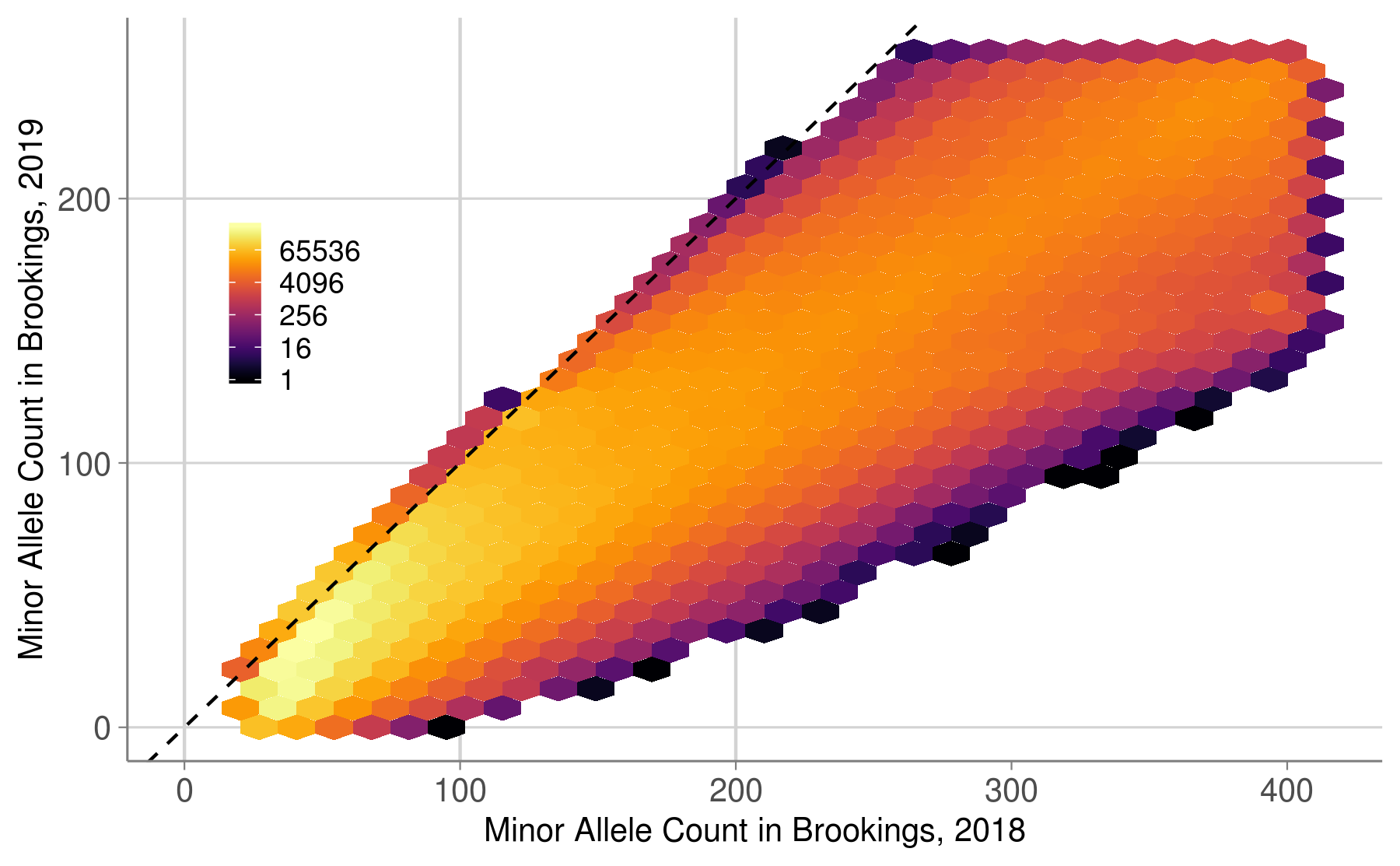
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Plants with SNP data & phenotypes | | Plants with phenotypes | | Plants Dead |
|  | **2018** | **2019** | **2018** | **2019** | **2018 to 2019** |
| **BRKG** | 408 | 254 | 412 | 254 | 154 |
| **KBSM** | 539 | 517 | 544 | 517 | 22 |
| **FRMI** | 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 |

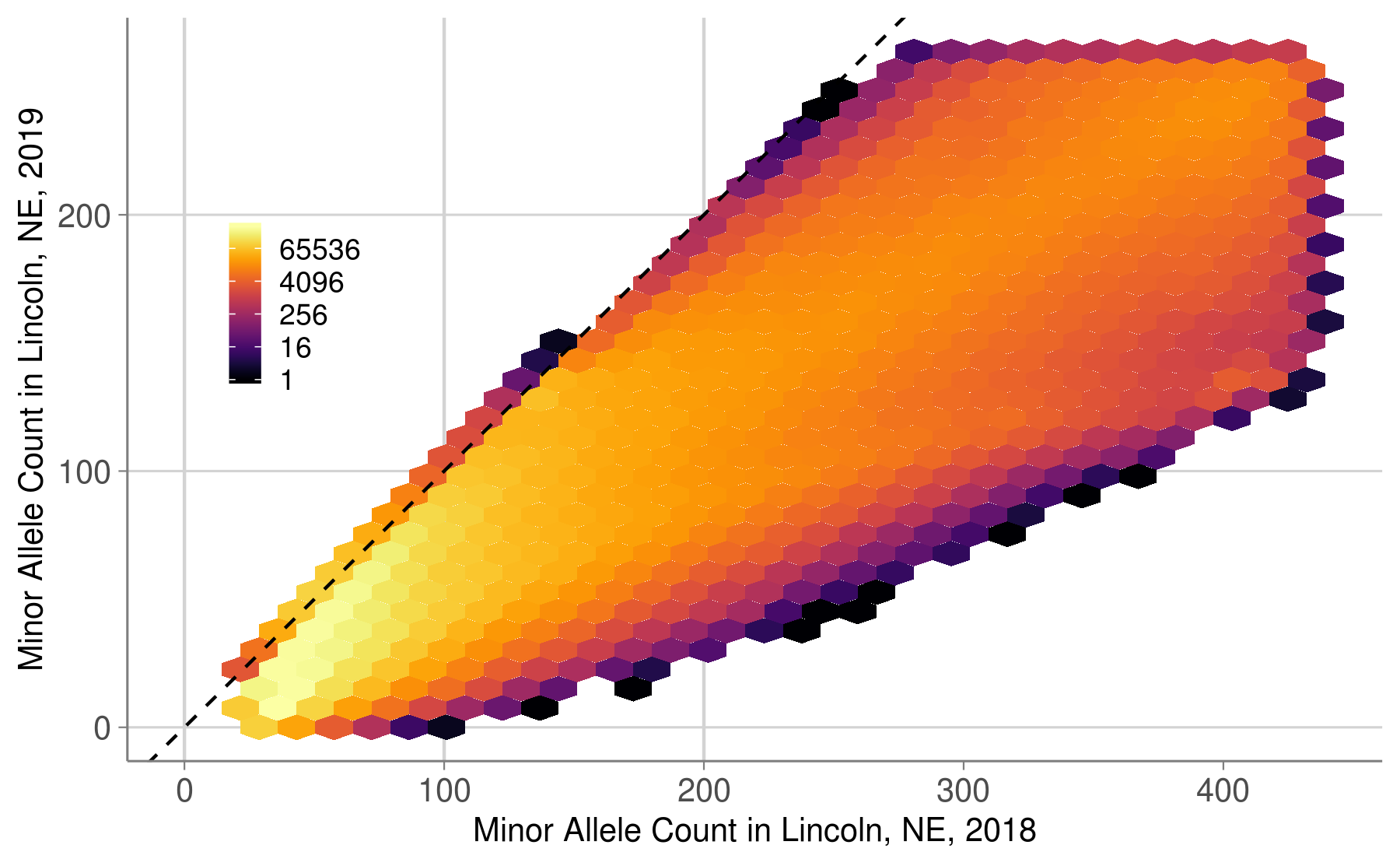
Here is a breakdown of the number of individuals we have for GWAS at each site, for phenotypes from 2018 and from 2019. The first two data columns have individuals with high quality SNP data and phenotypes. The next two have tetraploid plants with phenotypes – there is a lot of overlap, and only 5-6 plants that could have their libraries resequenced to help the GWAS data.

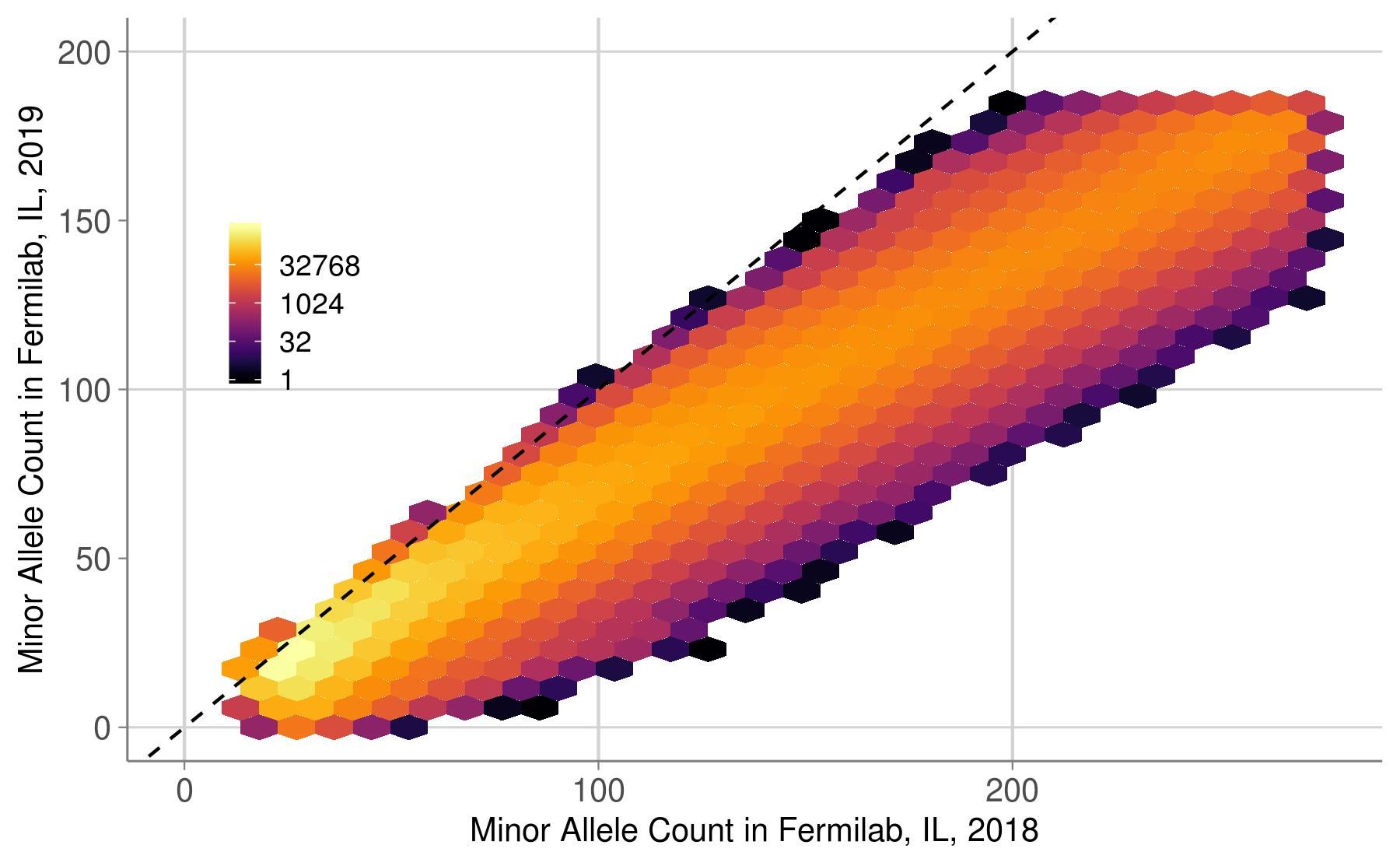
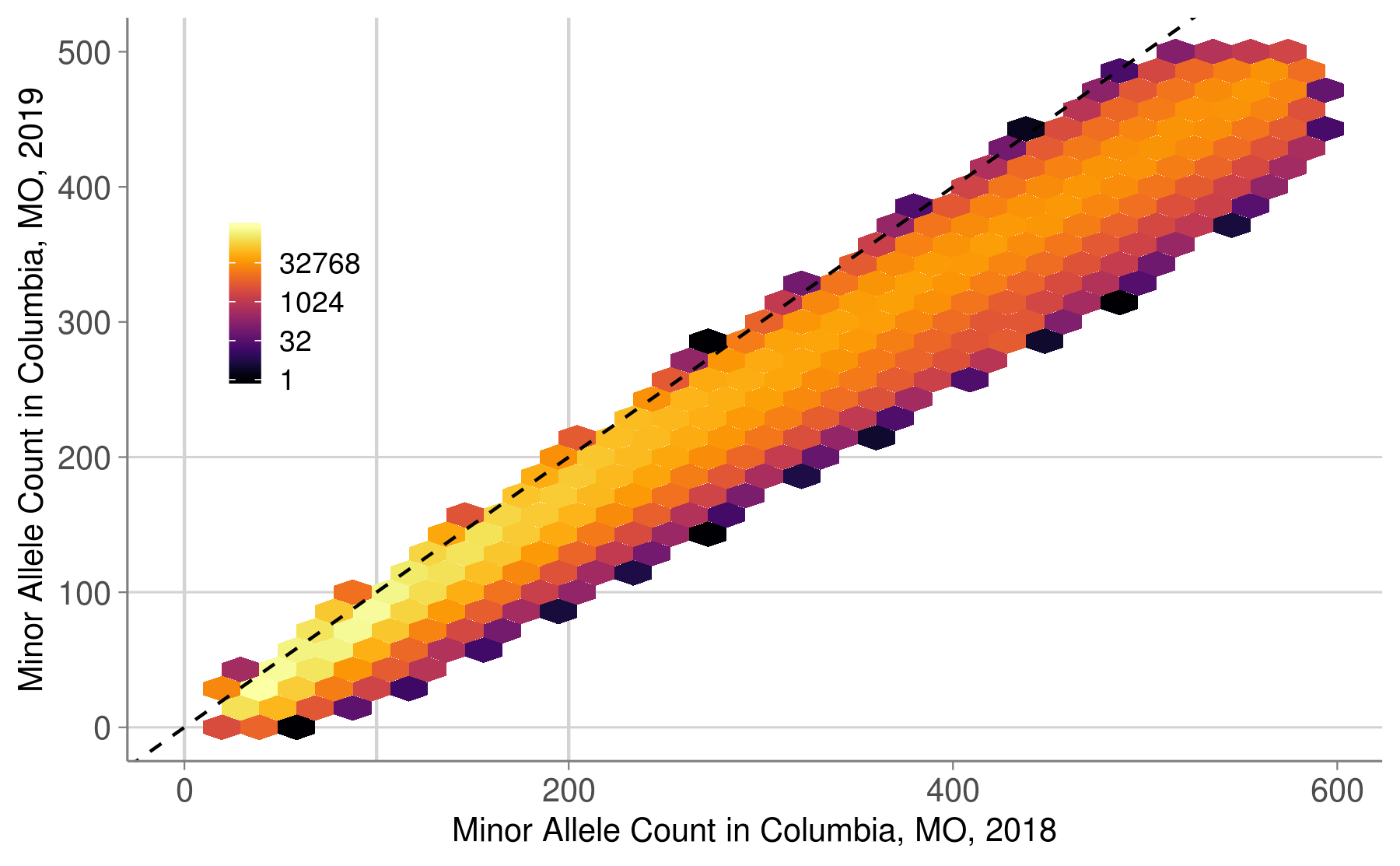
|  |  |
| --- | --- |
| Number of SNPs | Category |
| 12536071 | Can't be used for GWAS – these SNPs have less than 25 copies of the minor allele in the 631 individuals planted at the common gardens. |
| 10452584 | Can be used for GWAS at all 10 sites and both 2018 and 2019 data. |
| 1160326 | Can’t be used in LINC & BRKG in 2019 |
| 1155906 | Only at high enough frequency to be used at the big 3 sites |
| 546381 | Can only be used at PKLE & CLMB |
| 439918 | Can only be used at PKLE |
| 346189 | Can only be used at big 4 sites – KBSM CLMB PKLE STIL |
| 341773 | Can’t be used at KING OVTN FRMI |
| 331082 | Can’t be used at LINC in 2019 |
| 327364 | Can’t be used at KING |
| 315522 | Can only be used at big 4 and LINC 2018 |
| 305573 | Missing 2019 at 4 freezing sites – LINC FRMI CLMB BRKG |
| 285838 | Can only be used at PKLE x2yr & CLMB 2018 |
| 246340 | Missing 2019 at 3 freezing sites, not big site (BRKG) |
| 231650 | Missing KING x2yr and FRMI 2019 |
| 192819 | Missing KING x2yr and FRMI x2yr |
| …… | 4.67M SNPs have some other combination of Site x Year that is useable. |

Given that information, I then subdivided the SNPs from John’s new 33.9M SNP dataset. These are SNPs missing in 20% or less of individuals from each subpopulation. 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 slightly lower MAF for the other sites with more individuals. Here’s the breakdown of the site\*year combinations where SNPs can and can’t be used for GWAS:

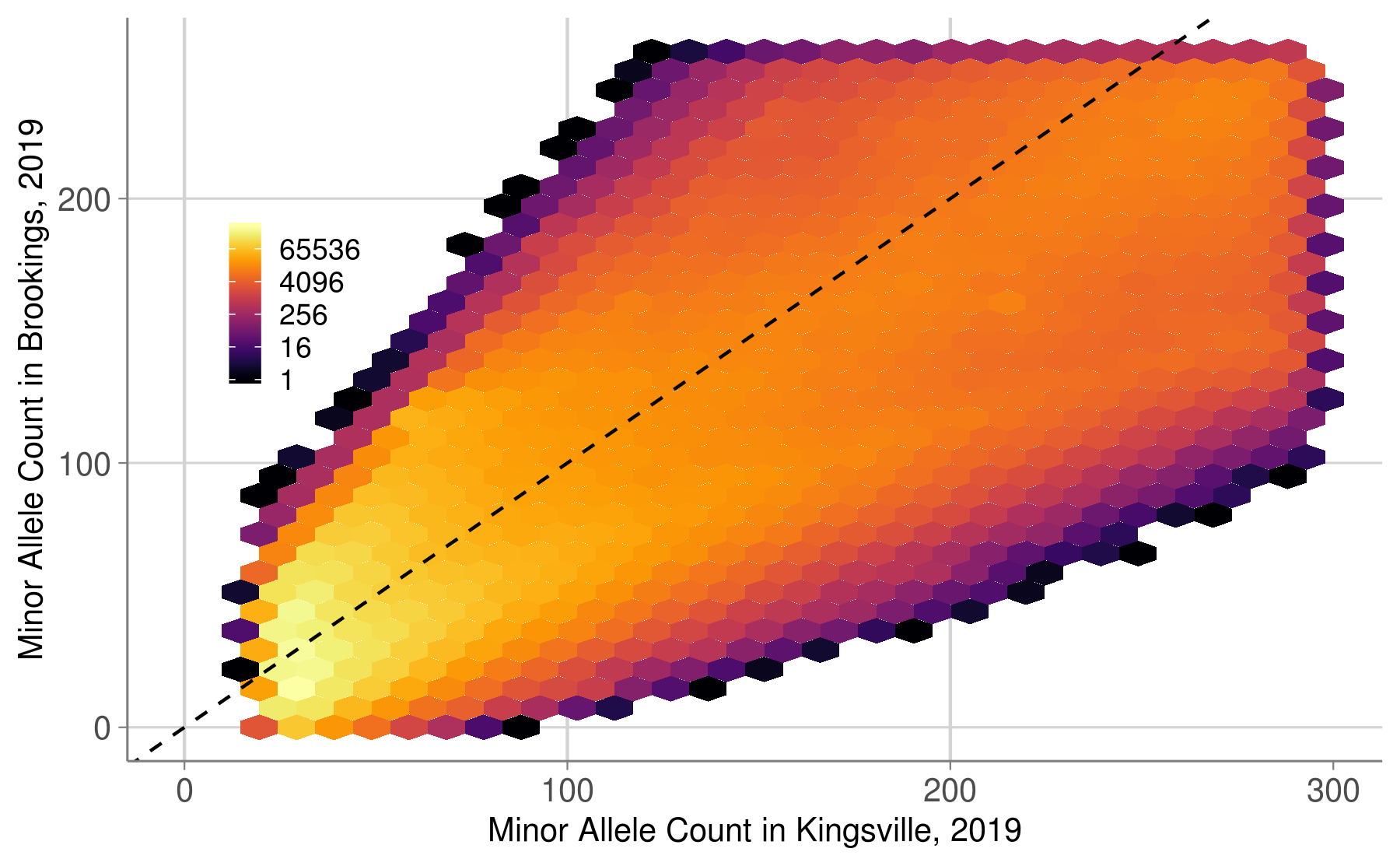
The minor allele count also lets us look at how allele frequency has shifted at northern sites between 2018 and 2019 (and also lets us compare allele frequencies between sites – think of them like SFS plots). Here are some examples of those plots for the four sites that lost 80 or more individuals. Note 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.

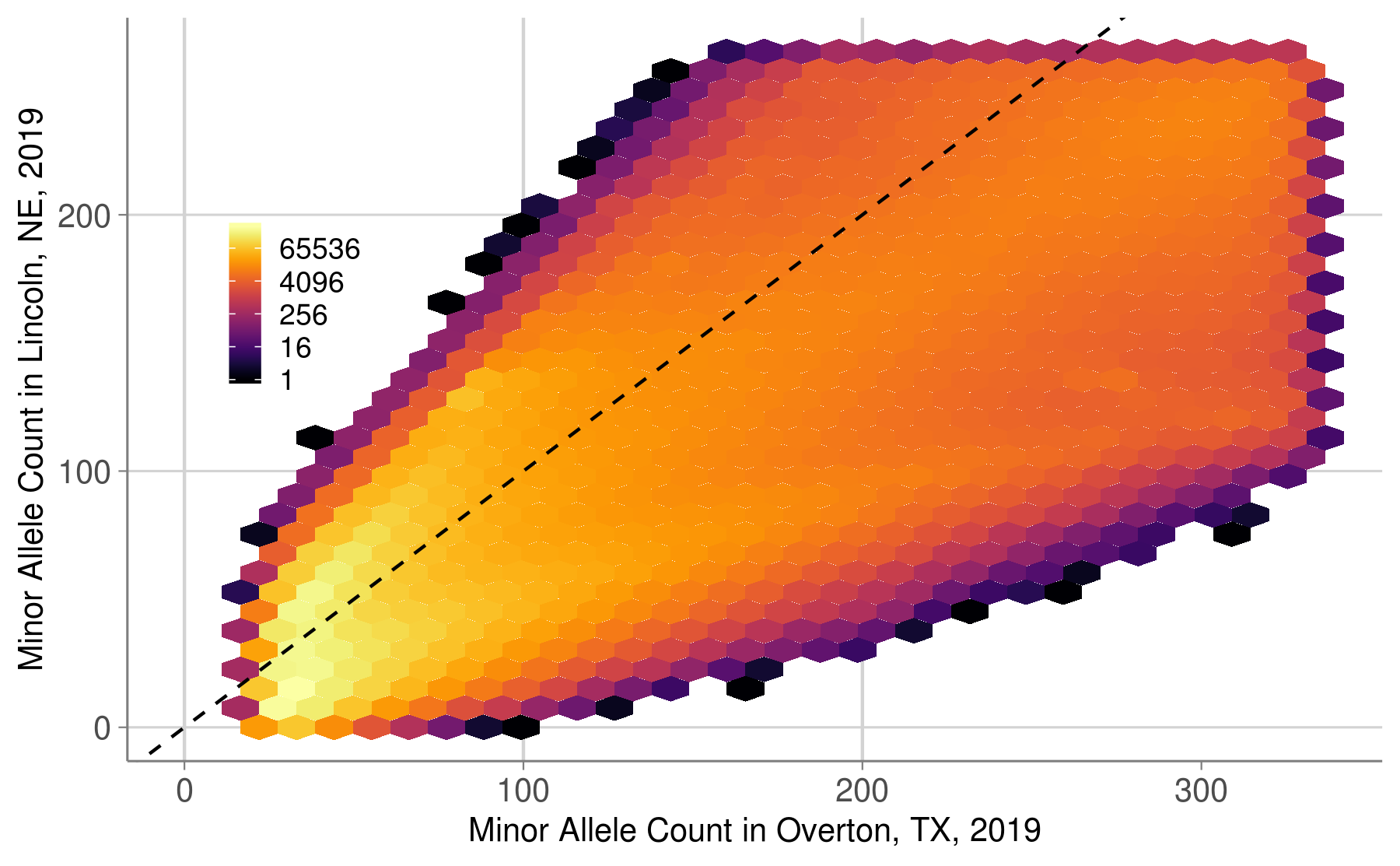




Note that I removed all SNPs with a minor allele count below 25 in the 2018 data, because we would not be able to use those SNPs for GWAS.

And here are a few examples of plots between sites, for 2019 data. 2018 data looks much more similar.





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.

