Variance and Covariance of Actual Relationships between Relatives at One Locus

2013 PLoS ONE 8(2):
e57003. doi:10.1371/journal.pone.0057003

MINISTERIO
DE EDUCACIÓN
Y CIENCIA

INIA
Instituto Nacional de Investigación
y Tecnología Agraria y Alimentaria

andres.legarra@toulouse.inra.fr

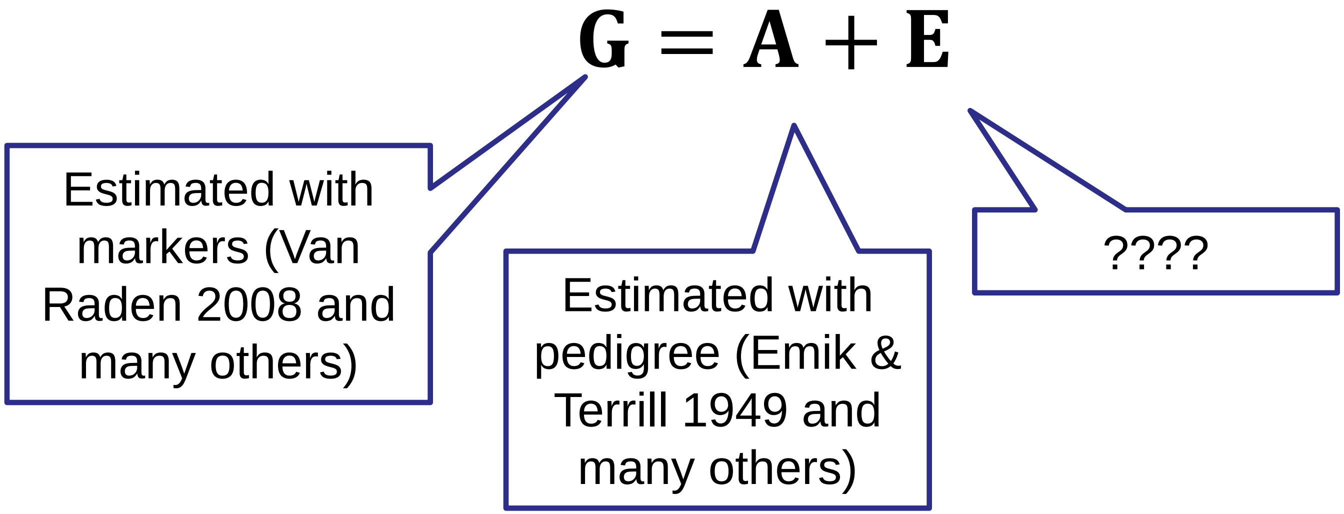
Luis Alberto Garcia-Cortés, Andrés Legarra,
Claude Chevalet, Miguel Angel Toro,

EAAP 2013

INRA
SCIENCE & IMPACT

Why

Realized relationships = expected relationships + deviations

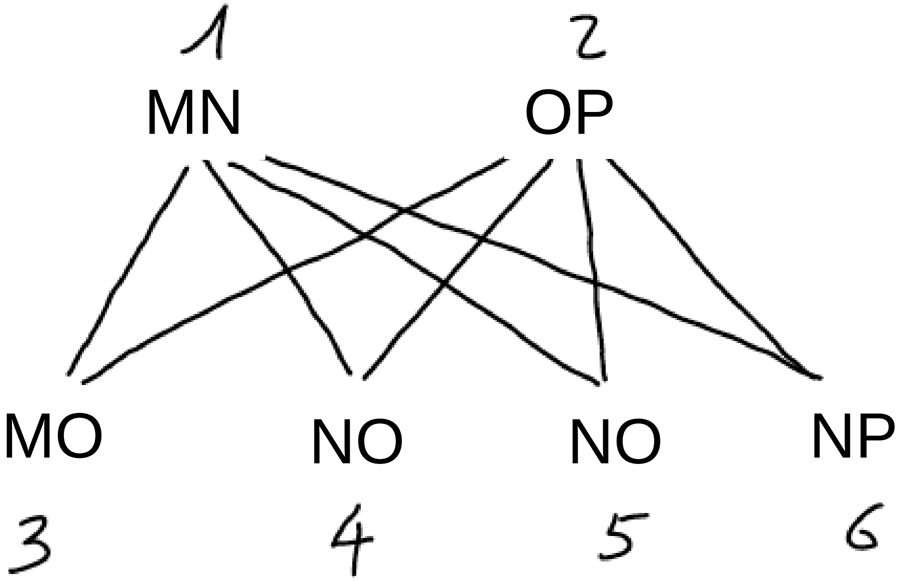


$$\mathbf{G} = \mathbf{A} + \mathbf{E}$$

Estimated with markers (Van Raden 2008 and many others)

Estimated with pedigree (Emik & Terrill 1949 and many others)

????



$$r_{34} = \frac{1}{4}$$

$$r_{35} = \frac{1}{4}$$

$$r_{36} = 0$$

$$r_{45} = \frac{1}{2}$$

Some sibs are more equal than others !!

(Invented) Full-sibs example

A_{ij}	G_{ij}
0.5	0.48
0.5	0.60
0.5	0.25
0.5	0.70
0.5	0.44
0.5	0.53

$$Var(G_{ij}) = Var(G_{ij} - A_{ij}) = 0.023$$

Can we compute this variance from pedigree *before* seeing the markers?

Why

- Increased accuracy of genomic selection relies on using a more accurate $\mathbf{G}$ instead of $\mathbf{A}$
  - But we don’t know *a priori* how much will they differ

- Small errors in $\mathbf{A}$ cumulate with generations !!
  - How many generations should one use?

- For both questions we need to know the extent of deviations in $\mathbf{E} = \mathbf{G} - \mathbf{A}$

Variation in actual relationship as a consequence of Mendelian sampling and linkage

W.G. HILL[1]* AND B.S. WEIR[2]

- Only consider two-individual coancestries and not a full pedigree
- Not a general formula but a collection of them
- They consider linkage, we consider one locus
- We provide a general formula for <u>all</u> pairs of individuals in a pedigree

What

Realized relationships = expected relationships + deviations

$$\mathbf{G} = \mathbf{A} + \mathbf{E}$$

What is $Var(\mathbf{E})$?

Let $\mathbf{E} = (\mathbf{e}_{:,1} \quad \mathbf{e}_{:,2} \quad \cdots \quad \mathbf{e}_{:,n})$

What is $Cov\begin{pmatrix} \mathbf{e}_{:,1} \\ \mathbf{e}_{:,2} \\ \cdots \\ \mathbf{e}_{:,n} \end{pmatrix}$?

Coancestry ’ $\mathbf{\Phi}$
$Var(\mathbf{E}) = 4Var(\mathbf{\Phi})$

Δ_i=Probability of identity state S_i

Father-son

$$\Delta_8 = 1$$

Full sibs

$$\Delta_7 = 1/4$$

$$\Delta_8 = 1/2$$

$$\Delta_9 = 1/4$$

K. LANGE AND J. S. SINSHEIMER

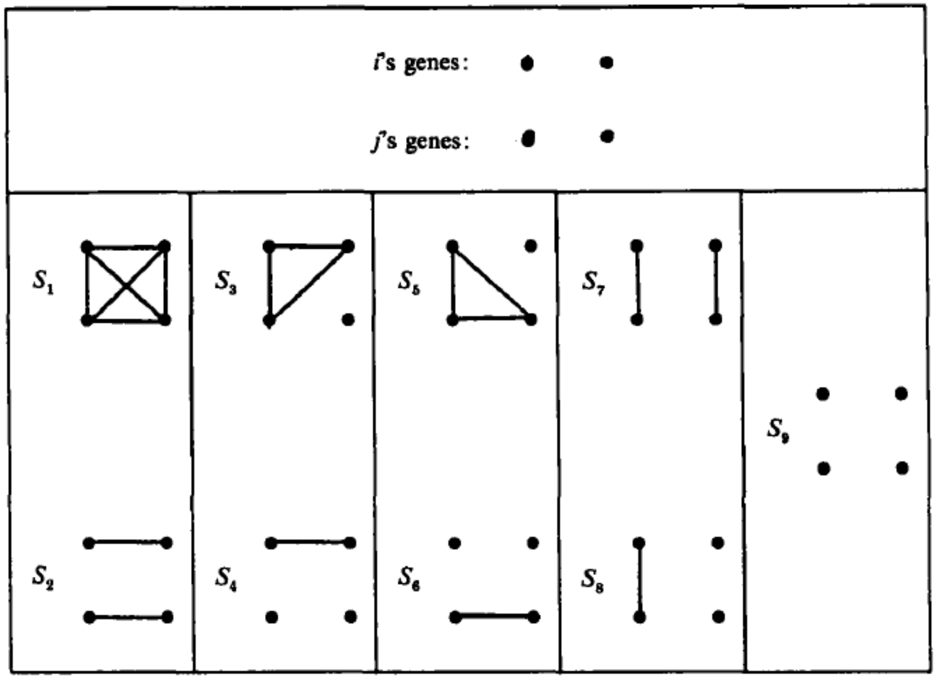


Fig. 1. Condensed identity states.

Coancestry ($= 0.5a_{ij}$) of i and j

$$\phi_{ij} = \Delta_1 + \frac{1}{2}(\Delta_3 + \Delta_5 + \Delta_7) + 1/4\,\Delta_8$$

$$E(i) = \Delta_i$$

$$Var(i) = \Delta_i(1-\Delta_i)$$

$$Cor(ij) = -\Delta_i\Delta_j$$

$$\phi_{ij} = \Delta_1 + \frac{1}{2}(\Delta_3 + \Delta_5 + \Delta_7) + \frac{1}{4}\Delta_8$$

$$\phi_{ij} = \begin{pmatrix} 1 & 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & \frac{1}{2} & \frac{1}{4} & 0 \end{pmatrix} \begin{pmatrix} \Delta_1 \\ \Delta_2 \\ \vdots \\ \\ \Delta_9 \end{pmatrix}$$

$$\begin{aligned} Var(\phi_{ij}) &= Var(W'\Delta) \\ &= W'\,Var(\Delta)\,W \\ &= \sum_{i=1}^{9} w_i^2\Delta_i(1-\Delta_i) - 2\sum_{j}\sum_{i>j} w_i w_j \Delta_i\Delta_j \end{aligned}$$

$$\begin{aligned} Var(\phi_{ij}) &= \sum_i w_i^2\Delta_i - \Big(\sum_i w_i\Delta_i\Big)^2 \\ &= \phi_{ij,ij} - (\phi_{ij})^2 \longrightarrow \text{regular coancestry} \end{aligned}$$

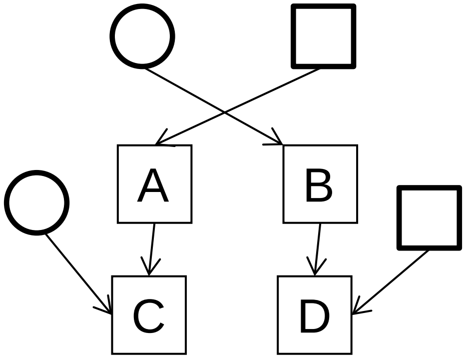
$\phi_{ij,ij} \longrightarrow$ Karigl (1981) generalized relationship

Covariance

- The covariance is a simple extension:

$$Cov(\Phi_{AB}, \Phi_{CD}) = \phi_{AB,CD} - \phi_{AB}\phi_{CD}$$

- This means that the Φ_{CD} is not independent of Φ_{AB} , example:



C and D can not share alleles if A and B do not

Generalized relationships

Karigl's 1981 generalized relationships

$$\phi_{AB,CD}$$

is the probability that

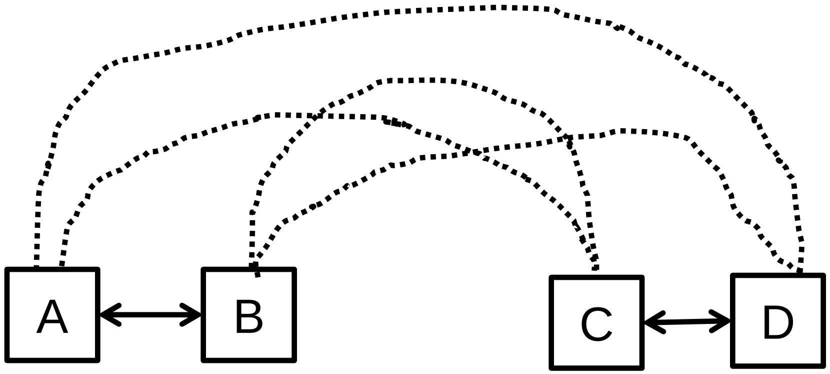
two genes drawn from A and B are IBD

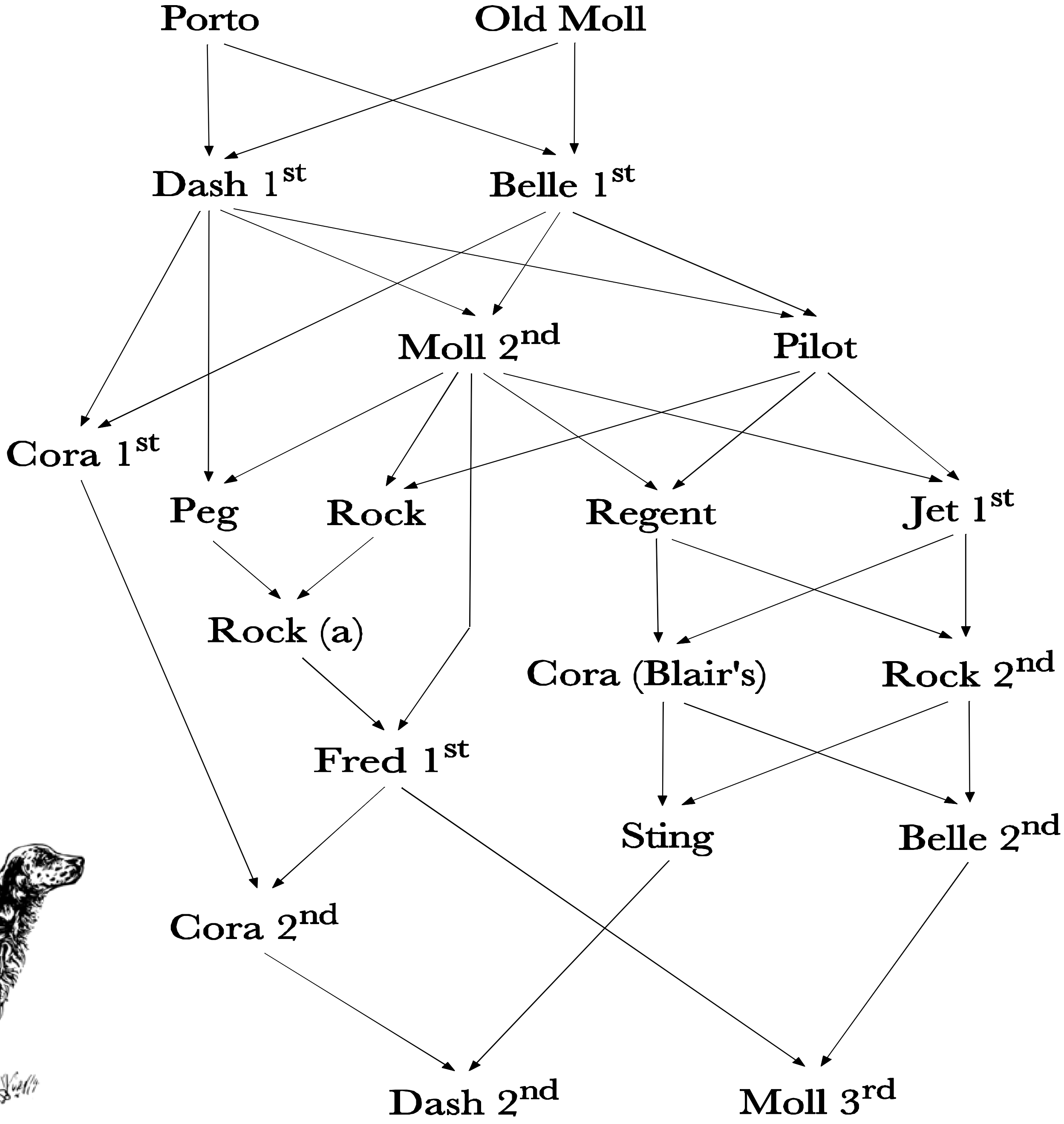
and at the same time

two genes drawn from C and D are IBD

Can be computed from pedigree

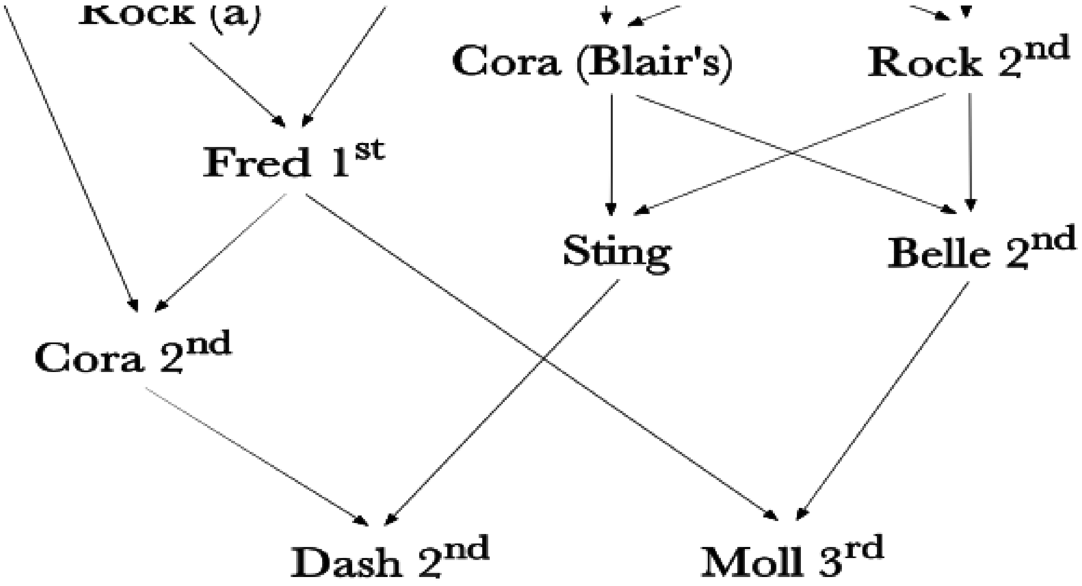
It will depend on the paths linking all four individuals





Porto
Old Moll
Dash 1st
Belle 1st
Moll 2nd
Pilot
Cora 1st
Peg
Rock
Regent
Jet 1st
Rock (a)
Cora (Blair's)
Rock 2nd
Fred 1st
Sting
Belle 2nd
Cora 2nd
Dash 2nd
Moll 3rd

http://www.clker.com/clipart-185658.html



Inbreeding of Dash 2nd

$F_D = 0,4297$

$V(F_D) = 0,2451$

$\phi_{cs} = 0,4297$

$V(\phi_{cs}) = 0,0922$

Coancestry of parents of Dash 2nd

$F_D = \phi_{CS}$, yet the variances are different !!

Extension to several loci

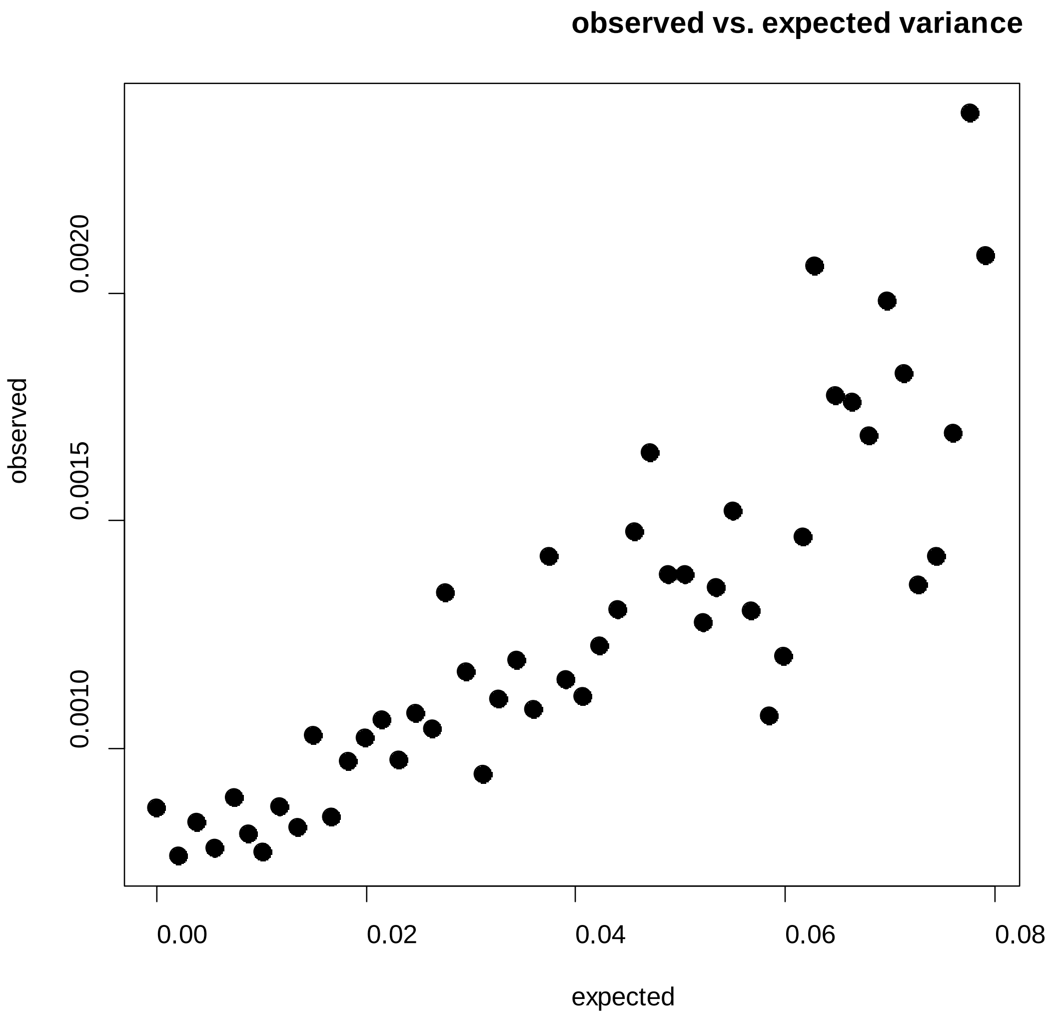
- If there were infinite loci, these variances are divided by ∞ and become 0
- Genome is not infinite, is composed of repetitive “chunks” due to linkage and finite number of animals
- Divide by « Me = equivalent number of loci » (Goddard 2009 and many others)

$$Cov(\Phi_{AB},\Phi_{CD}) = \frac{1}{Me}\left(\phi_{AB,CD} - \phi_{AB}\phi_{CD}\right)$$

Does it work?

- H. Wang, I. Misztal et al., in prep, chicken data, >4000 animals
- **E=G**-**A**
- empirical var(**E**) vs. predictions based on pedigree using $Cov(\Phi_{AB}, \Phi_{CD}) = \phi_{AB,CD} - \phi_{AB}\phi_{CD}$

Var(**G**-**A**) vs. our theory



- Pretty good fit
- From here, number of effective loci ~70

Conclusions

- $Var(\boldsymbol{E} = \boldsymbol{G} - \boldsymbol{A})$ can be computed from pedigree and a populational parameter (Me)
- Starting point for more developments

Acknowledgments

- project CGL2009-1327-C02-02 of the Ministerio de Ciencia e Innovacion (Spain)
- ANR project Rules & Tools (France)
- action X-Gen of the SelGen metaprogram (INRA)
- partially supported by the platform bioinformatics Toulouse Midi-Pyrenees

- *LAGC made the drawings*