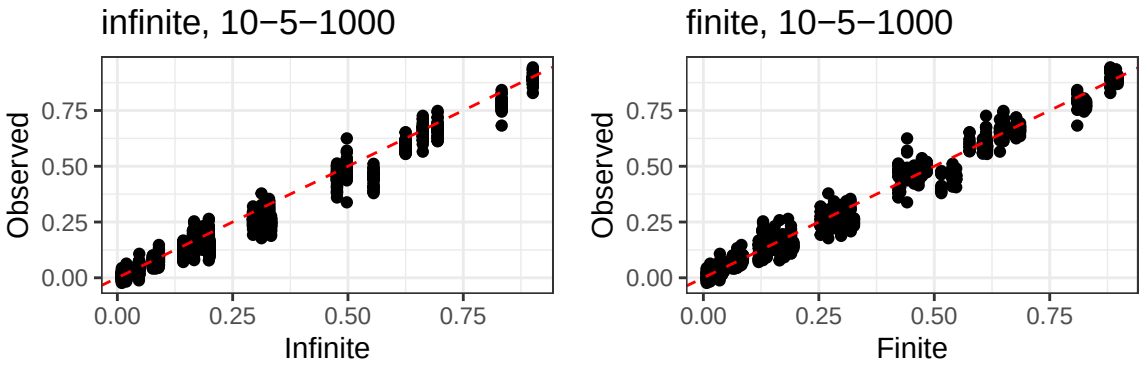



```
##
## Call:
## lm(formula = Observed ~ Infinite, data = df)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.118750 -0.007318  0.005204  0.010887  0.076113
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.006767   0.001245  -5.433 7.01e-08 ***
## Infinite     0.952861   0.003327 286.416  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0262 on 958 degrees of freedom
## Multiple R-squared:  0.9885, Adjusted R-squared:  0.9884
## F-statistic: 8.203e+04 on 1 and 958 DF,  p-value: < 2.2e-16


##
## Call:
## lm(formula = Observed ~ Finite, data = df)
##
## Residuals:
```

```
##       Min        1Q    Median        3Q       Max
## -0.123652 -0.007496  0.000700  0.012569  0.100631
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.003162   0.001271   2.488    0.013 *
## Finite      0.971534   0.003533 274.985   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.02728 on 958 degrees of freedom
## Multiple R-squared:  0.9875, Adjusted R-squared:  0.9875
## F-statistic: 7.562e+04 on 1 and 958 DF,  p-value: < 2.2e-16
```

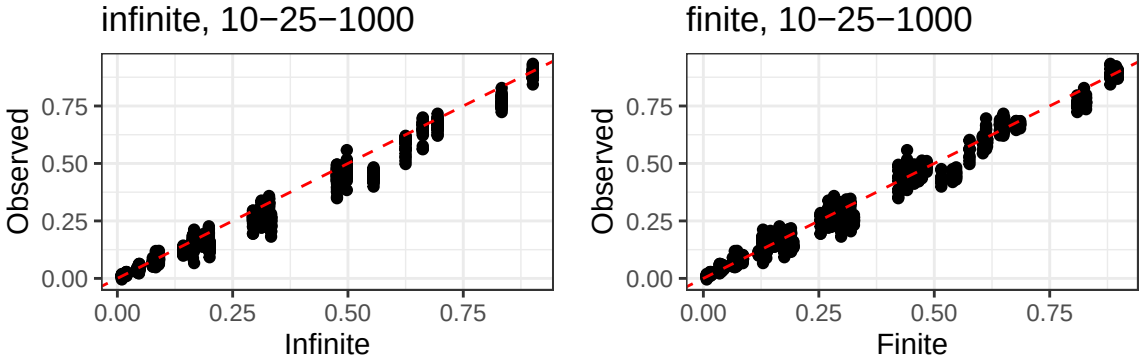
Effect of number of sample size (2, 10, 20)




```
##
## Call:
## lm(formula = Observed ~ Infinite, data = df)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.144980 -0.013908  0.004623  0.018603  0.156136
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.008010   0.001682  -4.764  2.2e-06 ***
## Infinite     0.958339   0.004492 213.348  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.03538 on 958 degrees of freedom
## Multiple R-squared:  0.9794, Adjusted R-squared:  0.9794
## F-statistic: 4.552e+04 on 1 and 958 DF,  p-value: < 2.2e-16


##
## Call:
## lm(formula = Observed ~ Finite, data = df)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
```

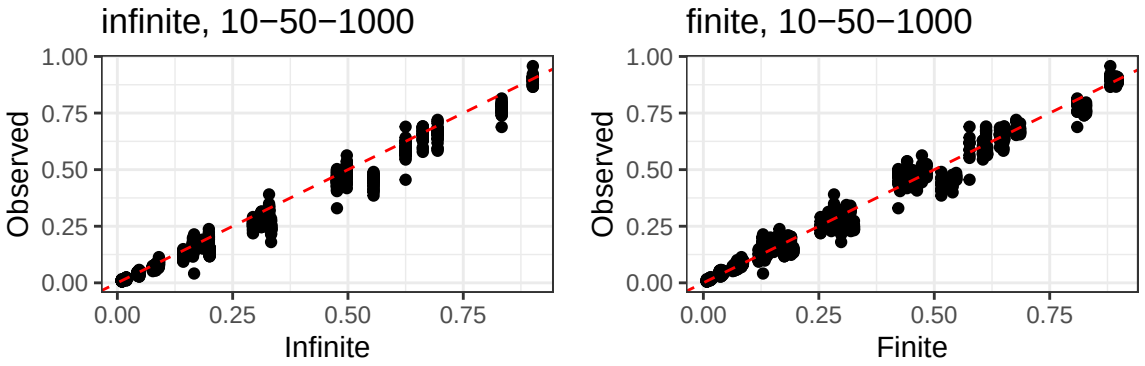
```
## -0.131061 -0.015890  0.001494  0.016441  0.191861
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.001944   0.001676    1.16    0.246
## Finite      0.977239   0.004661  209.65   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.03599 on 958 degrees of freedom
## Multiple R-squared:  0.9787, Adjusted R-squared:  0.9786
## F-statistic: 4.395e+04 on 1 and 958 DF,  p-value: < 2.2e-16
```




```
##
## Call:
## lm(formula = Observed ~ Infinite, data = df)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.127636 -0.008931  0.005794  0.014997  0.093354
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.006875   0.001458  -4.716 2.77e-06 ***
## Infinite     0.947874   0.003895 243.380  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.03068 on 958 degrees of freedom
## Multiple R-squared:  0.9841, Adjusted R-squared:  0.9841
## F-statistic: 5.923e+04 on 1 and 958 DF,  p-value: < 2.2e-16


##
## Call:
## lm(formula = Observed ~ Finite, data = df)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.109908 -0.008509  0.000702  0.014680  0.128603
##
```

```
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.002849   0.001419   2.009   0.0448 *
## Finite      0.967035   0.003944 245.177   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.03045 on 958 degrees of freedom
## Multiple R-squared:  0.9843, Adjusted R-squared:  0.9843
## F-statistic: 6.011e+04 on 1 and 958 DF,  p-value: < 2.2e-16
```

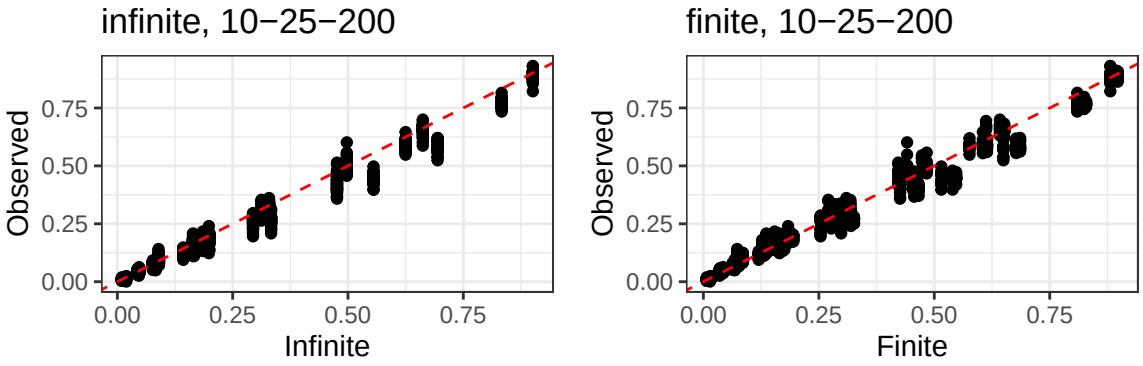



```
##
## Call:
## lm(formula = Observed ~ Infinite, data = df)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.137133 -0.008012  0.005160  0.012998  0.106174
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.006381   0.001429  -4.465 8.96e-06 ***
## Infinite     0.952035   0.003818 249.380  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.03007 on 958 degrees of freedom
## Multiple R-squared:  0.9848, Adjusted R-squared:  0.9848
## F-statistic: 6.219e+04 on 1 and 958 DF,  p-value: < 2.2e-16


##
## Call:
## lm(formula = Observed ~ Finite, data = df)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.127700 -0.008873  0.000291  0.012326  0.126554
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept) 0.003547   0.001447   2.451   0.0144 *
## Finite      0.970661   0.004024 241.220   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.03107 on 958 degrees of freedom
## Multiple R-squared:  0.9838, Adjusted R-squared:  0.9838
## F-statistic: 5.819e+04 on 1 and 958 DF,  p-value: < 2.2e-16
```

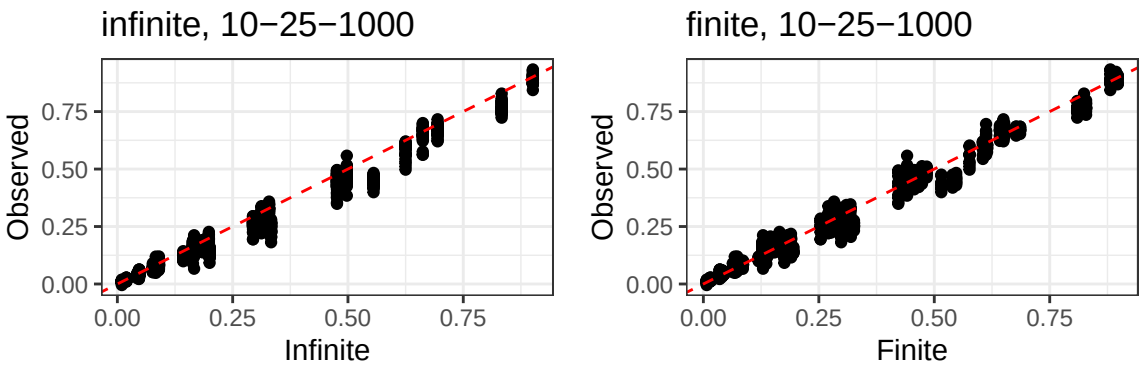
Effect of number of number of generations (2, 10, 20)




```
##
## Call:
## lm(formula = Observed ~ Infinite, data = df)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.119780 -0.008644  0.000644  0.014733  0.139635
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.001807   0.001473   1.227     0.22
## Infinite    0.924768   0.003934 235.045   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.03099 on 958 degrees of freedom
## Multiple R-squared:  0.983,  Adjusted R-squared:  0.9829
## F-statistic: 5.525e+04 on 1 and 958 DF,  p-value: < 2.2e-16
```

```
##
## Call:
## lm(formula = Observed ~ Finite, data = df)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.122757 -0.011248 -0.002926  0.016474  0.174168
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.011499   0.001502   7.657 4.65e-14 ***
```

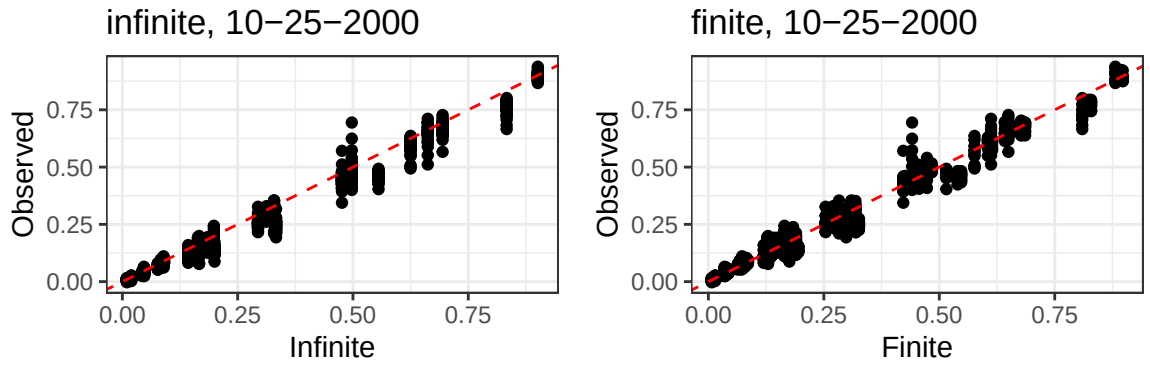
```
## Finite      0.942678   0.004176 225.753  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.03224 on 958 degrees of freedom
## Multiple R-squared:  0.9815, Adjusted R-squared:  0.9815
## F-statistic: 5.096e+04 on 1 and 958 DF,  p-value: < 2.2e-16
```




```
##
## Call:
## lm(formula = Observed ~ Infinite, data = df)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.127636 -0.008931  0.005794  0.014997  0.093354
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.006875   0.001458  -4.716 2.77e-06 ***
## Infinite     0.947874   0.003895 243.380  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.03068 on 958 degrees of freedom
## Multiple R-squared:  0.9841, Adjusted R-squared:  0.9841
## F-statistic: 5.923e+04 on 1 and 958 DF,  p-value: < 2.2e-16


##
## Call:
## lm(formula = Observed ~ Finite, data = df)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.109908 -0.008509  0.000702  0.014680  0.128603
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.002849   0.001419   2.009   0.0448 *
## Finite      0.967035   0.003944 245.177   <2e-16 ***
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.03045 on 958 degrees of freedom
## Multiple R-squared:  0.9843, Adjusted R-squared:  0.9843
## F-statistic: 6.011e+04 on 1 and 958 DF,  p-value: < 2.2e-16
```




```
##
## Call:
## lm(formula = Observed ~ Infinite, data = df)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.119837 -0.010052  0.005007  0.013721  0.228143
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.006884   0.001528  -4.506 7.44e-06 ***
## Infinite     0.950996   0.004081 233.004  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.03215 on 958 degrees of freedom
## Multiple R-squared:  0.9827, Adjusted R-squared:  0.9826
## F-statistic: 5.429e+04 on 1 and 958 DF,  p-value: < 2.2e-16


##
## Call:
## lm(formula = Observed ~ Finite, data = df)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.122378 -0.010102  0.000505  0.013982  0.263575
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.002968   0.001520   1.953   0.0511 .
## Finite      0.969851   0.004225 229.538   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 0.03262 on 958 degrees of freedom
## Multiple R-squared:  0.9821, Adjusted R-squared:  0.9821
## F-statistic: 5.269e+04 on 1 and 958 DF,  p-value: < 2.2e-16
```

Analysis

How close is the relationship between observed estimates of FST from the simulations and those predicted from the infinite island model? from the finite island model? The strength of the relationship between the observed FSTs and the estimated FSTs were consistently higher in the finite model. Higher loci number, sample size, and generation number each improved the fit, but it was always the same amount for finite and infinite.

Under what conditions is the observed variation in FST relatively small? For all simulations, Fst estimates grew less variable with higher parameters except for changes in generation number. Interestingly, this improvement was consistently larger for the infinite models.

Are there conditions under which the predictions of the infinite island and finite island model are similar enough that we don't need to worry about the added mathematical complexity of the finite island model? As FST estimates grew more consistent with higher parameters and this difference was larger for infinite models, the simulations where the models performed most similarly were with lower number of loci and lower sample size.