GENOME WIDE ASSOCIATION MAPPING OF AGRO-MORPHOLOGICAL AND DISEASE RESISTANCE TRAITS IN SUGARCANE

M. Gouy, Y. Rousselle, A. Thong Chane, A. Anglade, S. Royaert, S. Nibouche and L. Costet

Gouy et al. 2015 Euphytica



Marker assisted selection

Several QTLs studies for sugarcane

- Mainly based on biparental populations
  - Revealed many QTLs with small effect
  - QTL often dependant of the environment
  - not validated in other genetic background

MAS for sugarcane?

- Only one diagnostic marker available for *Bru*1 a brown rust resistance gene
- Costet *et al.* 2012; Glynn *et al.* 2013, Racedo *et al.* 2013

other approaches

Genome Wide Association Studies

Genomic selection

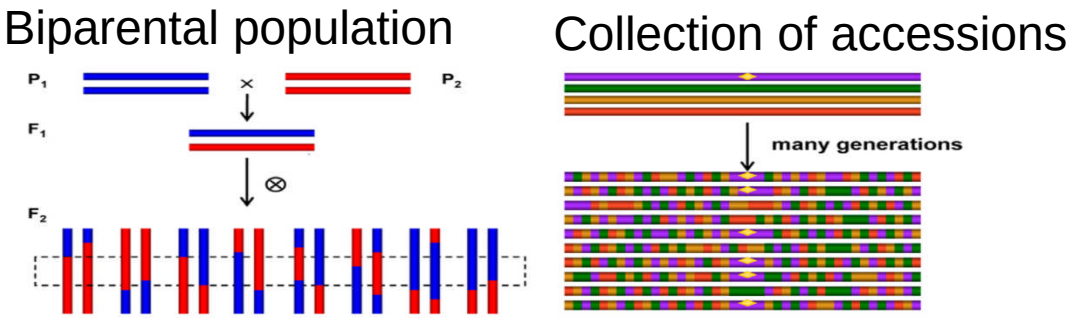


Genome Wide Association Studies (GWAS)

Molecular marker / trait association in collection of accessions

Avantages

- Based on available collections
- Multi traits phenotyping
- Allow the access to a greater allelic diversity
- Greater mapping resolution
  - Occurence of more events of recombinaison



Zhu *et al.* 2008



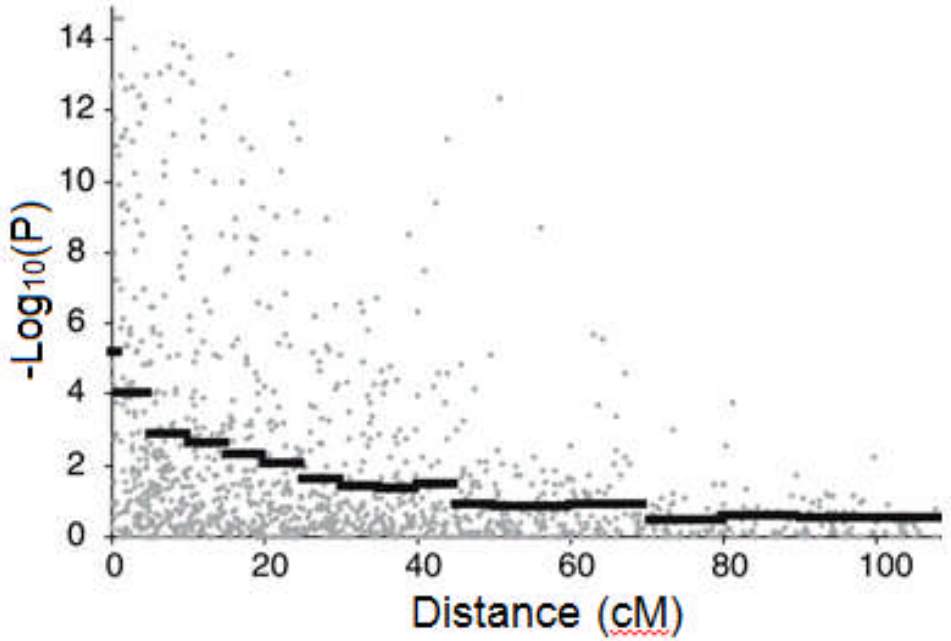
Prerequisites for GWAS

Knowledge the extension of the Linkage Disequilibrium

- it conditions the marker density needed

LD extent up to 5 cM (Jannoo *et al.* 1999, Raboin *et al.* 2008)

- 3000 to 6000 markers (Raboin *et al.* 2008)





Prerequisites for GWAS

Need to manage population structuration

- Induce detections of false associations

->But taking into account population structure will also induce false negatives

- Lost of QTLs (Breseghello et Sorrells, 2006, Zhao *et al.* 2011)



Objectives

- Evaluation of the impact of population structure on the phenotypic variability on a diversity panel of sugarcane cultivars

- Impact of population structuration management on identification of markers associated with 13 traits using an association mapping approach



Plant material

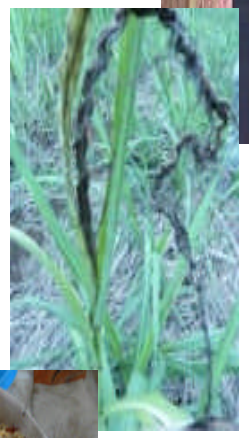
Panel

- 183 accessions

Phenotyping

- 13 traits
  - Yield components
    - Stalk number, diameter and height
    - sucrose content, Brix,
  - Disease and insect resistance
    - Rust, smut, Gumming, SCYLV, aphids
  - Fiber quality
    - Bagasse content and digestibility
  - Flowering rate









Genotyping

- DArT, AFLP et SSR
  - 3327 markers



BLUPs

Predict vectors of genetic values

$\hat{g}$ = BLUP

- $\hat{g}$ estimation using **mixed linear model** for gaussian distributed traits

- for non–Gaussian distributed traits $\hat{g}$ estimation using **Markov chain Monte Carlo (MCMC) routines to fit multi-response generalized linear mixed models** four disease-related traits and the flowering rate

R package MCMCglmm (Hadfield 2010)

QTL detection on $\hat{g}$



Description of population structuration

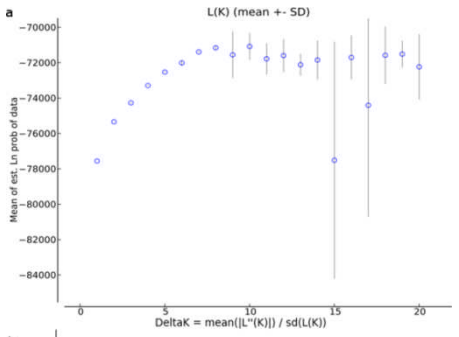
Subset of 820 independants DArT markers

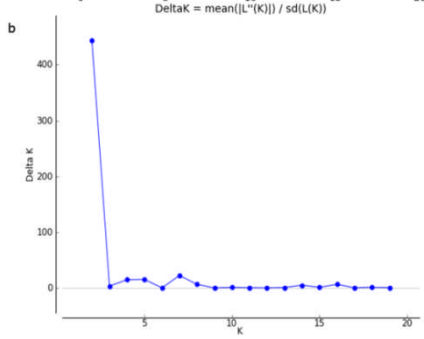
- Exact test of Fisher (P > 0.05)

Population structure

Bayesian methods STRUCTURE
(Pritchard *et al.* 2000)

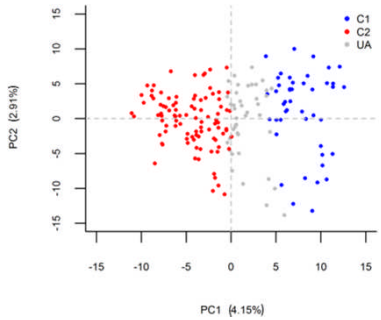
- Panel is most likely structured in 2 or 7 clusters (Evanno *et al.* 2005)
  - Clusters affiliation proportions
    - **Q2 matrix**
    - **Q7 matrix**





PCA
(Patterson *et al.* 2006)

- 18 significant PCs (Test de Tracy-Widom)



Familial relatedness

Kinship matrix of pairwise kinship coefficients

→kinship statistic cannot be computed for polyploids

- But genetic similarity is correlated with the coefficient of parentage based on pedigree data

(Lima et al. 2002;Plaschke et al. 1995; Tinker et al. 1993).

K = Similarity matrix using Jaccard's similarity coefficient
(Yu *et al.* 2006)



Impact of population structure on the phenotypic variability

- Q2→ R^2 [0 - 0.15] significant effect on 7 out 13 traits
- Q7→ R^2 [0.08 -0.36] significant effect on all traits
- PC → R^2 [0.16 -0.53] significant effect on all traits

Population structure explains a significant part of phenotypic variability



Models for QTL detection

- TASSEL (Bradbury *et al.* 2007)

$$\hat{g} = M\alpha + e$$ → **Naive GLM**

$$\hat{g} = M\alpha + Q\beta + e$$ → **GLM-Q2 or Q7**

$$\hat{g} = M\alpha + C\delta + e$$ → **GLM-PC**

$$\hat{g} = M\alpha + Q\beta + Zu + e$$ → **MLM-Q2 or Q7**

$$\hat{g} = M\alpha + C\delta + Zu + e$$ → **MLM-PC**

$$u \sim N(0, \mathbf{K}\sigma^2_g)$$

K = Similarity matrix using Jaccard index (Yu *et al.* 2006)



Type I error managment

Type I error (false positives)

- Due to multiple tests
  - False Discovery Rate (Benjamini et Hochberg 1995)
    - Fixed to10 %
    - 10 % of the significant marker-trait associations are false positives



Type I error managment

Erreur de type I (faux positifs)

- Due to multiple tests
- Due the genetic structure
  - Measure of inflation of test statistics (Devlin et Roeder 1999)
    - Inflation facteur **λ**

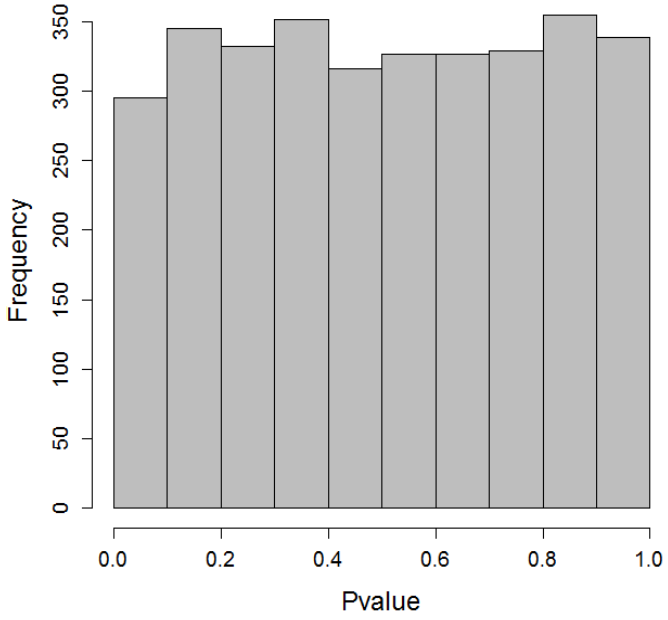
$$\lambda = \text{median}(F_{obs}) / \text{median}(F_{theo})$$

  - Structure controled if $\lambda < 1.05$ (Price *et al.* 2010)



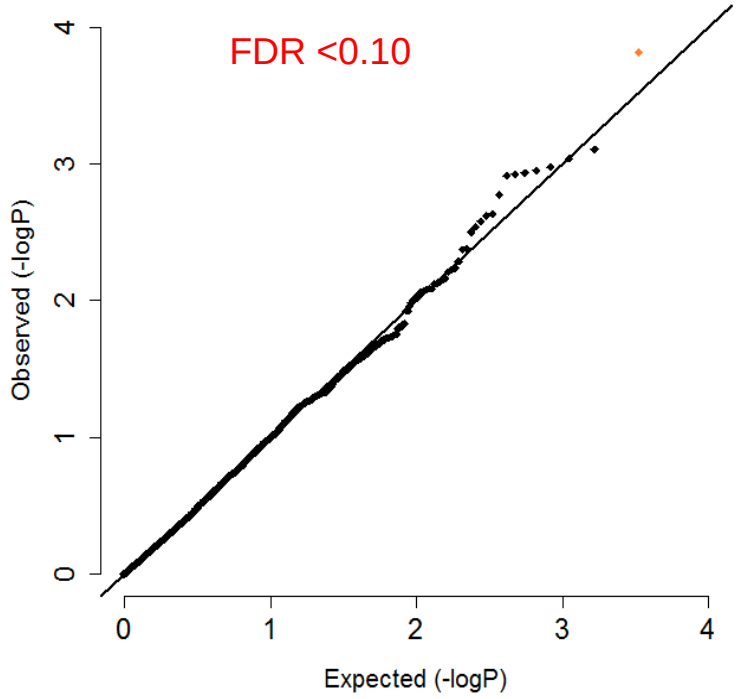
If no population structure

- Uniform distribution of the probabilities
  - Only few significant makers are expected



Lambda ~ 1

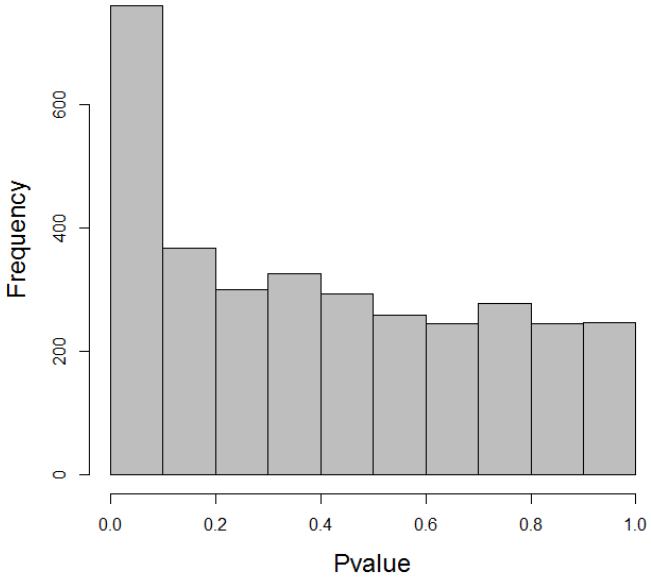
FDR <0.10

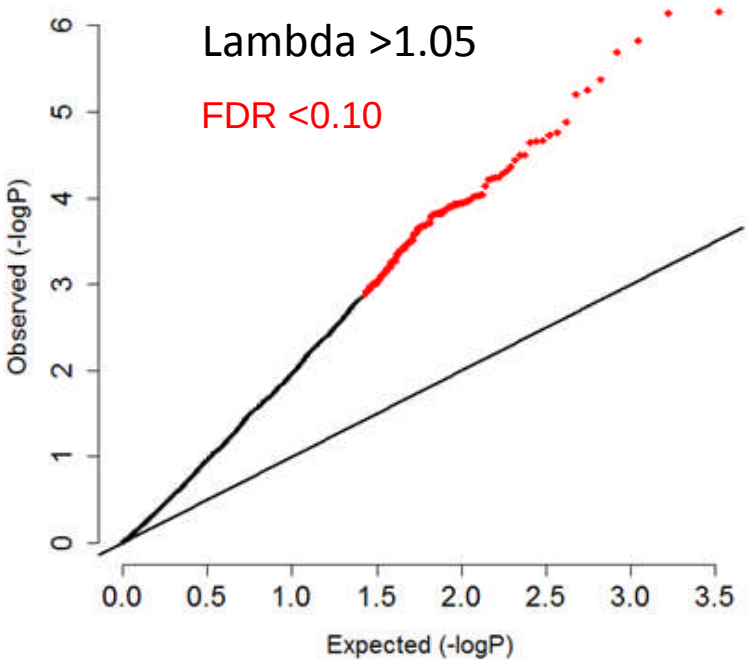




With population structure

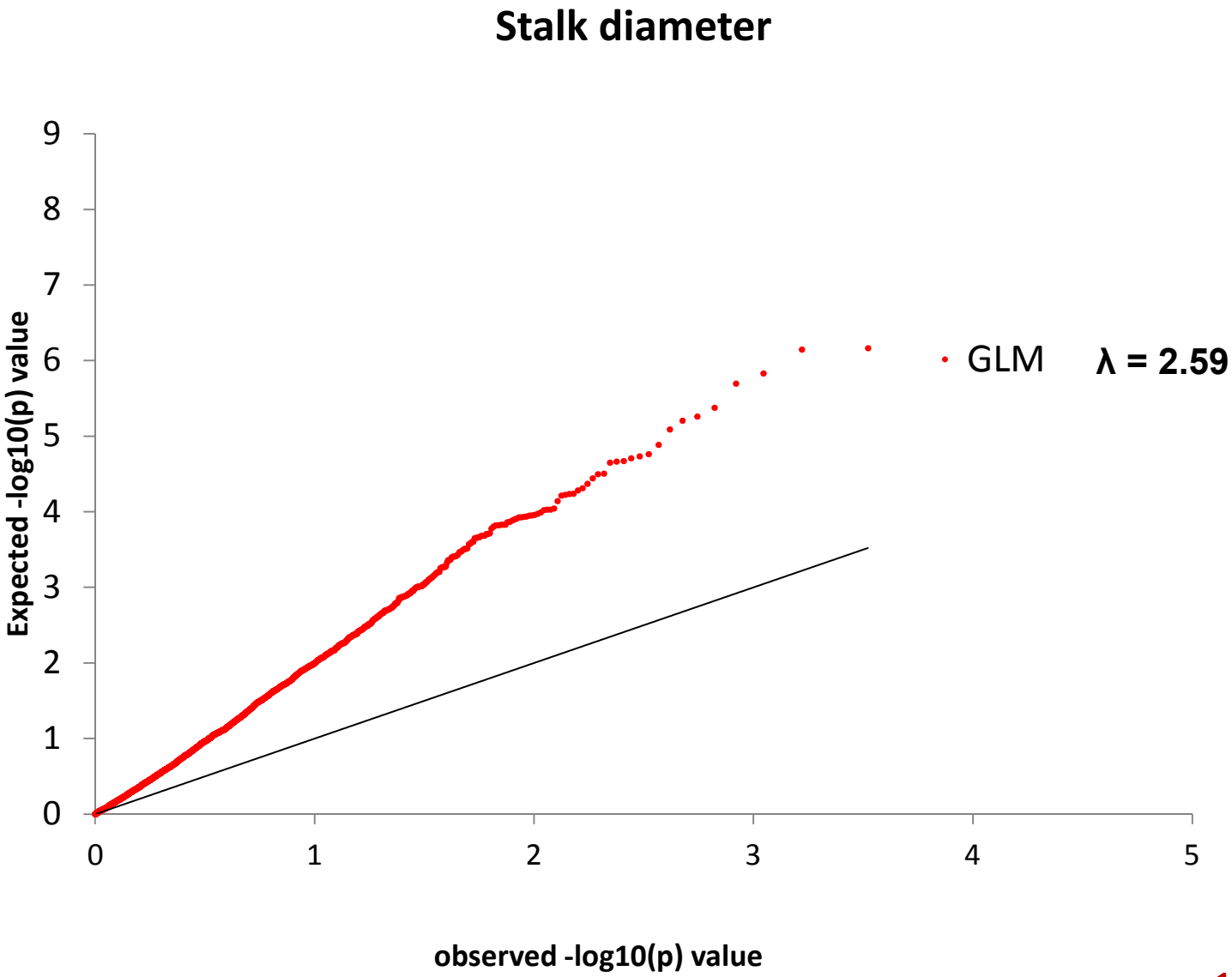
- High rate of false positives
  - Inflation of test statistics
  - Excess of low P values





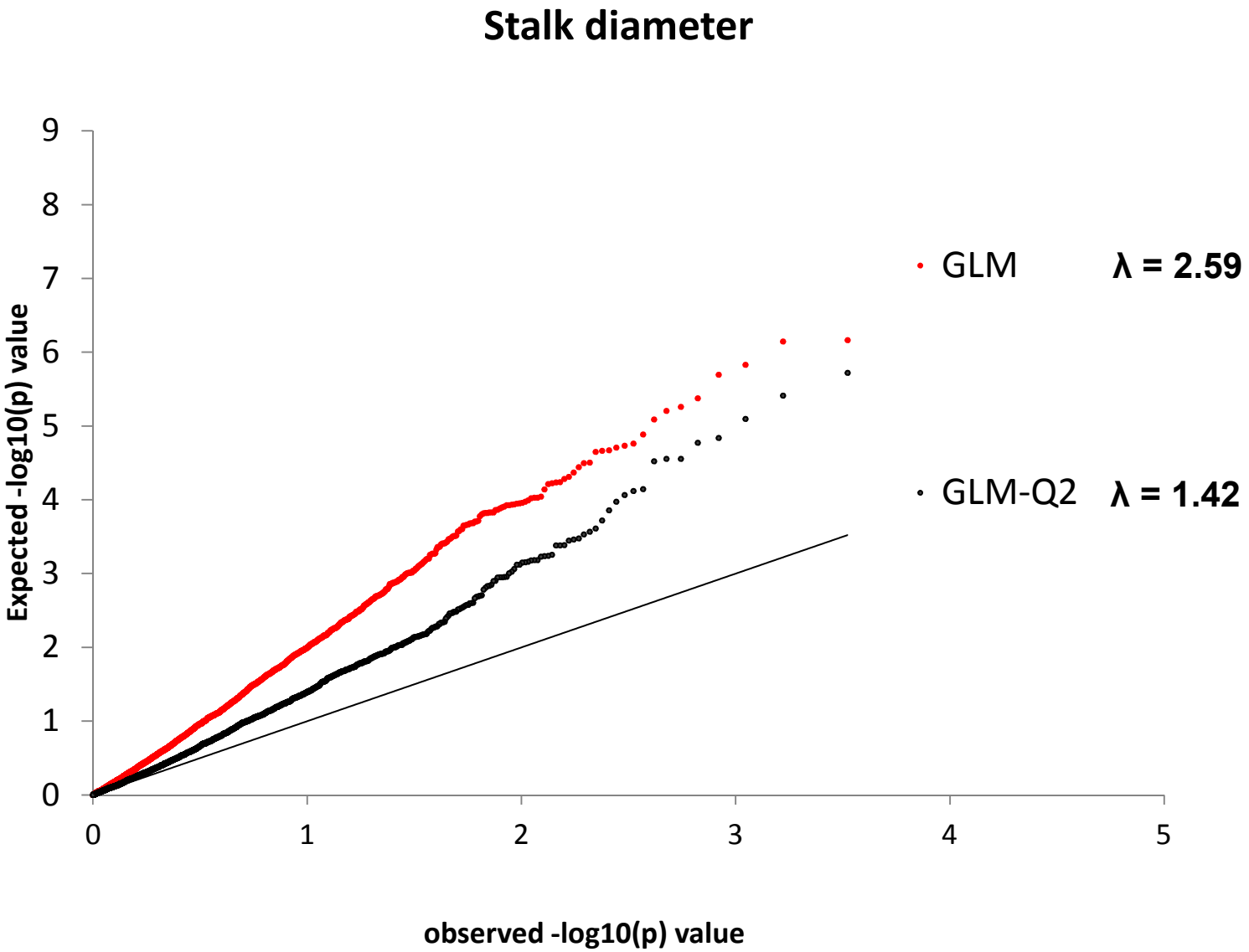


Test statistics inflation



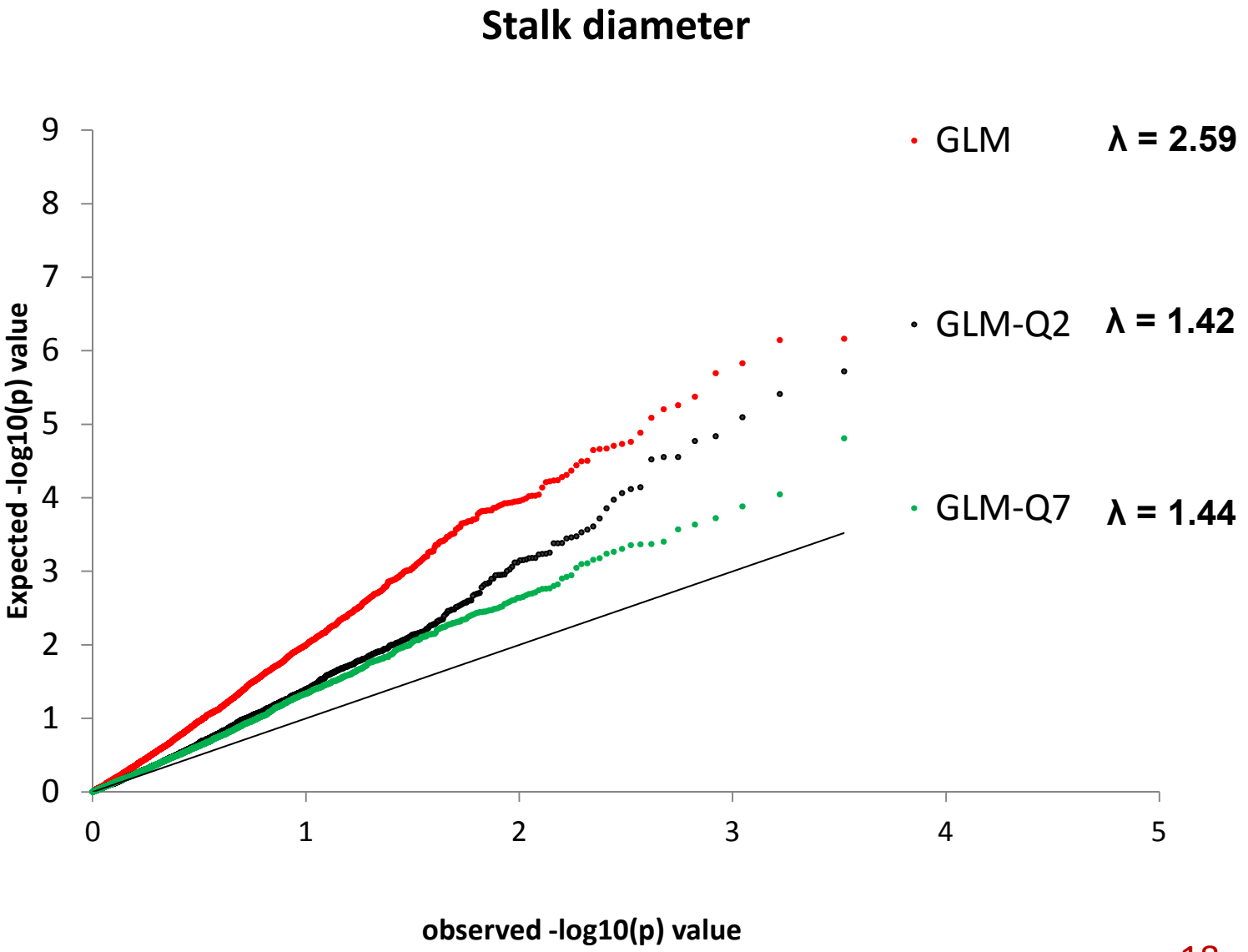


Test statistics inflation



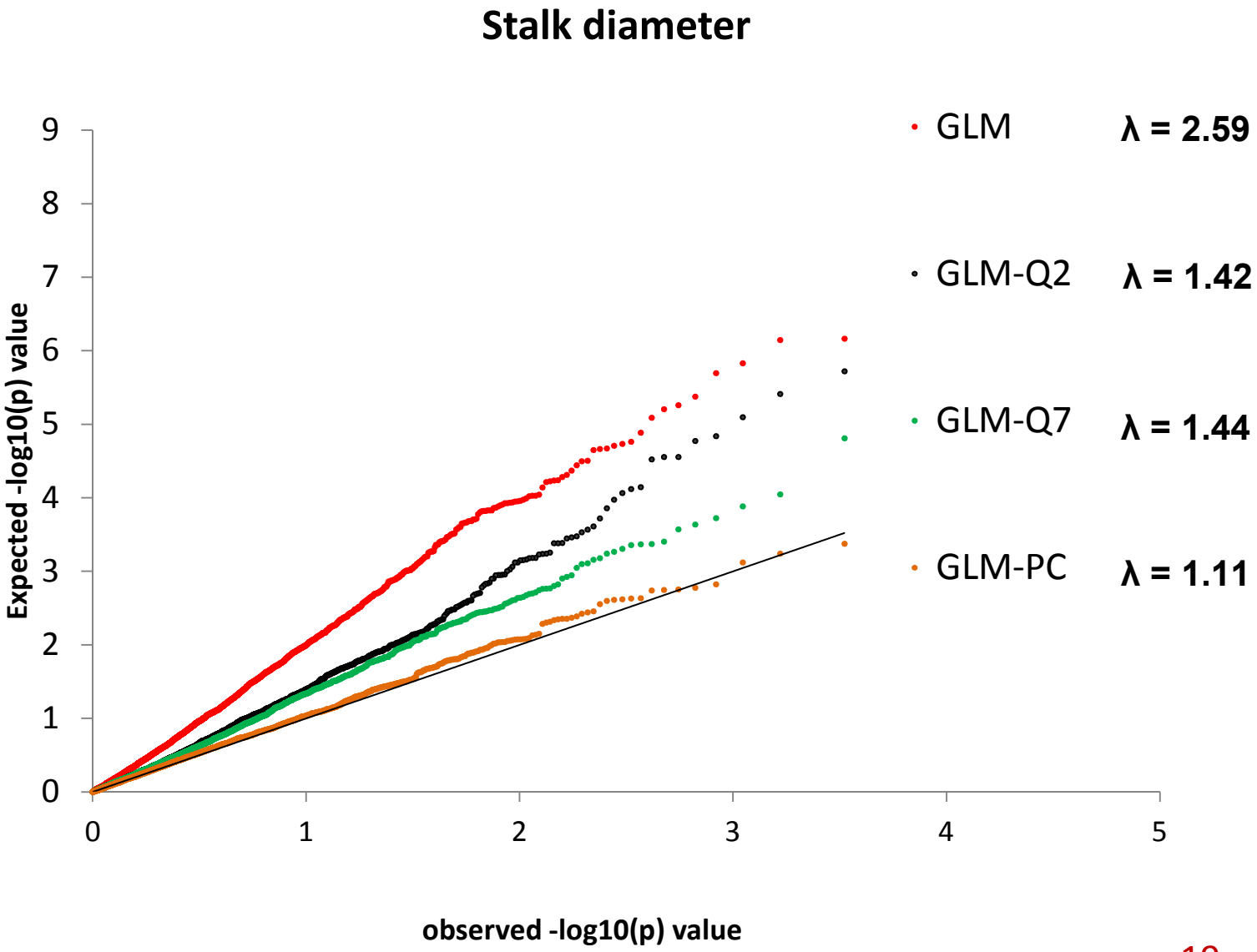


Test statistics inflation



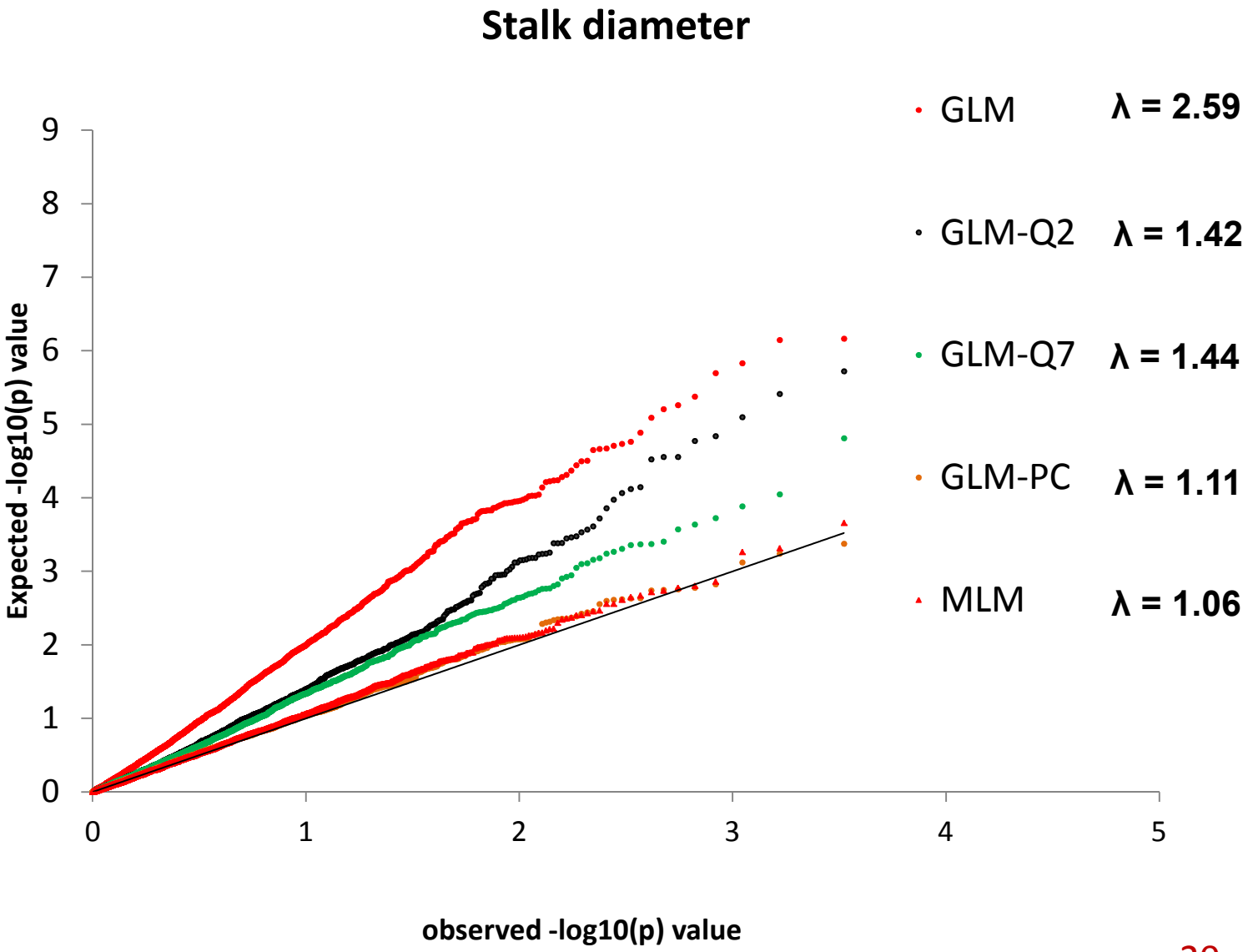


Test statistics inflation



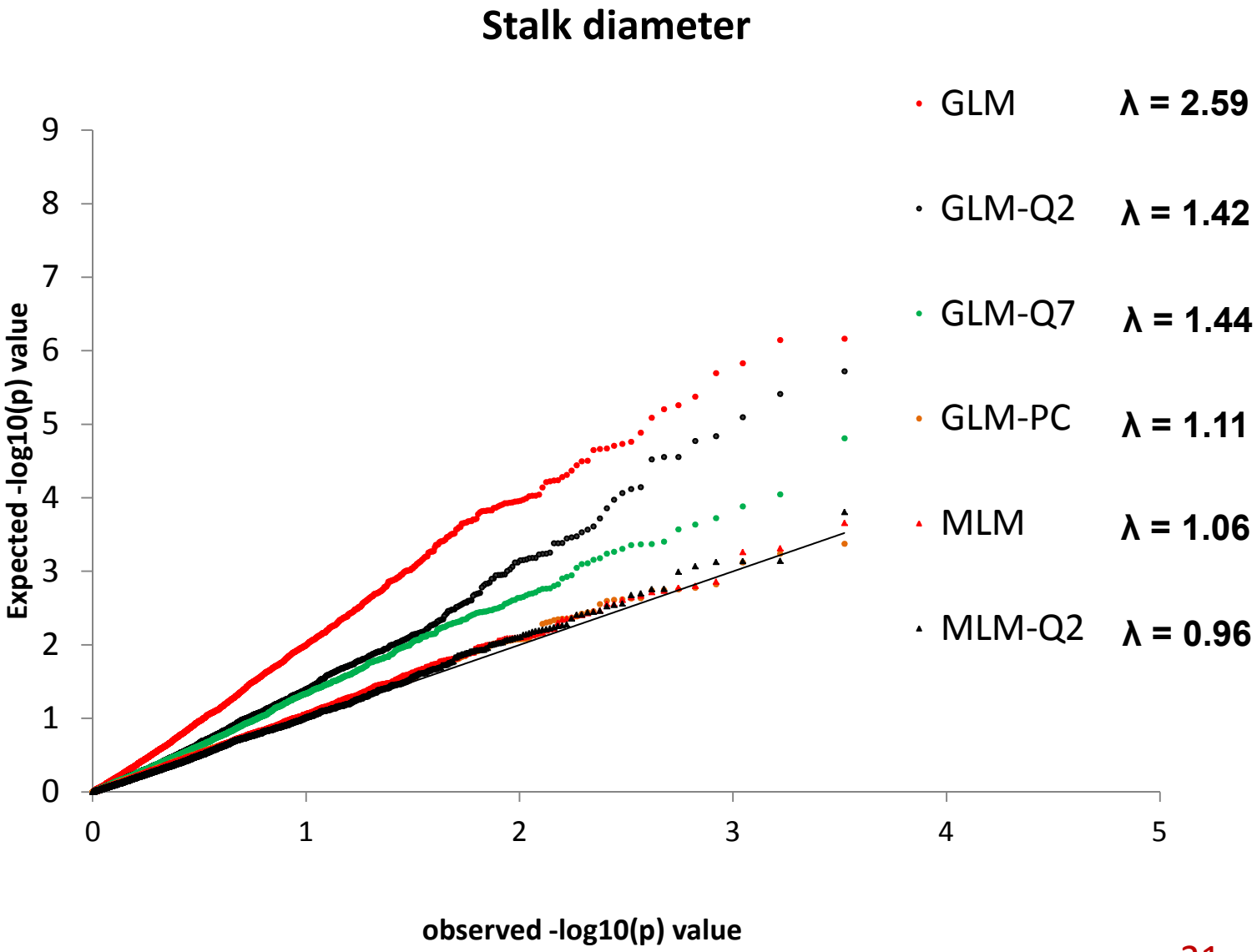


Test statistics inflation



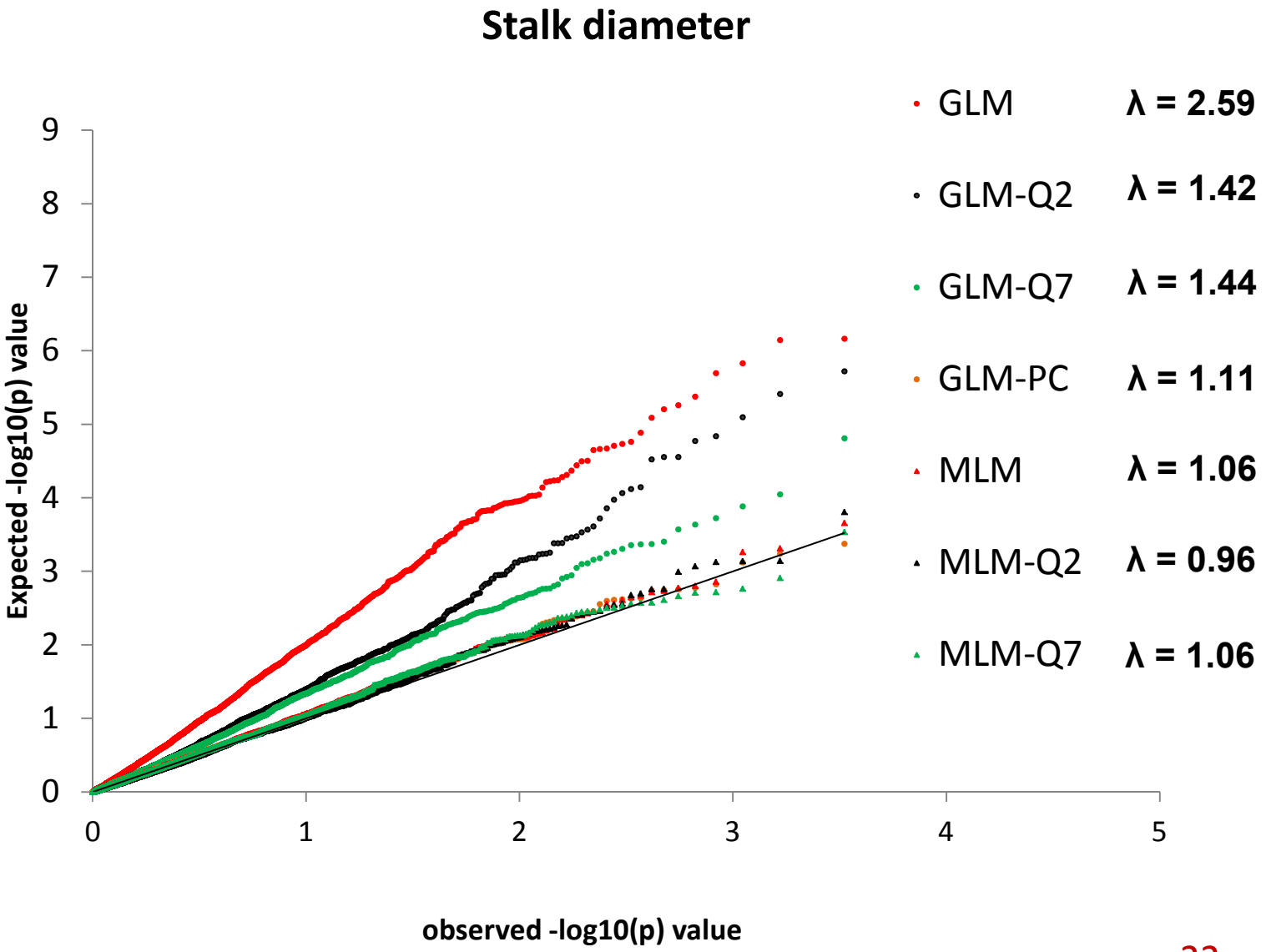


Test statistics inflation



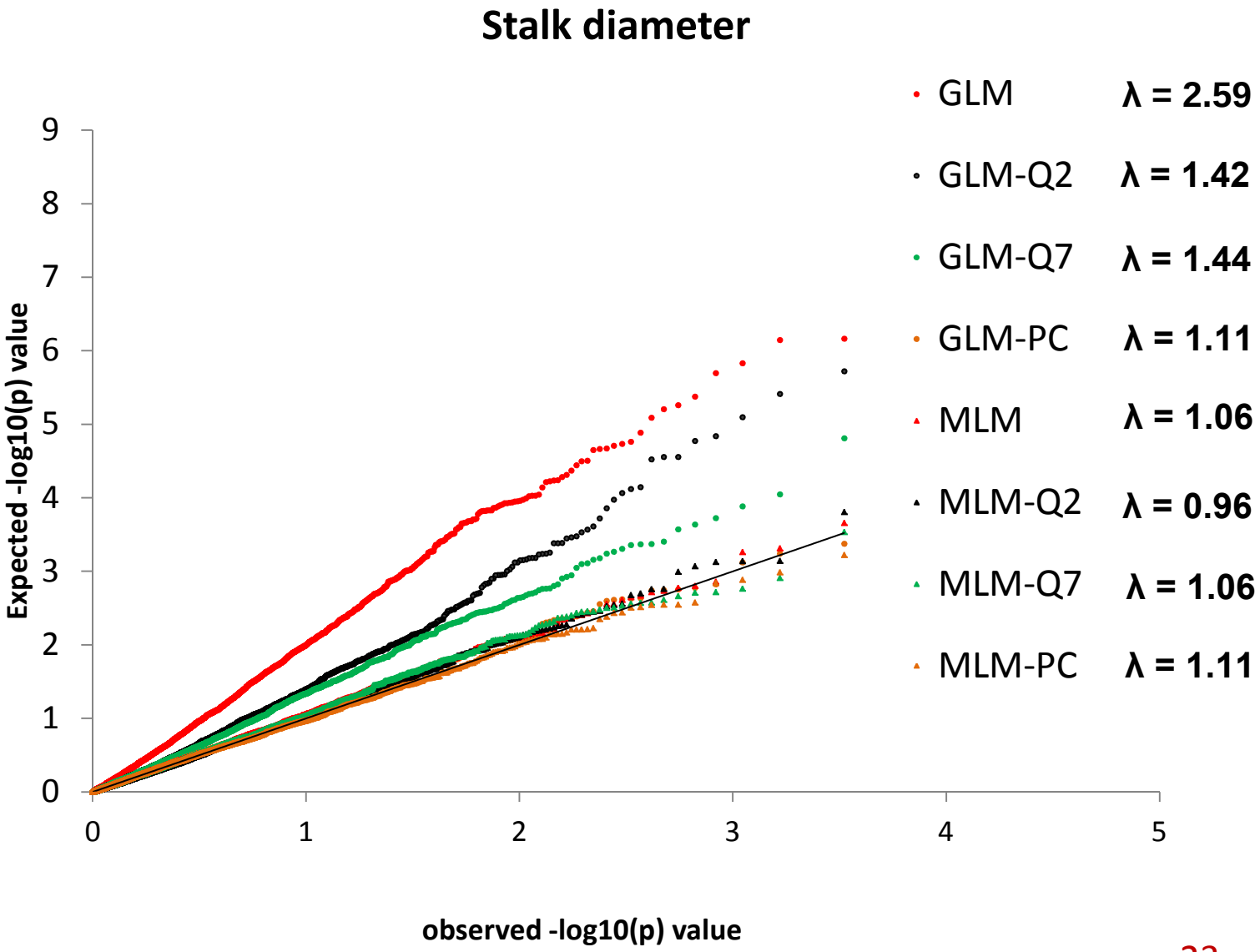


Test statistics inflation





Test statistics inflation





Inflation factors control (λ) result

General linear models

- No control if only the Q matrix
- better control using PC

Mixed models

- improved control
- MLM-PC > MLM-Q> MLM

Models	Number of traits with λ < 1.05
Naive	0/13
GLM-Q2	0/13
GLM-Q7	0/13
GLM-PC	5/13
MLM	5/13
MLM-Q2	7/13
MLM-Q7	7/13
MLM-PC	9/13



QTL detection results

Effect of genetic structure

- If not taken into account
  - Many markers detected
- If taken into account
  - Only few markers detected
- Detections using $\lambda < 1.05$
  - 3 traits
  - 11 significants markers

Models	Number of detected QTLs per traits
Naive	0 - 526
GLM-Q2	0 - 379
GLM-Q7	0 - 23
GLM-PC	0 - 10
MLM	0 - 2
MLM-Q2	0 - 2
MLM-Q7	0 - 2
MLM-PC	0 - 3



QTL detection results

- Maximum of signicant markers detected with GLM-PC
- QTL with small R^2
- Most of the QTLs associated with rust resistance
  - 6 markers linked with *Bru*1 ($P<4x10^{-03}$)

Traits	GLM-PC	MLM	MLM-Q2	MLM-Q7	MLM-PC	R^2
Sucrose yield	1	0	0	0	0	0.07
Flowering rate	2	0	0	0	1	0.07 - 0.11
Rust resistance	8	2	2	1	3	0.07 - 0.13

R12H16_PCR *Bru*1 diagnostic marker R^2 [0.46-0.52]



Discussion

GWAS is feasible for sugarcane

- Re-detection of *Bru*1 locus using random markers

High inflation of test statistics due to population structuration

- Controled by modeling parentage
  - Distance matrix could be a good alternative to kinship matrix

Only few QTLs detected

- lack of statistical power
- Traits control by numerous QTLs?
- Polyploid context, gene/allele expression depending of the genetic background?



Perspectives

Increasing the power of detection

- Modeling using current available data
- Increasing panel size ?
- Increasing marker density?
- Meta-analyse ?



Thanks







GWAS and sugarcane

- population structure not taken into account
  - McIntyre *et al.* (2005)
  - Raboin (2005)
  - Butterfield (2007)
    - Disease resistance

- population structure taken into account
  - Wei *et al.* (2006 & 2010)
    - Disease resistance and cane yield

Few studies - few traits

No diagnostic of the effect on QTL detection when genetic structure was taken into account



Trials

Trial	Unit area (m²)	Year	Sucrose yield	Stalk diameter	Stalk number	Stalk height	Bagasse content	Brix	Floraison rate	In-vitro NDF digestibility	SCYLV incidence	Brown rust score	Gumming Score	Smut incidence	Aphid AUIPC
La Mare	4.5	2010	■	■	■	■	■	■	■						
Vue-Belle	4.5	2010	■	■	■	■	■	■	■						
Bassin Martin	3.75	2007								■		■			
		2008													■
		2009								■	■				
		2010													
		2012											■		
Le Gol	4.5	2007												■	
		2008												■	
		2009												■	
Menciol	3.75	2012											■		

■ available data



BLUPs

Predict vectors of genetic values

$\hat{g}$ = BLUP

- $\hat{g}$ estimation using mixed linear model for gaussian distributed traits sucrose yield, stalk diameter, stalk number, stalk height, in vitro NDF digestibility, bagasse content, brix and aphids AUIPC

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}_1\mathbf{b} + \mathbf{Z}_2\mathbf{g} + \mathbf{Z}_3\mathbf{gl} + \mathbf{e}$$

Replications, locations year (fixe)

Blocs (random)

Genotypes (random)

GxE (random)

QTL detection on g

- for non−Gaussian distributed traits $\hat{g}$ estimation using Markov chain Monte Carlo (MCMC) routines to fit multi-response generalized linear mixed models four disease-related traits and the flowering rate

R package MCMCglmm (Hadfield 2010)



Quantative genetics

- **Genetic variance significantly higher than 0**
- **Broad sens heritability range from:**
  0.6 for sucrose yield to 0.83 stalk diameter
- **Coefficients of genetic variation (CVg) ranging from:**
  4.7 % for brix to 20.2 % for bagasse digestibility



Quantative genetics

- **Genetic variance significantly higher than 0**
- **Broad sens heritability from 0.6 for sucrose yield to 0.83 stalk diameter**
- **Coefficients of genetic variation (CVg) ranging from 4.7 % for brix to 20.2 % for bagasse digestibility**

Traits	Mean ± SEM (1)	σ^2_g (2)	H^2 (3)	CVg (4)
Sucrose yield	17.24 ± 0.19 ton/ha	9.51*	0.6	17.9
Stalk diameter	26.12 ± 0.12 mm	6.64*	0.83	9.9
Stalk number	9.13 ± 0.1 stalk/m²	3.06*	0.67	19.2
Stalk height	265 ± 1.45 cm	950*	0.76	11.6
Bagasse content	17.67 ± 0.08 %	1.77*	0.71	7.5
Brix	18.82 ± 0.08 %	0.8*	0.61	4.7
In-vitro NDF digestibility	8.68 ± 0.12 %	3.06*	0.78	20.2
Flowering rate	14.30 ± 0.78 %	-	-	-
SCYLV incidence	72.6 ± 1.7 %	-	-	-
Brown rust score	1.88 ± 0.05	-	-	-
Gumming score	0.41 ± 0.007	-	-	-
Smut incidence	5.42 ± 0.58 whip/m²	-	-	-
Aphid AUDPC	15.37 ± 0.33 %.day	0.305*	0.71	16.0

* P-value < 0.01



Impact of population structure on the phenotypic variability

- Q2→ R^2 [0 -0.15] significant effect on 7 out 13 traits
- Q7→ R^2 [0.08 -0.36] significant effect on all traits
- PC → R^2 [0.08 -0.36] significant effect on all traits

	R2		
Trait	**Q2**	**Q7**	**PC**
Sucrose yield	14.75**	18.65**	30.34**
Stalk diameter	15.38**	25.91**	41.88**
Stalk number	12.26**	24.25**	41.75**
Stalk height	6.56**	28.31**	44.38**
Bagasse content	4.16*	16.12**	29.63**
Brix	2.38*	35.73**	52.65**
In-vitro NDF digestibility	3.07*	11.40**	19.55*
Flowering rate	0.19	12.64**	24.01*
Yellow leaf virus incidence	1.53	12.04**	34.41**
Brown rust score	1.26	10.83**	16.14*
Gumming score	0.62	8.21*	23.86**
Smut incidence	0.04	11.46**	20.09**
Aphid AUIPC	0.11	9.48*	20.55**
*** P<0.05**			
**** P<0.01**			

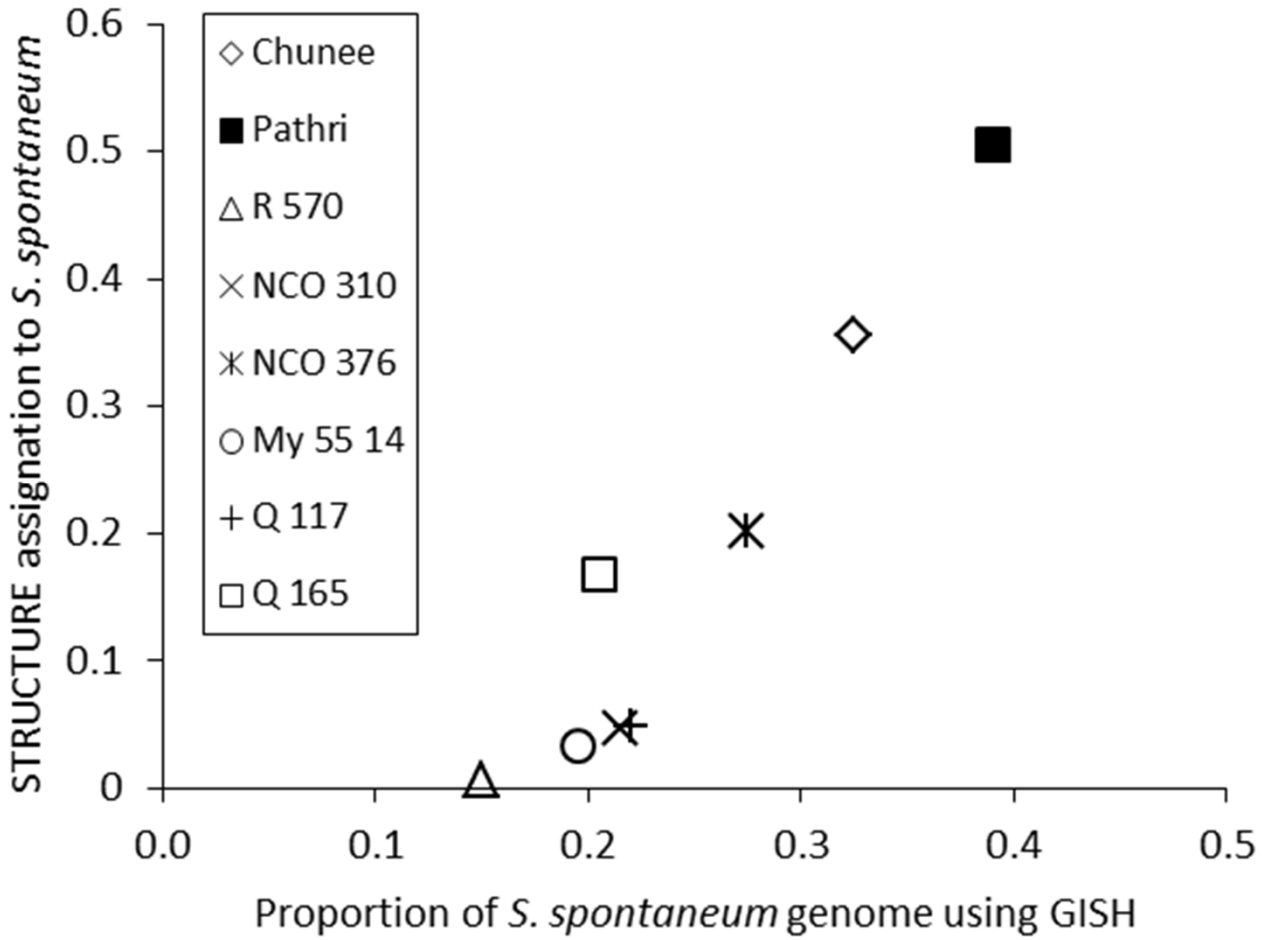


Markers

1406 AFLP
1892 DArT
29 SSR



Comparaison des assignations STRUCTURE/GISH



Annexe

Regroupement des centres de sélection:

C1 → Canal Point (Floride), Natal (AFS), Coimbatore (Inde), Fiji

C2 → Hawaï, Réunion, Maurice et Barbades

