

CIMMYT2013-2019

Monday, April 8, 2019

2:30 PM

- The project IDs from 2019: BW18GSBL-B04, BW18GSBL-B05 and BW18PVBL-B01 (GBS1539B)
- 1. The larger CIMMYT database from 53,525 lines (year 2013-2018) was used for SNP production.
- 2. The data from 2019:
 - 19 libraries: 3610 lines
 - 18 libraries: 3064 lines
 - 18 libraries: 2628 lines
- 3. The four VCF files were combined together: $53,525 + 9,302 = 62,827$
- 4. There were 38 redundant GIDs between 53,525 and 2019 lines (old one suffixed with "_1")
 - 2 GID16122
 - 2 GID2430154
 - 2 GID2465
 - 2 GID3822784
 - 2 GID3895
 - 2 GID41948
 - 2 GID4754390
 - 2 GID4755014
 - 2 GID4799764
 - 2 GID4902859
 - 2 GID4905617
 - 2 GID5398434
 - 2 GID6175067
 - 2 GID6332122
 - 2 GID6341870
 - 2 GID6387
 - 2 GID6680725
 - 2 GID6933502

2 GID6937850
2 GID7175970
2 GID7396039
2 GID7396550
2 GID7400914
2 GID7626085
2 GID7626320
2 GID7626414
2 GID7627366
2 GID7627435
2 GID7627560
2 GID7628468
2 GID7630047
2 GID7632196
2 GID7632217
2 GID7632219
2 GID7632233
2 GID7634384
2 GID775
2 GID7806808

5. Unfiltered data:
 - CIMMYT1319.vcf
 - CIMMYT1319.hmp.txt
6. Three different filtering methods were tested at each SNP locus.
SNPs passing at least one filter were recovered:
 - a. Inbred coefficient: $> 80\%$
 - b. Fisher Exact test ($p < 0.001$)
 - c. Chi Square (expected inbreeding of 96%)
 - Total SNPs: 91032
 - Total SNPs passing filters:
 - 14615 Chi2
 - 85281 Fisher
 - 74130 Inbred
 - SNPs passing one, two, or three filters:
 - 17388 1
 -

60787 2

11688 3

SNPs passing at least one filter: 89,863

7. Filtered:

- CIMMYT1319-SNPs-1filter-passed.vcf
- CIMMYT1319-SNPs-1filter-passed.hmp.txt

8. SNPs with >70% missing removed: remaining 38,361 SNPs.

9. Imputation was done with Beagle version 4.1:

- beagle_gt70_CIMMYT1319-SNPs-1filter-passed.vcf
- beagle_gt70_CIMMYT1319-SNPs-1filter-passed.hmp.txt