**Wheat Imputation PHG accuracy**

**To-do**

1. build interface to retrieve VCF from PHG (maybe BrAPI in future)

2. build feature to retrieve markers independent of project

**PHG Creation**

The wheat PHG was created in Eduard Akhunov's lab by Katie Jordan using 65 US wheat cultivars. The accessions were exome sequenced to give 1.39M markers. The FASTQ files were converted to gVCF then loaded into the PHG. When the Akhunov lab tested imputation accuracy they calculated 95% accuracy when using the same genotype protocol. The imputation accuracy when using a reduced coverage or GBS protocol gave an imputation accuracy of 85-90%.

<https://doi.org/10.1093/g3journal/jkab390>

**PHG Imputation**

The data in T3/Wheat was imputed using the VCF files for each genotype project. The procedure is described in step 3 of the [PHG Wiki User instructions](https://bitbucket.org/bucklerlab/practicalhaplotypegraph/wiki/UserInstructions/ImputeWithPHG_main.md)

**WheatCAP Goals**

1. Convert low density genotype data into a single high-density protocol

2. Convert historic data, including other protocols, into a single high-density protocol

**Accuracy testing**

Impute Illumina 90K using ImputePipelinePlugin with the following settings.

PHG software version 37  
/tassel-t-standalone/run\_pipeline.pl -ImputePipelinePlugin -imputeTarget pathToVCF -endPlugin

configParameters  
minReads=0.

Illumina 90K genotype protocol (34K markers, 1630 accessions)

Exome capture 2019 HapMap genotype protocol (7.4M markers, 329 accessions)

16 accessions

26 accessions  
1600 markers

130 accessions  
8538 markers

1. **Accuracy for different protocols (26 accessions in PHG)**

Compared genotype calls for overlapping markers between (9K and 2019 HapMap) and (90K and 2019 HapMap).

|  |  |  |
| --- | --- | --- |
| Protocol | Accuracy Raw | Accuracy imputed |
| 9K | 72% | 91% |
| 90K | 75% | 95% |

1. Accuracy for accessions in PHG vs not in PHG  
   Compared 90K imputation for 26 accessions in PHG vs 329 accession not in PHG

|  |  |  |
| --- | --- | --- |
| protocol | Accuracy in PHG | Accuracy not in PHG |
| 90K | 95% | 74% |

1. Accuracy for markers in PHG (minRead=1) vs not in PHG (minRead=0)

|  |  |  |
| --- | --- | --- |
| Protocol | Accuracy minRead=1 | Accuracy minRead=0 |
| 90K | 94% | 95% |