

Gloria-Soria (ddRAD58)

Gus Dunn

2014-12-29

Contents

1	Linkage disequilibrium	1
---	------------------------	---

1 Linkage disequilibrium

The mean bin-wise LD decays with physical distance to 0.2 (half of maximum value) near 708 bp and to the value of the 95th percentile value ($r^2 = 0.12$) near 10,570 bp of SNP-pair separation (Figure `ld_decay_by_physical_distance`). We identified SNP-pairs with outlier LD values within specific separation-bins using a custom MAP-based filter. From a total of 6,454,294 SNP-pairs considered, 24,372 (0.38%) were assigned corrected p-values ≤ 0.01 . These filtered SNP-pairs were compared against top SNPs from the selection analyses (supplemental table `ld_filtered_snps_interest_JOINED_ANNOS_grouped.xls`).

One hundred twenty five genes (52 with functional annotations) were identified within 1000 bp of at least one SNP identified in `Top10_InfectionOverall` (41 genes, 18 with annotations for `Top05_InfectionOverall`). Among the most frequently occurring annotations for these genes in the biological process (BP) domain were transmembrane transport; transcription, DNA-templated; ion transport; and metabolic process. In the molecular function (MF) domain, some of the most frequently occurring terms were zinc ion binding; DNA binding; transferase activity; and oxidoreductase activity. `Selected_PopPairwiseMSNB_Environm` had no overlap with the LD filtered SNP-pairs. `Top10_PopPairwiseOverlap_Infection` and `Top10_PopPairwiseOverlap_Environm` both contained only the same two SNPs in common with the LD filtered set. This SNP-pair ($r^2 = 0.9456$, adjusted p-value = 0.0017, genomic coordinates = Scaffold150:138,509;404,077) is separated by 265,568 bp, and each SNP is located near annotated genes. Scaffold150:138,509 is within GFUI009292, an ortholog of huckebein; Scaffold150:404,077 is 718 bp from GFUI009351, a predicted gene with no recognized protein domains but an ortholog predicted in *Glossina palpalis*. Sixteen genes were located within 1000 bp of at least one of the 19 SNPs in `Selected_PopPairwiseMSOT_Environm` that were also in the LD filtered set. Six of these had functional annotations assigned by Argot2 including small GTPase mediated signal transduction, glycogen biosynthetic process, carbohydrate metabolic process, and dorsal/ventral axis specification in the BP domain and nucleotide binding, hydrolase activity, transferase activity, and helicase activity in the MF domain.

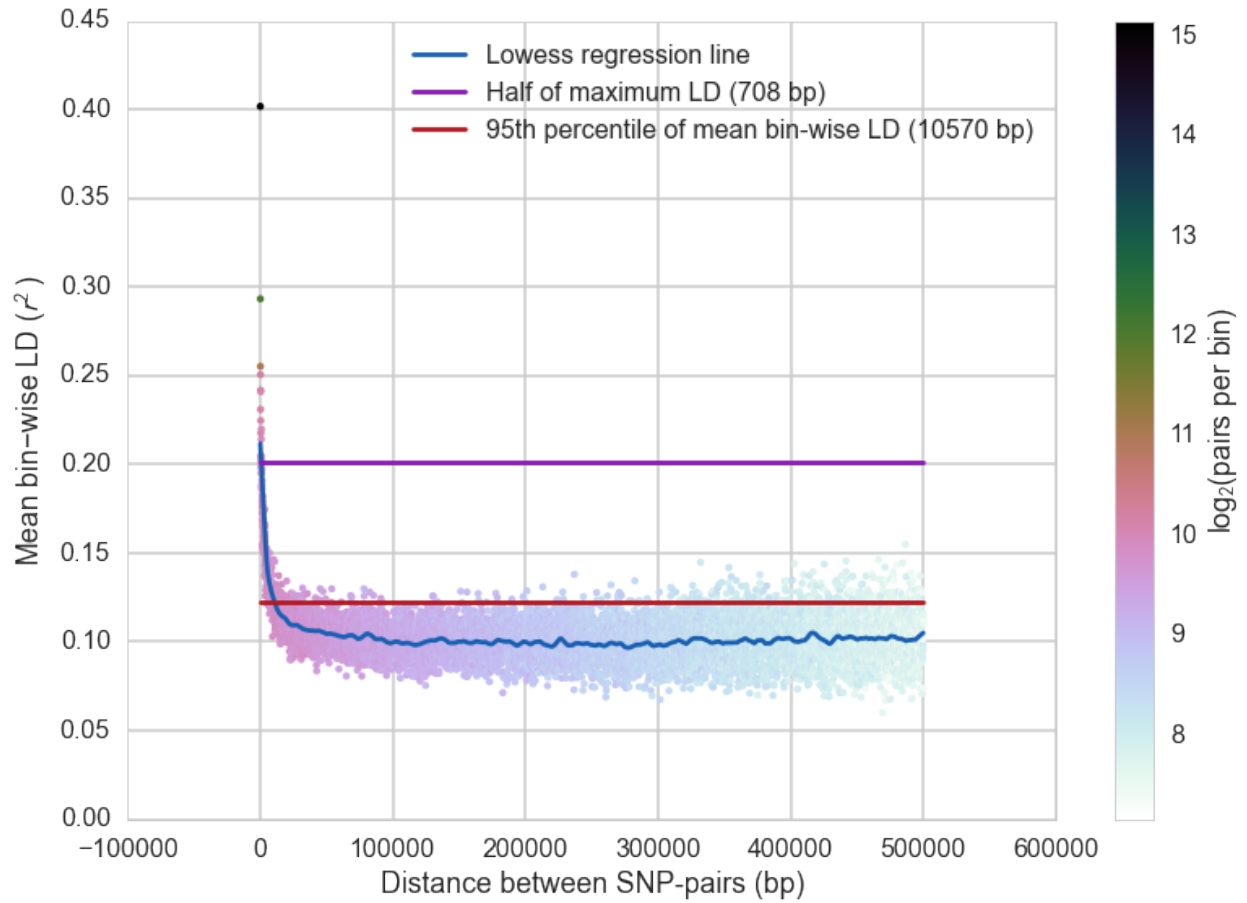


Figure 1: Linkage disequilibrium decay with physical distance. Each “dot” represents the mean LD for that set of binned SNP-pairs. The color of the dot illustrates the number of SNP-pairs contributing to the mean. The blue colored line is a lowess regression line of best fit. The purple colored line marks the LD value of one half the maximum while the brick colored line is the LD value at which 95% of the bins have mean LD values below this value.