EAAP Meeting 2007
August 26-29, 2007, Dublin, Ireland

Session 18 “Statistical analysis of genomics data”, Theater 9

Detection and Use of Single Gene Effects in Large Animal Populations

Accuracy of Estimation and Use of Gene Content

Nicolas Gengler*[1,2], Sven Abras[3], Catherine Verkenne[1], Maciej Szydlowski[1] & Robert Renaville[3]

[1]Animal Science Unit, Gembloux Agricultural University, Belgium
[2]National Fund for Scientific Research (FNRS), Brussels, Belgium
[3]Animal and Microbial Biology Unit, Gembloux Agricultural University, Belgium

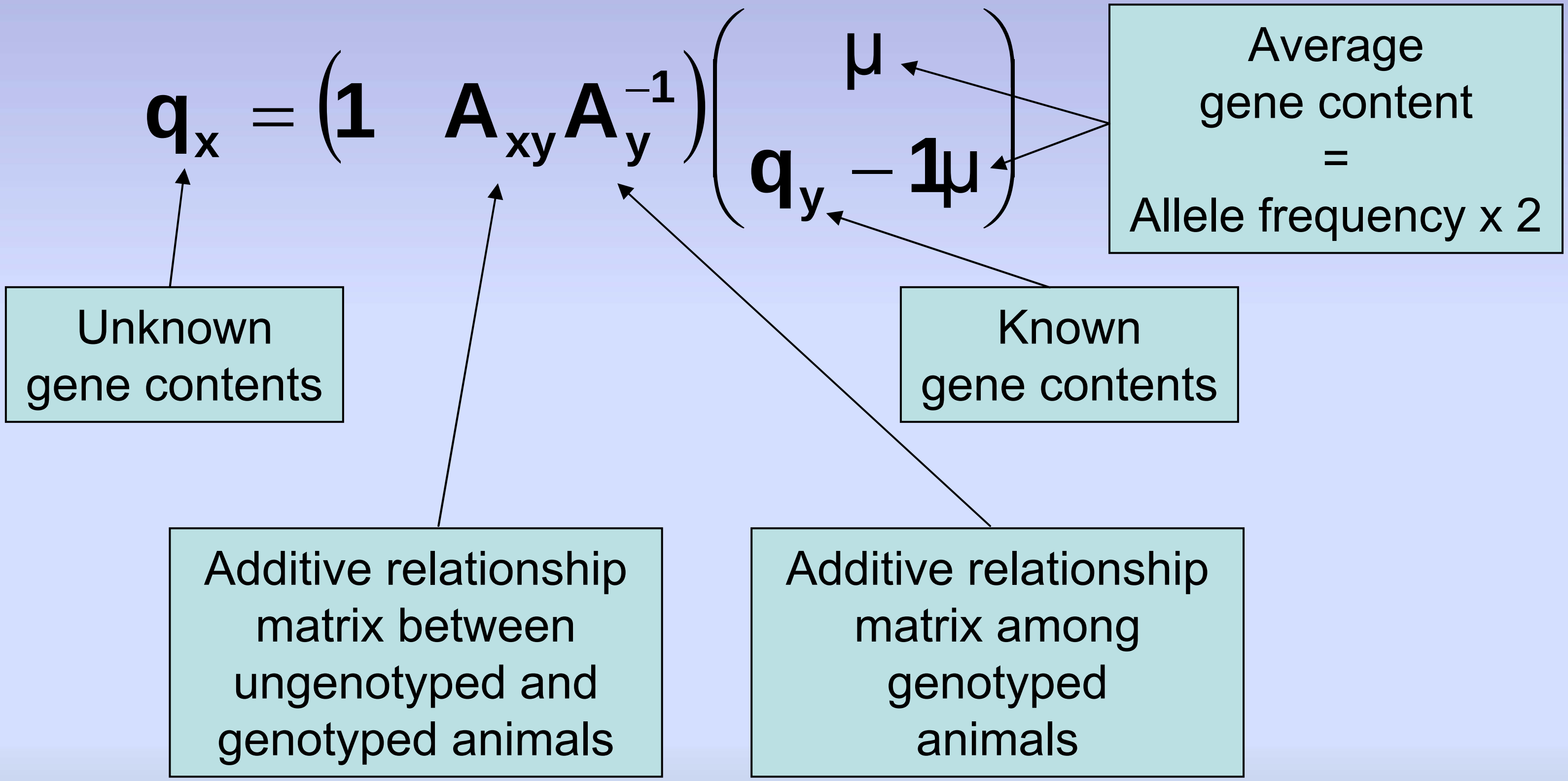
Accuracy of Estimation and Use of Gene Content

- Precise estimation of genetic values required to achieve genetic improvement of livestock
- Challenge of integrating molecular and phenotypic data
- Genotypes of animals have to be known
- However, in large animal populations, the vast majority of animals are not genotyped!

Accuracy of Estimation and Use of Gene Content

- Recently new method to estimate gene content
  - Gene content = number of copies of a particular allele in a genotype
- Present study evaluates accuracy and usefulness of this method
  - Based on its ability to detect correctly single gene effects
  - Here simulation in real-life situation of previously detected single gene effects

Conditional Expectation of Gene Contents for Ungenotyped Animals Given Molecular and Pedigree Data



$$\mathbf{q}_x = \begin{pmatrix}\mathbf{1} & \mathbf{A}_{xy}\mathbf{A}_y^{-1}\end{pmatrix}\begin{pmatrix}\mu \\ \mathbf{q}_y - \mathbf{1}\mu\end{pmatrix}$$

Average
gene content
=
Allele frequency x 2

Unknown
gene contents

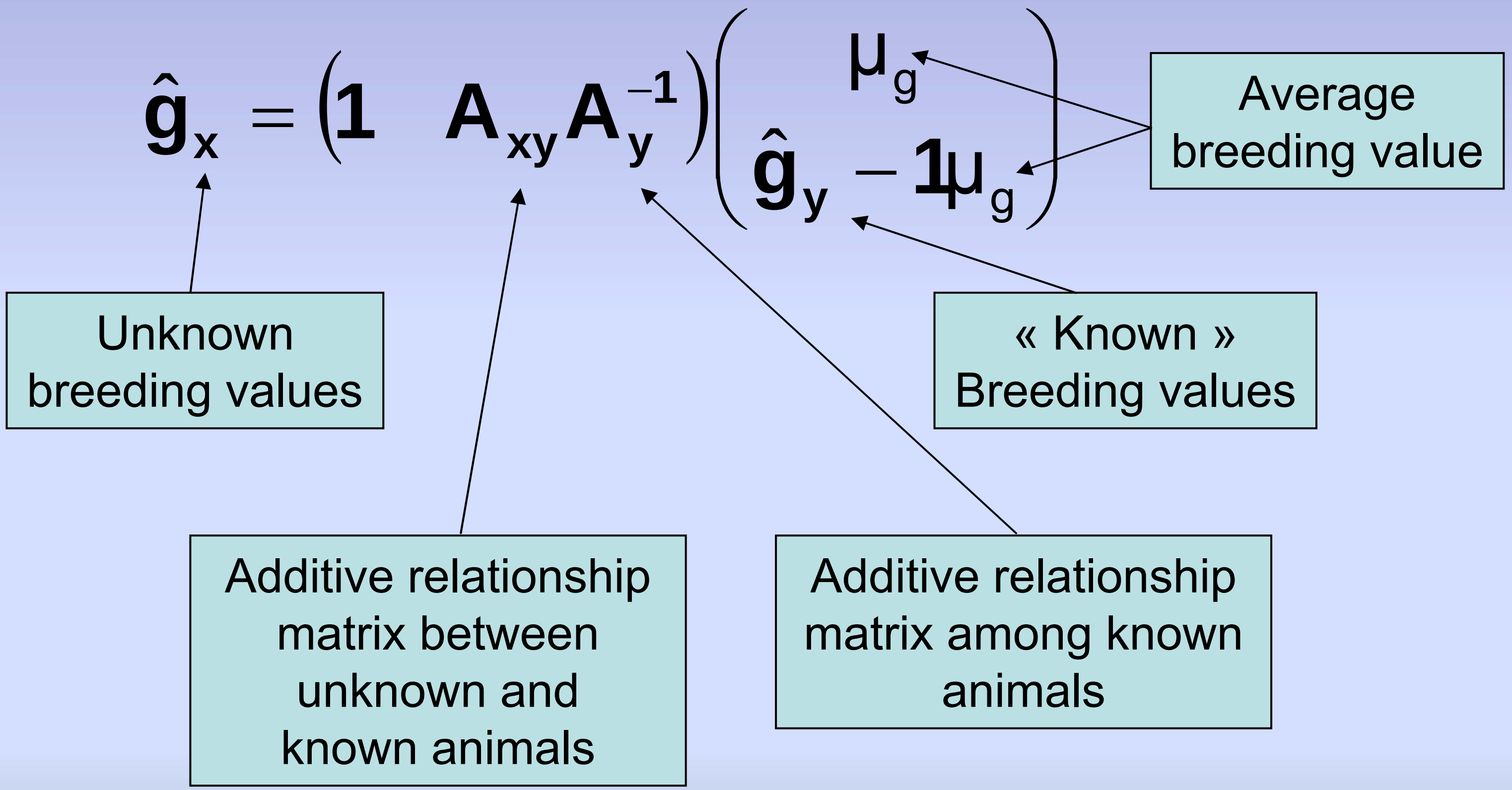
Known
gene contents

Additive relationship matrix between ungenotyped and genotyped animals

Additive relationship matrix among genotyped animals

Gengler et al. 2007, Animal 1: 21–28

Predicting Unknown Breeding Values Given Known Values and Pedigree Data



$$\hat{\mathbf{g}}_{\mathbf{x}} = \begin{pmatrix} \mathbf{1} & \mathbf{A}_{\mathbf{xy}}\mathbf{A}_{\mathbf{y}}^{-1} \end{pmatrix} \begin{pmatrix} \mu_g \\ \hat{\mathbf{g}}_{\mathbf{y}} - \mathbf{1}\mu_g \end{pmatrix}$$

Average breeding value

Unknown breeding values

« Known » Breeding values

Additive relationship matrix between unknown and known animals

Additive relationship matrix among known animals

Alternative derivation

Writing Breeding Values as Sum of Single Gene Effects

$$\hat{\mathbf{g}}_{\mathbf{x}} = \sum_i \hat{\alpha}_i \mathbf{q}_{\mathbf{x}i} = \begin{pmatrix} \mathbf{1} & \mathbf{A}_{\mathbf{xy}} \mathbf{A}_{\mathbf{y}}^{-1} \end{pmatrix} \begin{pmatrix} \sum_i \hat{\alpha}_i \mu_i \\ \sum_i \hat{\alpha}_i \mathbf{q}_{\mathbf{y}_i} - \mathbf{1} \sum_i \hat{\alpha}_i \mu_i \end{pmatrix}$$

where $\hat{\mathbf{g}}_{\mathbf{y}} = \sum_i \hat{\alpha}_i \mathbf{q}_{\mathbf{y}_i}$ and $\mu_g = \sum_i \hat{\alpha}_i \mu_i$

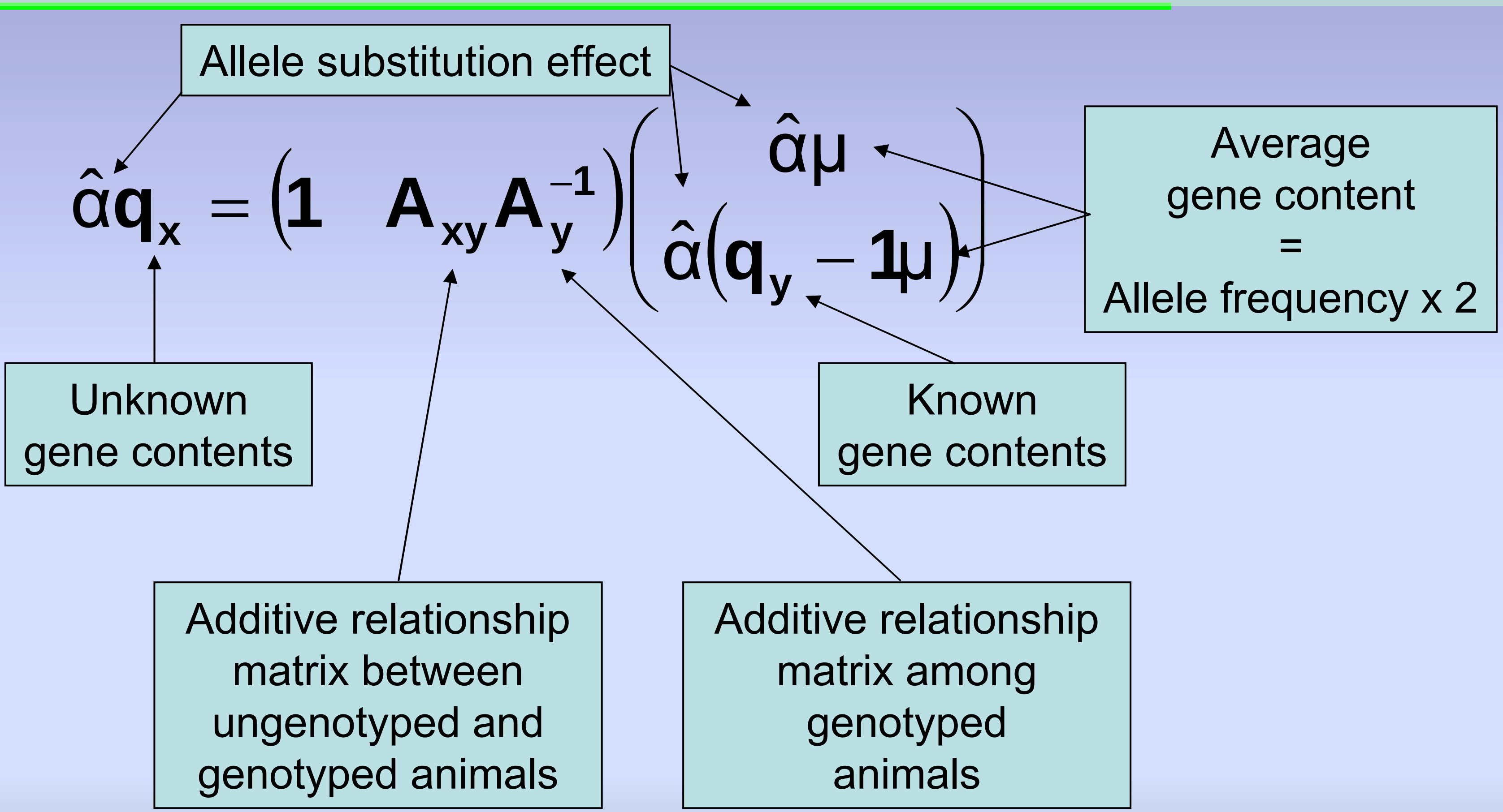
Allele substitution effects

Gene contents

Average
gene contents
=
Allele frequency x 2

Alternative derivation

Under Certain Hypotheses as Normality of Contributions of Single Gene Effects



$$\hat{\alpha}\mathbf{q}_x = \begin{pmatrix}\mathbf{1} & \mathbf{A}_{xy}\mathbf{A}_y^{-1}\end{pmatrix}\begin{pmatrix}\hat{\alpha}\mu \\ \hat{\alpha}(\mathbf{q}_y - \mathbf{1}\mu)\end{pmatrix}$$

Allele substitution effect

Average
gene content
=
Allele frequency x 2

Unknown
gene contents

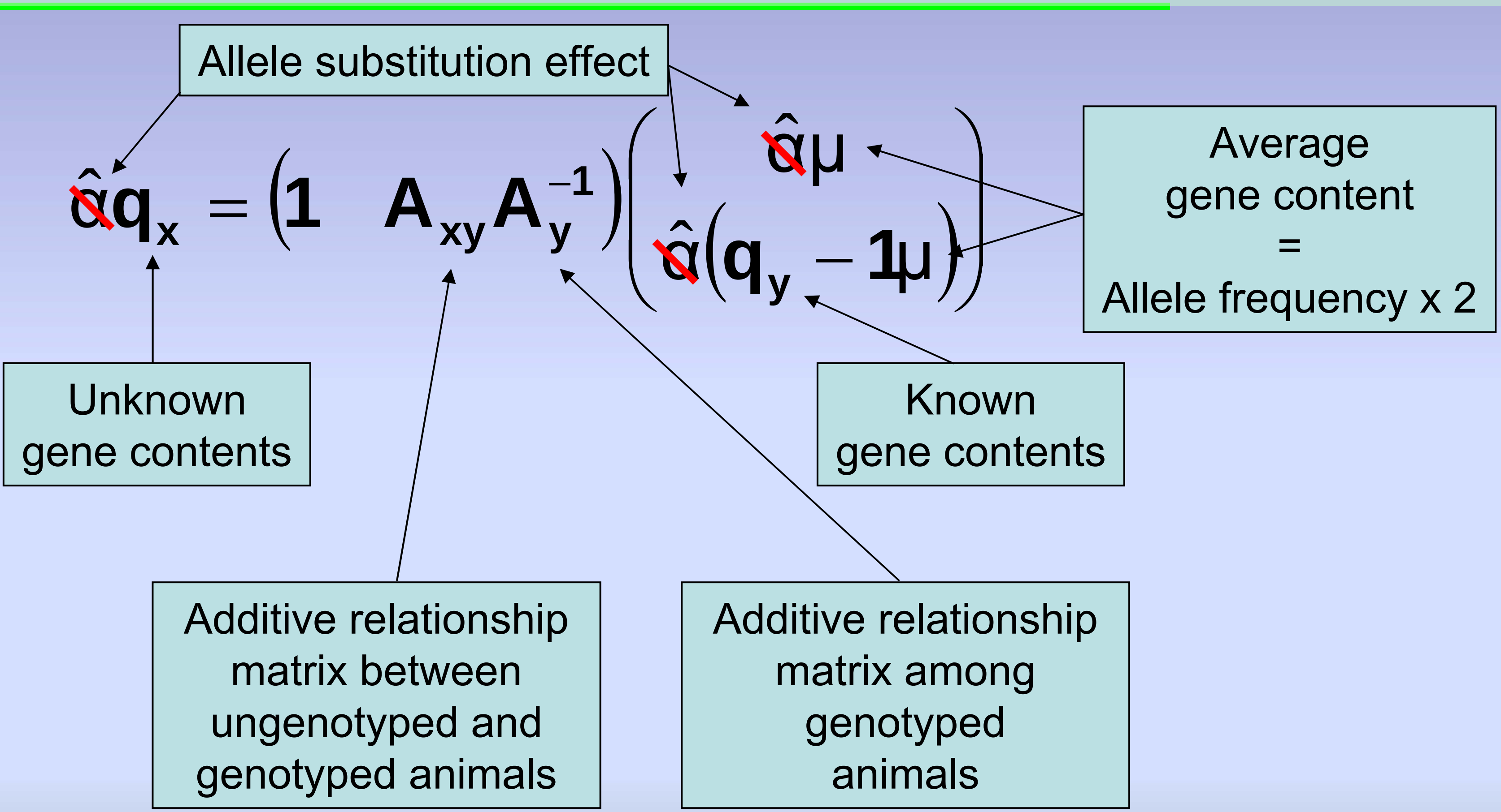
Known
gene contents

Additive relationship
matrix between
ungenotyped and
genotyped animals

Additive relationship
matrix among
genotyped
animals

Alternative derivation

Dividing on Both Sides by Allele Substitution Effect



$$\cancel{\hat{\alpha}}\mathbf{q}_x = \begin{pmatrix} \mathbf{1} & \mathbf{A}_{xy}\mathbf{A}_y^{-1} \end{pmatrix} \begin{pmatrix} \cancel{\hat{\alpha}}\mu \\ \cancel{\hat{\alpha}}(\mathbf{q}_y - \mathbf{1}\mu) \end{pmatrix}$$

Allele substitution effect

Average gene content = Allele frequency x 2

Unknown gene contents

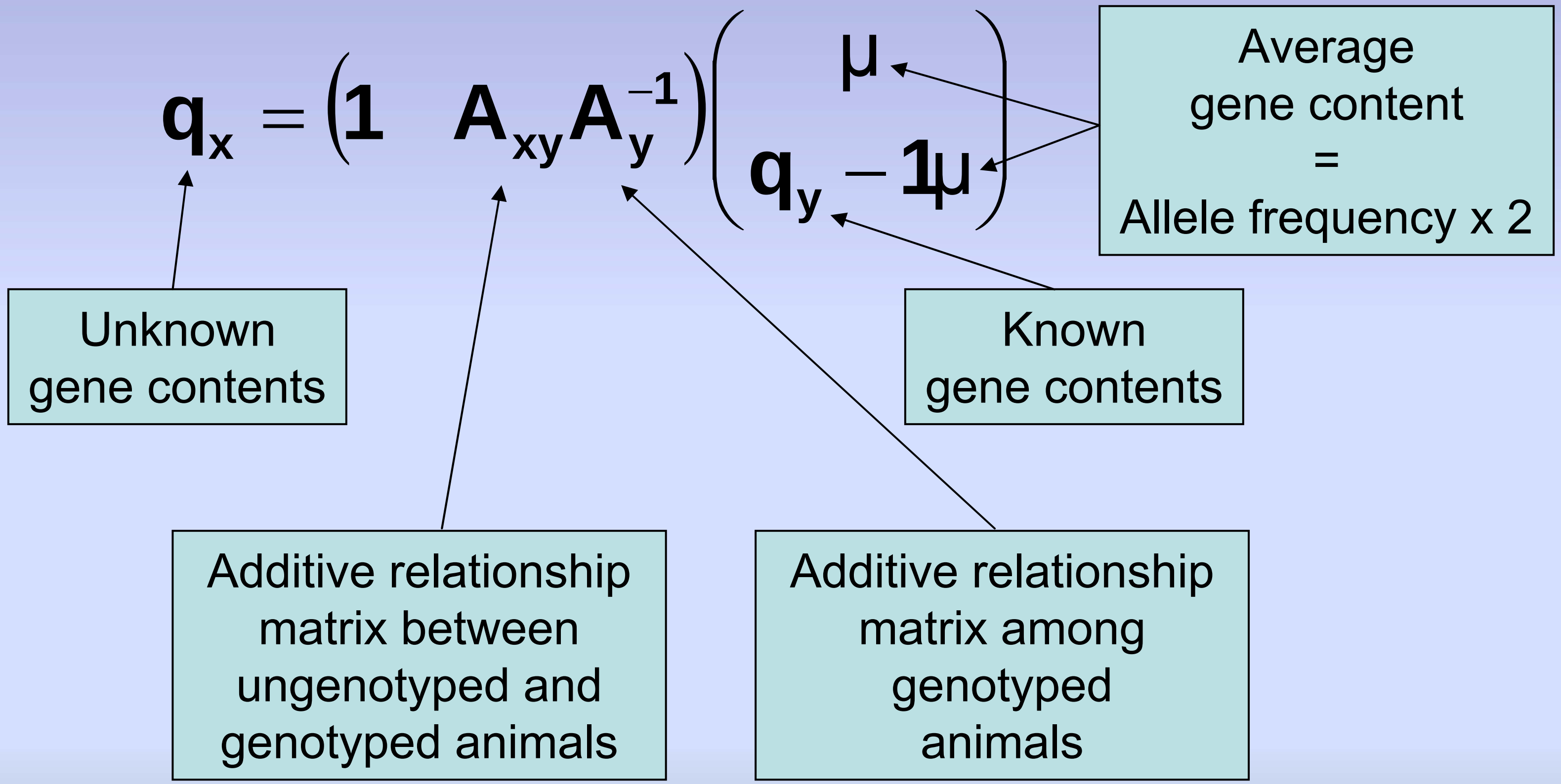
Known gene contents

Additive relationship matrix between ungenotyped and genotyped animals

Additive relationship matrix among genotyped animals

Alternative derivation

Alternative Derivation of New Method



$$\mathbf{q}_x = \begin{pmatrix} \mathbf{1} & \mathbf{A}_{xy}\mathbf{A}_y^{-1} \end{pmatrix} \begin{pmatrix} \mu \\ \mathbf{q}_y - \mathbf{1}\mu \end{pmatrix}$$

Average
gene content
=
Allele frequency x 2

Unknown
gene contents

Known
gene contents

Additive relationship matrix between ungenotyped and genotyped animals

Additive relationship matrix among genotyped animals

Comparison with MCMC and Iterative Peeling (Animal, 1: 21-28)

- Correlations in simulation study (only 12% genotyped) with simulated gene content:

	Estimation method			
	MCMC	Iterative peeling	Israel and Weller (1998)	New method
P(B) = 0.4	0.52	0.52	0.38	0.50
P(B) = 0.2	0.42	0.40	0.38	0.47

- MCMC and Iterative peeling less practical for large pedigree with sparse data than proposed method

Material and Methods

- Holstein population
  - 12,858,741 first lactation TD records
  - 1,656,599 cows with records
  - 2,755,041 animals in pedigree
- Single gene
  - Bovine transmembrane growth hormone-receptor (*GHR*)
- Model
  - Classical mixed inheritance (TD) model, with regression on gene content

Material and Methods

- First: regular estimation of single gene effects (α)
- Then: creation of 15 validation data sets

  1. Bi-allelic (A,B) gene simulated with same allele frequency than observed in real data
  2. Production records modified $y^* = y + d$ where:
     - If AA: $d = \alpha$
     - If AB: $d = 0$
     - If BB: $d = - \alpha$
  3. Only 961 genotyped bulls considered known as in real data, others estimated using new method

- Finally: allelic substitution effect re-estimated

Results

	Substitution effect				
	F_A (%)	Milk (g/day)	Fat (g/day)	Protein (g/day)	SCS
Value used in simulation	23.3	295	-8.14	-1.83	-0.022
Relative to phenotypic SD (%)	N/A	6.5	-4.6	-1.4	-1.3
-- Statistics --					
Mean	23.3	306	-7.87	-2.84	-0.027
SD	4.4	113	4.87	2.33	0.021
Relative bias (%)	0.2	3.7	3.3	-55.3	-22.4
Relative standard error (%)	18.8	38.5	60.0	139.7	96.4
t test values (d.f. = 14)	N/A	2.597	1.667	0.716	1.037
P – values	N/A	0.021	0.118	0.489	0.317

$$\frac{\sum_{i=1}^{n}(\hat{\theta}_i-\theta)}{n}\,\frac{100}{|\theta|}$$

$$\left[\frac{\sum_{i=1}^{n}(\hat{\theta}_i-\theta)^2}{n-1}\right]^{1/2}\frac{100}{|\theta|}$$

Results

	Substitution effect				
	F_A (%)	Milk (g/day)	Fat (g/day)	Protein (g/day)	SCS
Value used in simulation	23.3	295	-8.14	-1.83	-0.022
Relative to phenotypic SD (%)	N/A	6.5	-4.6	-1.4	-1.3
-- Statistics --					
Mean	23.3	306	-7.87	-2.84	-0.027
SD	4.4	113	4.87	2.33	0.021
Relative bias (%)	0.2	3.7	3.3	-55.3	-22.4
Relative standard error (%)	18.8	38.5	60.0	139.7	96.4
t test values (d.f. = 14)	N/A	2.597	1.667	0.716	1.037
P – values	N/A	0.021	0.118	0.489	0.317

Results

		Substitution effect			
	F_A (%)	Milk (g/day)	Fat (g/day)	Protein (g/day)	SCS
Value used in simulation	23.3	295	-8.14	-1.83	-0.022
Relative to phenotypic SD (%)	N/A	6.5	-4.6	-1.4	-1.3
-- Statistics --					
Mean	23.3	306	-7.87	-2.84	-0.027
SD	4.4	113	4.87	2.33	0.021
Relative bias (%)	0.2	3.7	3.3	-55.3	-22.4
Relative standard error (%)	18.8	38.5	60.0	139.7	96.4
t test values (d.f. = 14)	N/A	2.597	1.667	0.716	1.037
P – values	N/A	0.021	0.118	0.489	0.317

Results

	Substitution effect				
	F_A (%)	Milk (g/day)	Fat (g/day)	Protein (g/day)	SCS
Value used in simulation	23.3	295	-8.14	-1.83	-0.022
Relative to phenotypic SD (%)	N/A	6.5	-4.6	-1.4	-1.3
---------- Statistics ----------					
Mean	23.3	306	-7.87	-2.84	-0.027
SD	4.4	113	4.87	2.33	0.021
Relative bias (%)	0.2	3.7	3.3	-55.3	-22.4
Relative standard error (%)	18.8	38.5	60.0	139.7	96.4
t test values (d.f. = 14)	N/A	2.597	1.667	0.716	1.037
P – values	N/A	0.021	0.118	0.489	0.317

Conclusions

- New method to approximate gene content potentially very useful
  - To fill in easily missing gene content (SNP panels)
  - To detect single gene effects
  - To allow integration of single gene effects in genetic evaluations
    - Method virtually same limitations as evaluation models (Equivalent mixed model see Animal, 1: 21-28)
- Results shown here indicate limitations in detection of single gene effects
  - Only possible as long as they are sufficiently large!
- Method could be improved
  - Gametic prediction model (Poster 17 in this Session 18)

Acknowledgments: Study supported through FNRS grants F.4552.05 and 2.4507.02, RW-DGA project D31-1112, semen provided by SEMEX and data provided by CDN

Thank you for your attention

Presenting author's e-mail:

gengler.n@fsagx.ac.be