

Preliminary exploration of geographic trends in *P. falciparum* relatedness on the Colombian coast

Introduction

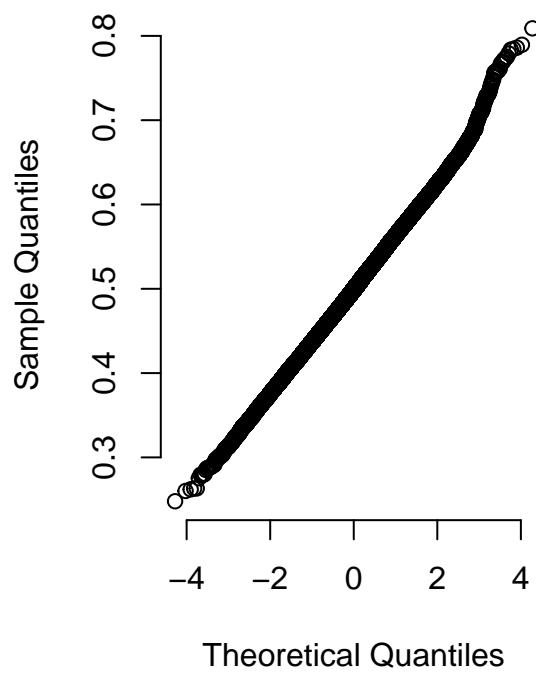
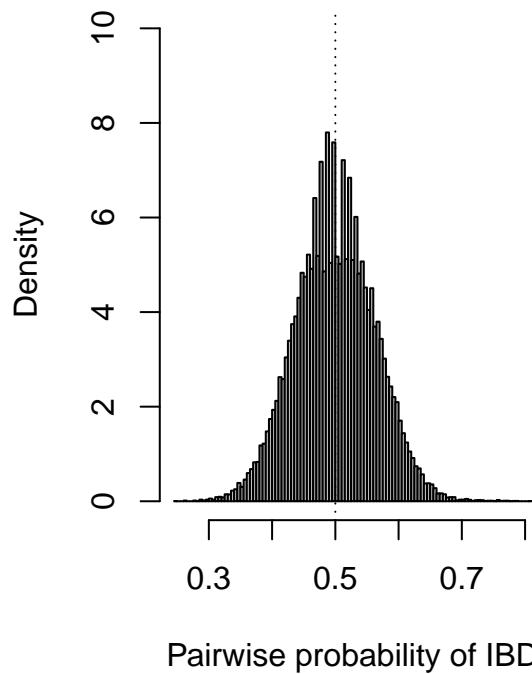
In the following I have quantified the trend in parasite relatedness (based on identity by descent, IBD, estimated under a hidden Markov model, HMM, described elsewhere) with inter-state distance (km), with a view to assessing the potential for comparison with the Thai-Myanmar border, where, on average, the log-odds of relatedness decrease by 0.02 with every kilometer between collection sites and week between collection dates. It is a preliminary analysis based on the following steps, which may be erroneous.

- I have used lon/lat for Tadó, Buenaventura, Guapi and Tumaco to represent data from states Choco, Valle, Cauca and Nariño, respectively.
 - I have assumed the names of SNPs 1 to 250 are the SNP IDs in Additional file 1 of Echeverry et al., with the same order, but including only SNPs with a cross in the column labelled ‘250 most informative SNPs’.
 - I based chromosome number and position on the SNP ID. Positions will thus differ to 3d7, but distances between SNPs (required under the HMM) should be roughly equivalent.
 - I have treated heteroallellic calls as missing.
 - I have treated calls labelled ‘–’ as missing.
-

Pairwise identity by state

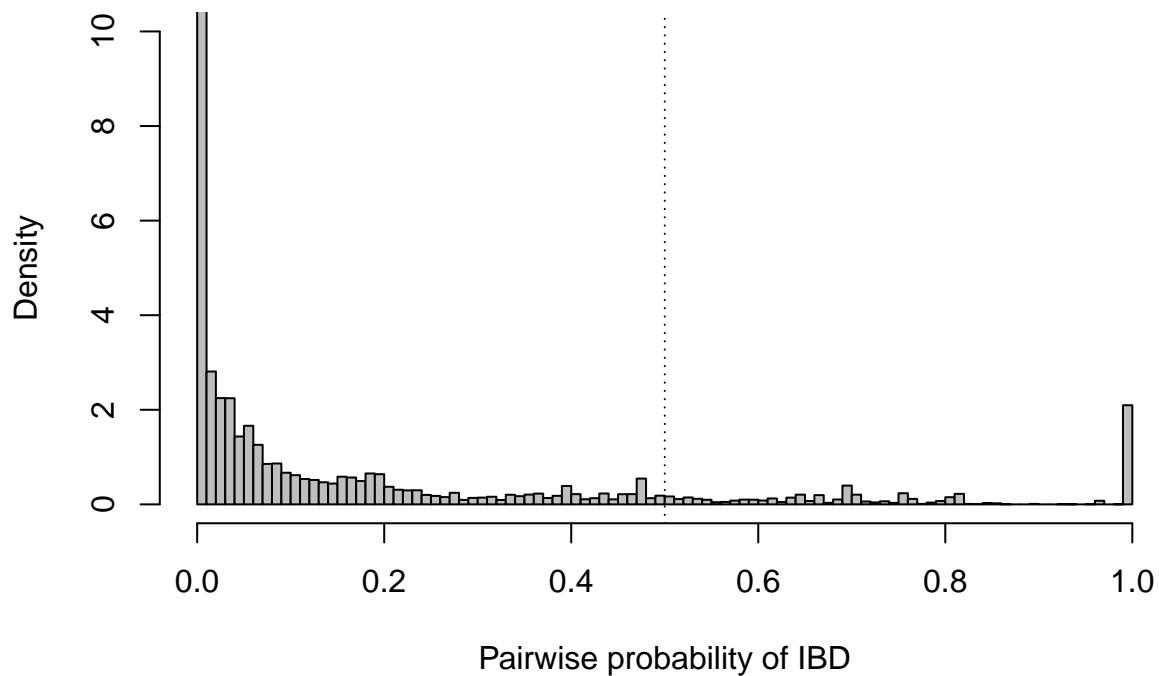
The empirical density of proportion identity by state (IBS) for all pairwise sample comparisons appears to be Guassian:

Normal Q-Q Plot



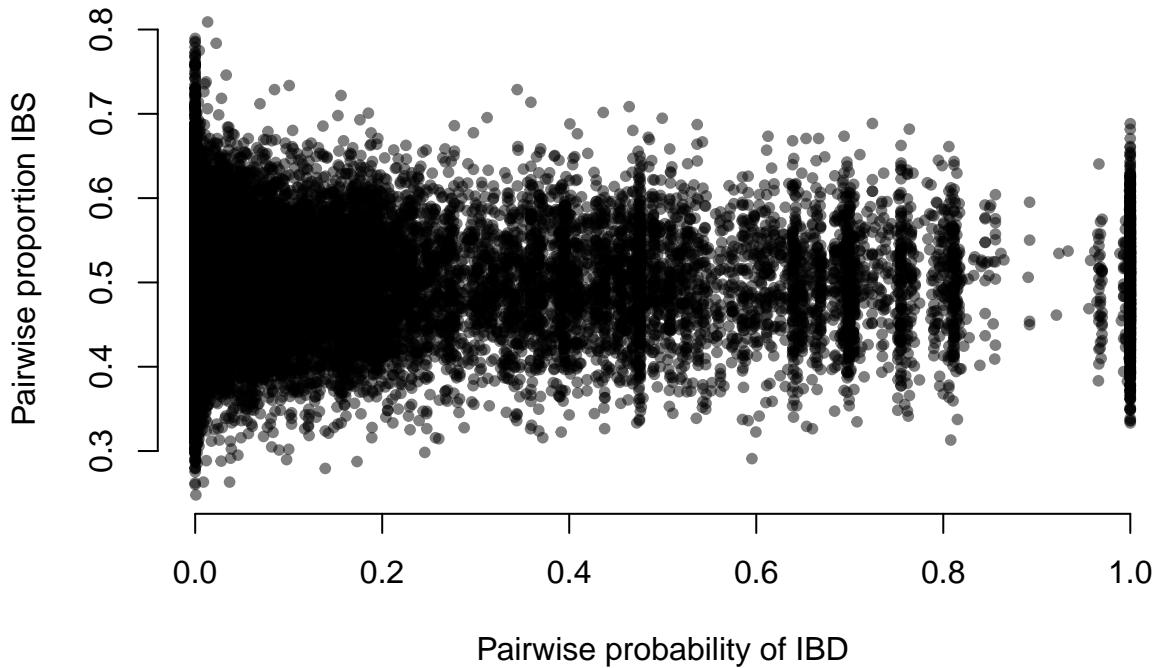
Pairwise identity by descent

The empirical density of probability IBD for all pairwise sample comparisons is positively skewed with some highly related parasites:



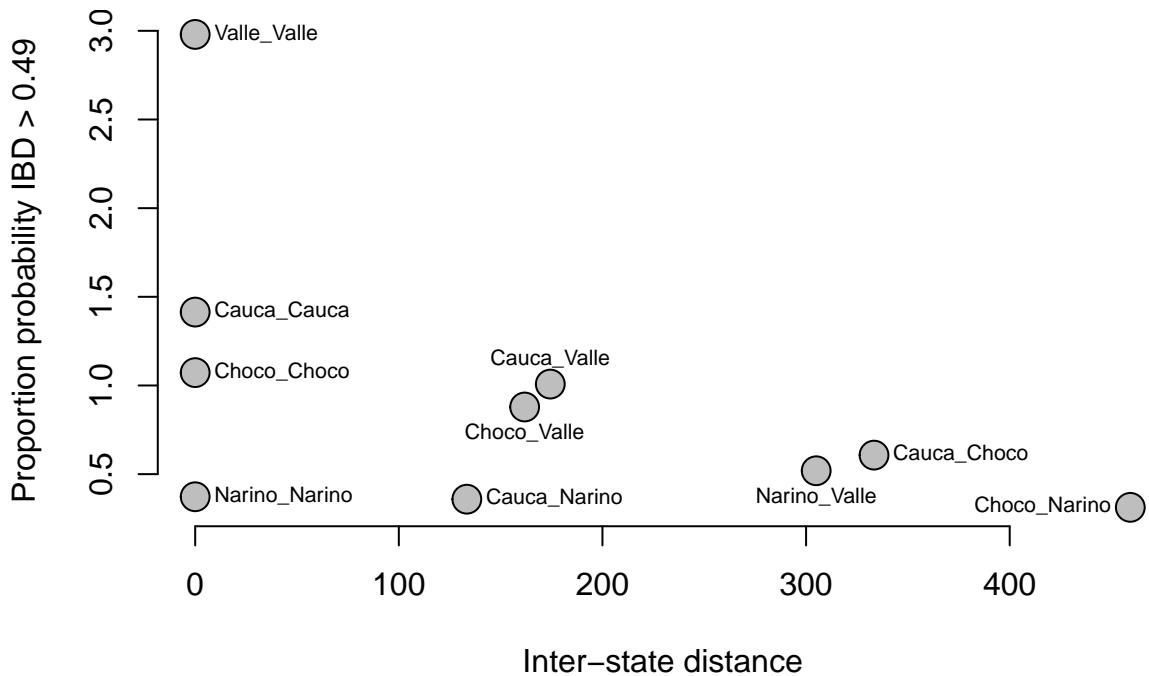
IBD versus IBS

IBD and IBS do not appear to be correlated:



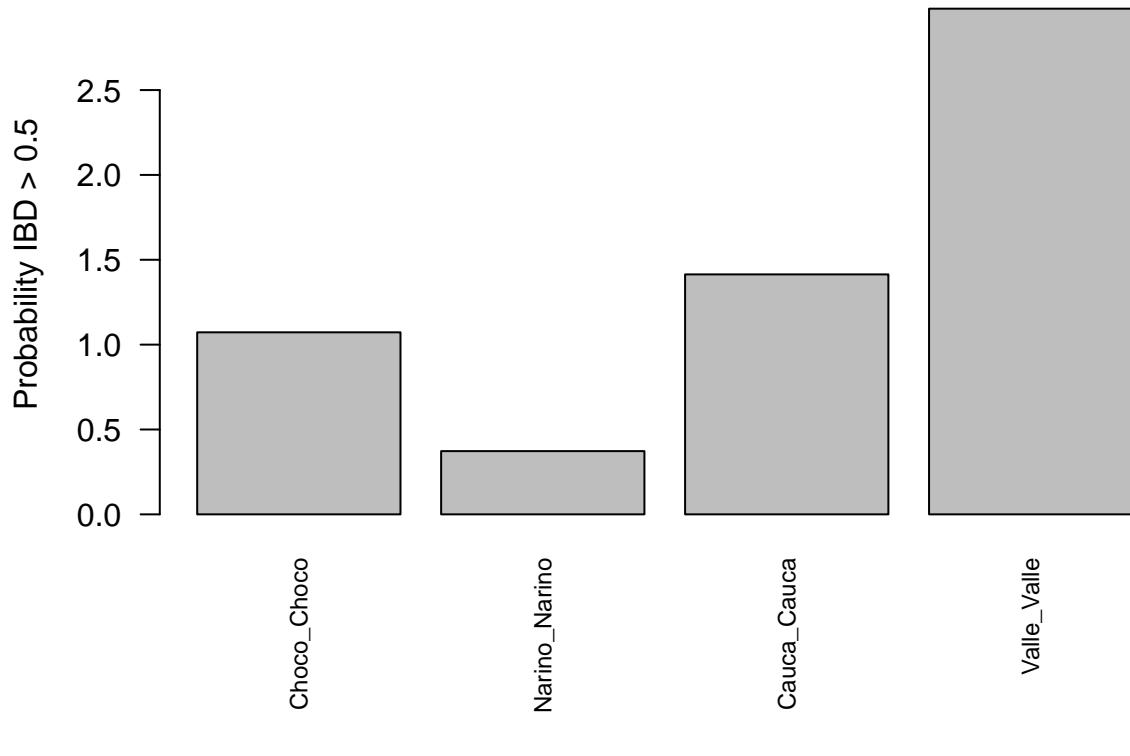
Proportions of related comparisons correlate with distance

The proportion of parasite comparisons with $\text{IBD} > 0.49$ correlate with distance, but there is a lot of within-state variability, and Nariño appears to have particularly low within and across state IBD:



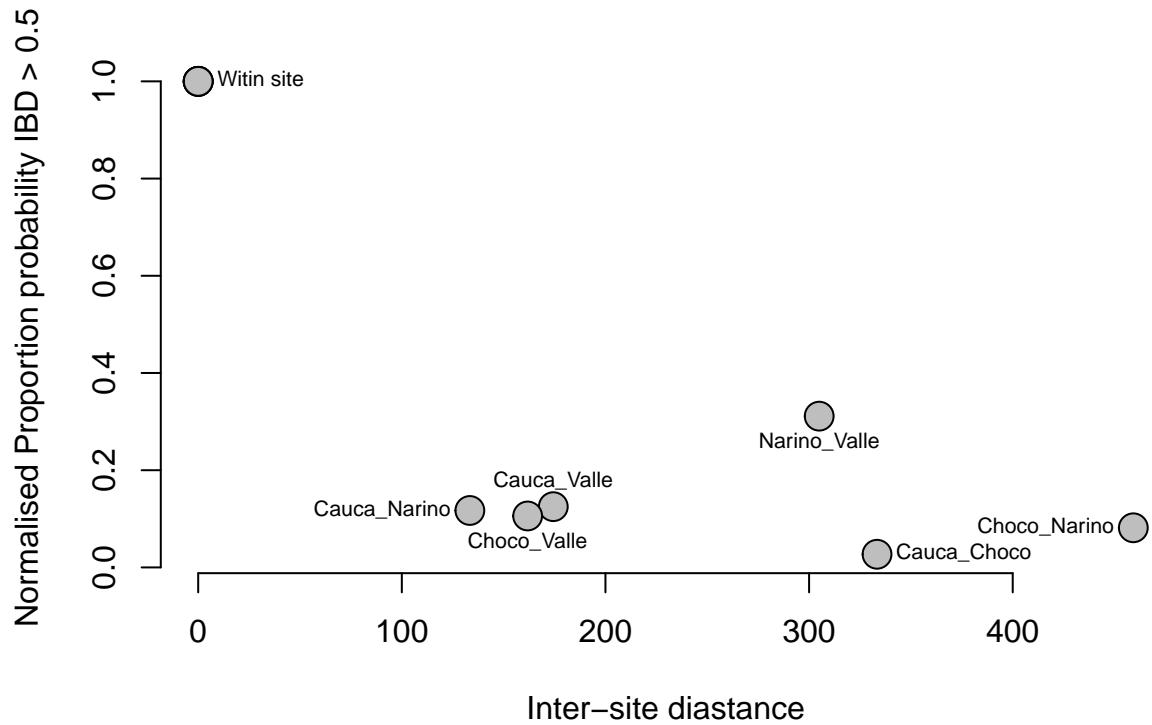
Within-state variability ordered by transmission

Plot of within-state relatedness in order of decreasing transmission (order based on Figure 1B of Echeverry et al., such that an increase is expected from left to right). Again, Nariño appears to have particularly low within-state IBD:



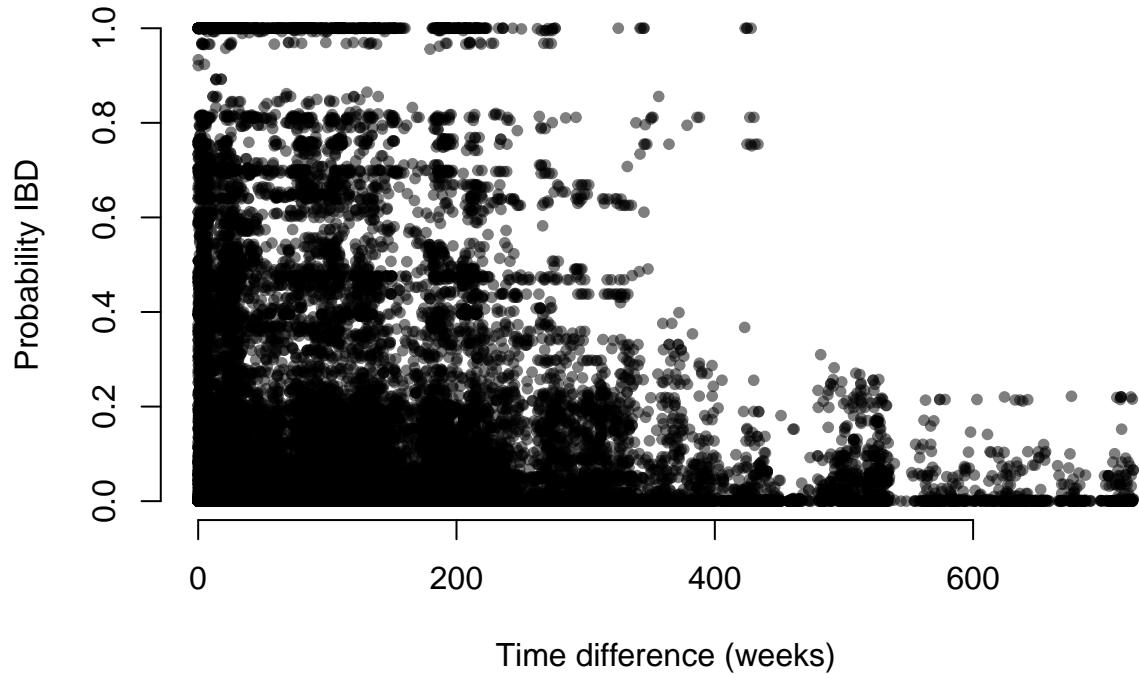
Normalised proportions of related comparisons correlate with distance

Correlation between normalised proportions and distance appears to be driven by within versus across comparisons:



Plot of IBD probability against time

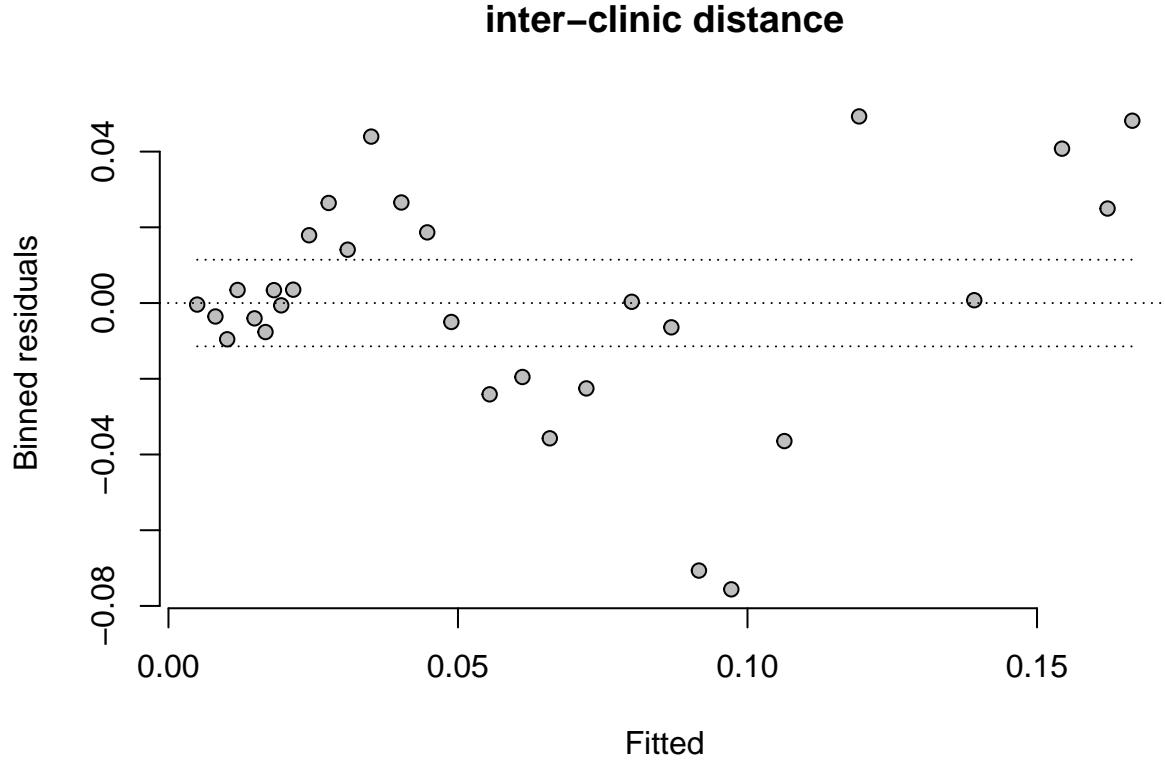
Relatedness is also negatively associated with time between sampling dates:



Regression results

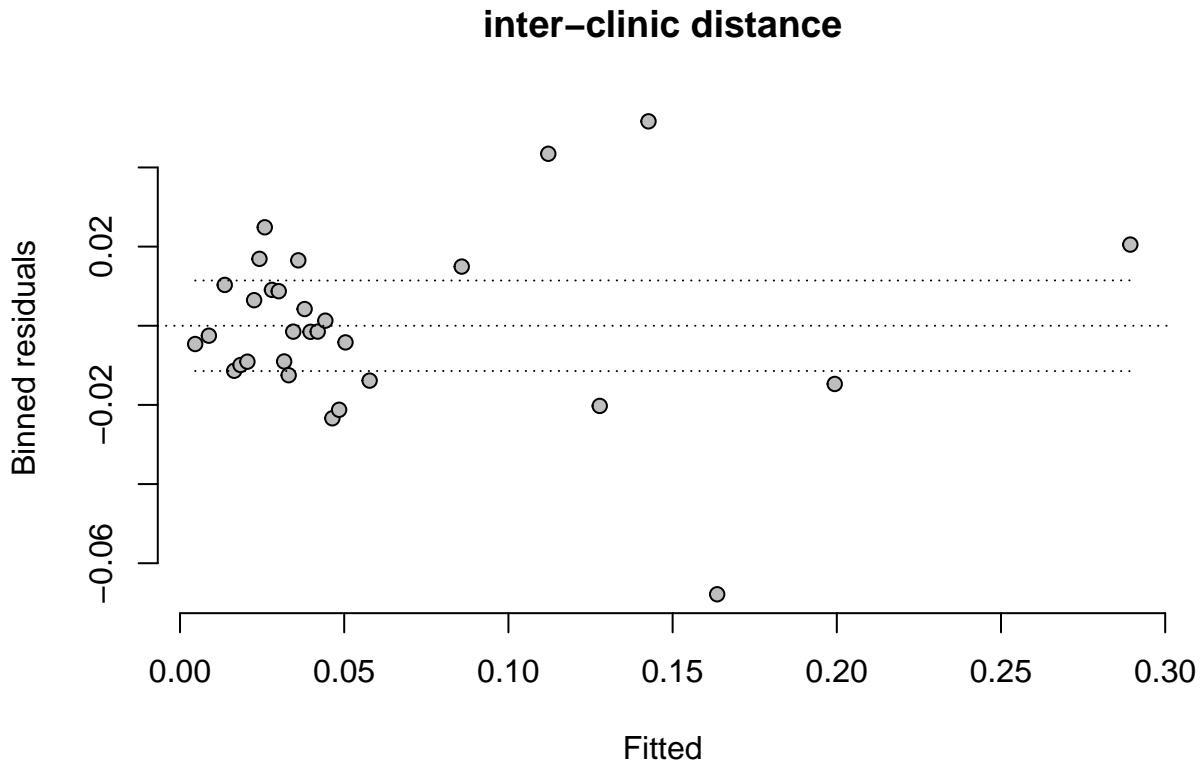
Before accounting for inter-state variability, regression coefficients suggest significant decrease of IBD with time and distance, but binned residuals suggest a poor fit:

```
##              Estimate Std. Error   z value Pr(>|z|)
## (Intercept) -1.602671e+00 3.193747e-02 -50.1815152 0.000000e+00
## geo_dist     -4.568728e-03 2.265126e-04 -20.1698661 1.801256e-90
## time_dist    -4.070412e-03 2.646802e-04 -15.3786038 2.278072e-53
## geo_dist:time_dist -1.090763e-06 1.688669e-06 -0.6459303 5.183245e-01
```



After accounting for within-state variance, fit is better, but there is no-longer evidence of a geographic trend:

```
##              Estimate Std. Error   z value Pr(>|z|)
## (Intercept) -2.9782505567 7.349423e-02 -40.523596 0.000000e+00
## ChocoTRUE    0.4444650620 1.077342e-01   4.125570 3.698177e-05
## NarinoTRUE   1.4217708804 7.660380e-02  18.560056 6.763433e-77
## ValleTRUE    0.9704019128 1.308184e-01   7.417930 1.189650e-13
## CaucaTRUE    2.1104239592 8.551377e-02  24.679348 1.782297e-134
## geo_dist     0.0003240168 2.849868e-04   1.136954 2.555576e-01
## time_dist   -0.0024893840 2.700234e-04  -9.219143 2.994833e-20
## geo_dist:time_dist -0.0000090501 1.564414e-06 -5.784977 7.252215e-09
```

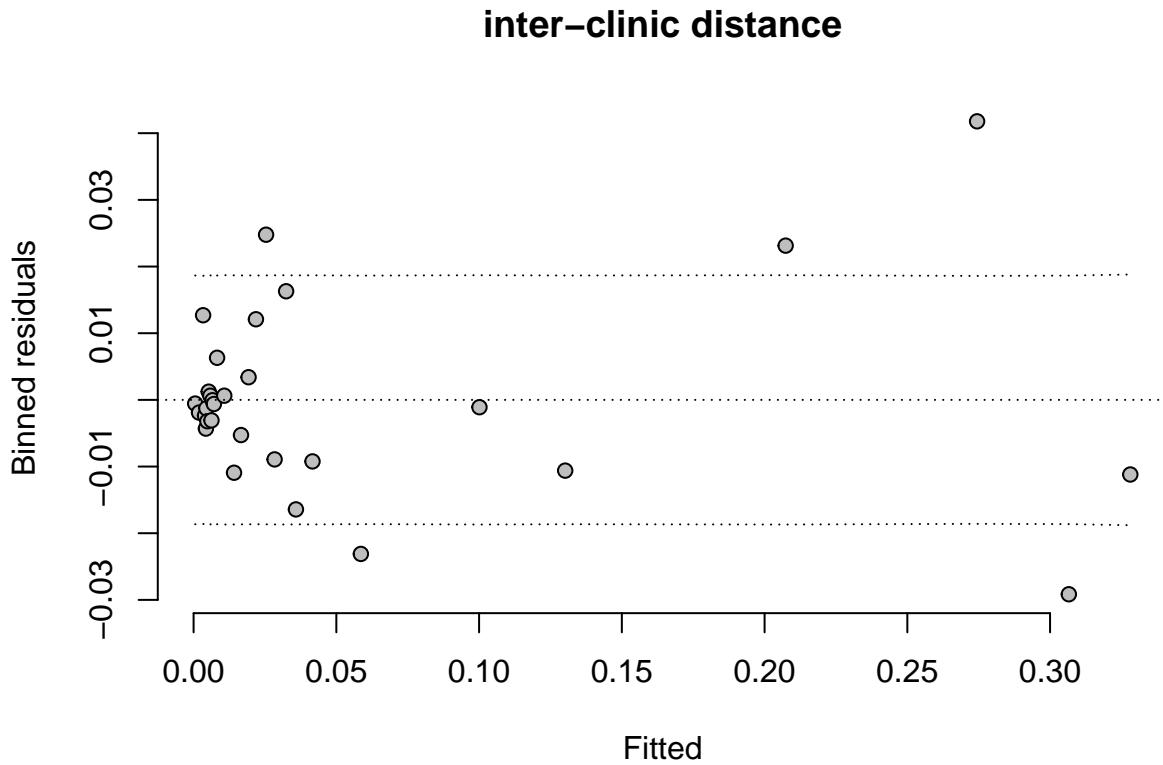


Removal of data from Nariño, however, further improves fit and recovers the trend, which is of the same order of weeks:

```

##                               Estimate   Std. Error     z value    Pr(>|z|)
## (Intercept)           -1.297119e+00 2.885484e-01 -4.495326 6.946340e-06
## ChocoTRUE            -5.322893e-01 2.871569e-01 -1.853653 6.378880e-02
## ValleTRUE             -1.512586e-01 2.994437e-01 -0.505132 6.134661e-01
## CaucaTRUE              6.051906e-01 2.870808e-01  2.108085 3.502367e-02
## geo_dist             -1.029421e-02 1.678077e-03 -6.134529 8.541211e-10
## time_dist            -1.083365e-02 8.875202e-04 -12.206650 2.864419e-34
## geo_dist:time_dist  2.116967e-05 7.171841e-06  2.951776 3.159517e-03

```



Conclusion

In summary, based on data from Choco, Valle and Cauca, there appears to be a decrease in the log-odds of relatedness of 0.01 with every kilometer and week between samples. This is less than that found on the Thai-Myanmar border, but the spatio-temporal equivalence is the same. Data from Nariño does not appear to follow the trend. It remains to be seen if this is a robust conclusion (this is a very rough preliminary analysis).