UPOV
International Union for the Protection of New Varieties of Plants

E

Working Group on Biochemical and Molecular Techniques and DNA-Profiling in Particular

BMT/16/17

Sixteenth Session
La Rochelle, France, November 7 to 10, 2017

Original: English
Date: October 13, 2017

USE OF GBS FOR LUCERNE VARIETY DISTINCTNESS

Document prepared by experts from France

Disclaimer: this document does not represent UPOV policies or guidance

The Annex to this document contains a copy of a presentation “Use of GBS for lucerne variety distinction” to be made at its sixteenth session of the Working Group on Biochemical and Molecular Techniques and DNA-Profiling in particular (BMT).

[Annex follows]

Use of GBS for lucerne variety distinction

Julier, B.[1], Barre, P.[1], Lambroni, P.[1], Delaunay, S.[1], Lafaillette, F.[2], Thomasset, M.[3], Gensollen, V.[4]

[1] P3F, INRA, Lusignan, [2] GEVES, Erdre-en-Anjou, [3] BioGEVES, Surgères, [4] GEVES, Montpellier, France

Context of variety registration

2 tests:

- VCU testing
- DUS testing

⟷

Lucerne:

important genetic progress

high proportion of refusal

Difficulty:

- Huge within-variety variation in lucerne (outbreeder, tetraploid)
- Large reference collection (230 varieties) + 30 candidate varieties

Question:

- Could we separate lucerne varieties by using a high number of molecular markers?

Technology : Genotyping by Sequencing (GBS)

Context of variety registration

- Reference collection ~600 varieties, with seeds ~230 varieties
- Registered varieties
  - Europe: 555
  - USA: 167
  - Canada: 280
  - China: 77
  - Argentina: 715
  - ...

INRA SCIENCE & IMPACT

Use of molecular markers to structure the collections

- Not so efficient in previous studies on lucerne (RAPD, AFLP, SSR)
- New technologies to obtain a high number of markers

An emerging technology : Genotyping by Sequencing (GBS) → Elshire et al. 2011

Used on bulks of heterozygous individuals (= lucerne variety) → Raineri et al. 2012

Encouraging results on perenial ryegrass Reasonable costs → Byrne et al. 2013

INRA SCIENCE & IMPACT

Questions

- Method: is the evaluation of allele frequency on bulks of individuals accurate ?
  - On 3 varieties, comparison of allelic frequencies on 4 bulks of 100 individuals each and on 40 individuals that were separately genotyped
- Structure of genetic diversity
  - Is it possible to differentiate among 20 varieties?
  - Is the genetic distance correlated to the phenotypic distance?

INRA SCIENCE & IMPACT

List of 20 varieties

Name	Flower Colour				Dormancy	Verticillium	Ditylenchus	Colletotrichum	Type	Breeder	Registration year
	C1	C2	C3	C4							
Arpège	1	8	2	1	4	7	7	4	N	Florimond Desprez	2004
Artémis	1	8	2	1	4	6	8	9	N	Barenbrug	2010
Barmed *	1	8	2	2	7	5	5	8	S	Barenbrug	2002
Capri	1	7	2	1	4	7	7	4	N	Florimond Desprez	1995
Daphné	1	7	2	2	4	5	7	4	N	Florimond Desprez	1996
Dorine	1	7	3	1	6	3	5	3	S	Barenbrug	2001
Europe	1	8	2	1	4	5	3	RT	N	Florimond Desprez	1961
Fado	1	8	2	1	4	6	8	6	N	Florimond Desprez	2012
Félicie *	1	7	2	2	4	7	8	7	N	GIE Grass	2009
Franken Neu	1	5	1	6	4	3	4	RT	N	Schmitz Ernst	1980
Galaxie *	1	7	2	2	4	5	7	7	N	Semunion	2007
Greenmed	1	7	2	2	1	4	7	RT	N	AgriObtentions	2011
Juuriu	5	1	1	4	no	no	no	RT	N	Estonian Crop Research Institute	1993
Luzelle	1	6	2	3	3	3	6	RT	N	INRA	1993
Meldor	1	6	2	2	6	4	5	4	S	INRA	1995
Midi	1	6	2	2	5	4	5	RT	S	Tourneur	2002
Milky Max	1	7	2	2	4	7	8	6	N	GIE Grass	2015
Orca	1	1	8	1	4	2	4	RT	N	Carneau	1966
Verdor	1	7	2	2	6	5	5	3	S	Barenbrug	2011
Vernal	1	4	4	4	2	no	no	RT	N	Wisconsin and Utah Exp. Stn.	1953

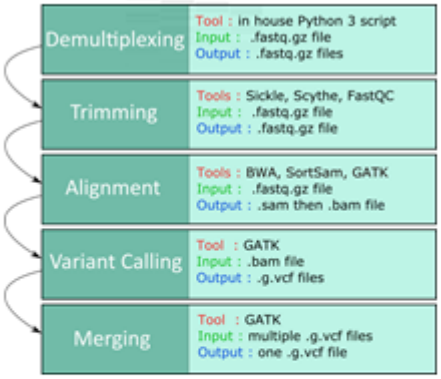
* 3 varieties studied at the individual level
2 very close: Félicia et Galaxie, 1 different: Barmed

INRA SCIENCE & IMPACT

GBS sequencing

- Sowing, leaflet sampling
- DNA extraction on individual plants and bulks
- Building of librairies at INRA Lusignan
  8 lanes, single-end, 100 pb, 10 bulks + 15 individuals/lane
- Sequencing at INRA Genopole Toulouse

Bioinformatics



Each read is attributed to a genotype or a bulk

Cleaning: length, quality

Alignment on *M. truncatula* genome

SNP detection

Merging the filesof all genotypes

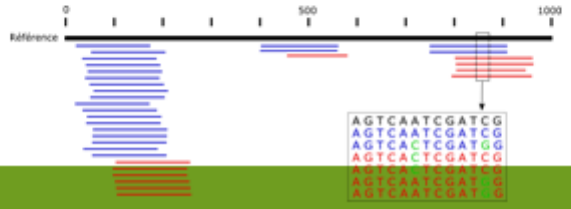
Protocole GNU/Linux

- Scripts : BASH, Python 3.4, AWK
- Programmes :
  - GATK, BWA, Sickle/Scythe,...
  - Instructions within script BASH

Bioinformatics

- Between 3.7 and 4.6 millions reads after trimming et mapping
- For the individuals, the dosage of each SNP is estimated (from 0 to 4 for each allele).
- For the bulks, the SNP are coded depending on their frequency in the bulk :

$$\frac{\text{number of reads carrying the allele}}{\text{total number of reads at this SNP}}$$

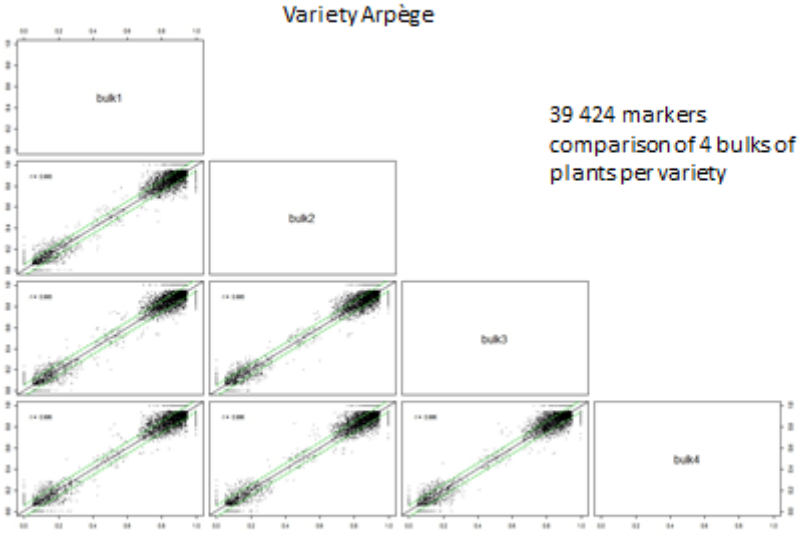


Data analysis

- Repeatability of allelic frequency determined on the 4 bulks
- Comparison of allelic frequency calculated on the average of 40 individual genotypes and on bulks
- Differentiation with GBS markers:
  - Principal component analysis (PCA) on allelic frequencies,
  - F_{ST} and Nei distances among varieties
- Correlation with phenotypic distances
  - Calculation of a GAIA distance between 2 varieties: sum on all traits of a weighted difference
  - Correlation and Mantel test of GAIA distance with F_{ST} and Nei distances

Repeatability of allelic frequency determined on the 4 bulks

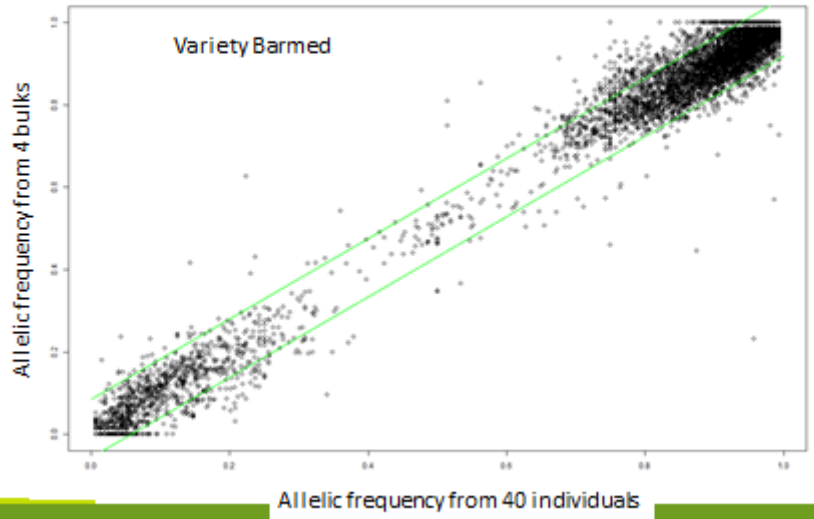
Variety Arpège



39 424 markers
comparison of 4 bulks of 100 plants per variety

INRA

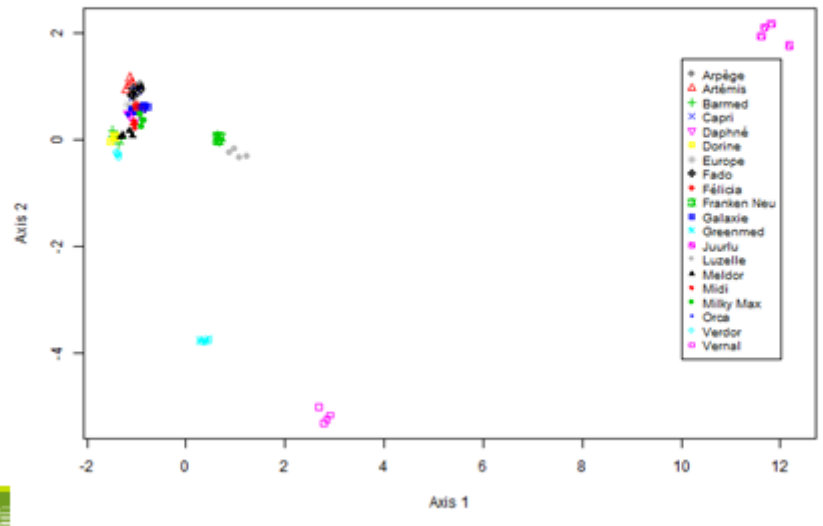
Comparison of allelic frequency calculated on the average of 40 individual genotypes and on bulks



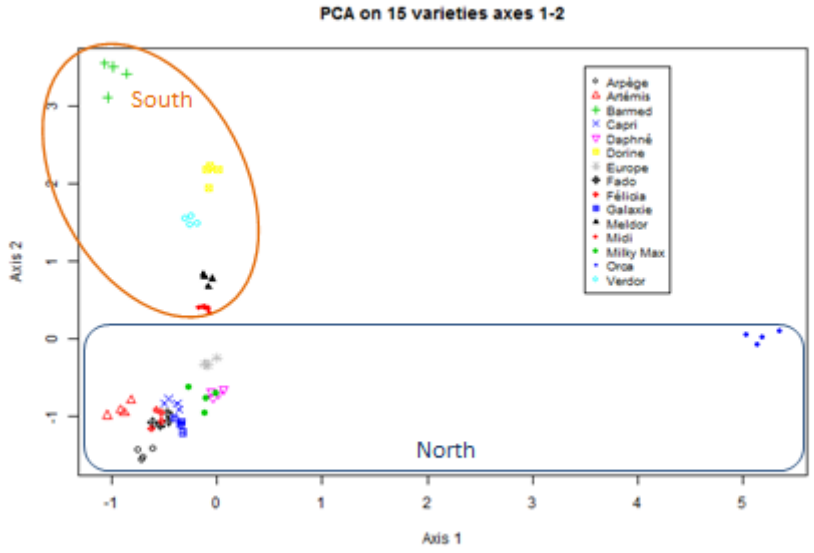
INRA

Distinctness of 20 varieties for which 4 bulks have been genotyped

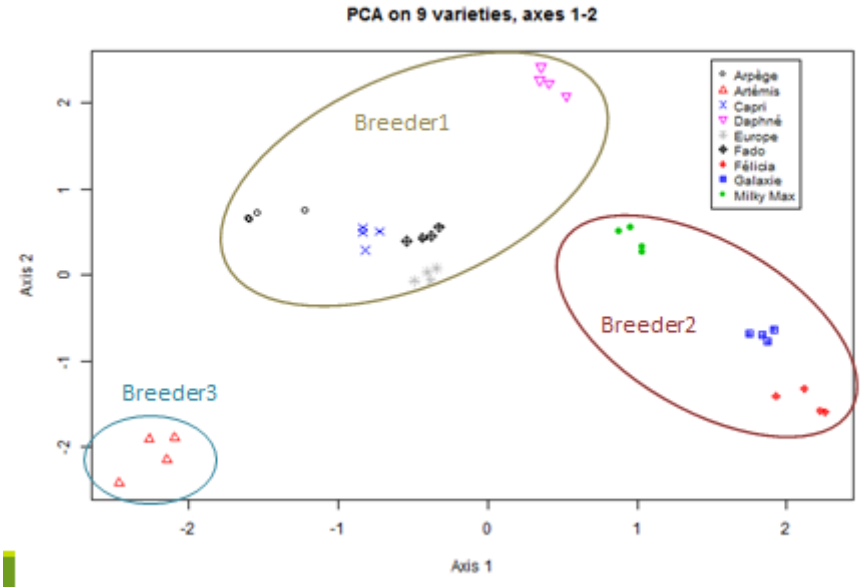
PCA on 20 varieties



Zoom on the 15 French varieties



Zoom on the 9 North varieties



Test of significance

- F_{ST} between populations (significant at P < 0.95)

	Arpège	Artémis	Barmed	Capri	Daphné	Dorine	Europe	Fado	Félicia	Franken Neu	Galaxie	Greenmed	Juurlu	Lusalle	Maldor	Midi	Milky Max	Orca	Verdor
Artémis	0.079																		
Barmed	0.093	0.093																	
Capri	0.079	0.079	0.090																
Daphné	0.081	0.084	0.094	0.080															
Dorine	0.090	0.089	0.094	0.095	0.087														
Europe	0.079	0.079	0.088	0.075	0.079	0.084													
Fado	0.079	0.080	0.092	0.079	0.080	0.089	0.077												
Félicia	0.084	0.084	0.095	0.081	0.084	0.090	0.079	0.082											
Franken Neu	0.090	0.091	0.100	0.087	0.090	0.095	0.085	0.090	0.091										
Galaxie	0.083	0.084	0.094	0.081	0.083	0.089	0.080	0.081	0.081	0.090									
Greenmed	0.116	0.116	0.120	0.114	0.114	0.116	0.113	0.115	0.117	0.115	0.116								
Juurlu	0.189	0.192	0.203	0.189	0.192	0.199	0.190	0.189	0.190	0.171	0.188	0.192							
Lusalle	0.107	0.106	0.115	0.103	0.108	0.111	0.102	0.105	0.107	0.106	0.108	0.125	0.175						
Maldor	0.088	0.089	0.094	0.085	0.087	0.092	0.084	0.087	0.089	0.094	0.088	0.117	0.197	0.110					
Midi	0.081	0.081	0.088	0.077	0.081	0.083	0.075	0.080	0.081	0.085	0.081	0.111	0.190	0.102	0.084				
Milky Max	0.079	0.081	0.092	0.079	0.078	0.085	0.077	0.079	0.080	0.087	0.078	0.111	0.188	0.104	0.085	0.078			
Orca	0.095	0.096	0.104	0.092	0.093	0.097	0.089	0.094	0.096	0.099	0.094	0.123	0.197	0.114	0.098	0.090	0.091		
Verdor	0.088	0.087	0.091	0.085	0.086	0.085	0.083	0.087	0.087	0.096	0.087	0.114	0.200	0.109	0.089	0.081	0.082	0.097	
Vermil	0.118	0.120	0.127	0.117	0.119	0.123	0.117	0.119	0.119	0.111	0.118	0.130	0.165	0.125	0.122	0.115	0.114	0.126	0.121

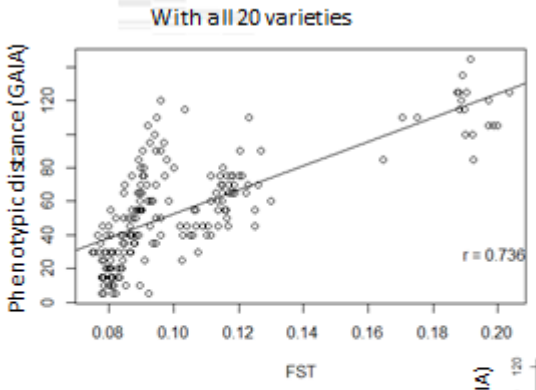
Correlation with phenotypic distances

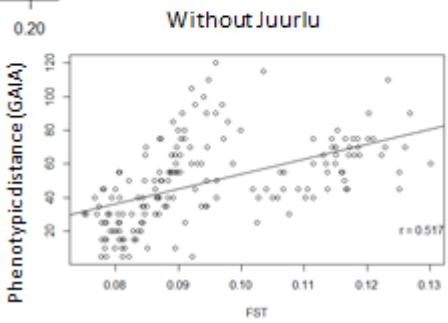
- 2 methods to test difference among varieties:
  - CoyD: calculated for each quantitative trait (+ χ^2 for qualitative traits)
  - GAIA: distance between a pair of varieties resulting from the addition of the weightings of all traits
- CoyD and χ^2 presently used in DUS testing; GAIA will be used from 2018 onwards on lucerne

GAIA distances

	Anpilgo	Antarès	Barmed	Capri	Daphné	Dorine	Europe	Fado	Fâlicia	Franken Neu	Galaxie	Greenmed	Juuru	Luzelle	Maldor	Midi	Milky Max	Orca	Verdor
Antarès	30																		
Barmed	60	45																	
Capri	5	25	55																
Daphné	15	35	60	20															
Dorine	55	65	35	50	40														
Europe	25	45	55	30	30	40													
Fado	15	20	60	10	25	55	30												
Fâlicia	20	25	70	15	35	60	40	5											
Franken Neu	80	80	80	75	55	40	50	70	75										
Galaxie	15	25	50	10	20	40	30	5	15	65									
Greenmed	55	75	75	60	45	50	65	65	65	80	55								
Juuru	115	145	125	120	100	105	115	135	125	110	125	85							
Luzelle	45	55	60	40	30	45	45	40	55	40	30	45	110						
Maldor	50	65	35	45	40	5	40	55	55	35	40	45	105	40					
Midi	40	55	35	35	20	15	30	45	55	35	30	60	100	25	20				
Milky Max	20	25	70	15	25	70	40	5	10	75	15	75	125	45	65	45			
Orca	110	120	115	105	95	75	85	100	90	60	90	110	120	75	85	75	90		
Verdor	35	50	25	30	30	10	30	40	50	50	30	55	105	45	10	10	50	95	
Vermel	65	90	90	70	65	70	45	75	70	40	75	60	85	55	65	70	70	70	75

Correlation of GAIA distance with phenotypic distances





Conclusion

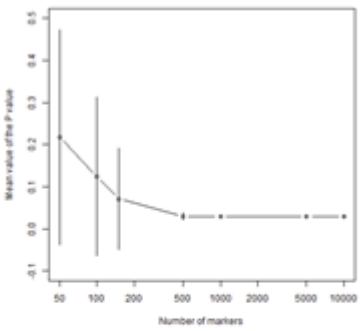
- Efficiency of GBS to genotype individuals or bulks on a heterozygous tetraploid species
- Accuracy of the estimation of allelic frequencies on bulks
- Significant difference between each pair of varieties
- Structure of the diversity is consistent to our knowledge of the varieties
- Good correlation between genetic distance and phenotypic distances

Conclusion

- Use of GBS markers to help DUS testing?
  - To structure the collection of varieties
    - A candidate variety is tested phenotypically with the varieties with small genetic distance
    - Less field plots / experiments
    - More precise trials
  - Use of molecular markers only
    - Good correlation with phenotypic traits
    - No effect of environment

INRA SCIENCE & IMPACT 21

Conclusion



- More studies needed
  - To optimize GBS protocole number of markers / cost
  - To estimate genetic distances
    - between candidate varieties that were not distinct in DUS tests,
    - between seed lots of the same variety, etc
  - More varieties to be considered: reference collection

INRA SCIENCE & IMPACT 22

[End of Annex and of document]