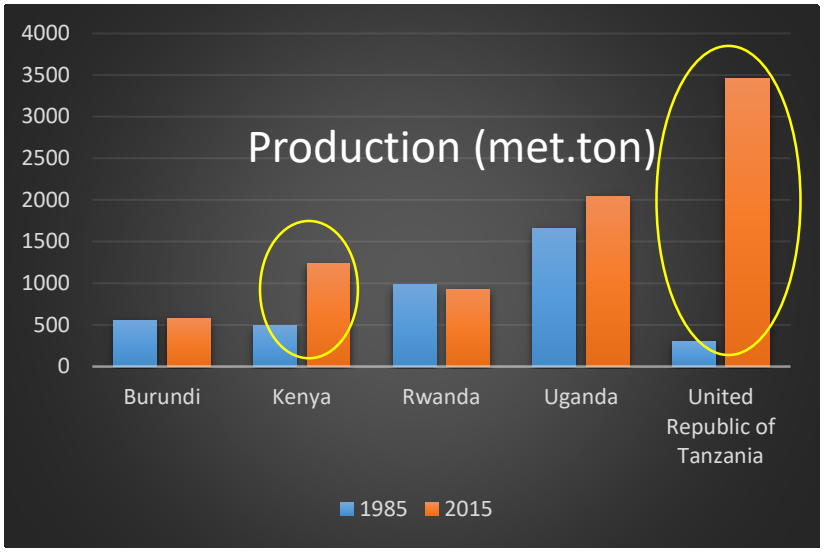
GENETIC VARIABILITY, HERITABILITY AND GENOME WIDE ASSOCIATION STUDY (GWAS) OF CONTINUOUS STORAGE ROOT FORMATION AND BULKING TRAITS IN SWEETPOTATO (*Ipomoea batatas*)

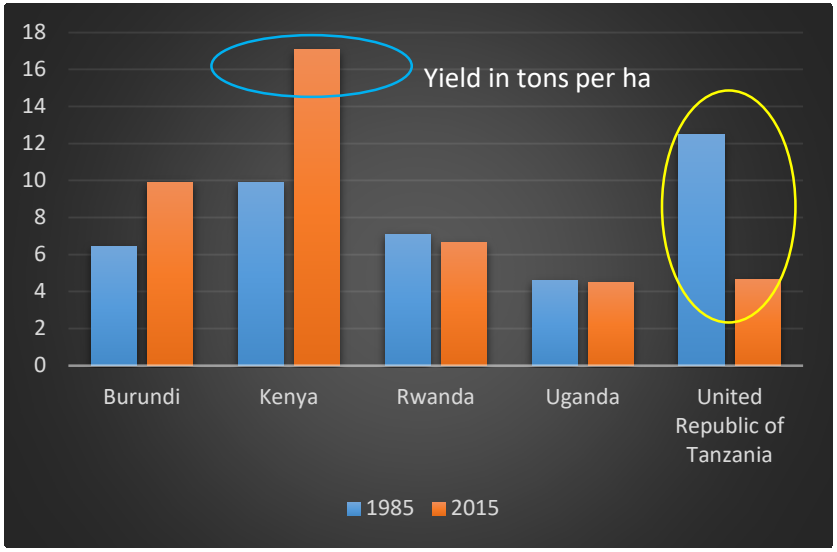
By Astere Bararyenya

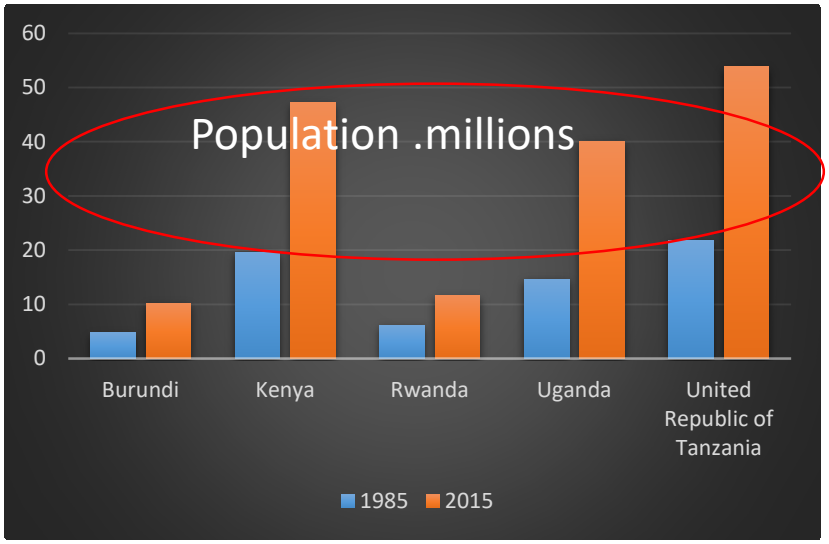
- Supervisors:
  - Prof. Phinehas Tukamuhabwa (Mak University)
  - Dr. Robert O. Mwanga (CIP)
  - Dr. Yao Nasser (ILRI-BecA) & Dr. Mercy Kitavi (CIP)

SWEETPOTATO BREEDER'S MEETING
4-9 JUNE 2018

BACKGROUND







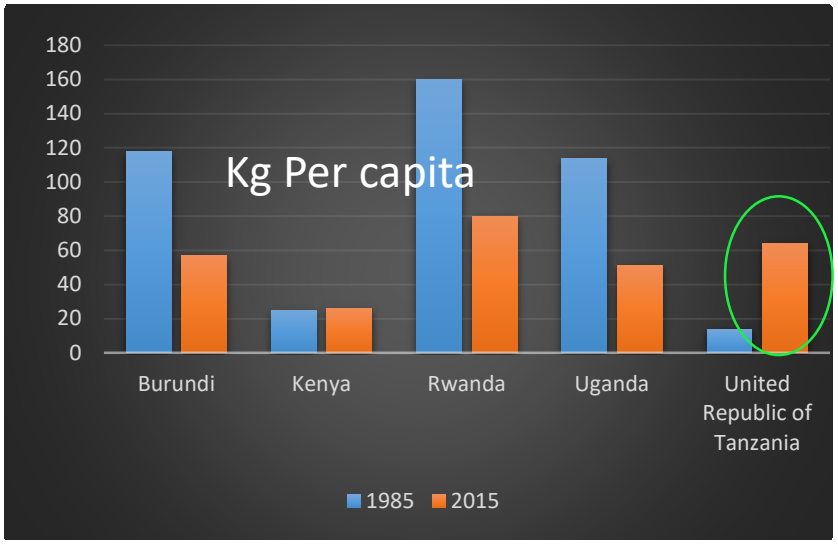


Fig2 (a, b, c, d). Information of the development of sweetpotato production in Eastern African countries in 1985 and 30 years later in 2015 (FAO, 2015)

BACKGROUND. CON'T

Key yield indicators	Average period range	Extreme ranges	Source
Yield	12t/ha	3-70t/ha	(Van Vugt, D., 2017, Kukimura, H., et al., 1990)
Maturity period	12 to 16 weeks	12-21 weeks	(Ravi, V., et al.,2009)
Storage initiation period	35 to 60 DAP	7 to 112 DAP	(Wilson, 1982)
Max SR number	49-56 DAP	30 to 112 DAP	Wilson and Lowe (1973)

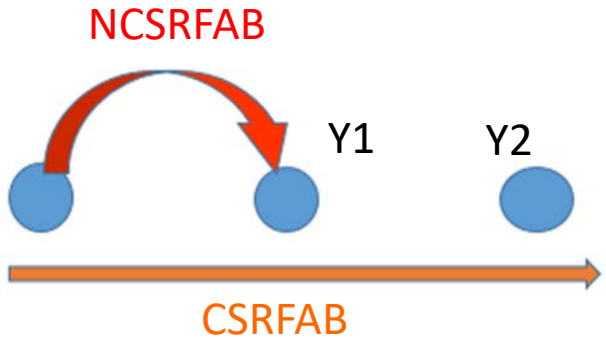
Uganda has many SP varieties (landraces >800, and introductions, >200), genetic variability is important for systematic breeding (Adebisi *et al.* 2001, Engida *et al*. 2007)

BACKGROUND. CON'T

Unmature plant

Maturity signs~senescence

- In Uganda, 90% (n=350) of farmers have no knowledge of the maturity periods (Bashaasha *et al.*, 1995) of their various varieties.
- Maturity period vary with cultivar and Environment
- Senescence signs not always true



- yields are ~ speed and the duration of the period of initiation and bulking

PROBLEM STATEMENT

- 27 varieties have been released (Mwanga et al., 2011; Ssemakula et al., 2013)
- Farmers still prefer their local varieties (low yielding and susceptible to disease)
- These local varieties are adapted to continuous harvesting, a main practice of farmers
- variety selection based on one-time harvest
- No breeding information to understand this common practice in small-scale farmers

JUSTIFICATION

- CSRFAB traits are relevant under high population growth context and low food source.
- Understanding the mode of inheritance of CSRFAB varieties will increase adoption and production in small-scale farmers.
- Identification of SNP markers associated with CSRFAB in sweetpotato genotypes will facilitate acceleration of breeding through molecular assisted selection

OBJECTIVES

OVERALL OBJECTIVE:

Contribute to sustainable improved food security among small-scale sweetpotato farmers

SPECIFIC OBJECTIVES:

1. Identify genetic variability and growth patterns associated with CSRFAB in sweetpotato.
2. Determine the inheritance patterns and breeding values of CSRFAB traits for future breeding decisions for high yielding sweetpotato varieties
3. Identify SNP markers associated with CSRFAB in sweetpotato.
4. Convert identified SNPs into Kompetitive allele specific PCR (KASP) assays for SNP marker validation
5. Discover QTL associated with CSRFAB traits in F1 sweetpotato population

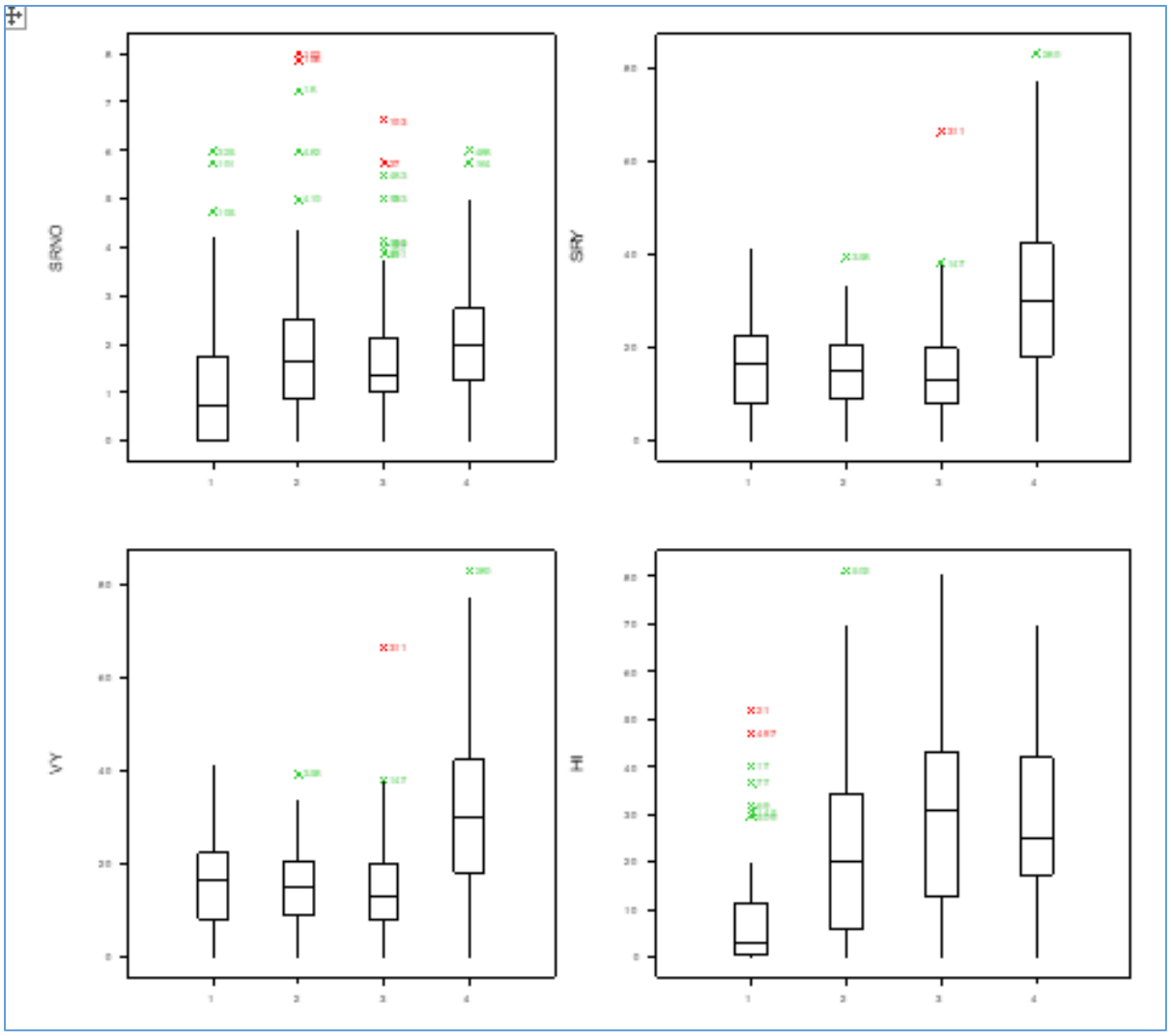
MATERIALS AND METHODS

Study 1: Identify genetic variability and growth patterns associated with CSRFAB in sweetpotato

- **Study area:** Two sites
- **Test Genotypes**: 130 diverse clones
- **Design:** Repeated measurements (4 waves), genotypes arranged in RCBD, 2 Replications
- **Period:** Second rains (2016) & first rains (2017)
- **Data collection:** (i) growth parameters (ii) yield and yield component parameters (iii) Diseases (SPVD) and weevil damage

DATA ANALYSIS

- Genstat 11 edition were used
- The general model used is Y = Fixed Effect model + Random effect model
- Variance components extracted and used to estimate broad-sense heritabilities for the assessed traits
- ANOVA with treatments considered fixed effects, and summary statistics of mean squares generated to test the significance of the different sources of variation
- Phenotypic correlations of the different traits were estimated using means of the datasets generated



General observations

- genotypes increase SRN at 4th harvest
- More genotypes displays high yields at 4th harvest
- genotypes have more vine yield at 4 harvest
- HI tends to be stable and high as HT increases

Fig 3: Variability and distribution of CSRFAB across the four harvesting times at 3MAP, 4MAP, 5MAP and 6MAP in 2016 & 2017 combined

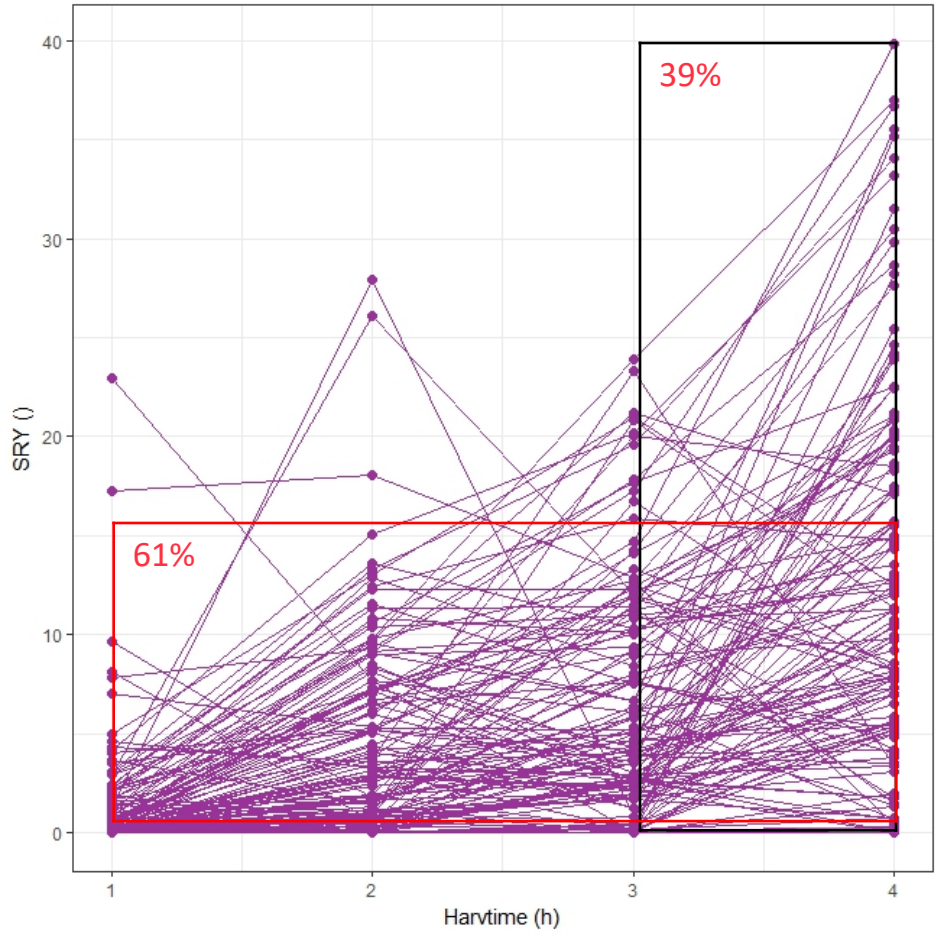


Fig4a. show the storage yield trend over 4 harvesting ting of 130 sweetpotato genotypes screened for continuous storage root formation and bulking.
1= 3MAP, 2=4MAP, 3=5MAP, 4=6MAP

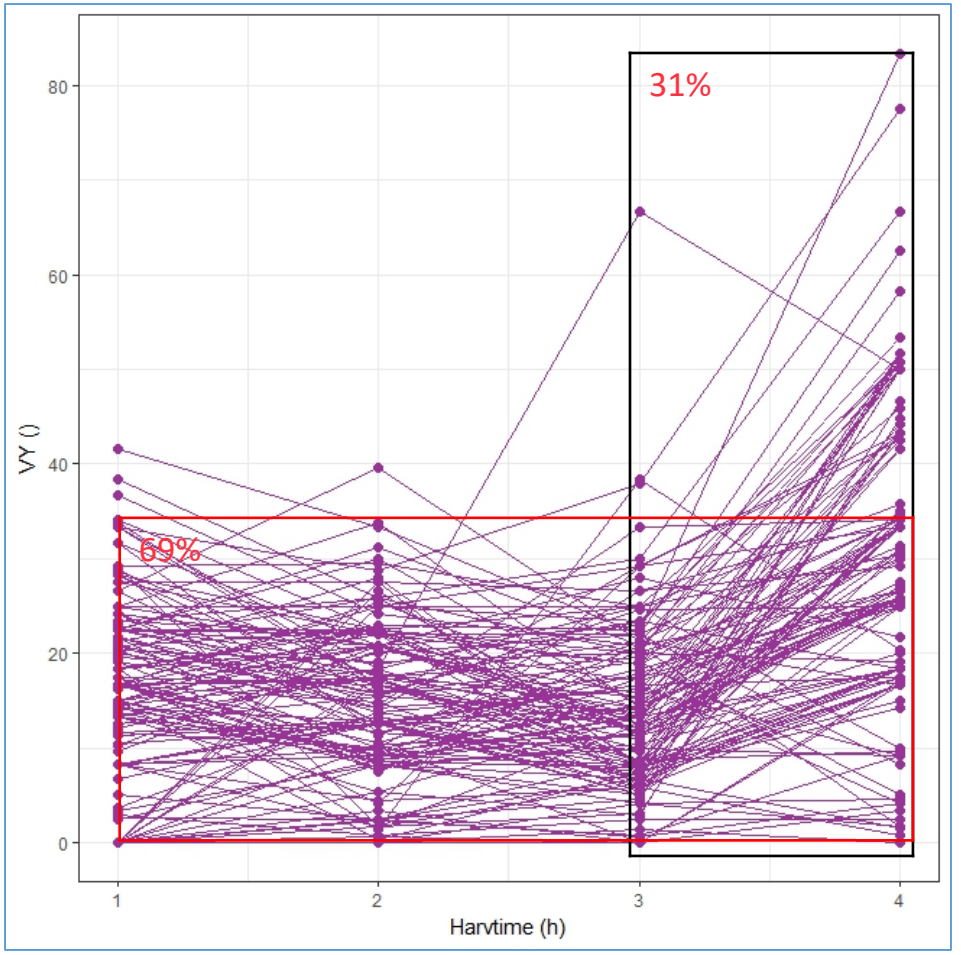
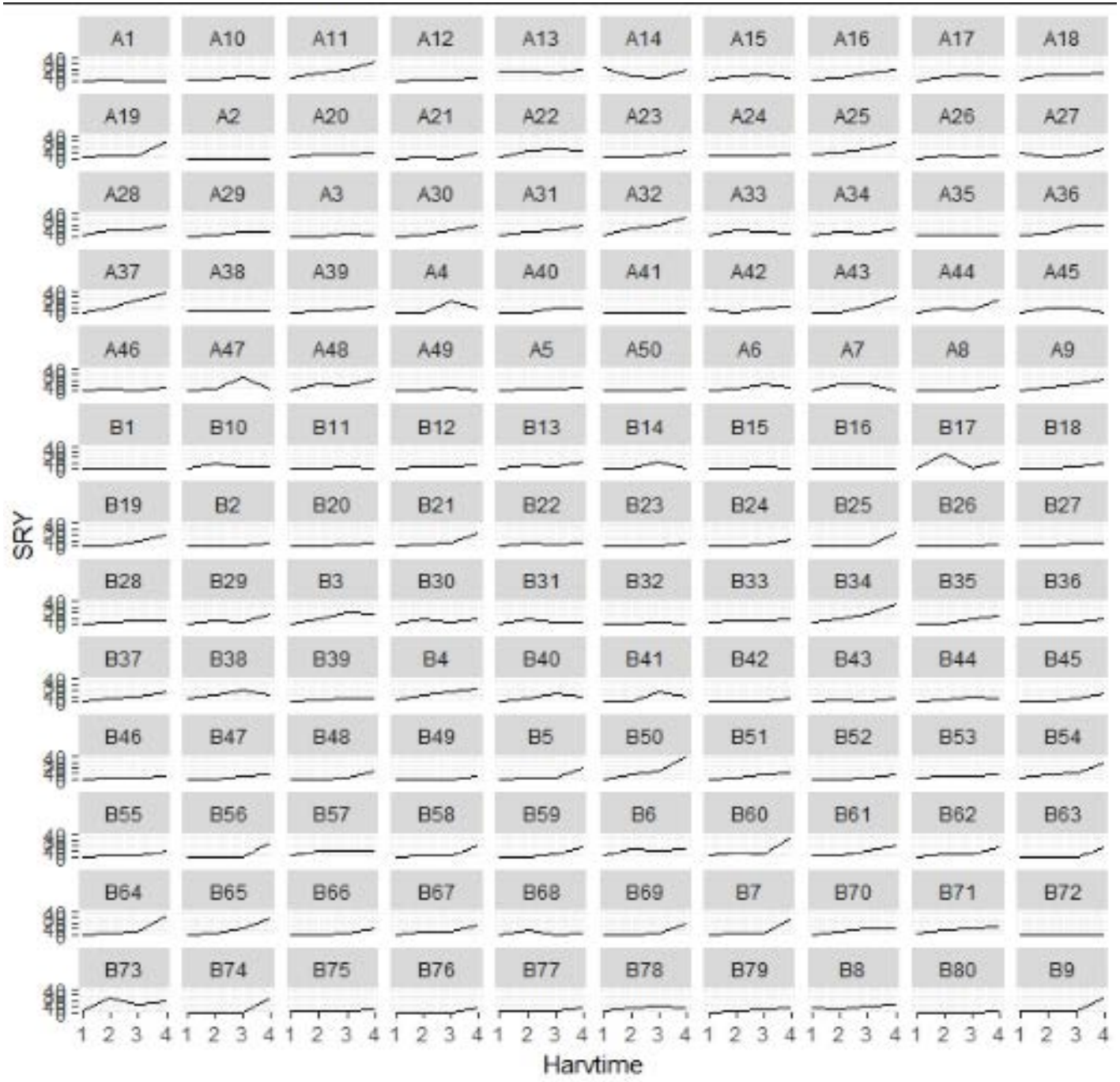
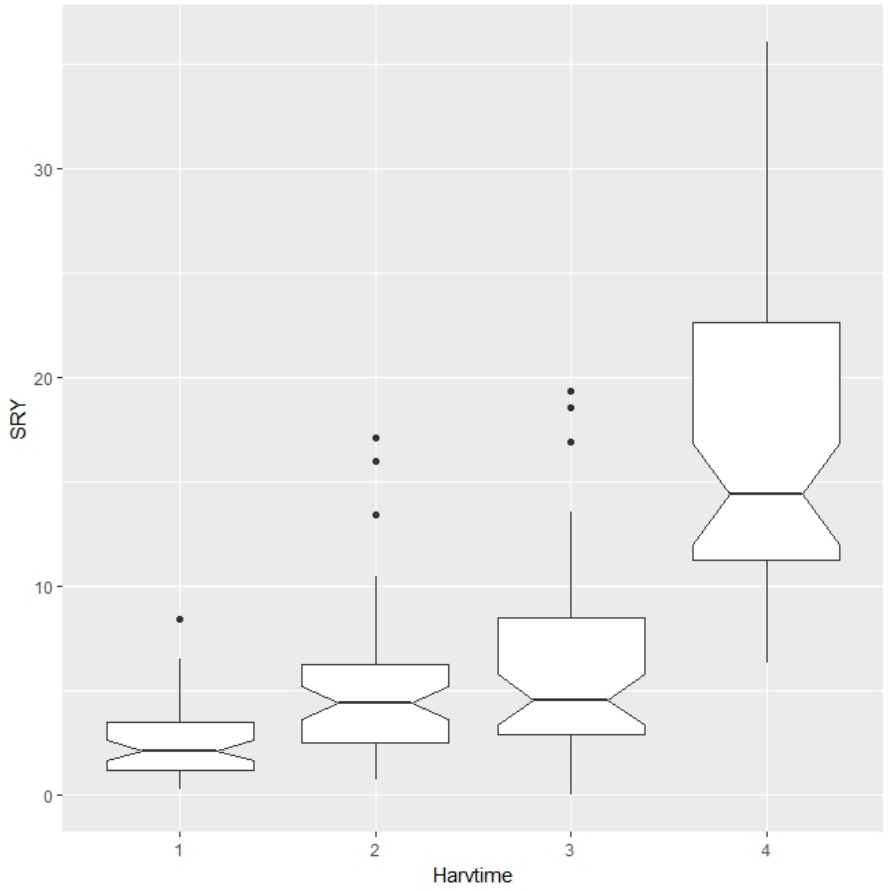
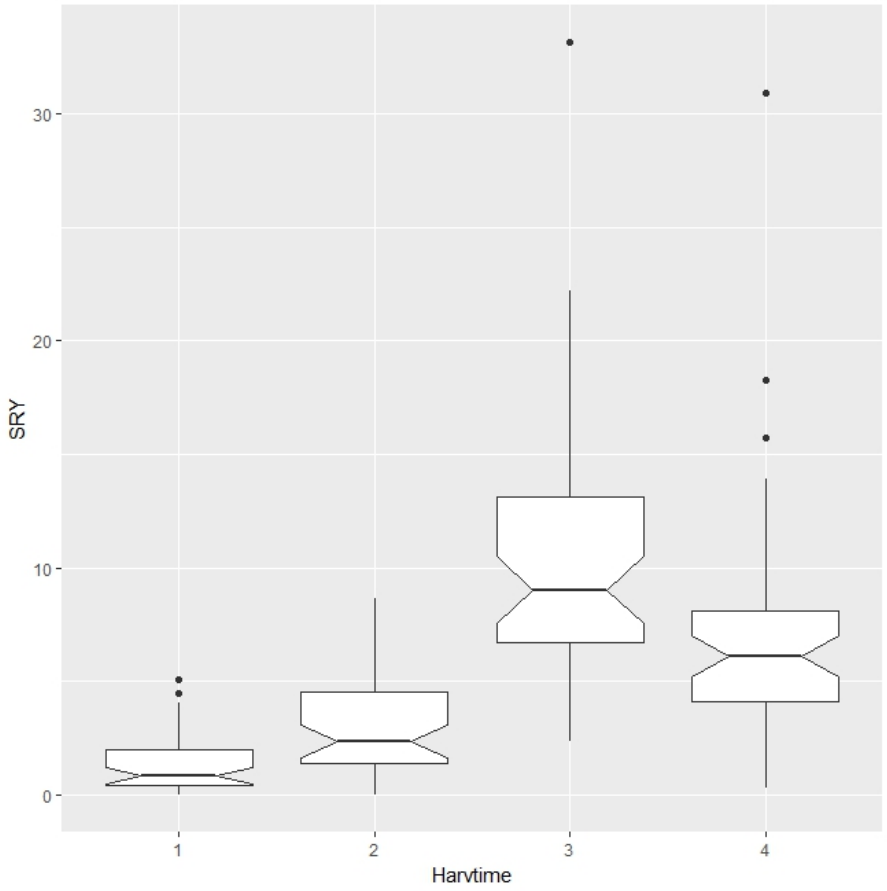


Fig4b. show the Vine yield trend over 4 harvesting ting of 130 sweetpotato genotypes screened for continuous storage root formation and bulking.
1= 3MAP, 2=4MAP, 3=5MAP, 4=6MAP



Genotype growth behavior. Each yield genotype where plotted for 4 harvesting time.





- Fifty-one clones (39%) exhibited continuous storage root formation and bulking properties.
- High bulking speed
- Yield increase as harvest time increases

- 81 clones (61%) are determinate in bulking and had pick at 5 MAP
- On average yield is reduced as harvest time increase

Table 3. shows the model fit and factor significances across the location and season. Strong linear and quadratic effect across treatments and their interactions was observed in most of studied parameters

Fixed term	DDF	SRN_P	SRYLD	VY	HI	CRW	CRN
Seas	1	297.81***	264.12***	2676.87***	257.64***	92.55***	180.46***
Loc	1	329.64***	0.91NS	214.84***	4.12*	21.47***	404.33**
ENTRY	129	1731.91***	827.88***	772.14***	951.76***	642.87***	1269.26***
Lin	1	91.85***	729.43***	9.41**	2589.96***	755.63***	441.95***
Quad	1	11.25***	14.63***	2.91NS	368.28***	10.07**	0.41NS
HT	1	1.89NS	0.92NS	50.4***	122.89***	0.05NS	0.04NS
Seas.Lin	1	2.8NS	52.99***	108.48***	784.95***	7.23**	7.69**
Seas.Quad	1	24.52***	9.07**	3.81NS	425.06***	4.46*	23.32***
Loc.Lin	1	15.19***	72.62***	8.88**	891.87***	0.07NS	1.12NS
Loc.Quad	1	12.18***	7.96**	10.68**	590.24***	2.81NS	1.18NS
ENTRY.Lin	129	180.41**	369.61***	166.82*	250.51***	235.48***	209.64***
ENTRY.Quad	129	127.35NS	125.54NS	124.39NS	187.7***	125.77NS	141.79NS
Seas.Loc.ENTRY	126	278.39***	388.37***	288.88***	219.3***	183.49***	198.57***
Seas.Loc.ENTRY.Lin	381	527.02***	1079.23***	468.13**	400.62NS	584.97***	444.23**
Seas.Loc.ENTRY.Quad	372	403.87NS	452.72**	321.58NS	330.11NS	289.48NS	303.86NS

Wald stat significance for storage root number per plant, storage yield (tons/ha), vine yield (tons/ha), harvest index, storage root diameter (mm), storage root length (mm), commercial root number and weight of one hundred and thirty (130) sweetpotato genotypes across two locations (Namulonge and Serere) and two seasons (2016B and 2017A) in Uganda were strongly significant in most cases.

Table4. shows the overall growth mean trend over four harvesting time points across location and season, the % of change between 4 (recommended harvesting period in the study areas) and 6MAP (extended harvesting time) and their respective growth coefficients (intercept. Linear and quadratic slopes).

Location	Season	Traits	3MAP	4MAP	5MAP	6MAP	% of change	a	b1	b2	R2 Lin	R2 Quad
NaCRRI	2016B	SRN_P	1.19	1.82	1.7	2.11	41	-1.2	1.5	-0.2	0.81	0.81
NaCRRI	2017A	SRN_P	2	2.09	2.51	1.67	-84	-2.2	2.7	-0.4	0.39	0.93
NaSARRI	2016B	SRN_P	0.46	0.64	0.42	1.02	60	-0.2	0.25	-0	0.73	0.74
NaSARRI	2017A	SRN_P	0.99	1.6	2.03	2.33	30	-1.1	1.24	-0.1	0.95	1
NaCRRI	2016B	SRY	1.35	4.53	6.75	12.8	600	-0	-0.5	0.6	0.94	0.99
NaCRRI	2017A	SRY	4.12	4.42	7.74	6.92	-82	-4	4.69	-0.5	0.83	0.92
NaSARRI	2016B	SRY	0.45	0.51	0.36	5.1	474	2.44	-2.6	0.6	0.56	0.83
NaSARRI	2017A	SRY	1.35	6.29	17.6	27.6	994	0.93	-3	1.69	0.92	0.99
NaCRRI	2016B	VY	26.5	22.6	24.6	24.6	0	-18	24.9	-3.4	0.45	0.78
NaCRRI	2017A	VY	25.5	27.4	30.6	26.4	-417	-24	30.7	-4.2	0.54	0.93
NaSARRI	2016B	VY	17	10.9	10.5	0.32	-1014	-15	20.3	-3.5	0.02	0.79
NaSARRI	2017A	VY	58	48.6	62.8	72.7	991	-33	46.1	-5.2	0.7	0.81

Table 5. Pearson's correlation coefficients of CRW, HI, NCRW, NOCR, NONCR, NOSR_P, SRDIA, SRLG and SRY in the 130 sweetpotato genotypes across two locations (Namulonge and Serere) and two seasons (2016B and 2017A) (N=4160) in Uganda

	CRW	HI	NCRW	NOCR	NONCR	NOSR_P	SRDIA	SRLG	SRY	VY
CRW	-									
HI	0.6467***	-								
NCRW	0.2249*	0.2809*	-							
NOCR	0.6526***	0.5617**	0.3299*	-						
NONCR	0.093NS	0.215*	0.4953**	0.2798*	-					
NOSR_P	0.4898**	0.5083**	0.4925**	0.7844***	0.6944***	-				
SRDIA	0.6889***	0.5749**	0.2607*	0.5935***	0.185NS	0.4931***	-			
SRLG	0.5265**	0.4016**	0.1934NS	0.4236***	0.024NS	0.2832*	0.4918***	-		
SRY	0.9919***	0.654***	0.3013*	0.6666***	0.1235NS	0.5174***	0.6935***	0.5352***	-	
VY	0.3506*	0.1895NS	0.0799NS	0.2219*	-0.044NS	0.1072NS	0.2824**	0.2213*	0.3584*	-

High positive correlation significance was observed between SRY and CRW (.9919), NOSR_P and NOCR (0.7844), NOSR_P and NONCR (.6944), SRY and SRDIA (.6935), SRDIA and CRW (.6889), SRY and NOCR (.6666), SRY and HI (.654), NOCR and CRW (.6526).

The high correlation between traits implies simultaneously selection of the traits for CSRFAB

Table 6: Means of most CSRFAB genotypes (tons/ha) Accros location and season

Name	Code	3MAP	4MAP	5MAP	6MAP	Change 5-6MAP	%increase
Huarmayano	B80	3.9	2.4	9.1	36.1	27.0	297
KML956	A15	5.0	8.4	19.3	33.6	14.3	74
KML872	A32	2.2	5.0	12.9	33.2	20.3	158
NASPOT1	A24	4.4	6.7	5.2	31.0	25.8	497
MSD380	B4	3.7	10.5	9.2	30.2	21.0	229
PAL94SilkOMOYAKA	A9	3.0	6.5	9.4	29.7	20.3	216
PAL133TEGERERE	A8	2.0	1.5	6.9	29.2	22.3	322
KYABAFURUKI	B73	5.8	16.0	2.6	27.8	25.1	954
MPG1128	A11	2.1	10.0	7.7	26.4	18.7	243
ARA209	B2	3.4	4.0	10.4	26.2	15.8	152

These 10 genotypes are potential sources of genes for breeding for CSRFAB

STUDY 2: DETERMINE THE INHERITANCE PATTERNS AND BREEDING VALUES OF CSRFAB TRAITS FOR FUTURE BREEDING DECISIONS FOR HIGH YIELDING SWEETPOTATO VARIETIES

- **Locations**: Namulonge , Serere
- **Test Genotypes**: 280 genotypes
- **Traits:** Resistance and non-resistance to whitefly
- **Season**: 1st rains (2017) & 2nd rains (2017)
- **Mating design**: North Carolina II design method

TABLE. SUMMARY OF CROSSING, MATING DESIGN AND SEED OBTAINED AT NAMULONGE

		NCSRFAB					CSRFAB					
		A2	A12	A17	A19	A24	A18	A22	A30	A33	A41	
	Code	EJUMULA	TANZANIA	SILK(1254)	SPK004	NASPOT1	RAK819	MAGABALI	NASPOT 7	MSK 1040	UKEREWE	
NCSRFAB	B1	269	8	0	135	532	2	138	216	60	1	RESISTO
	B2	156	60	17	29	50	0	19	10	54	50	ARA209
	B54	250	31	0	697	232	18	0	200	81	7	NK1081L
	B56	450	39	0	109	78	135	10	731	55	88	NEW KAWOGO
	B73	0	129	20	106	85	0	3	40	133	17	KYABAFURUKI
	B80	790	0	0	114	974	45	0	78	0	414	HUARMAYANO
CSRFAB	B3	1	150	NS	48	43	0	2	52	64	0	HMA496
	B9	13	13	7	134	8	11	44	23	1	34	MARY
	B17	0	27	8	17	10	66	2	39	35	0	MBR536
	B43	0	26	2	6	0	17	0	3	0	4	WAGABOLIGE
	B44	60	56	2	39	60	54	21	17	15	18	MUGANDE
	B49	2	8	9	22	21	22	0	19	44	0	ARA22

- 5 seedling per cross were multiplied for experiment set up
- 59/100 successful crosses

DATA COLLECTION AND ANALYSIS

- Data generated from both the seedling and clonal trials will be subjected to parent-offspring analysis procedures that will enable quantification of heritabilities
- Mixed linear models (MLM) for analysis (Felipe, 2015)

Matrix notation for MLD

Fixed effect:

mean + Crosses+ Female+male+ Female*Male+ Crosses*Female+Crosses*Male

Random:

Replication + Replication * Crosses.

DATA NOT ANALYSED

STUDY 3: IDENTIFY LOCI ASSOCIATED WITH PHENOTYPIC TRAITS FOR CSRFAB SP GENOTYPES

- **Locations**: Namulonge, Serere
- **Season**: November, 2016 to May, 2017
- **Test Genotypes** & **Traits:** 280 diverse F1 clones - Disease resistance, agronomic and morphological traits
- **Experimental design**: Repeated measurements
- **Data collection:** (i) growth parameters (ii) yield and yield component parameters (iii) Diseases (SPVD) and weevil damage

STUDY3. Discover SNP markers associated with CSRFAB in sweetpotato

- Genomic DNA extracted (Dellaporta *et al.,* 1983) and GBS sequencing libraries prepared

- Sequencing performed using Illumina Hiseq2000 (Swarts *et al.,* 2014)

- TASSEL-GBS pipeline (Glaubitz *et al.,* 2014) used to process the FASTQ sequence data into SNP calls based on *Ipomoea Trifida* reference genome

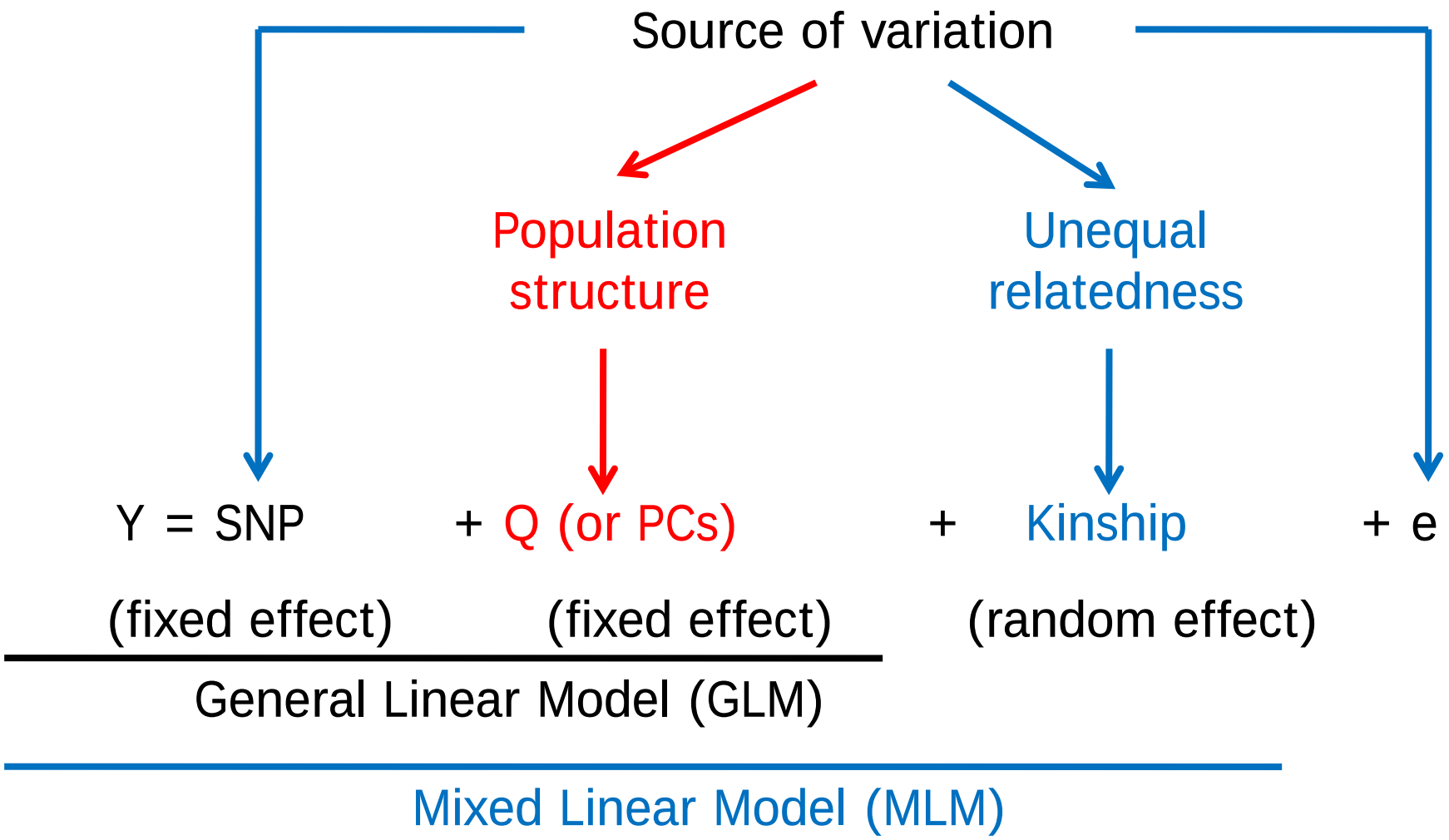
DATA ANALYSIS

- Both the phenotypic means and average performance means best linear BLUPs extracted (R core team, 2013)

- Non-segregating and uninformative sites, imputed data were filtered at (MAF) = 0.01 (TASSEL v 5.2.9)

- This filtered dataset used in estimating PCA and kinship, both of which were important for subsequent statistical analyses

GWAS MODELS


Source of variation
Population structure
Unequal relatedness
Y = SNP
+ Q (or PCs)
+ Kinship
+ e
(fixed effect)
(fixed effect)
(random effect)
General Linear Model (GLM)
Mixed Linear Model (MLM)

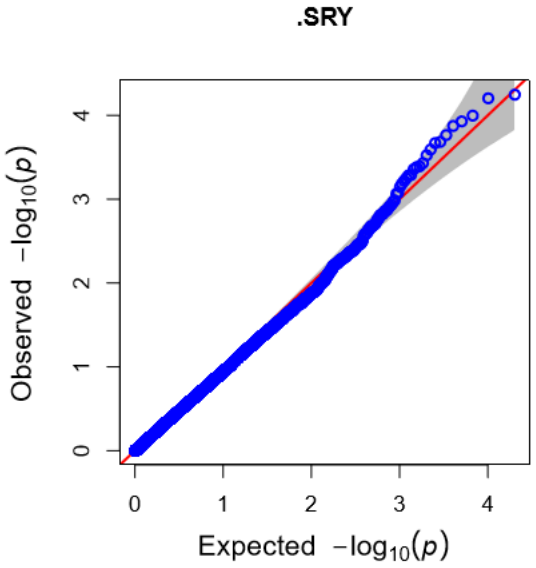
EVALUATIONS OF THE ASSOCIATION

- Evaluations of the association mapping based on the quantile–quantile (Q–Q) plot, under the null hypothesis that there is no association between a SNP and the phenotype

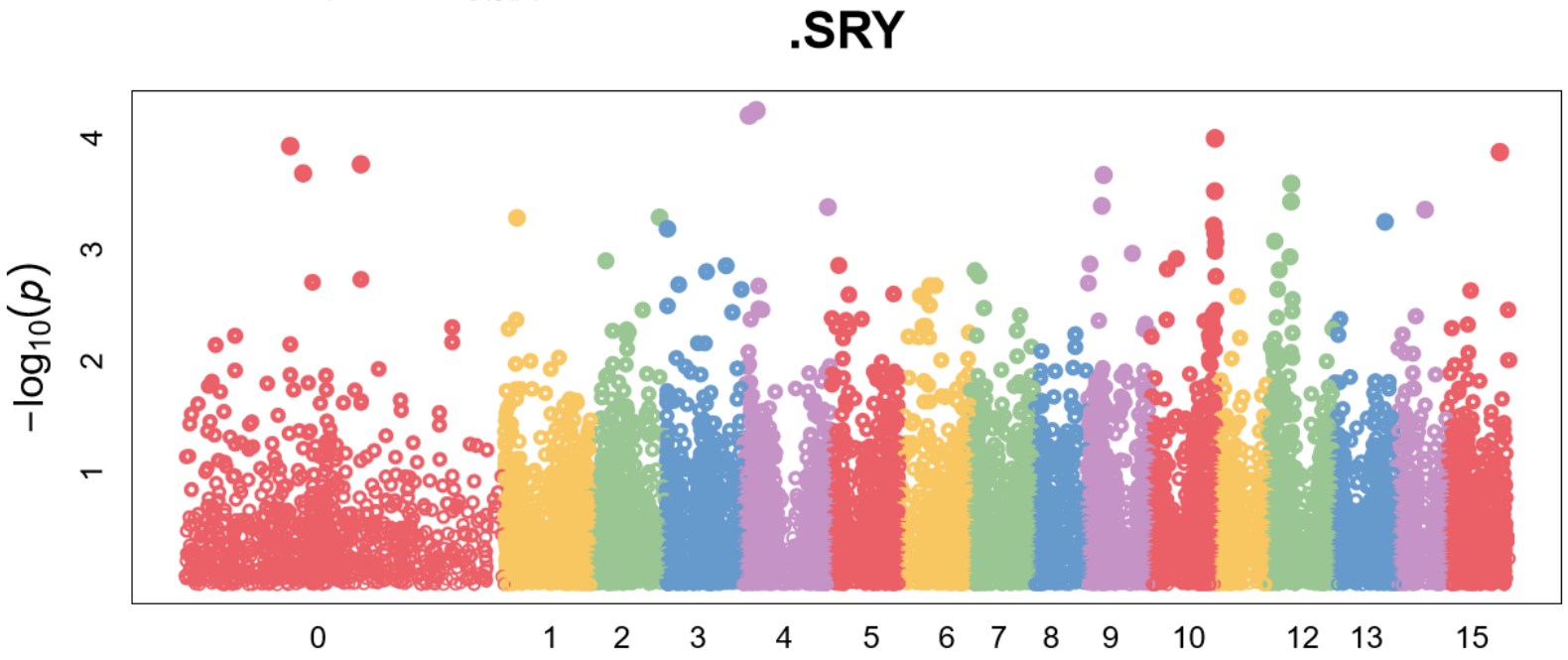
- SNPs with P values less than the 5 % Bonferroni threshold were considered to be significantly associated with phenotypes

- Chromosome-wise association signals were visualised from Manhattan plots generated using the qqman package of R software (R core team, 2013)

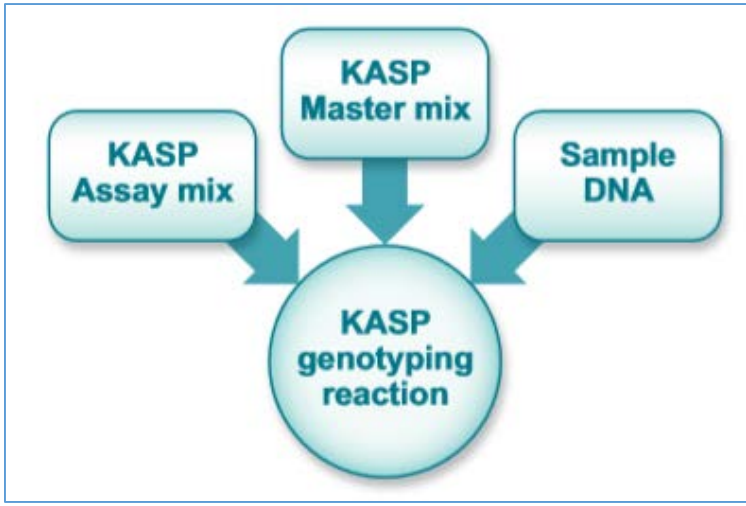
SOME PRELIMINARY RESULTS



SNP	Chromoso	Position	P.value	Rsquare
7569303	5	5124199	5.67E-05	0.238967
7565444	5	2471646	6.26E-05	0.237455
7571201	11	24234037	0.000101	0.230292
7570133	1	39160001	0.000118	0.227909
7566285	16	19672328	0.000134	0.226046
7571641	1	65355185	0.000172	0.222308
7555231	1	43954845	0.000207	0.219558



Study 4. Convert identified SNPs into KASP



1. Assay components
   - SNP-specific KASP Assay mix (2 specific forward+1 Common reverse)
   - universal KASP Master mix (reference dye, taq polymerase, free nucleotides and $MgCl_2$)
   - DNA samples,
2. PCR, followed by an end-point fluorescent read.

- for homozygous SNP, only one of the two possible fluorescent signals will be generated.
- For heterozygous, a mixed fluorescent signal will be generated.

1) Assay components:

KASP uses three components: test DNA with the SNP of interest; KASP Assay mix containing two different, allele-specific, competing forward primers with unique tail sequences and one reverse primer; the KASP Master mix containing FRET cassette plus Taq polymerase in an optimised buffer solution.

A) KASP Assay mix

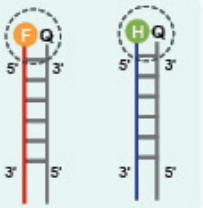
Allele-specific forward primers:

allele-1 5' C 3'

allele-2 5' A 3'

Reverse primer: 3' 5'

B) KASP Master mix



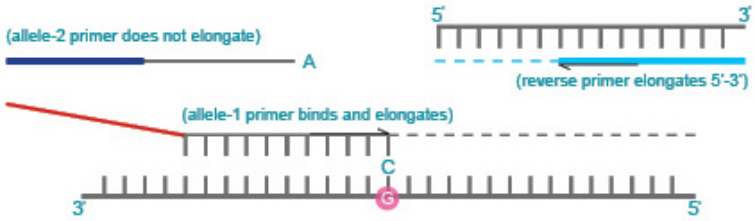
C) DNA template (sample)

5' 3'

3' 5'

[G/T]

2) Denatured template and annealing components – PCR round 1:



In the first round of PCR, one of the allele-specific primers matches the target SNP and, with the common reverse primer, amplifies the target region.

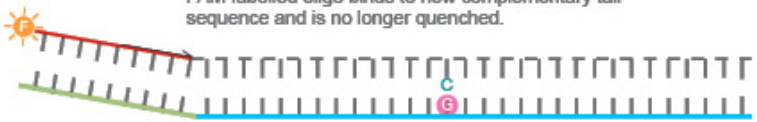
Legend

- Allele-1 tail FAM-labelled oligo sequence
- Allele-2 tail HEX-labelled oligo sequence
- Common reverse primer
- F FAM dye
- H HEX dye
- Target SNP
- Q Quencher

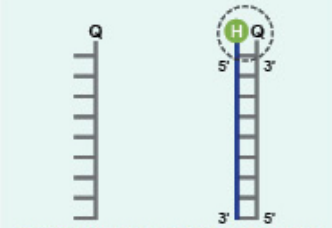
3) Complement of allele-specific tail sequence generated – PCR round 2:

(Reverse primer binds, elongates and makes a complementary copy of the allele-1 tail.)

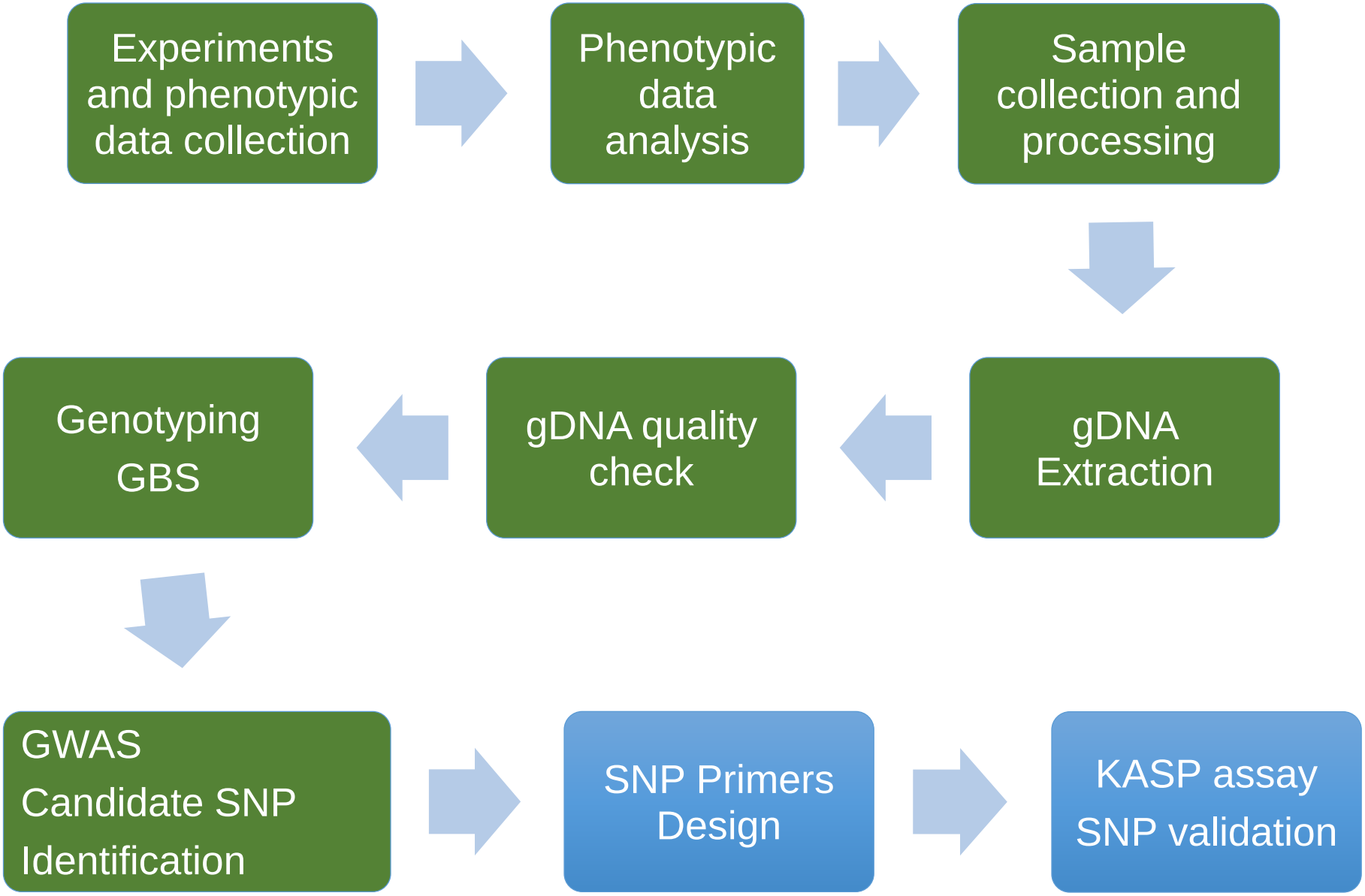
4) Signal generation – PCR round 3:



In further rounds of PCR, levels of allele- specific tail increase. The fluor labelled part of the FRET cassette is complementary to new tail sequences and binds, releasing the fluor from the quencher to generate a fluorescent signal.



❑ WORK PLAN


Experiments and phenotypic data collection
Phenotypic data analysis
Sample collection and processing
gDNA Extraction
gDNA quality check
Genotyping GBS
GWAS Candidate SNP Identification
SNP Primers Design
KASP assay SNP validation

ACKNOWLEDGEMENT

- Donors
  - CIP
  - BecA- ILRI hub
- Makerere University
- ISABU
- Supervisors
- Audience