Genomic Selection in Timothy (*Phleum pratense*

Kovi Mallikarjuna Rao

Norwegian University of Life Sciences, Ås, Norway

IFTBC -2019

Overview

- **Background**
  - *Phleum pratense* (Timothy) and its importance
  - Current breeding program in Timothy
  - Applications of genomic selection in forage species

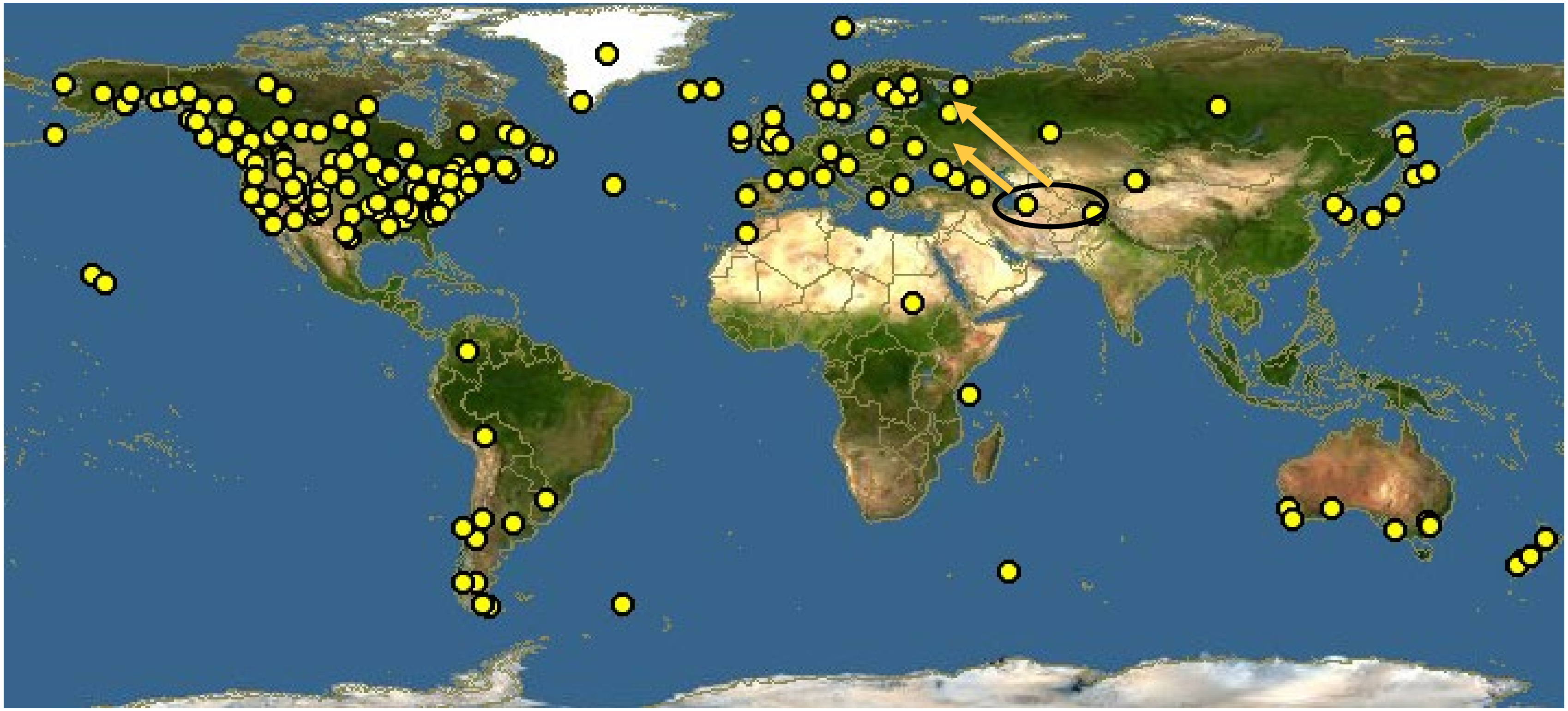
- **Outline of the GenSelTim Project**
  - Developing high density molecular markers in timothy
  - Accuracy of Genomic selection models in predicting GEBV in full-sib families of timothy
  - Cross validation of the genomic selection prediction models

- **Prediction accuracies for different traits**
- **Conclusions and future work**

Timothy (*Phleum pratensis*)

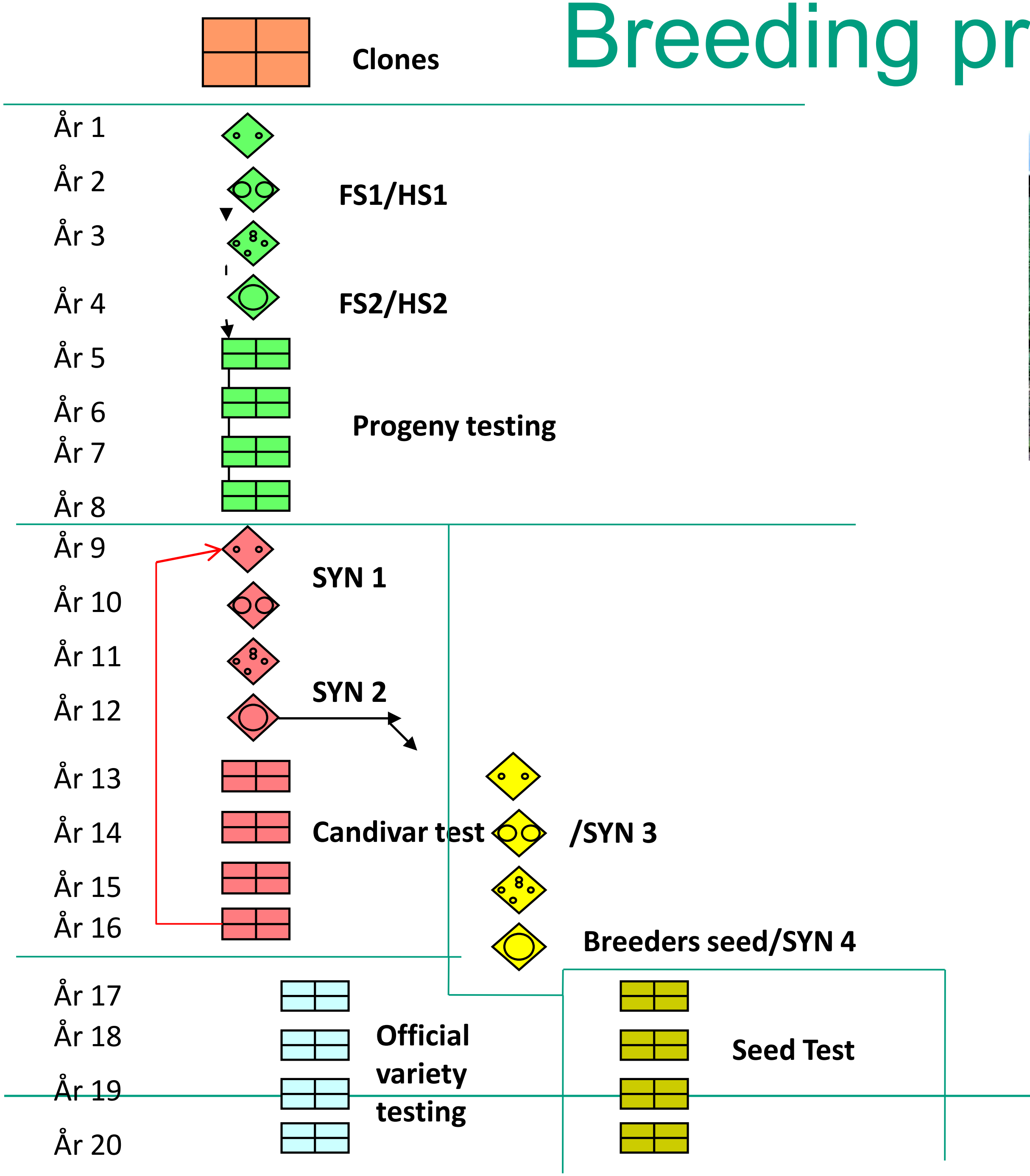
- Perennial, hexaploid (2n = 6x = 42) species
- Important crop for pasture and silage.
- High nutrition.
- Winter hardiness
- Economically most important forage crop. Timothy seeds constitute more than 50% seed sold for establishing leys for herbage production in Norway

Distribution of Timothy (*Phleum pratensis*)



- Dogan (1991) describes the Mediterranean and Western Asia as the center of origin for the genus *Phleum*
- First migration into Europe was of an ancestor of diploid *P.alpinum* subsp. Rhaeticum during the penultimate Riss glaciation 130,000-150,000 year B.P.

Breeding program



Clones

År 1

År 2

FS1/HS1

År 3

År 4

FS2/HS2

År 5

År 6

Progeny testing

År 7

År 8

År 9

SYN 1

År 10

År 11

SYN 2

År 12

År 13

År 14

Candivar test

/SYN 3

År 15

År 16

Breeders seed/SYN 4

År 17

År 18

Official variety testing

Seed Test

År 19

År 20

Genomic Selection

- **Genomic Selection;** unlike marker-assisted selection (MAS), GS predicts breeding values of lines using genome-wide marker profiling and allows selection of lines prior to field-phenotyping, thereby shortening the breeding cycle.
- Reduce breeding cycles and enhance genetic gain

- **GEBVs;** Prediction of the genetic merit of an individual based on its genome.

The term '**GS**' was first introduced by **Haley** and **Visscher** at the 6th World Congress on Genetics Applied to Livestock Production at Armidale, Australia in 1998.

GS was first propounded by **Meuwissen *et al* (2001)** : Seminal paper '***Meuwissen et al (2001) Prediction of total genetic value using genome-wide dense marker maps. Genetics 157: 1819-29."***

GS studies in forage crops

Fè *et al. BMC Genomics* (2015) 16:921
DOI 10.1186/s12864-015-2163-3

BMC Genomics

RESEARCH ARTICLE Open Access

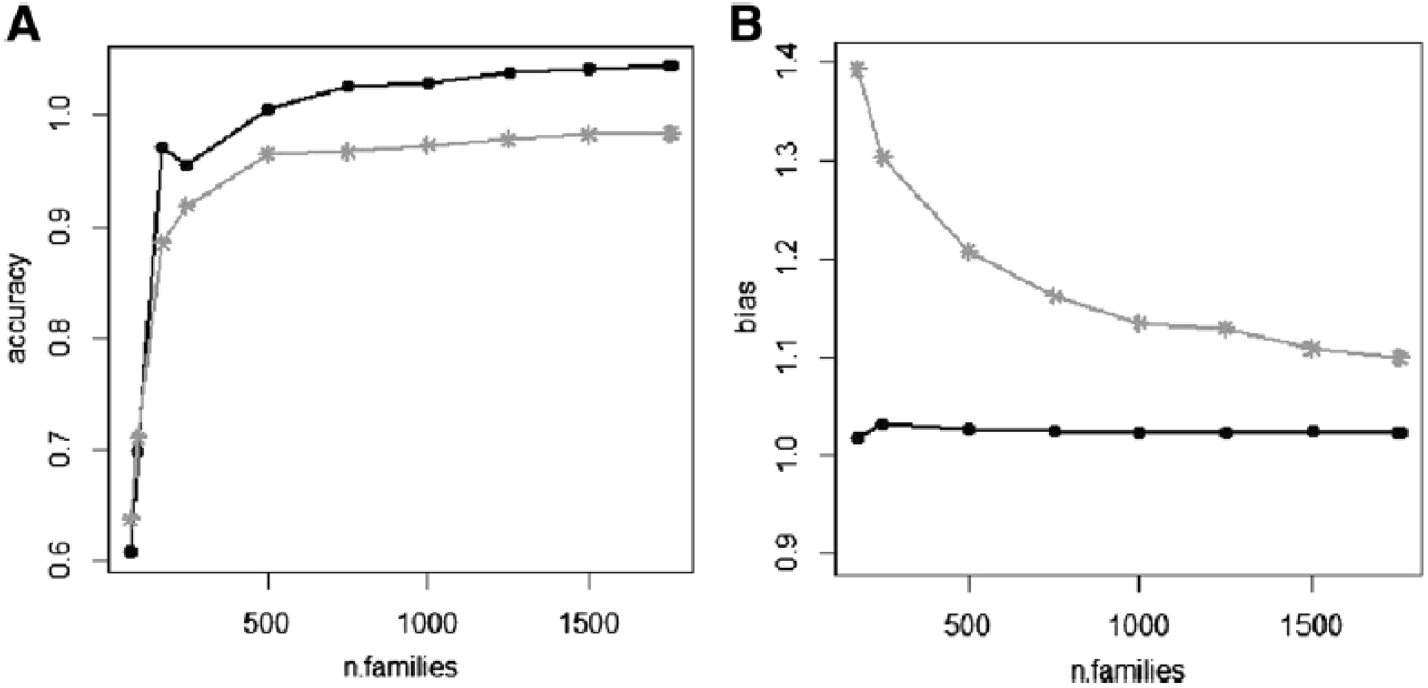


Fig. 7 Accuracies (**a**) and bias (**b**) with different population sizes: *k-fold* (black) and *pp-fold* scheme (grey)

Fig. 6 GEBV vs. corrected mean phenotypes: **a** within F_2s (*k-fold*); **b** predicting SYNs from F_2s. Legend: blue line = plot diagonal; red line = linear regression

structure and related families in the breeding population. While comprehensive Genome Wide Association analysis is difficult in species with extremely low LD, we did identify variants proximal to genes with a known role in flowering time (e.g. CONSTANS and Phytochrome C).

Keywords: Genomic selection, Perennial ryegrass, Heading, Flowering, GWAS, Lolium, CONSTANS, Phytochrome, Accuracy

FIG
ave
dep

s having low
high average

Cardinal points for success of GS

1. Population type & size of training population
2. Genotyping Platforms & marker densities.
3. Availability of HD genome wide markers.
4. Appropriate statistical methods for accurate GEBVs.
5. G x E.
6. Linkage disequilibrium

GenSelTim Project

Breeding populations and traits

- 847 2nd generation full-sib families (FS-2) available as remnant seeds at Graminor.

- FS-2 familes were progeny tested (2003-2013) in field trails at two locations in Southern Norway- Bjørke (low land) and Løken (High land).

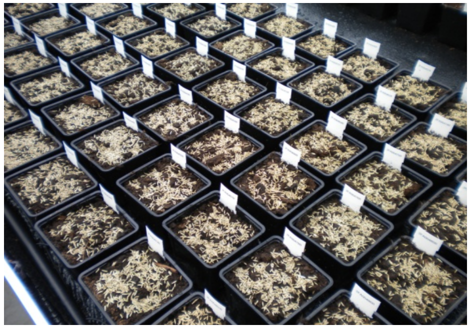
Breeding populations and traits

- Yield data available for 3 harvest years (2-3 cuts per year).

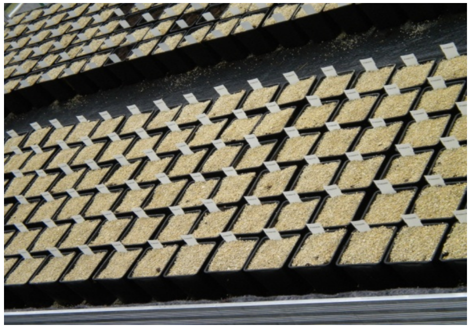
- Forage quality data (crude protein, invitro digestability, NDF, water soluble carbohydrates, various estimates of energy concentration) is available for each of 3 cuts of 630 samples at Bjørke.

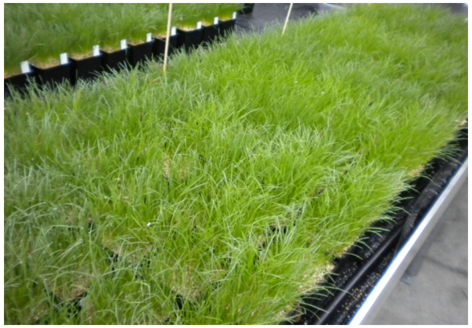
- 213 new FS-2 families sown in 2015 at Bjørke and Løken and complete yield and forage quality data will be available in the fall of 2018.

WP1: Developing high density molecular markers in timothy

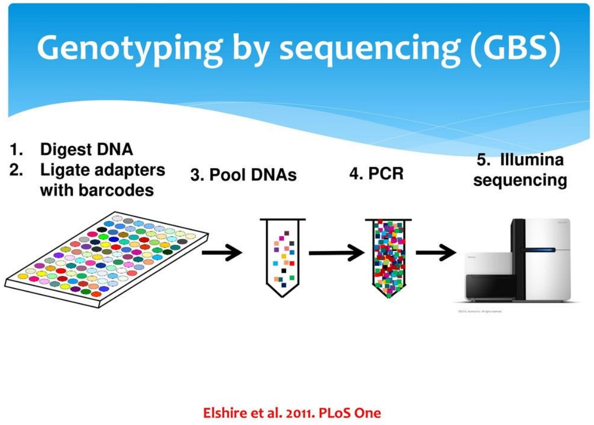


Seed





Discarded



Genotyping by sequencing (GBS)

1. Digest DNA
2. Ligate adapters with barcodes

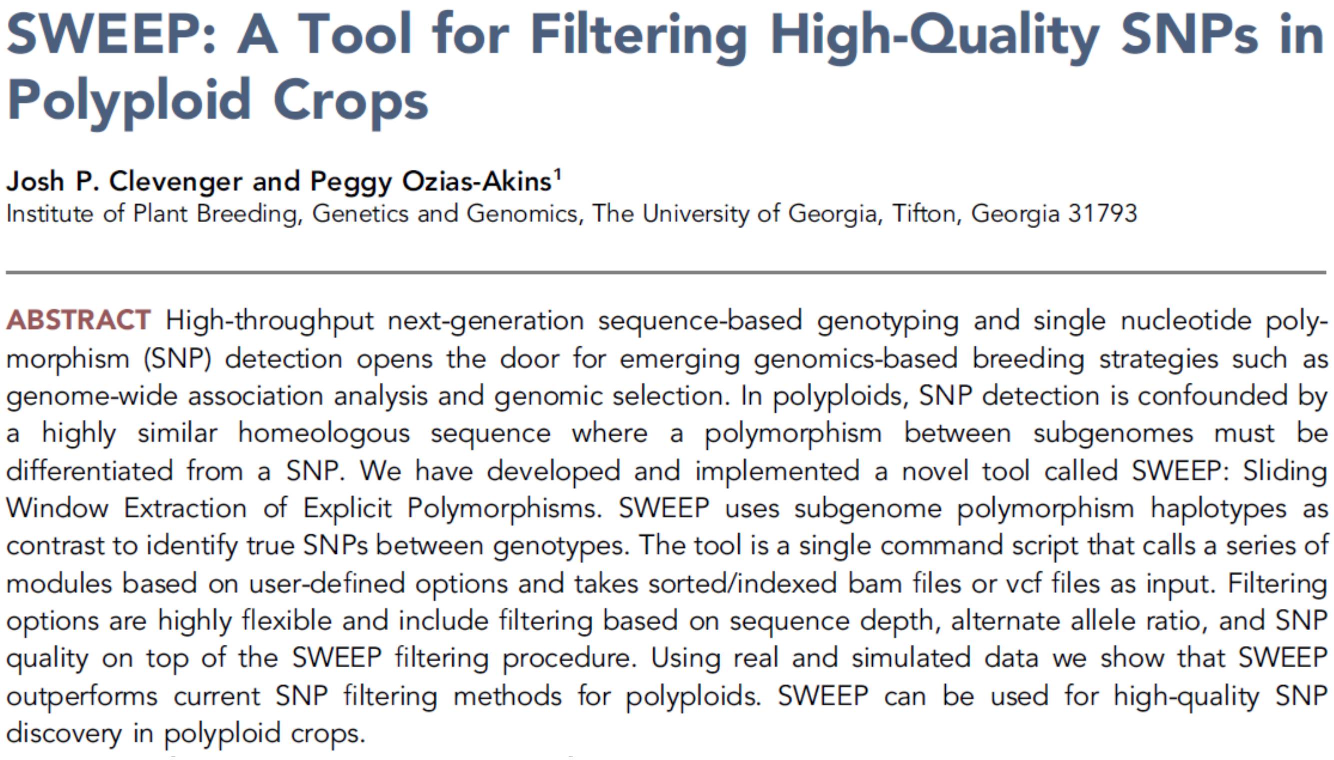
3. Pool DNAs

4. PCR

5. Illumina sequencing

Elshire et al. 2011. PLoS One

Effic
genc

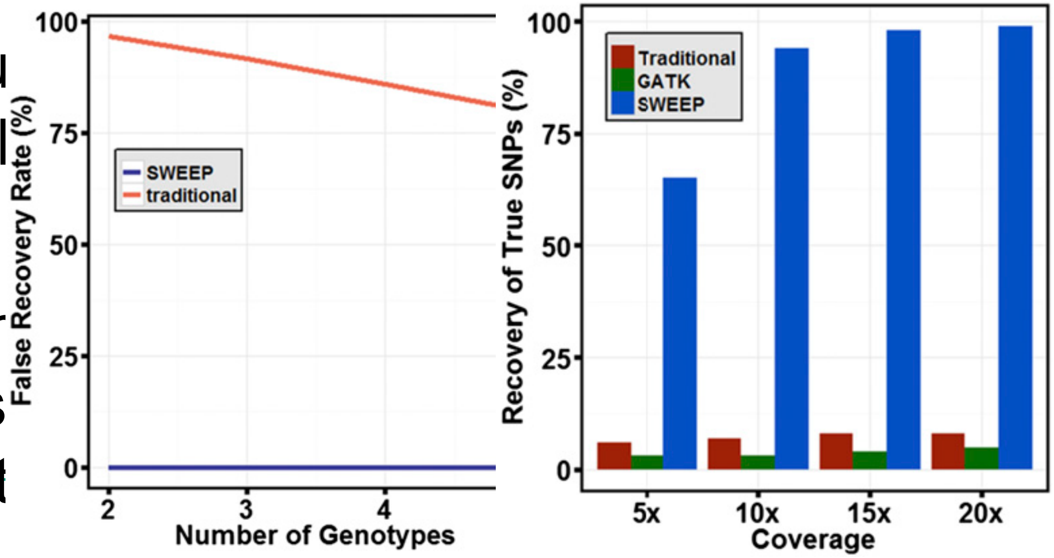


SWEEP: A Tool for Filtering High-Quality SNPs in Polyploid Crops

Josh P. Clevenger and Peggy Ozias-Akins[1]
Institute of Plant Breeding, Genetics and Genomics, The University of Georgia, Tifton, Georgia 31793

ABSTRACT High-throughput next-generation sequence-based genotyping and single nucleotide polymorphism (SNP) detection opens the door for emerging genomics-based breeding strategies such as genome-wide association analysis and genomic selection. In polyploids, SNP detection is confounded by a highly similar homeologous sequence where a polymorphism between subgenomes must be differentiated from a SNP. We have developed and implemented a novel tool called SWEEP: Sliding Window Extraction of Explicit Polymorphisms. SWEEP uses subgenome polymorphism haplotypes as contrast to identify true SNPs between genotypes. The tool is a single command script that calls a series of modules based on user-defined options and takes sorted/indexed bam files or vcf files as input. Filtering options are highly flexible and include filtering based on sequence depth, alternate allele ratio, and SNP quality on top of the SWEEP filtering procedure. Using real and simulated data we show that SWEEP outperforms current SNP filtering methods for polyploids. SWEEP can be used for high-quality SNP discovery in polyploid crops.

- SNP ca timothy
  due to
- To simu otypes,
  the UN pipeline
  will be
- Further u esigned to filter
  high qual
- Finally, th will be calculated
  for SNPs ate GWAFFs
  (Byrne et



GBS results

- SNP calling by UNEAK pipeline
- In total, 3,194,960 SNPs detected.
- Filtering:
- MAF 0.05
- SNP covering at least 10% of 918 samples to be retained: **874,163** filtered SNPs
- Sample must have at least 10% of SNPs present to be retained: 874,163 filtered SNPs in 912 filtered taxa

WP2: Accuracy of GS models in predicting GEBVs

- Detecting GEBV by different GS models

- The SNP effects will be estimated on the basis of five different statistical models:

➢ RR–BLUP, Bayes A (Meuwissen et al. 2001)

➢ Bayes Cπ(Habier et al. 2011)

➢Improved Bayesian LASSO (BLASSO) approach (Legarra et al. 2011)

➢GBLUP (Clark and Van der Werf, 2013).

Traits and model tested

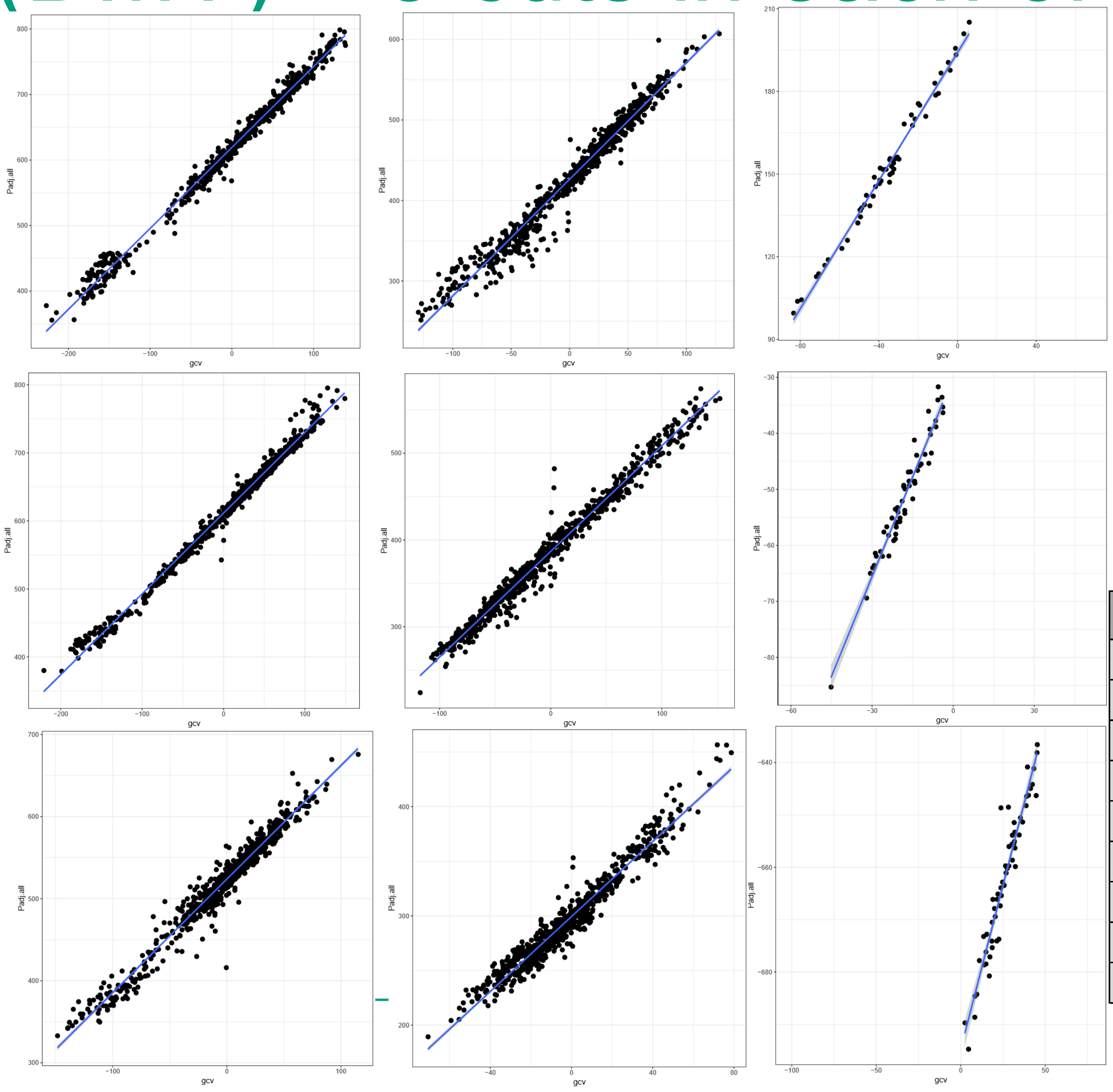
- All the yield related traits
- Spring growth, powdery mildew and leaf spot traits.
- Heading date has problem due to inconsistent phenotype recordings.
- gBLUP modeling was tested

GEBVs and Predictive ability

- Genomic estimated breeding values (GEBVs)
- gfull : breeding values calculated from the variance components
- gcv: breeding values calculated based on cross validation.

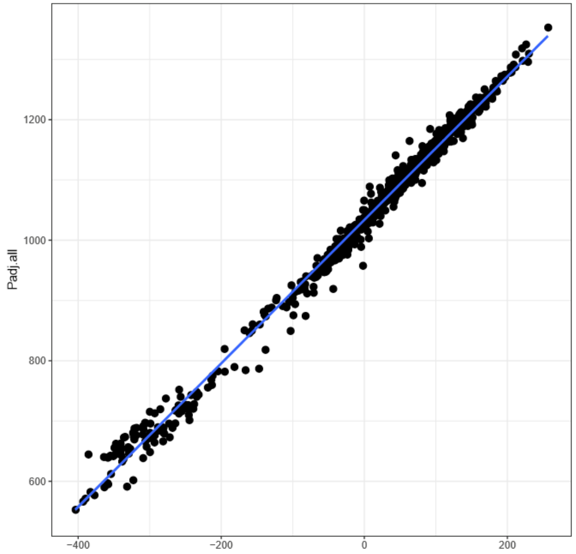
- PA: correlation between observed phenotypes and genomic estimated breeding values.

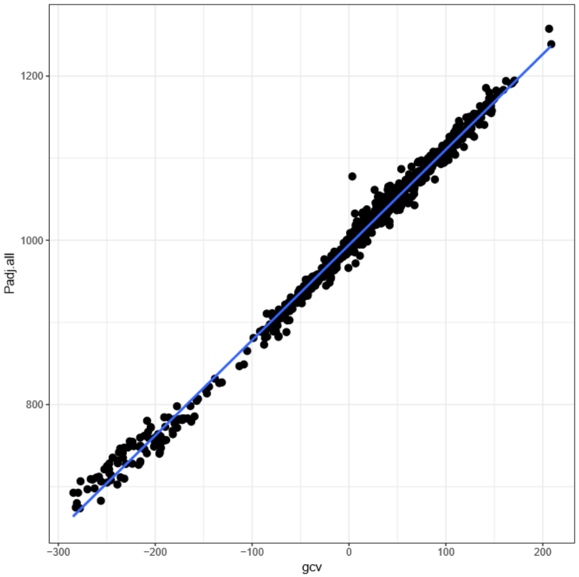
Predictive ability for dry matter yield (DMY) – 3 cuts in each of 3 years

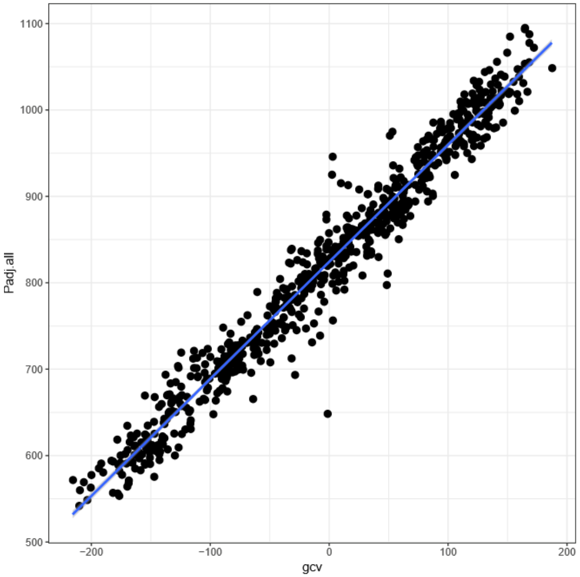


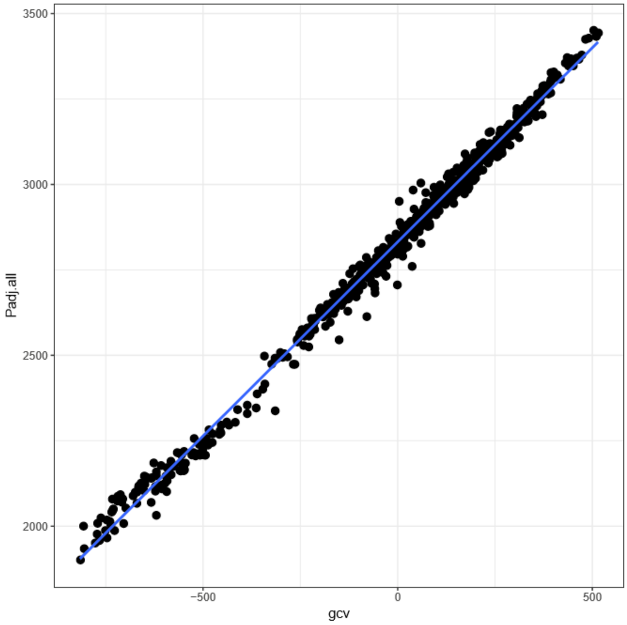
Trait	Bias	Accuracy	h2
KGDM101	1.235	0.994	0.318
KGDM102	1.448	0.983	0.240
KGDM103	1.162	0.993	0.366
KGDM201	1.187	0.995	0.348
KGDM202	1.217	0.990	0.321
KGDM203	1.185	0.973	0.277
KGDM301	1.384	0.981	0.221
KGDM302	1.716	0.972	0.167
KGDM303	1.251	0.959	0.266

Predictive ability for sum of dry matter yield (SUMDM)









Trait	Bias	Accuracy	h2
SUMDM1	1.136	0.997	0.356
SUMDM2	1.161	0.996	0.352
SUMDM3	1.355	0.982	0.252
SUMDM	1.136	0.997	0.343

WP3: Additional validation to cross-check the GS model efficiency

- 216 new FS-2 families sown in plots at Bjørke and Løken in 2015.

- Phenotype data for yield and forage quality traits will be availble by end of 2018.

- Data will be compared to the prediction models (GBLUP) to predict the GEBV and to ensure the prediction accuracy is high enough to rely upon.

Conclusions

- Genomic selection is reliable in Timothy in prediciting GEBVs particularly for the traits where we have many phenotypic observations, e.g. the yield related traits.

- The yield data GEBV predictions looks quite good with high correlations to the observed phenotype, a low bias and slightly high heritabilities.

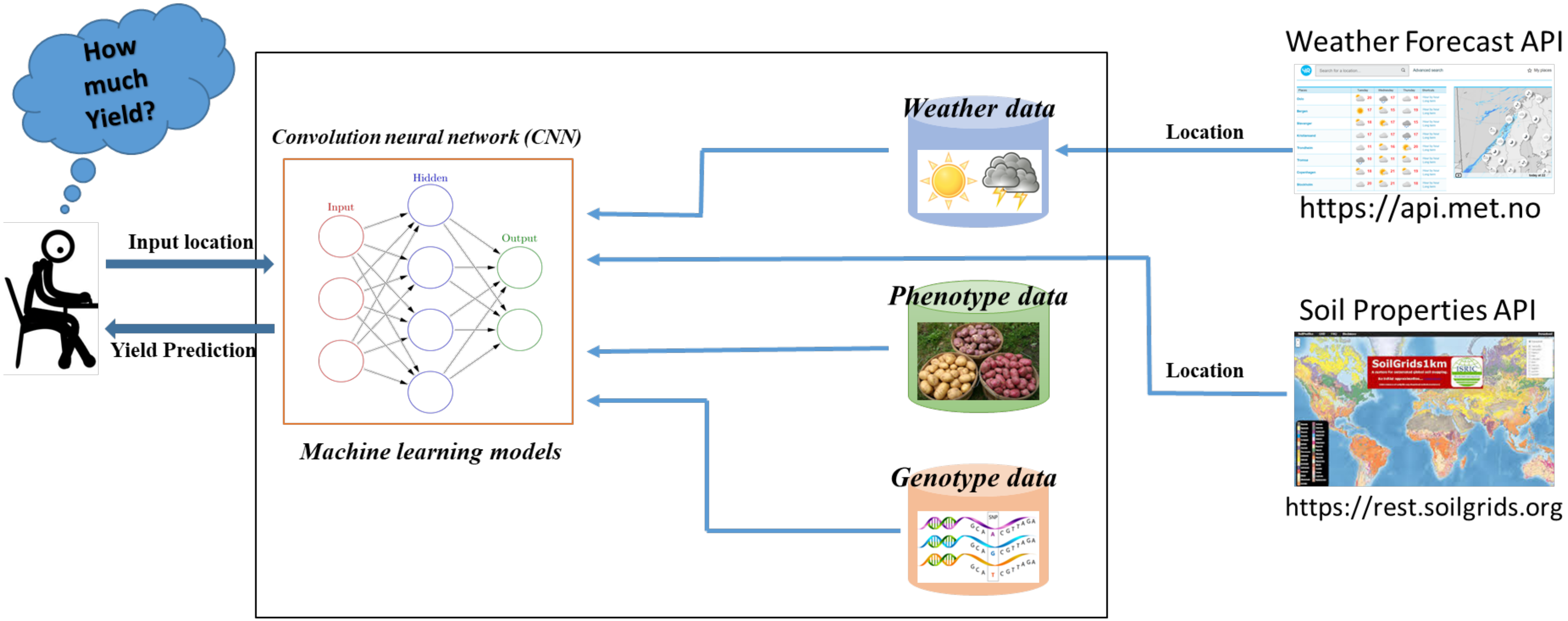
- For other traits like spring growth, powdery mildew and leaf spot traits the analysis failed because of too few phenotypic observations.

Ongoing and future work

- GS for forage quality

- Compare the predictive GEBVs of the validation population (213 FS-2 families) to the actual phenotype data collected from 2015-2018
- Developing draft genome sequence of timothy – Oxford nanopore long read sequencing technology: PromethION

Future work


How much Yield?
Input location
Yield Prediction
Convolution neural network (CNN)
Hidden
Input
Output
Machine learning models
Weather data
Phenotype data
Genotype data
Location
Location
Weather Forecast API
https://api.met.no
Soil Properties API
SoilGrids1km
ISRIC
https://rest.soilgrids.org

TEAM

NMBU-IPV

Prof. Odd Arne Rognli

Dr. Kovi Mallikarjuna Rao

Graminor

Dr. Petter Marum

Dr. Muath Alsheikh

Dr. Helga Amdahl

Aarhus University

Dr. Torben Asp

Other Colloborators

Dr. Crossa Jose , CIMMYT, Mexico

Prof. Theodorus Meuwissen, NMBU


Thank
You