DNA fingerprinting for estimating varietal adoption

Introduction and overview

CIMMYT MR

Overview of studies

Author	**Labarta et al**	**Hareau et al**	**Maredia et al**	**Maredia et al**	**Maredia et al**	**Stevenson et al**
Crop	**Cassava**	**Potato**	**Cassava**	**Lentil**	**Wheat**	**Maize**
Multiplication	Clonal	Clonal	Clonal	Self-pollinated, with cross-pollination vectored by insects	Self-pollinated (> 95%)	Cross-pollinated
Area (ha)	500k + 30k					
Region	Vietnam & Colombia	Yunnan, China	Ghana	Bihar, India	Bihar, India	Uganda
Sample #	3500 + 434	141 out of 615 HH	917 from 495 plots	880	3,162	416
Markers	SNP	SSR	GBS >> 56,849 SNPs	GBS	GBS	
Cost	US$ 15-20	US$ 50-70 >> US$ 10-20	US$ 30	US$ 50	US$ 50	
Conclusion	Clear identification as vegetatively propagated crop	Why genotyping if visual identification is 97% accurate	11 varietal cluster	No results yet	No results yet	All samples lost
Objective	Confirming adoption	Confirming adoption	Identification	Identification	Identification	Identification

Author	**Stevenson et al**	**Yamano et al**	**Veettil et al**	**Aw-Hassan**	**Maredia et al**	**Kosmowski et a**
Crop	**Maize**	**Rice**	**Rice**	**Lentil**	**Beans**	**Sweet potato**
Multiplication	Cross-pollinated	Self-pollinated	Self-pollinated	Self-pollinated, with cross-pollination vectored by insects	Partially cross-pollination (10-50%)	Clonal
Area (ha)						
Region	Uganda	Bangladesh	India	Bangladesh	Zambia	Ethiopia
Sample #	550	1,289	2,797	Samples from dealers	402	259
Markers	140 SNPs; 10,000 DArT	6k SNP chip	6k SNP chip	ISSR & SSR	66 SNP markers	DArT
Cost					US$ 34	
Conclusion	Different resolution between different number of samples	Distinct allele (Sub1) is easy to recognize, a variety not; what is same what is different	Role of reference library; What is same what is different	8 out of 9 samples matched with reference samples	4-71% of datapoints (or samples) were identified as IVs	Accurate variety identification; large reference library
Objective	Identification	Varieties with a particular trait	Identification	Identification	Identification	Identification

Overview of studies

- **Propagation**
  - **Clonal**: cassava, potato, sweet potato
  - **Self-pollinated** (>98%): wheat, rice, barley, chick peas, pea, groundnuts
  - **Partially cross-pollinated** (5 - 50%): sorghum, lentils, beans, pigeon peas, faba beans, cowpeas
  - **Cross-pollinated**: maize, pearl millet

- **Sample size**: 141 to 3,500 – farmers/dealers; random/clustered in village/clustered in field.

- **Molecular markers**: 60 to 56,849; type of markers.

- **Methodological challenges**: seed *versus* leaf samples, degeneration, type of markers, availability of reference samples

- **Purpose**: use of trait or improved varieties, variety identification

Overview of studies

- **Propagation**
  - **Clonal**: cassava, potato, sweet potato
  - **Self-pollinated** (>98%): wheat, rice, barley, chick peas, pea, groundnuts
  - **Partially cross-pollinated** (5 - 50%): sorghum, lentils, beans, pigeon peas, faba beans, cowpeas
  - **Cross-pollinated**: maize, pearl millet

- **Sample size**: 141 to 3,500 farmers/dealers; random/clustered in village/clustered in field.

- **Molecular markers**: 60 to 56,849; type of markers.

- **Methodological challenges**: seed *versus* leaf samples, degeneration, type of markers, availability of reference samples

- **Purpose**: use of trait or improved varieties, variety identification

A good basis to assess what works, what does not work

CIMMYT

“Same study” – different questions

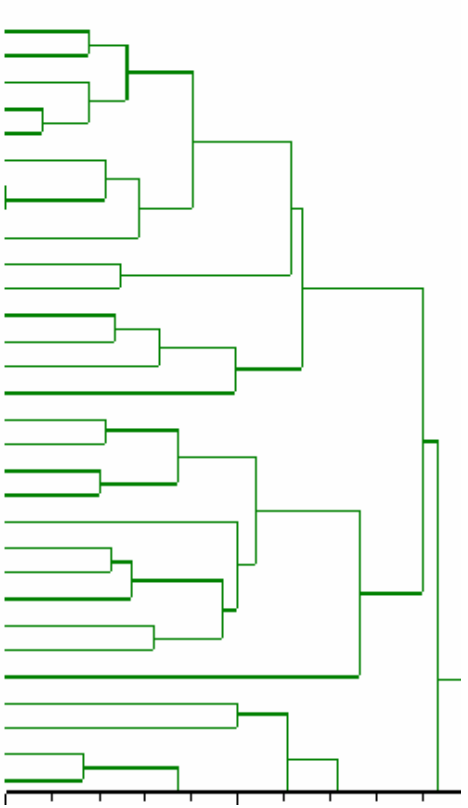
- “Adoption” or “use”?
- How many farmers use
  - Improved varieties?
  - CGIAR/NARS/private sector varieties?
  - Project varieties?
  - A trait?
- How many plants in a field are the actual variety?
- What are relevant questions and why? Why and when are we asking it? Is the methodology appropriate?

“Same study” – different questions

- “Adoption” or “use”?
- How many farmers use
  - Improved varieties? – Reference varieties
  - CGIAR/NARS/private sector varieties? – Reference varieties
  - Project varieties? - after what time frame and seed distribution?
  - A trait? e.g. Sub1 (easier because monogenic) – drought tolerance (very difficult because polygenic)
- How many plants in a field are the actual variety?
  - This is not about establishing adoption or conscious use; this is about establishing seed purity and outcrossing. Sampling within field variation <> more fields.
- What are relevant questions and why? Why and when are we asking it? Is the methodology appropriate?

Need to further define the methodology

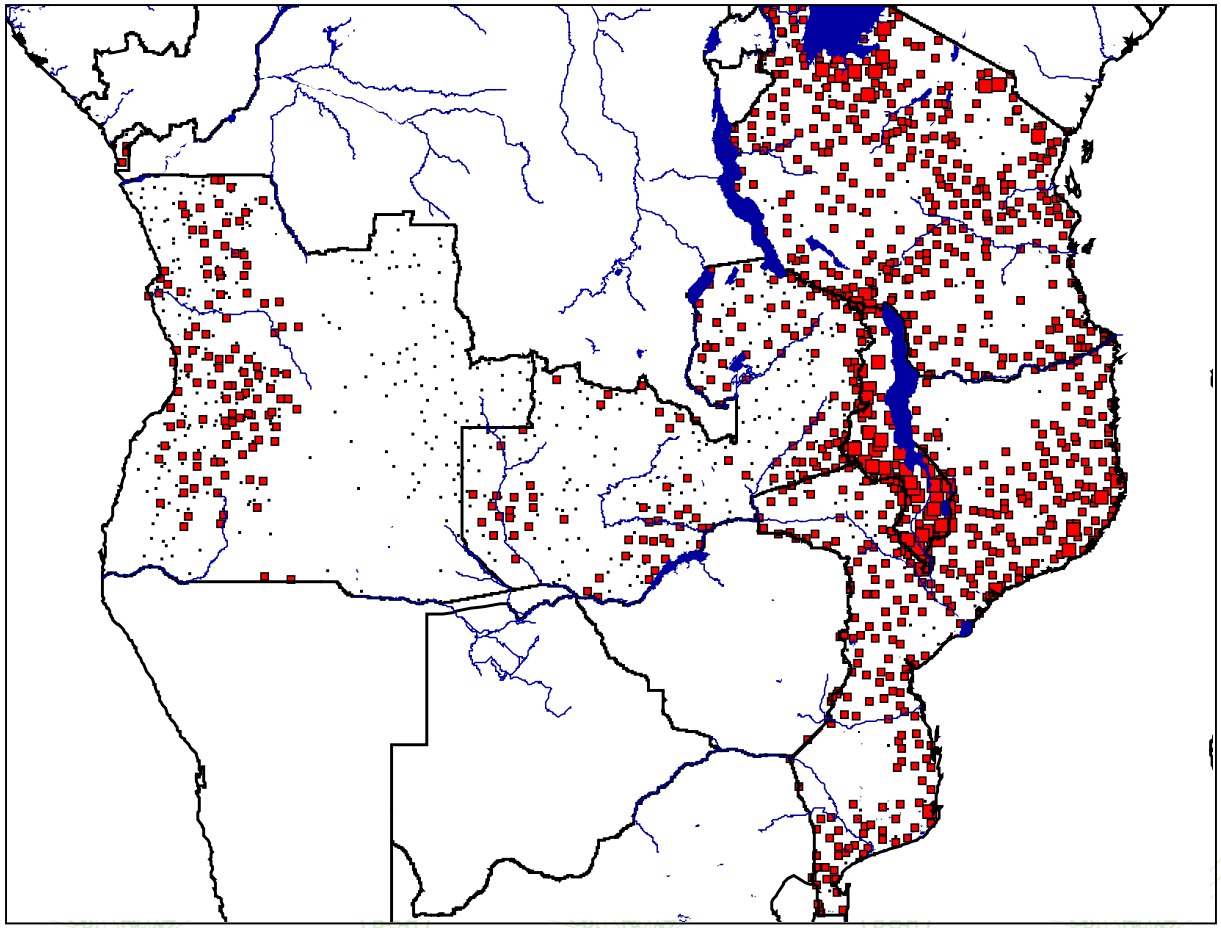
- Sample size 141 to 3,500 out of ??? fields in a district, province, country
  - Predict the confidence interval to set the sampling frequency.
- Molecular markers: 60 to 56,849
  - Optimize the confidence interval of the conclusion <> costs.
- What to we declare to be the same/different?
  - Easier for: clonal, self-pollinated crops, mono-genic traits (e.g Sub1)
  - Difficult for (partially) cross pollinated crops (100%? 90%? 50%?) and polygenic traits (visual, molecular)
  - When does “similar” imply “descending from” or “benefit”?
- Random sampling
  - Seed distribution/sale is not random.



CIMMYT

Seed distribution is not uniform

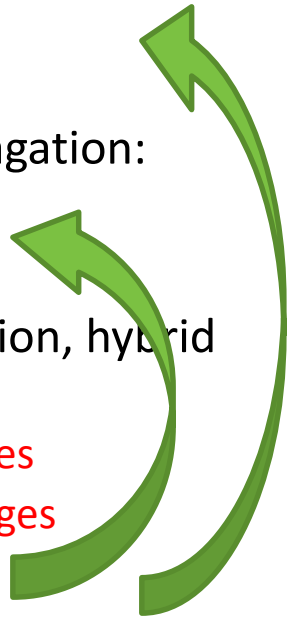
- Predicting maize seed distribution in southern Africa (2004)



CIMMYT

Consider impact pathway

- Reference variety (-ies) or trait
- Formal dissemination: geography & volume over time
- Informal dissemination by using grain as seed, or vegetative propagation: volume over time
- (Mostly unintentional) changes: mixtures, cross pollination, selection, hybrid segregation
  - Clonal/self-pollinated/large field sizes/formal sector >> slower changes
  - Cross-pollinated/small field sizes/informal sector >> more rapid changes
- Change of seed and/or variety: undesirable variety or degenerated variety or loss of seed



Excellence in Breeding Platform (former Genetic Gains Platform)

"Breeders need to work with socio-economists on questions of impact assessment"

- What are relevant/most important questions to ask and why?
- Appropriate methodology = f (purpose, propagation, impact pathway, confidence interval)
- Appropriate conclusion - that stay within the original hypothesis & confidence intervals
- Lots of differently framed use/impact studies are academically interesting but are they of value? >> Aligned studies become more powerful
- Agreeing on the caveats (genetics/socioeconomics) >> Best practices for use/impact studies

CIMMYT