I’ve been looking into the genomic distribution of SNPs that changed dramatically in allele frequency in the four northern sites that had major plant death between 2018 and 2019 - the “freezing” sites.

**I thought genome regions enriched for SNPs that changed in allele frequency at the freezing sites could contain good candidates for freezing tolerance genes.** These methods should be complementary to GWAS on winter survival, and (I think) agnostic to population structure. Think of them like pilot data for a field selection experiment.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Plants with SNP data & phenotypes | | Plants with phenotypes | | Plants Dead |
|  | **2018** | **2019** | **2018** | **2019** | **2018 to 2019** |
| **BRKG** | 408 | 254 | 412 | 254 | 154 |
| **KBSM** | 539 | 517 | 544 | 517 | 22 |
| **FRMI** | 272 | 185 | 273 | 185 | 87 |
| **LINC** | 433 | 262 | 437 | 262 | 171 |
| **CLMB** | 585 | 496 | 591 | 496 | 89 |
| **STIL** | 465 | 451 | 469 | 451 | 14 |
| **OVTN** | 343 | 332 | 346 | 332 | 11 |
| **TMPL** | 384 | 373 | 388 | 373 | 11 |
| **PKLE** | 617 | 615 | 623 | 615 | 2 |
| **KING** | 303 | 294 | 305 | 294 | 9 |

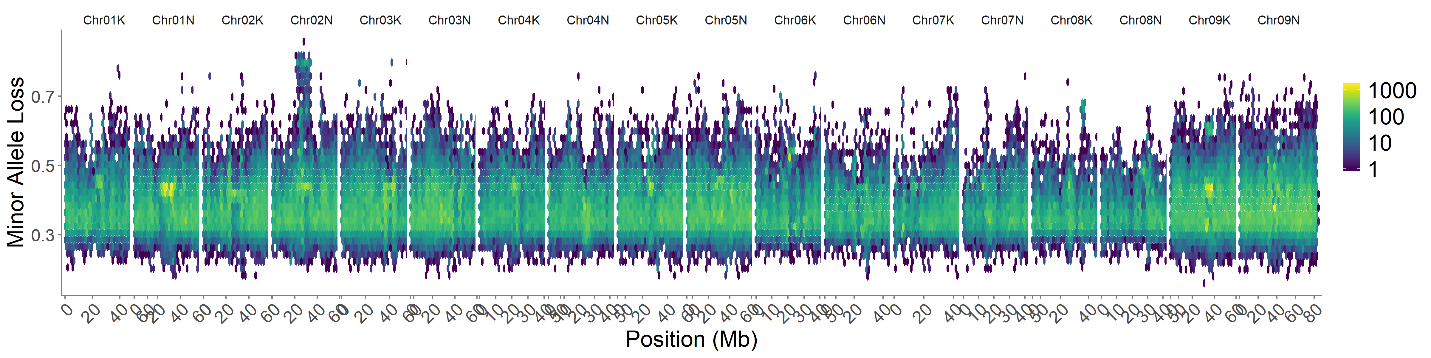
As a reminder, here is the breakdown of the number of individuals we have for GWAS at each site, for phenotypes from 2018 and from 2019. The last column shows the number of plants dead between 2018 and 2019, and the four sites with lots of death in winter 2018 – the “freezing” sites – are highlighted.

I took the 1% tail of SNPs that had changed the most in minor allele count for two cases:

1. The minor allele count had changed the most at one or more freezing site;
2. The allele had been ‘lost’ for GWAS purposes for 2019 data, i.e. the minor allele count had started above 25 at all sites, was in the 1% tail of allele count changes, and had dropped below a minor allele count of 25 at one or more freezing site.

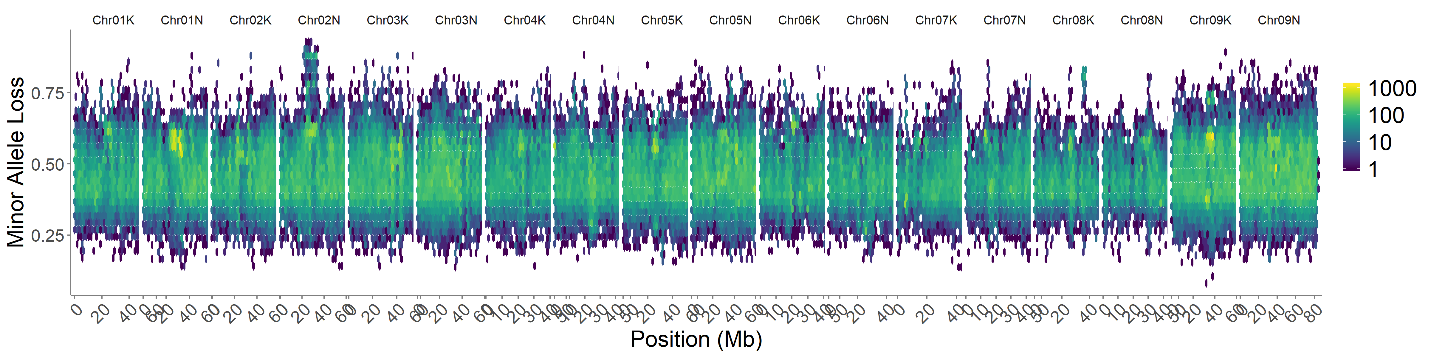
*{I used un-imputed SNP minor allele counts from John’s new 33.9M SNP dataset. Instead of a minor allele frequency, I used a minor allele count of 25 or more (except for FRMI where I used 15). This was the same as a MAF of 5% at BRKG, FRMI, and LINC, and a MAF of 2.5% for CLMB.}*

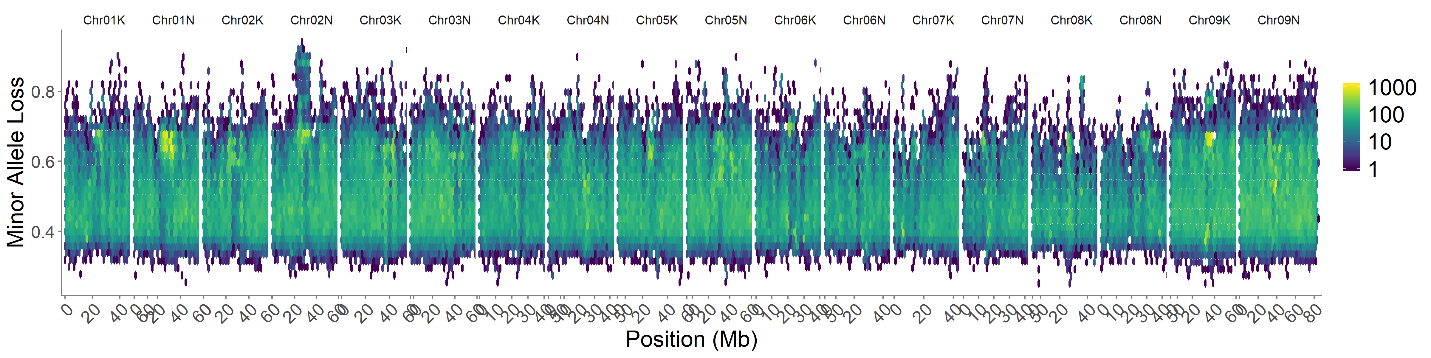
I then plotted the distribution of these SNPs across the genome. The y-axis is the fraction of the minor allele lost, here on average across the four sites:

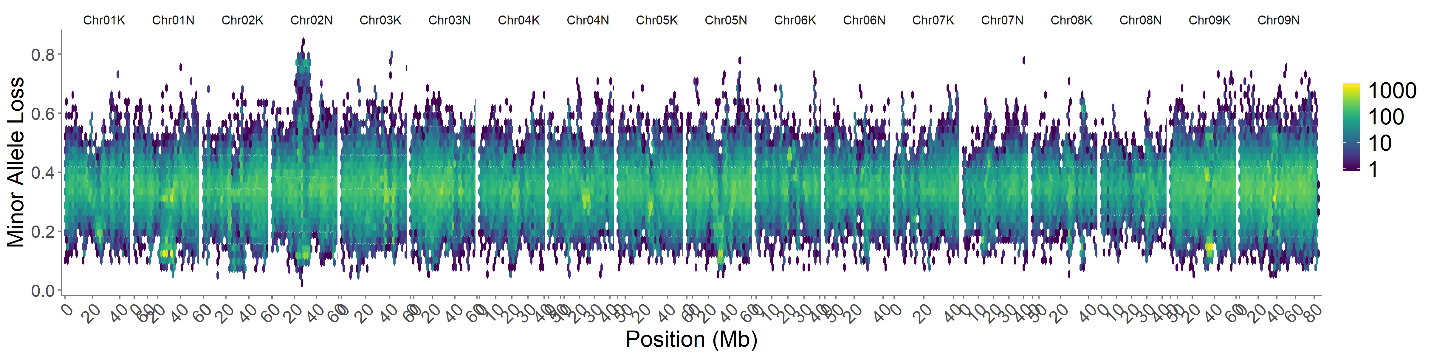
Four Site Average, Analysis (1)

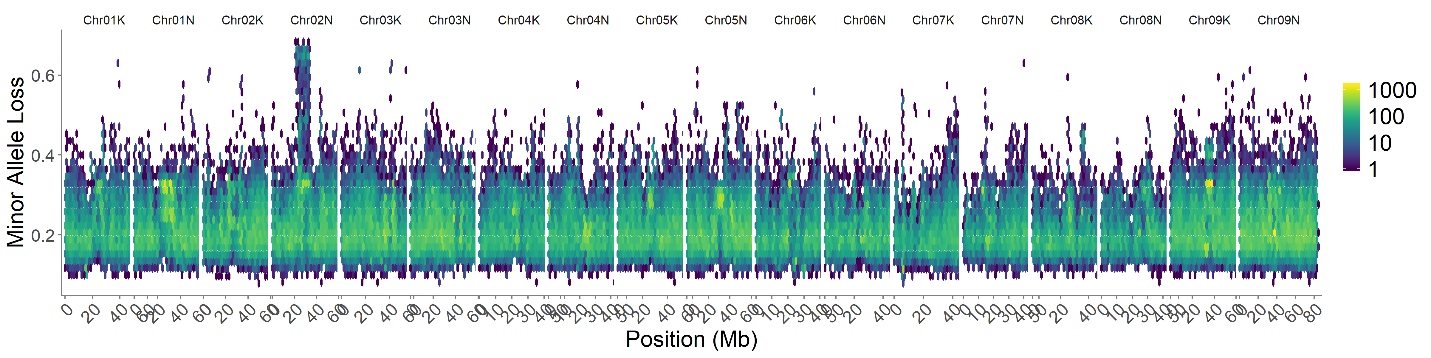
I’m guessing that one type of centromere region of Chr02N has been lost at high frequency at freezing sites, as well as some regions on Chr07K, Chr07N, and most of one type of Chr09K and Chr09N.

If you look at allele frequency changes where the allele had changed the most just at that site (Analysis (1)), for the four sites, you see quite similar patterns. BRKG and LINC look the most similar, which perhaps makes sense as they had the most plant death.

BRKG Analysis (1):

LINC Analysis (1):

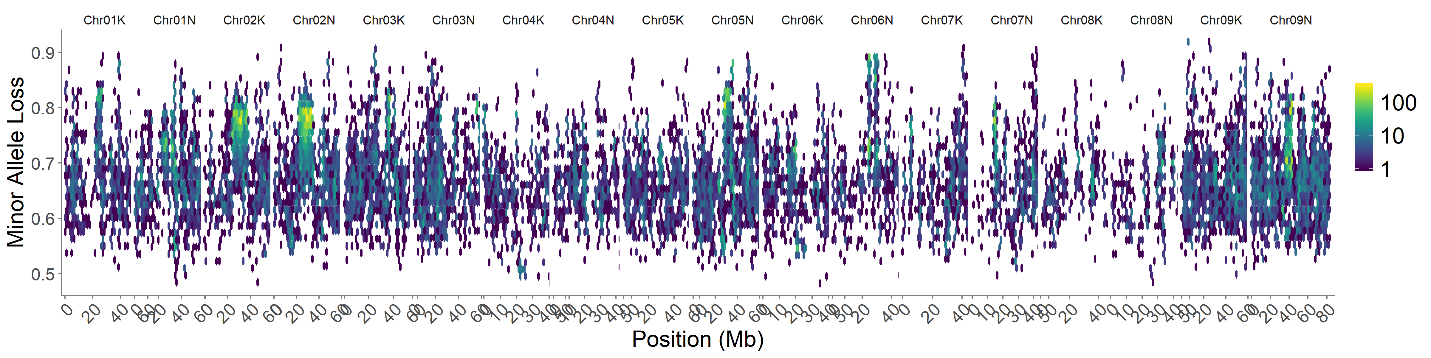
FRMI Analysis (1):

CLMB Analysis (1):

The Chr02N signal comes up at all four sites. There are a large number of repeated signals.

Here’s the change in allele frequency for each site for just the alleles “lost” for GWAS purposes. We’re probably most concerned about these SNPs for GWAS purposes - SNPs with a count that drops below 25 in the 2019 data.

Four Site Average, Analysis (2):



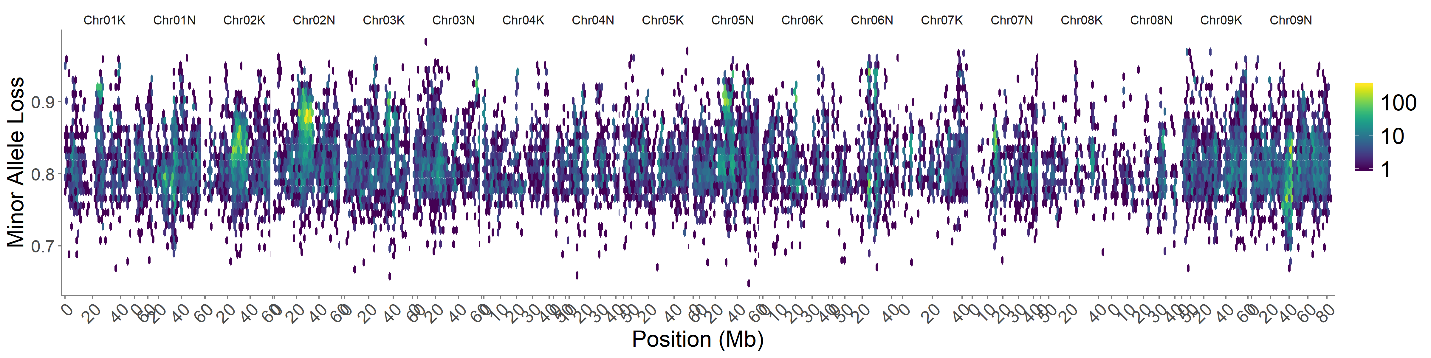
Here it looks like (I’m guessing) the centromeric region on Chr02K, Chr02N, Chr05N, Chr06N, and Chr09N are repeatedly lost for one chromosome type.

Of the 2.3M SNPs that were lost for GWAS purposes at one or more of the freezing sites, most (8%) are on Chr09N, Chr02N (7%), Chr05N (6.7%), and Chr09K (6.5%), twice as many as the least (Chr08K, 4%).

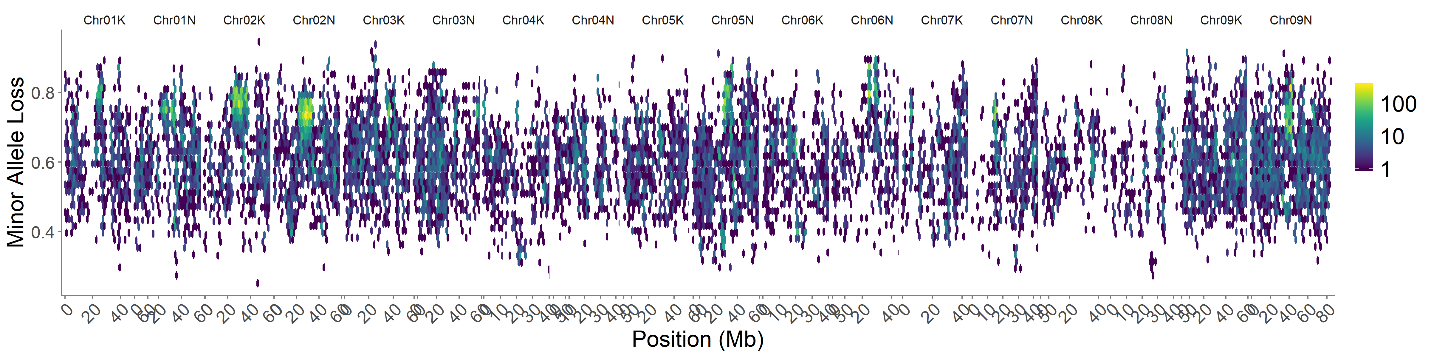
BRKG Analysis (2):



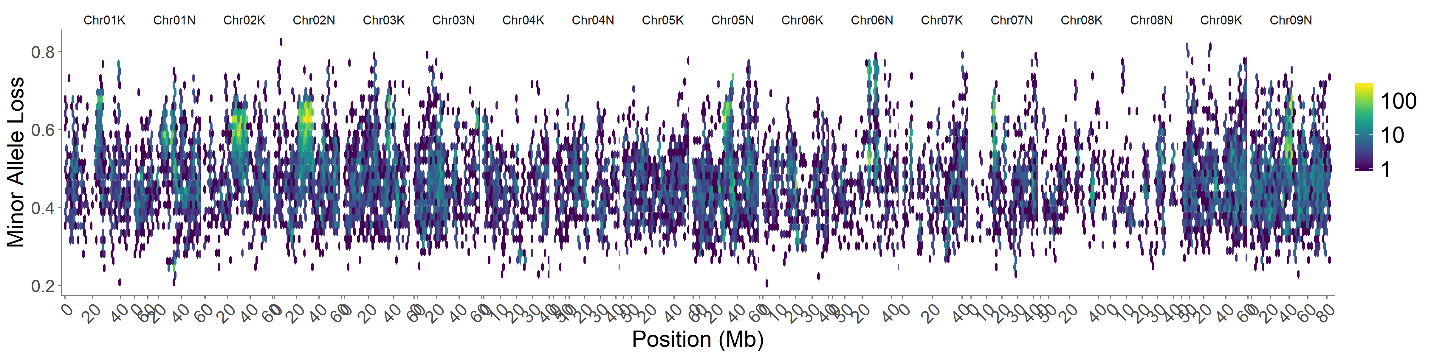
LINC Analysis (2):



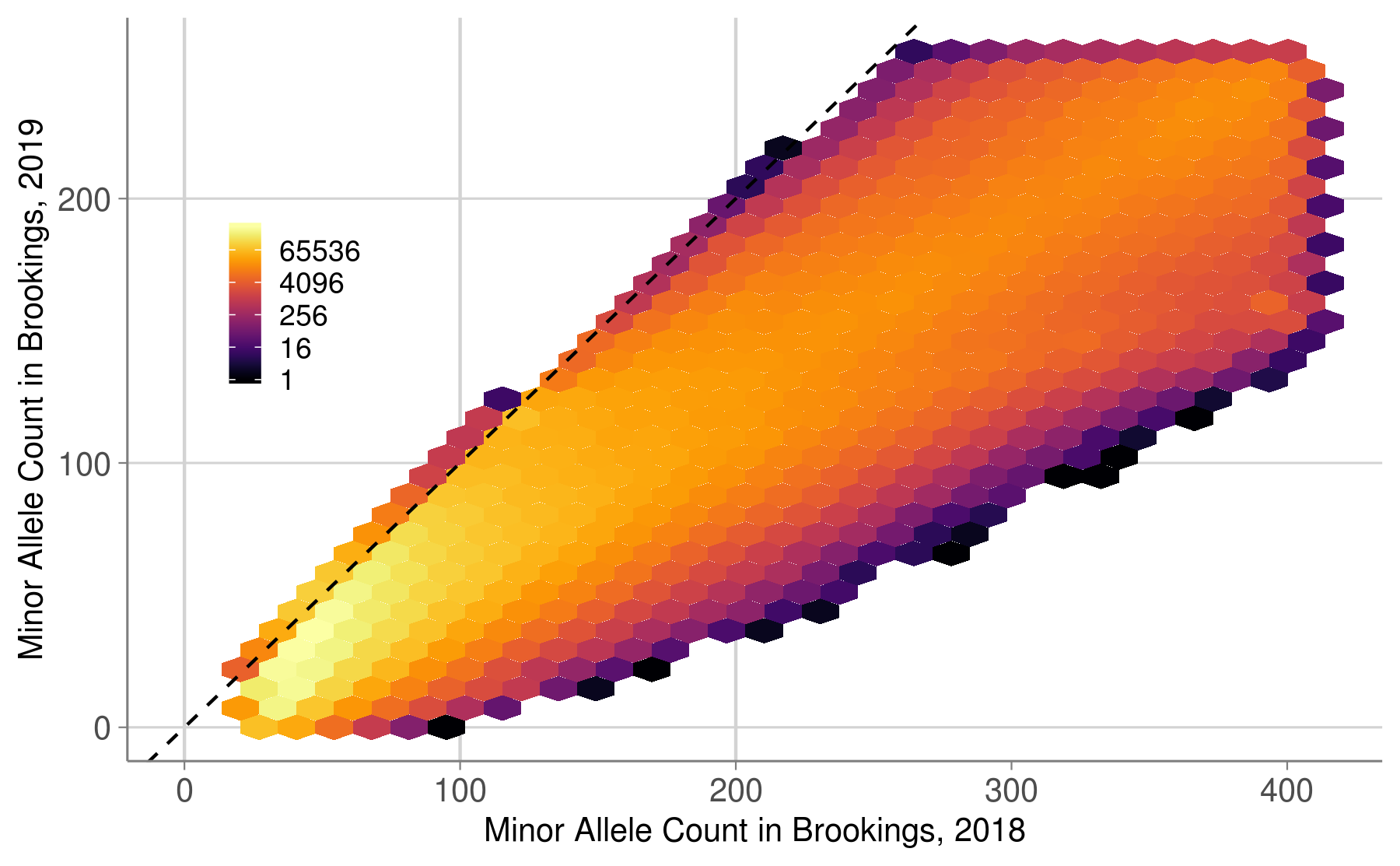
FRMI Analysis (2):

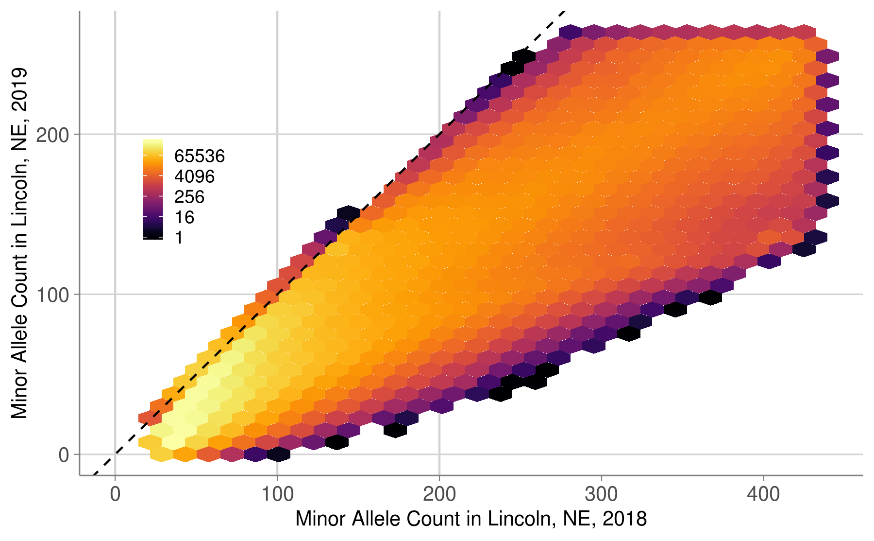


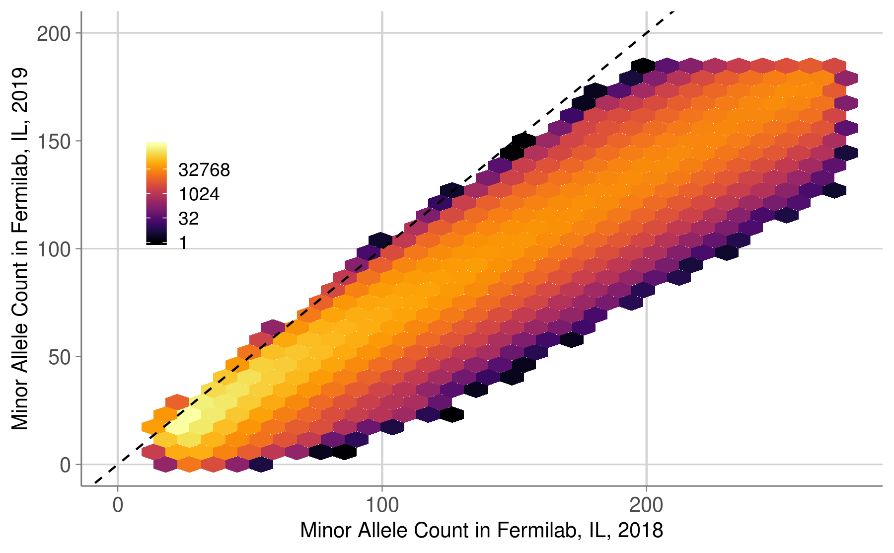
CLMB Analysis (2):



Just as a reminder, I already looked at how allele frequency has shifted at northern sites between 2018 and 2019 (this also lets us compare allele frequencies between sites – think of them like SFS plots). Here are some examples of those plots for the four sites that lost 80 or more individuals. Note that I removed all SNPs with a minor allele count below 25 in the 2018 data, because we would not be able to use those SNPs for GWAS. We’re probably most concerned about SNPs with a count that drops below 25 in the 2019 data. And 2018 is always on the x-axis for these year to year comparison plots.







Note that I removed all SNPs with a minor allele count below 25 in the 2018 data, because we would not be able to use those SNPs for GWAS.

