

GOBII Seminar 3/13/2017

haplotype graphs

machine learning

cost of sequencing dropping

divergence of two lines 1% (less in wheat?)

throw out most (95%) of sequence used for genotyping (only include genes/exom?)

requires known parents

crossovers or meiosis ~28 in maize

$15,000/28 = 536$ variants/crossover

error rate 1%, 5 errors/crossover

score sites (requires prior species knowledge) or score haplotypes (new bioinformatics)

\$1M to develop cross species pipeline

practical haplotype graph

why do we work with SNPs (phasing can be hard)

pangenome can be accurately represented as a graph

Practical haplotype graph replaces pipeline in GBS, rAmpSeq, and low coverage

GOBII B4R TASSEL

Graph DB

works only for common alleles not rare alleles

rAmpSeq

use gatk in docker, use off the shelf tools as possible