

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

To-do list

- Add CIs to proportions
- Improve model fit
- How might we consolidate the two FST stories
- Tabulate regression results
- Re-add sensitivity to the threshold

Summary of FST analyses

At the city level Using hudson's estimator, I recover insignificant p-values that do not vary in any systematic way with distance. P-values generated by permutation of city labels suggest insignificance, as do confidence intervals generated using an independent method of bootstrapping over SNPs. The point estimates agree almost perfectly with estimates generated using Weir and Hill's estimator. They do not agree with estimates generated using my Weir and Cockerham's estimator however, which as expected under theoretical results reported in literature were demonstrably sensitive to unbalanced sample sizes. Of those that can be compared at the state, estimates did not agree with those previously published.

At the state level To explore the discrepancy of my method, I reestimated FST at the state level, and compared them with estimates generated using a independent package, DivRsity, which apparently uses Weir & Cockerham, 1984 estimator. Estimates generated using Hudson's and Wier and Hills estimator almost perfectly and with estimates generated using DivRsity. My Wier and Cockerham estimates don't exactly match DivRsity, suggesting either my Weir and Cockerham estimator is wrong or DivRsity includes a sample size correction. In either case, DivRsity estimates do not agree with GENALEX6.

For skype with Diego:

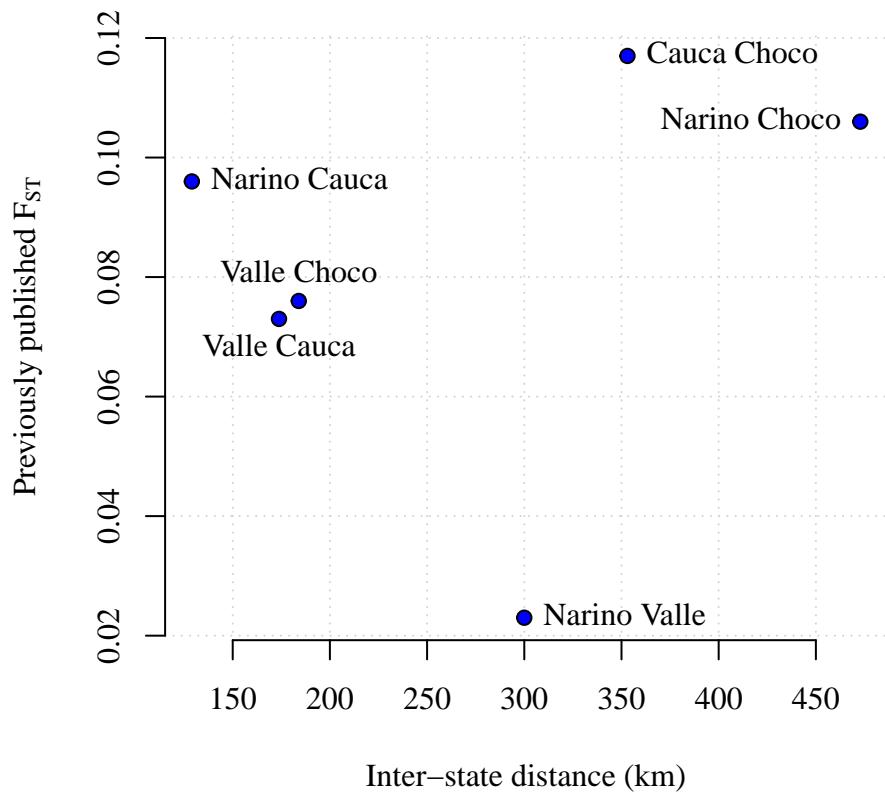
- Why “As expected, the lowest differentiation between the provinces was found between Valle-Nariño ($FST = 0.023$) while the highest was for Cauca–Chocó ($FST = 0.117$) (Additional file 4).”
- Confirm missing ‘-’
- I have treated het calls as missing data since all the samples were previously categorized as single-genotype
- Apologise for not seeing the lon/lat
- Which estimator does GENALEX 6 use?

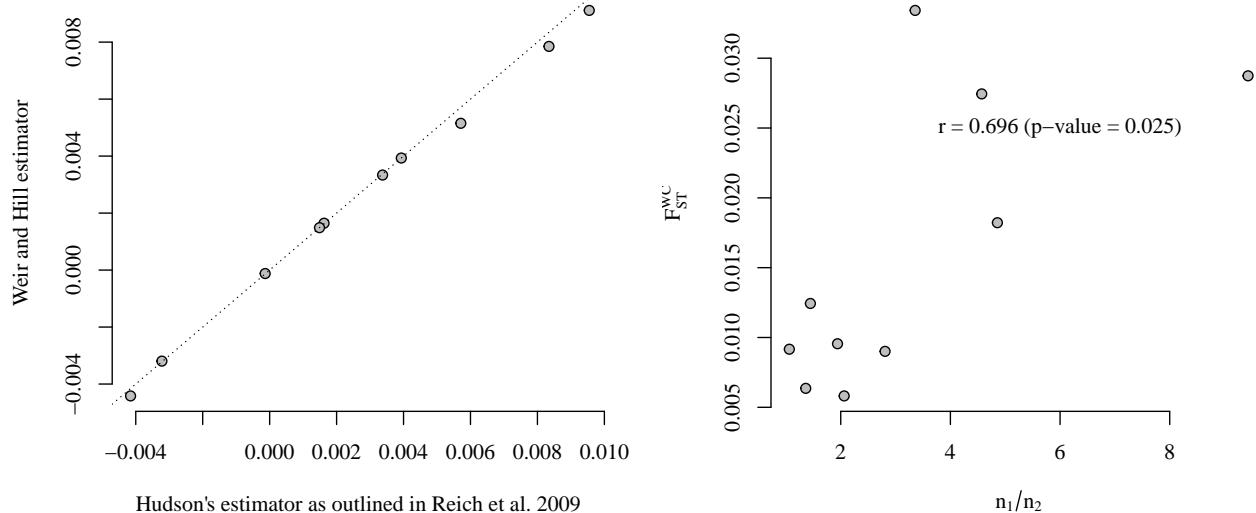
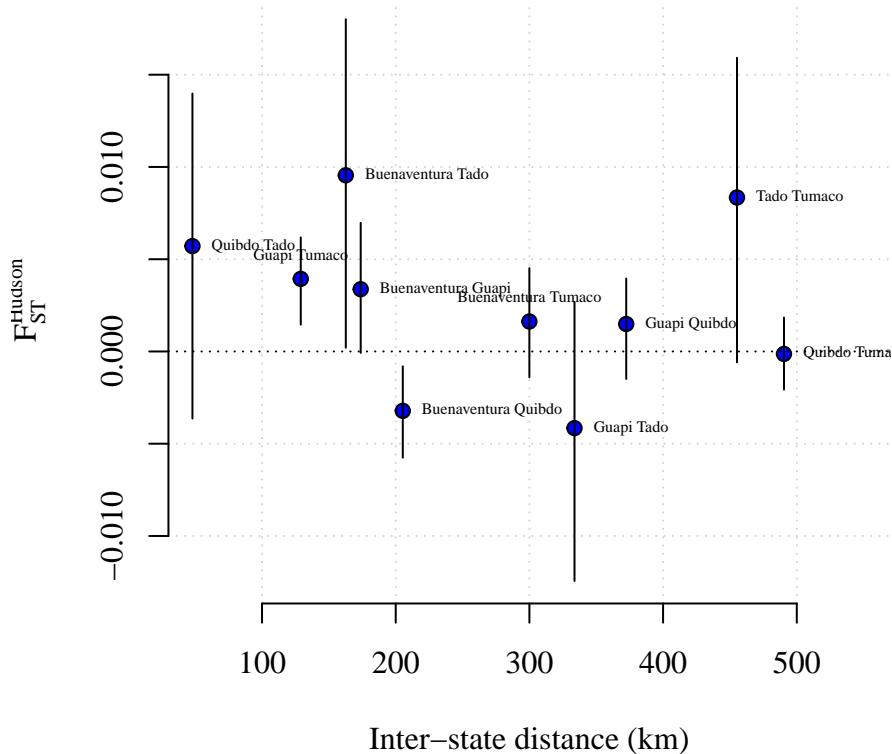
Introduction

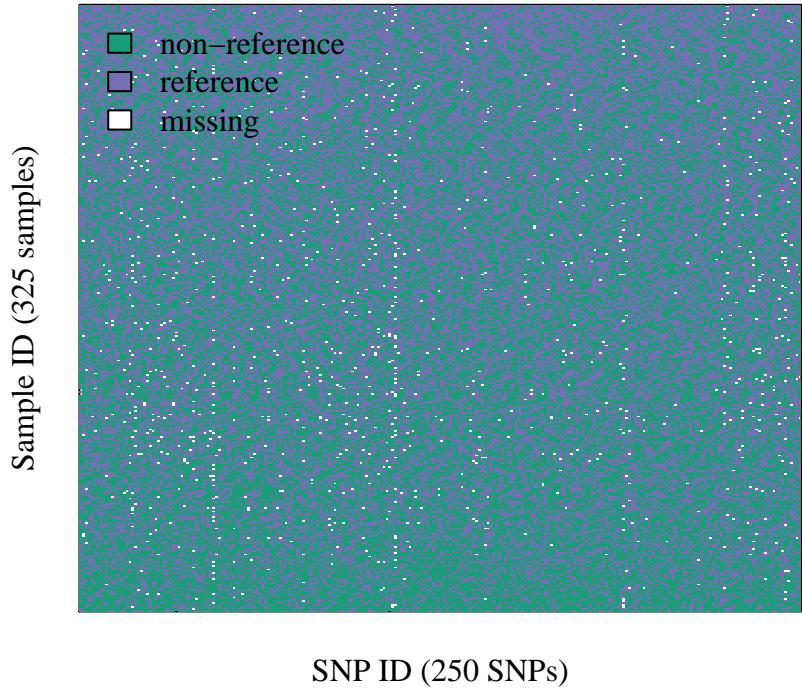
In the following I have quantified the trend in parasite relatedness (based on identity by descent, IBD, estimated under a hidden Markov model, HMM, described elsewhere) with inter-state distance (km), with a view to assessing the potential for comparison with the Thai-Myanmar border, where, on average, 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, in which I have treated calls labelled ‘-’ as missing.

	1993	1994	1997	1999	2000	2001	2002	2003	2004	2005	2006	2007	Total
Tumaco	0	0	0	2	2	10	11	59	0	23	0	25	132

	1993	1994	1997	1999	2000	2001	2002	2003	2004	2005	2006	2007	Total
Guapi	0	0	0	1	1	0	0	66	0	0	0	0	68
Buenaventura	4	1	0	5	0	0	0	0	12	15	10	0	47
Quibdo	0	0	2	0	6	1	0	0	14	6	13	22	64
Tado	0	0	0	0	0	12	2	0	0	0	0	0	14
Total	4	1	2	8	9	23	13	125	26	44	23	47	325

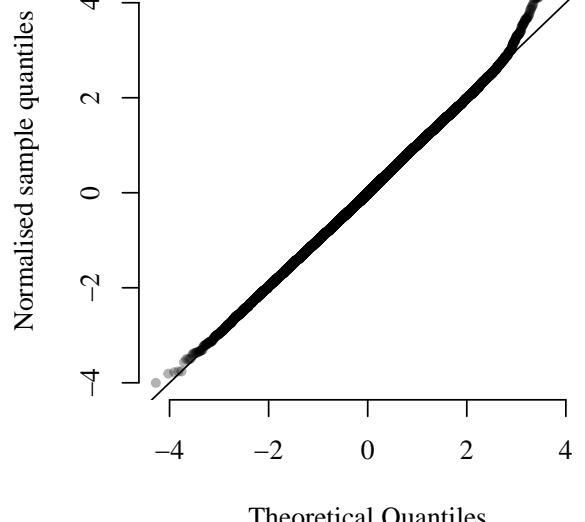
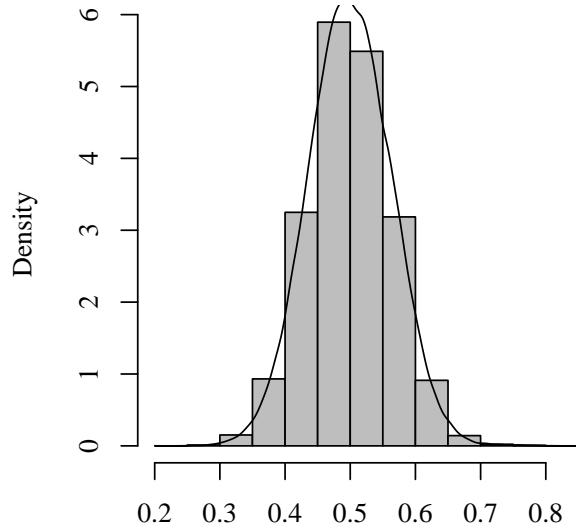


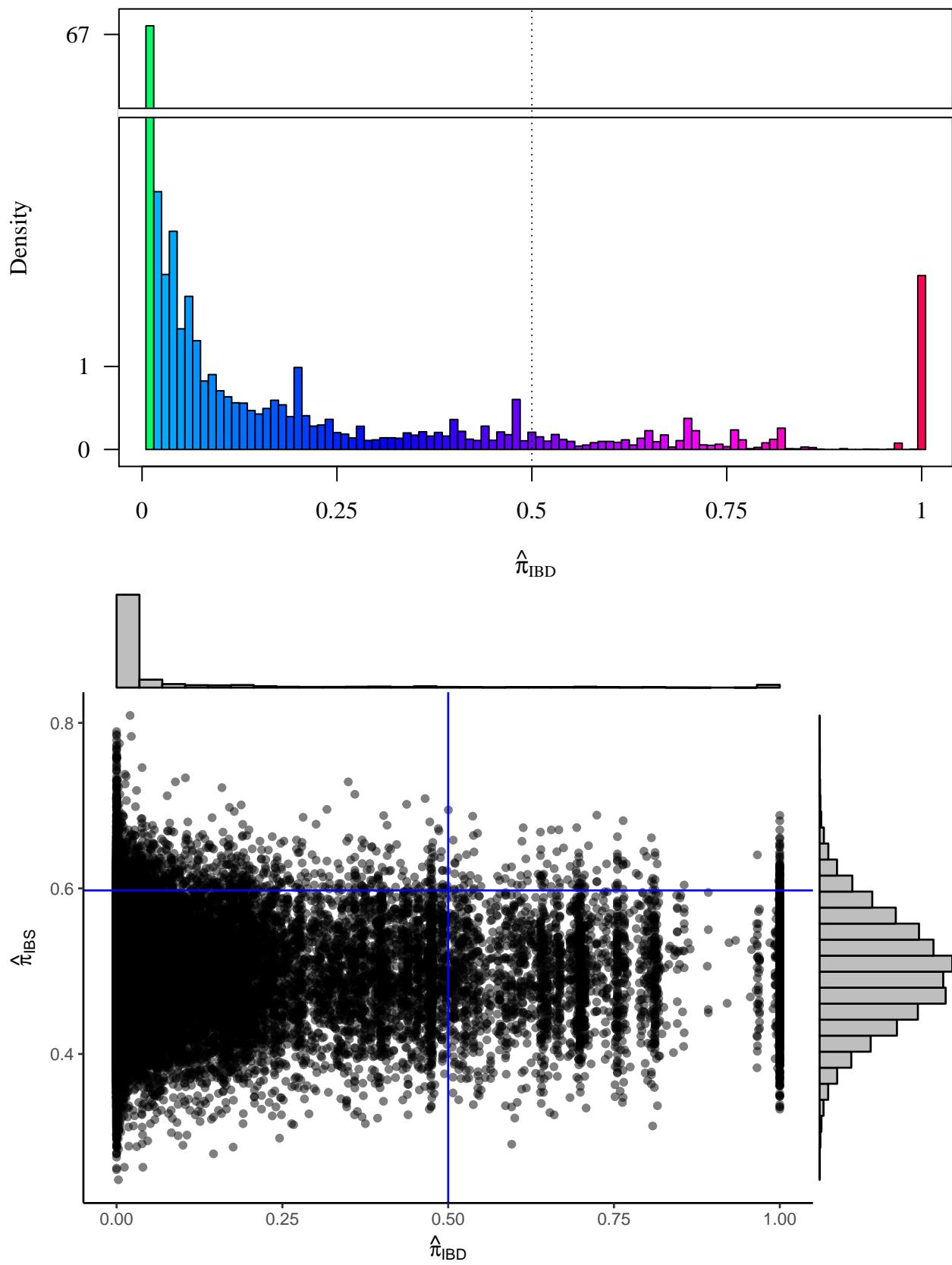




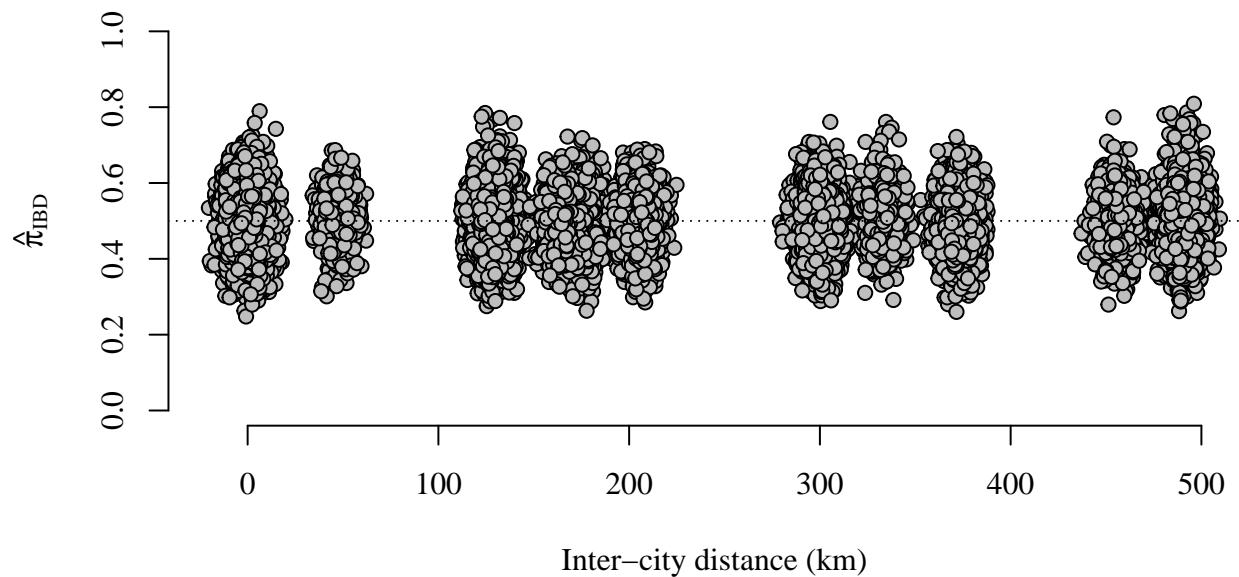
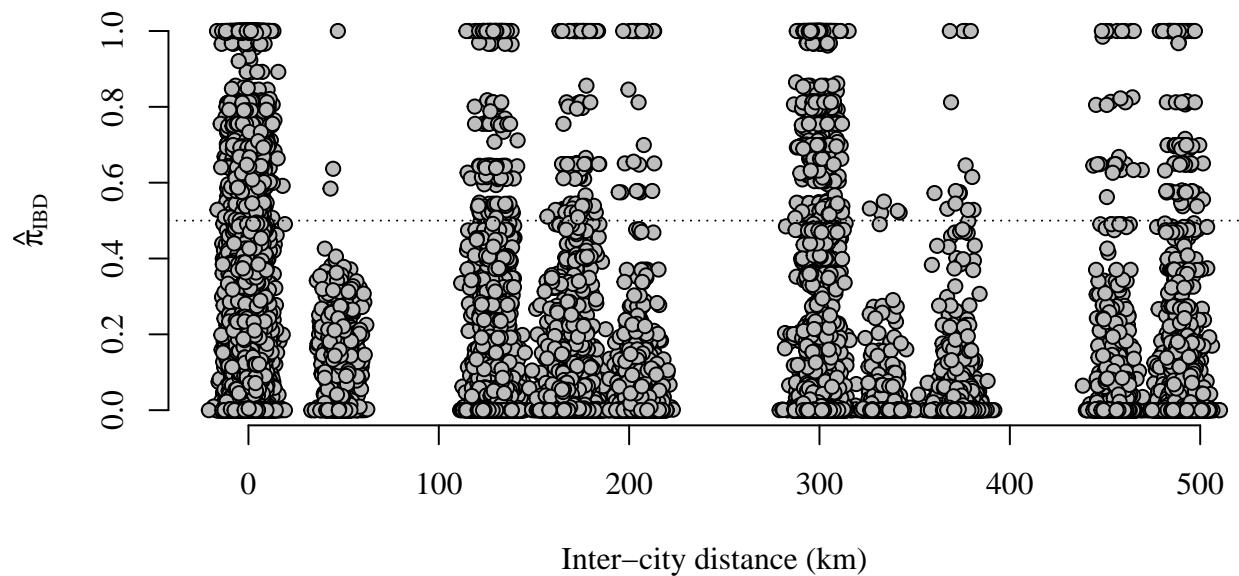
Pairwise identity by state

The empirical density of $\hat{\pi}_{IBS}$ for all pairwise sample comparisons appears to be Gaussian, but with very slightly lighter tails on the left and heavier tails on the right. The empirical density $\hat{\pi}_{IBD}$ for all pairwise sample comparisons is positively skewed with some highly related parasites. $\hat{\pi}_{IBS}$ and $\hat{\pi}_{IBD}$ do not appear to be correlated.

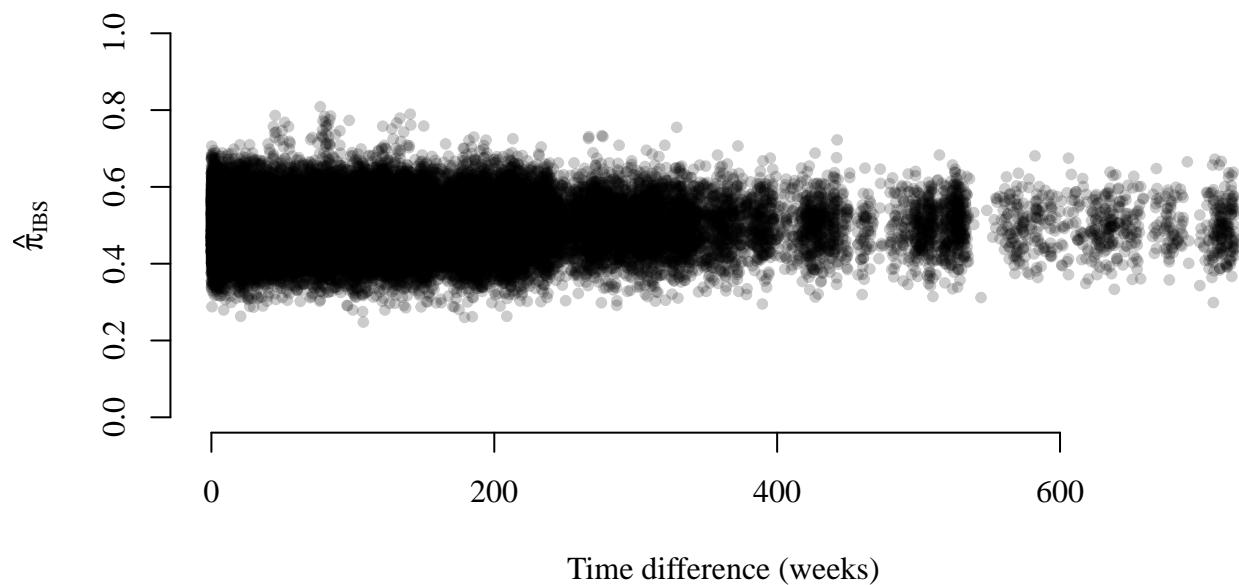
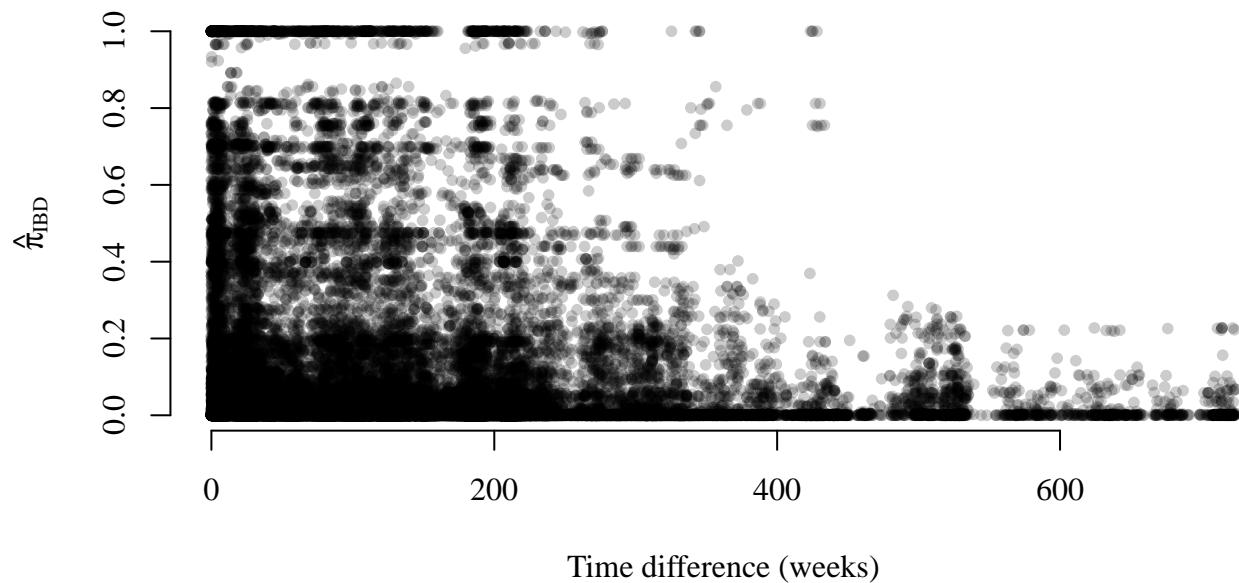


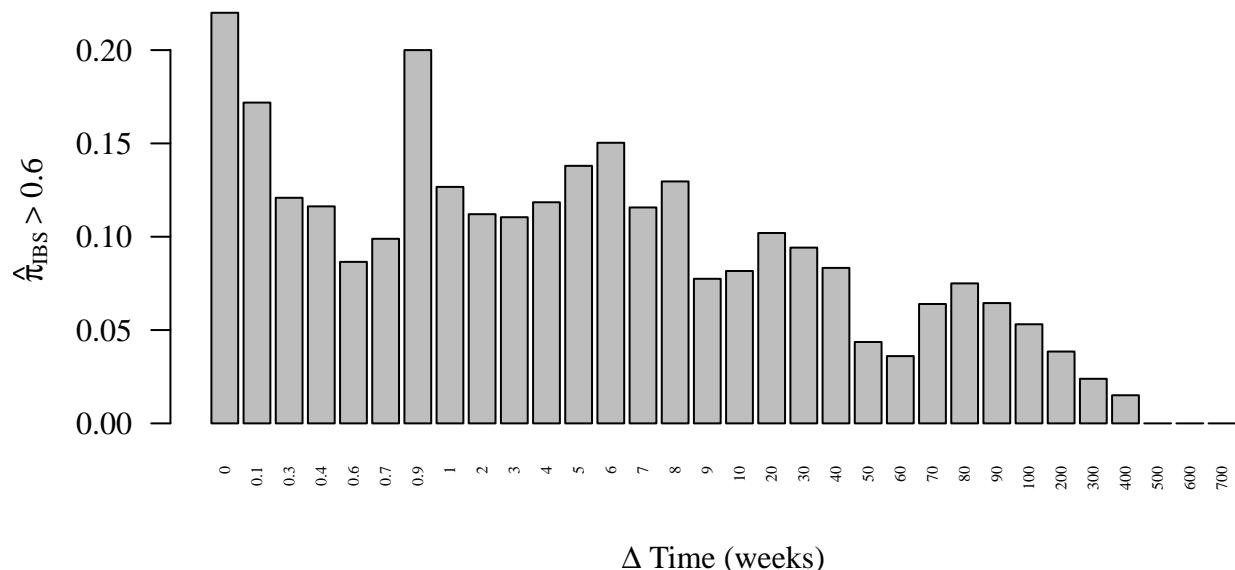
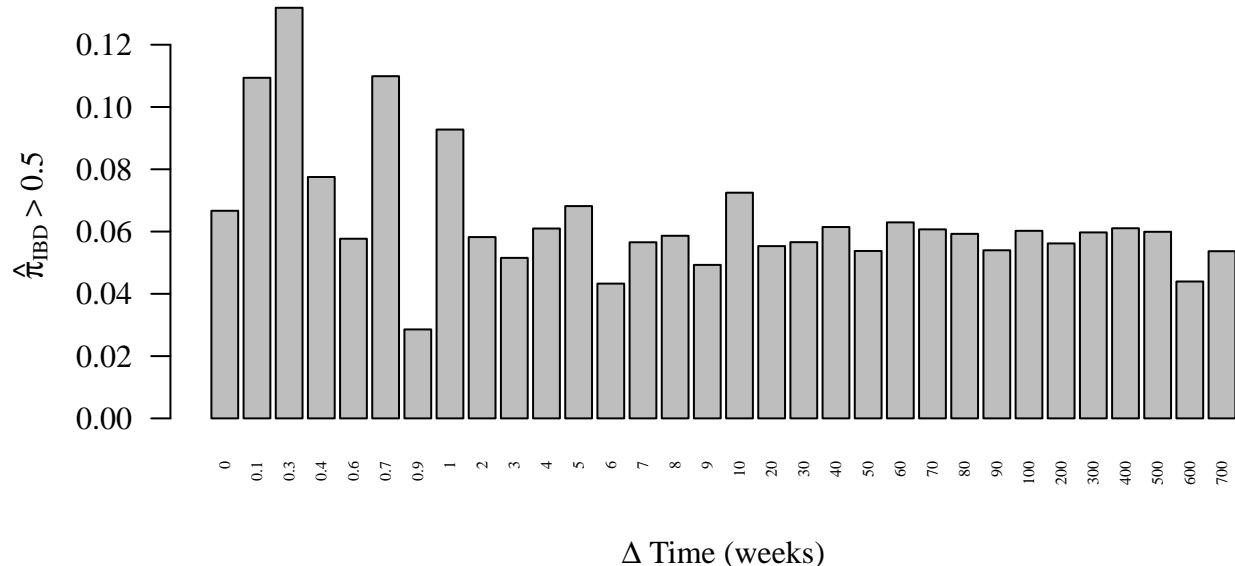


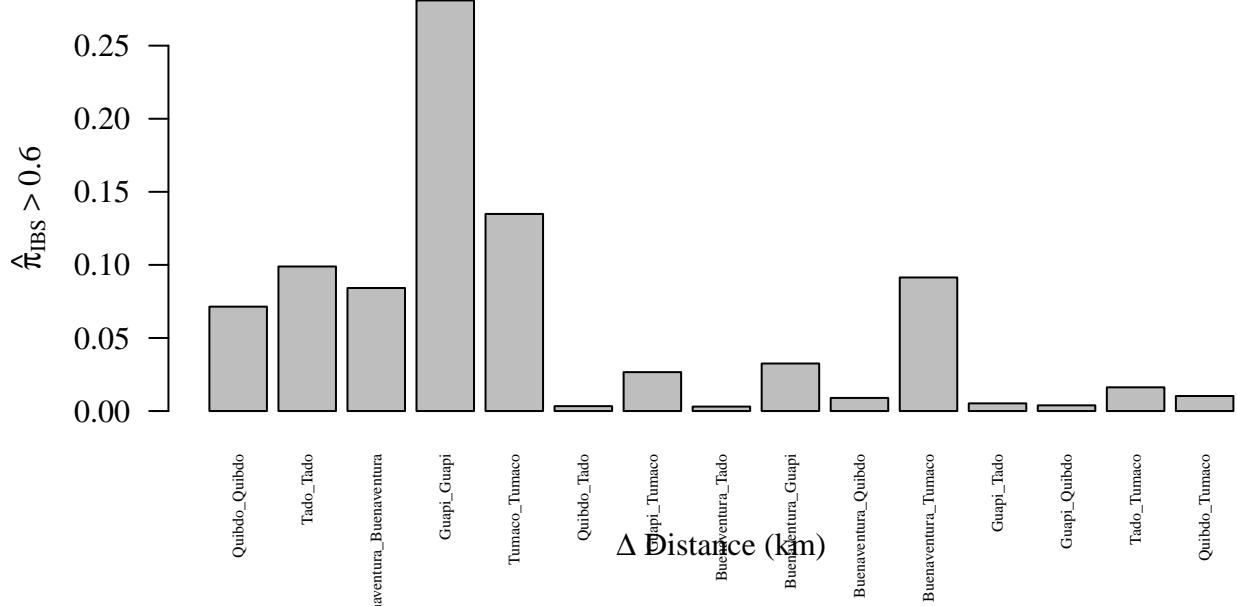
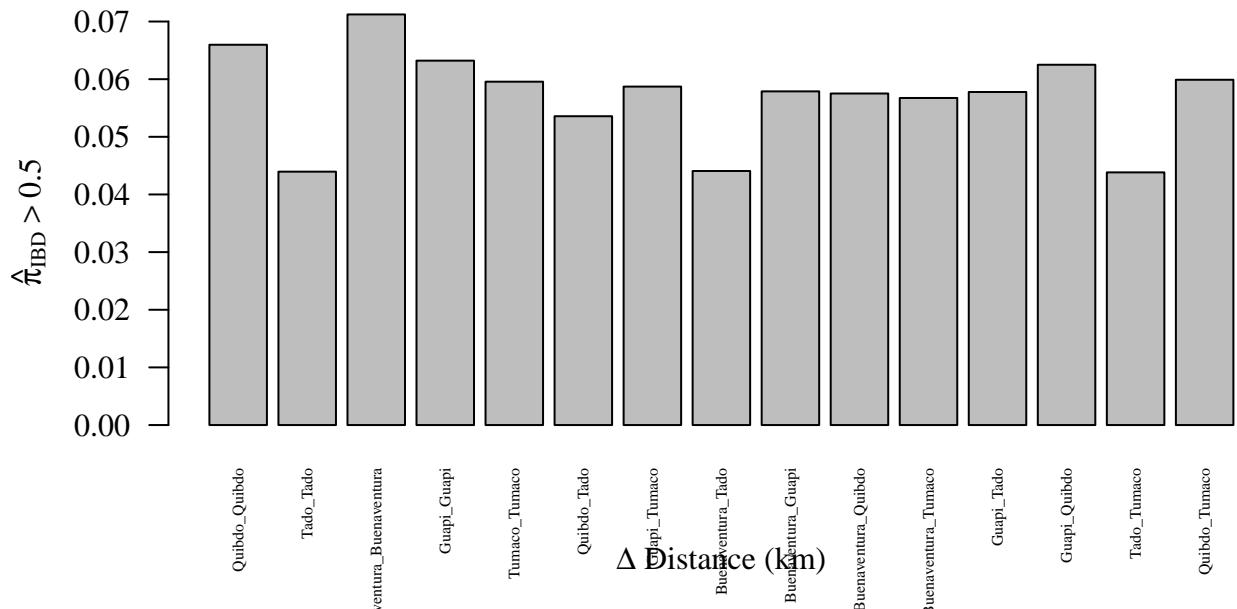
Plot of IBD and IBS against distance

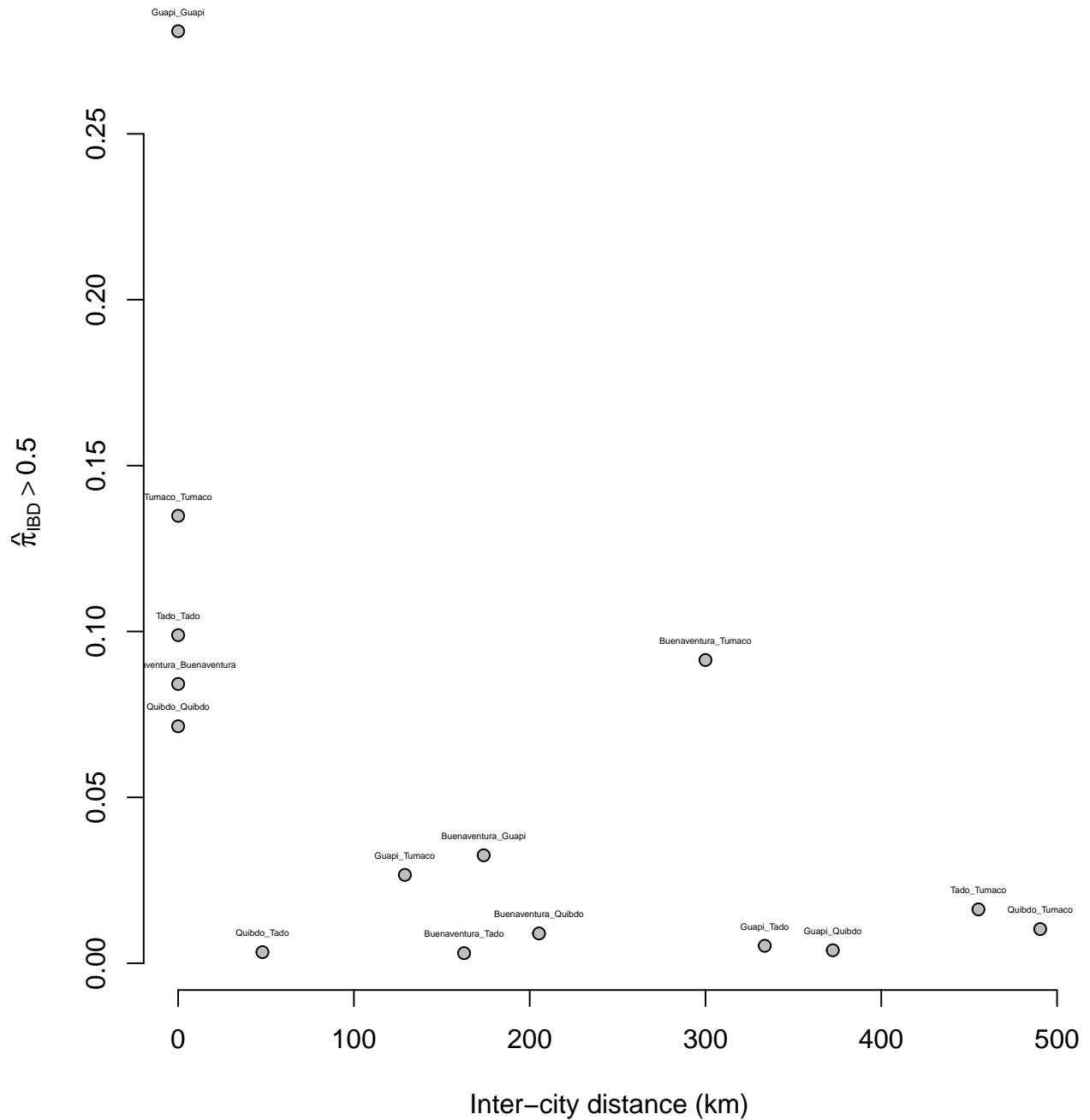


Plot of IBD and IBS against distance









Regression results