Animal breeding

Prof. Pekka Uimari

University of Helsinki

Department of Agriculral Sciences

Finnish dairy breeds

Ayrshire, Ay (66%)

Production: 8600 kg, 4.2% fat, 3.4% protein

Holstein, Hol (33%)

Production:
9400 kg, 3.9%, 3.3%

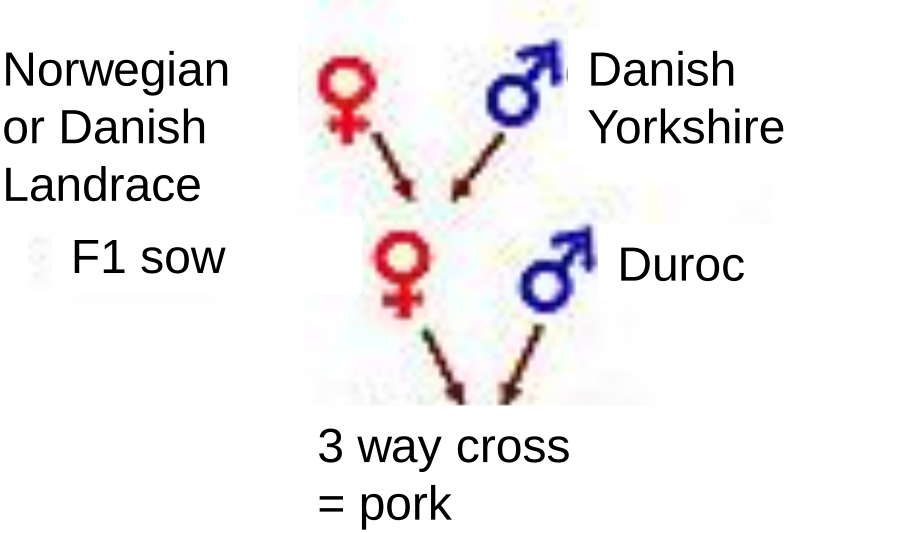
Finncattle, Sk (1%)

Production:
6200 kg, 4.3%, 3.4%

An average cow produces 30 l milk per day and 1 kg and butter

Average milk production in Finland
1930: 2000 kg
1970: 4000 kg
1990: 5600 kg
2010: 8700 kg

Two breeding programs for pigs in Finland



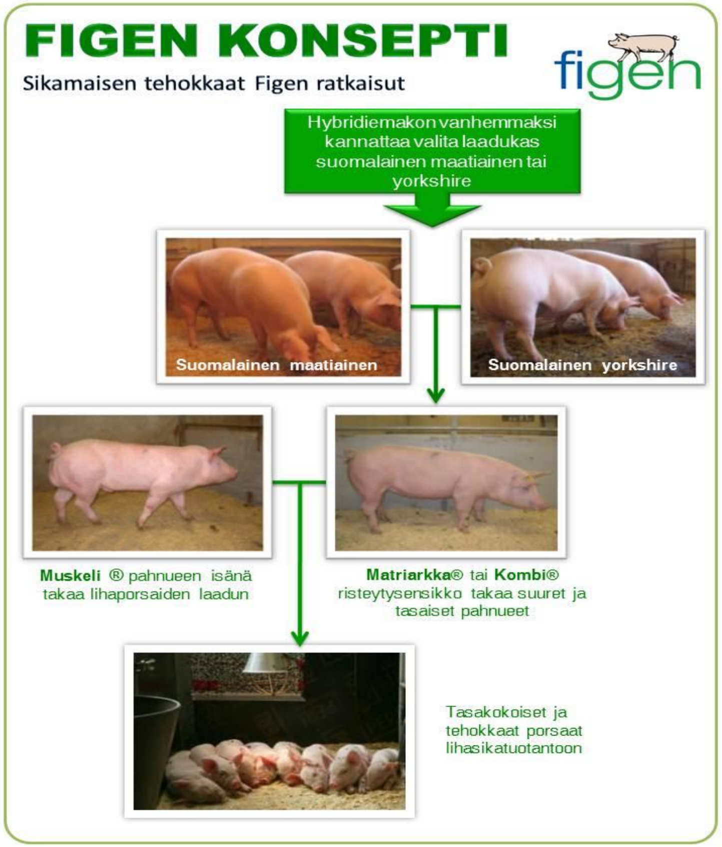
Norwegian or Danish Landrace

Danish Yorkshire

F1 sow

Duroc

3 way cross = pork



FIGEN KONSEPTI

Sikamaisen tehokkaat Figen ratkaisut

Hybridiemakon vanhemmaksi kannattaa valita laadukas suomalainen maatiainen tai yorkshire

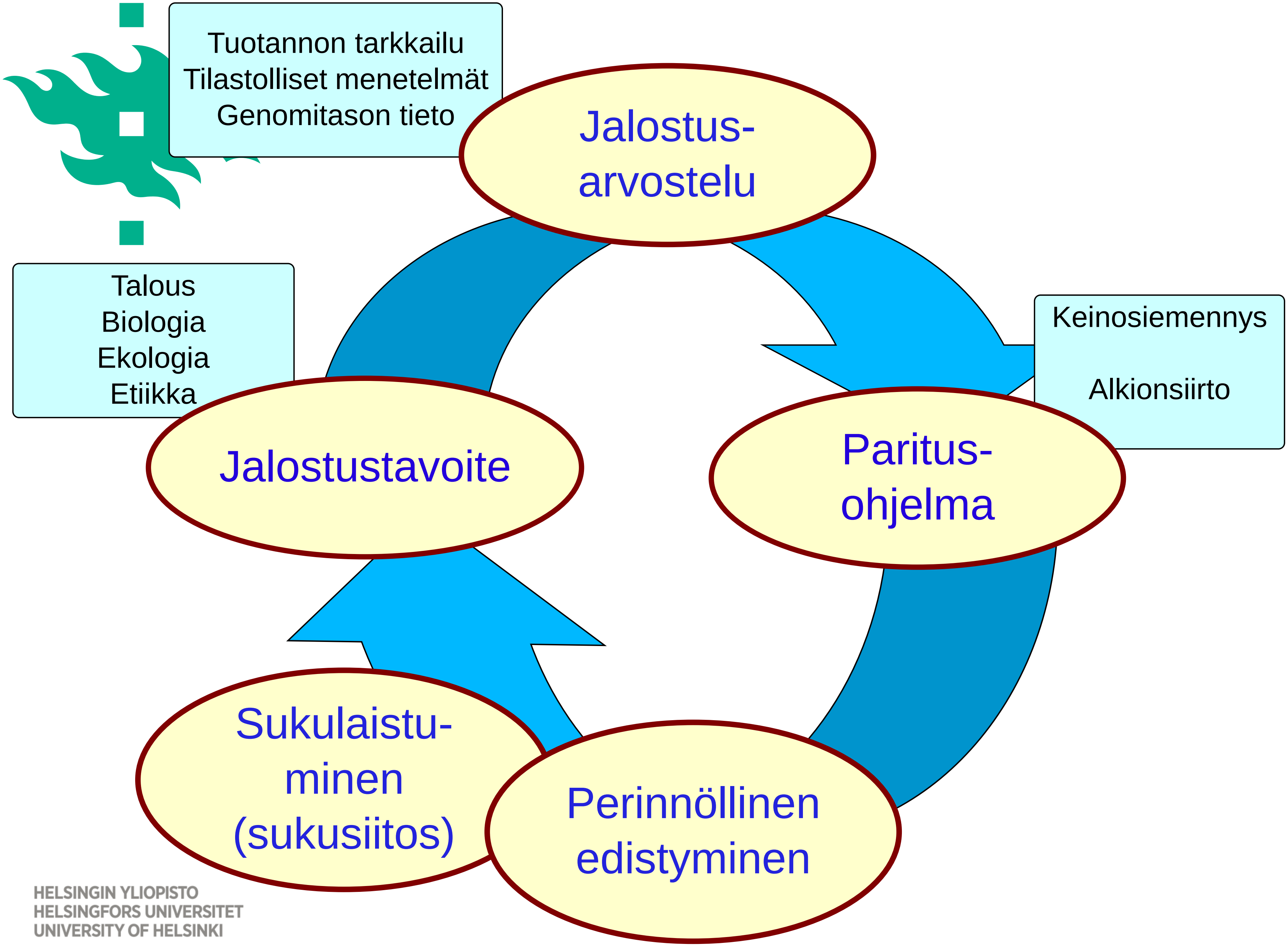
Suomalainen maatiainen

Suomalainen yorkshire

Muskeli ® pahnueen isänä takaa lihaporsaiden laadun

Matriarkka® tai **Kombi**® risteytysensikko takaa suuret ja tasaiset pahnueet

Tasakokoiset ja tehokkaat porsaat lihasikatuotantoon



Talous
Biologia
Ekologia
Etiikka

Jalostustavoite

Tuotannon tarkkailu
Tilastolliset menetelmät
Genomitason tieto

Jalostus-
arvostelu

Keinosiemennys

Alkionsiirto

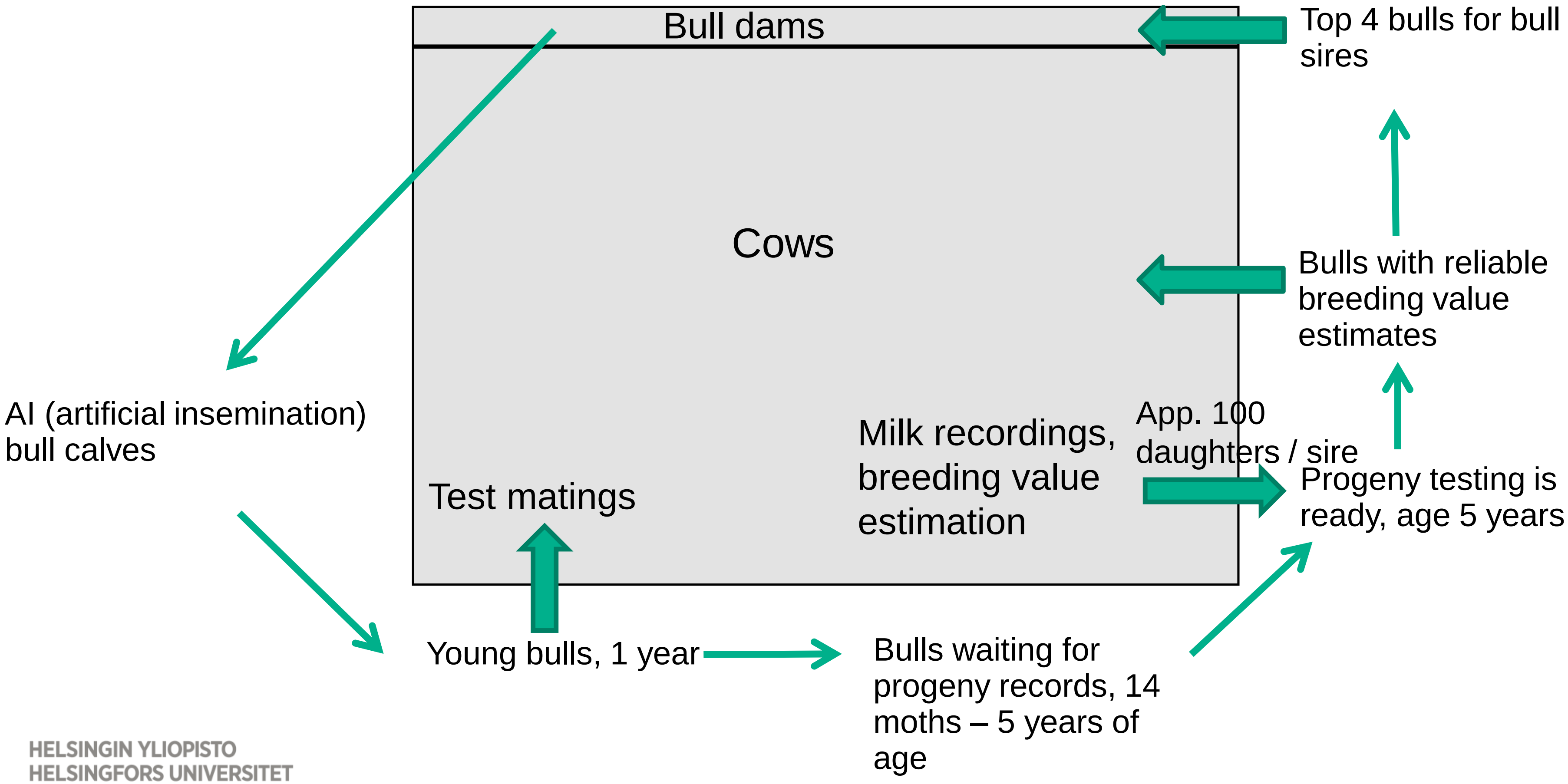
Paritus-
ohjelma

Perinnöllinen
edistyminen

Sukulaistu-
minen
(sukusiitos)

HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

Traditional breeding program for dairy cows



Bull dams

Cows

Top 4 bulls for bull sires

Bulls with reliable breeding value estimates

AI (artificial insemination) bull calves

Test matings

Milk recordings, breeding value estimation

App. 100 daughters / sire

Progeny testing is ready, age 5 years

Young bulls, 1 year

Bulls waiting for progeny records, 14 moths – 5 years of age

Traditional breeding value estimation

- Recorded measurements (phenotypes) such as milk production (dairy cows), litter size (pigs), heath records etc.

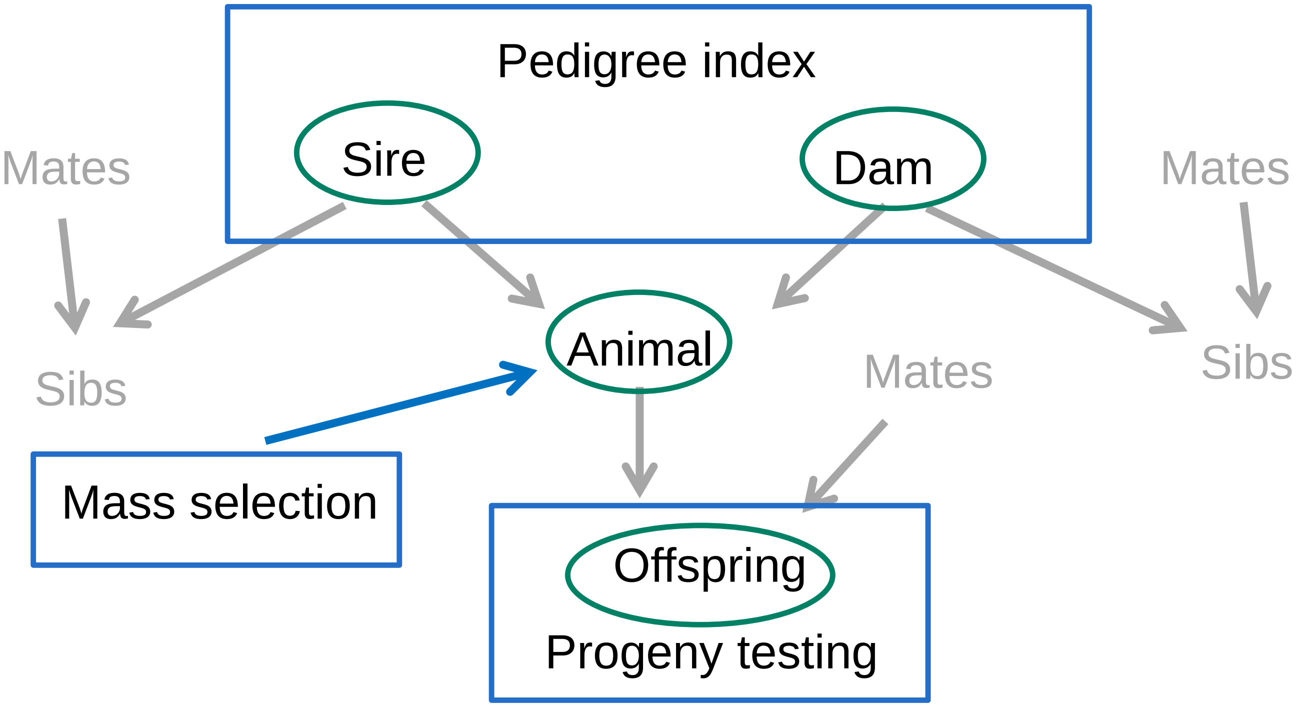
Phenotypes

Pedigree information

Statistical models (BLUP) and programs

Estimated breeding values for all animals (with records and without)

Information sources



Pedigree index

Sire

Dam

Mates

Mates

Animal

Sibs

Sibs

Mates

Mass selection

Offspring

Progeny testing

Traditional breeding value estimation

- Requires a large number of records
- Accuracy depends on heritability of the trait h^2
- Repeated measurements increase accuracy
- The number of offspring is important especially for traits that are sex specific such as milk production and litter size

Traditional breeding value estimation, steps

- Collection of phenotypic values
- Registry of pedigree → relationship between animals
- Linear mixed model:
  - Systematic environmental effects
    - Sex, age, herd, season, parity
  - Random effects
    - Permanent environment
    - Litter
    - Maternal effect
    - Animal itself (polygenic effect of the animal) ← relationship/kinship matrix
  - Best linear unbiased prediction (BLUP) → Estimated breeding values EBV
  - Standardization of the EBVs → Indices
  - Publication of the results

Estimation of heritability

- Proportion of the phenotypic variation that is explained by genetic effects

  $$h^2 = V_A / V_P$$

- Offspring on parent regression
  - $h^2 = 2^* b_{OP}$
- Twin data, correlation coefficient
  - $h^2 = 2(r_{MZ} – r_{DZ})$
  - MZ: monozygotic ("identical")
  - DZ: dizygotic

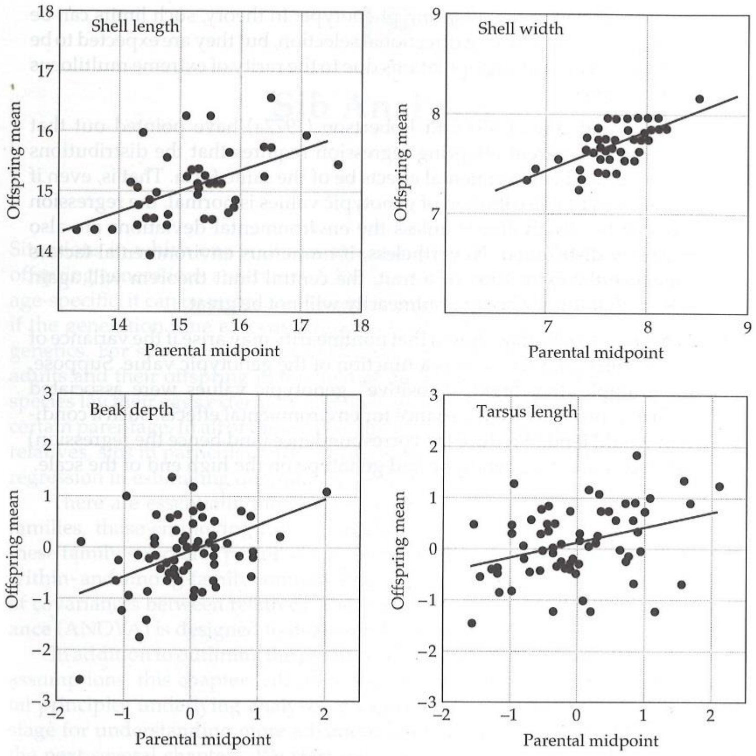
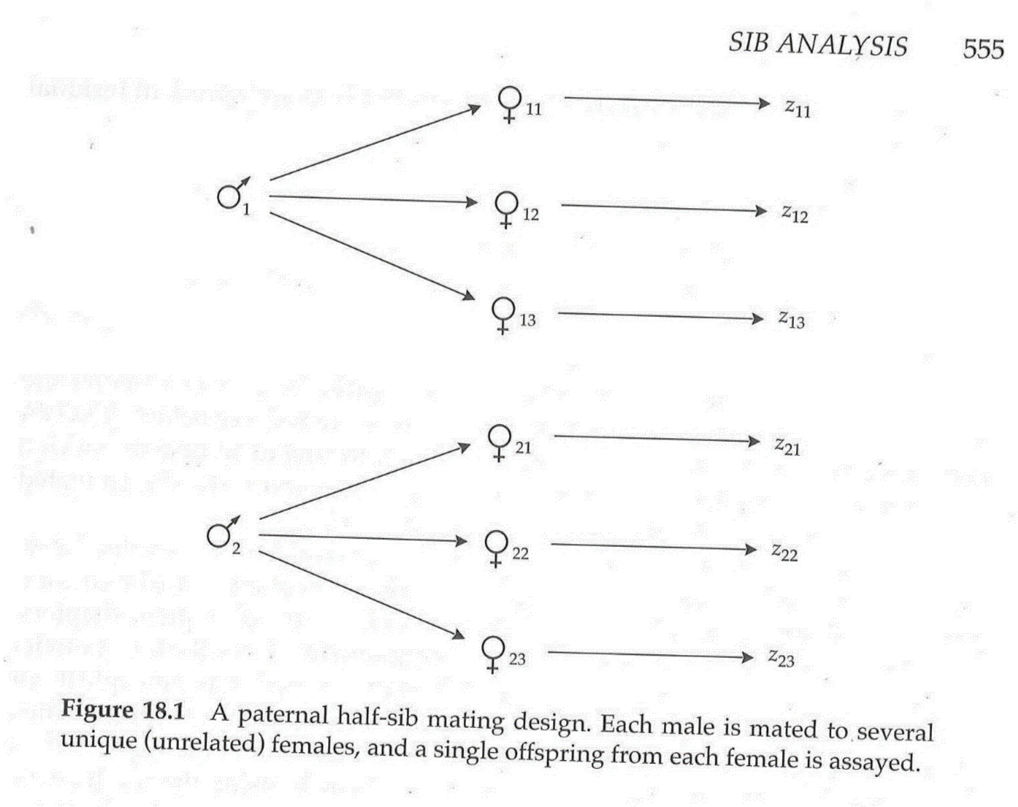


Figure 17.4 Examples of midparent-offspring regressions. **Top panels:** Land snail *Partula taeniata* (Murray and Clarke 1968). **Bottom panels:** Song sparrow *Melospiza melodia* (Smith and Zach 1979).

Genetics and analysis of quantitative traits, Lynch & Walsh

Estimation of heritability

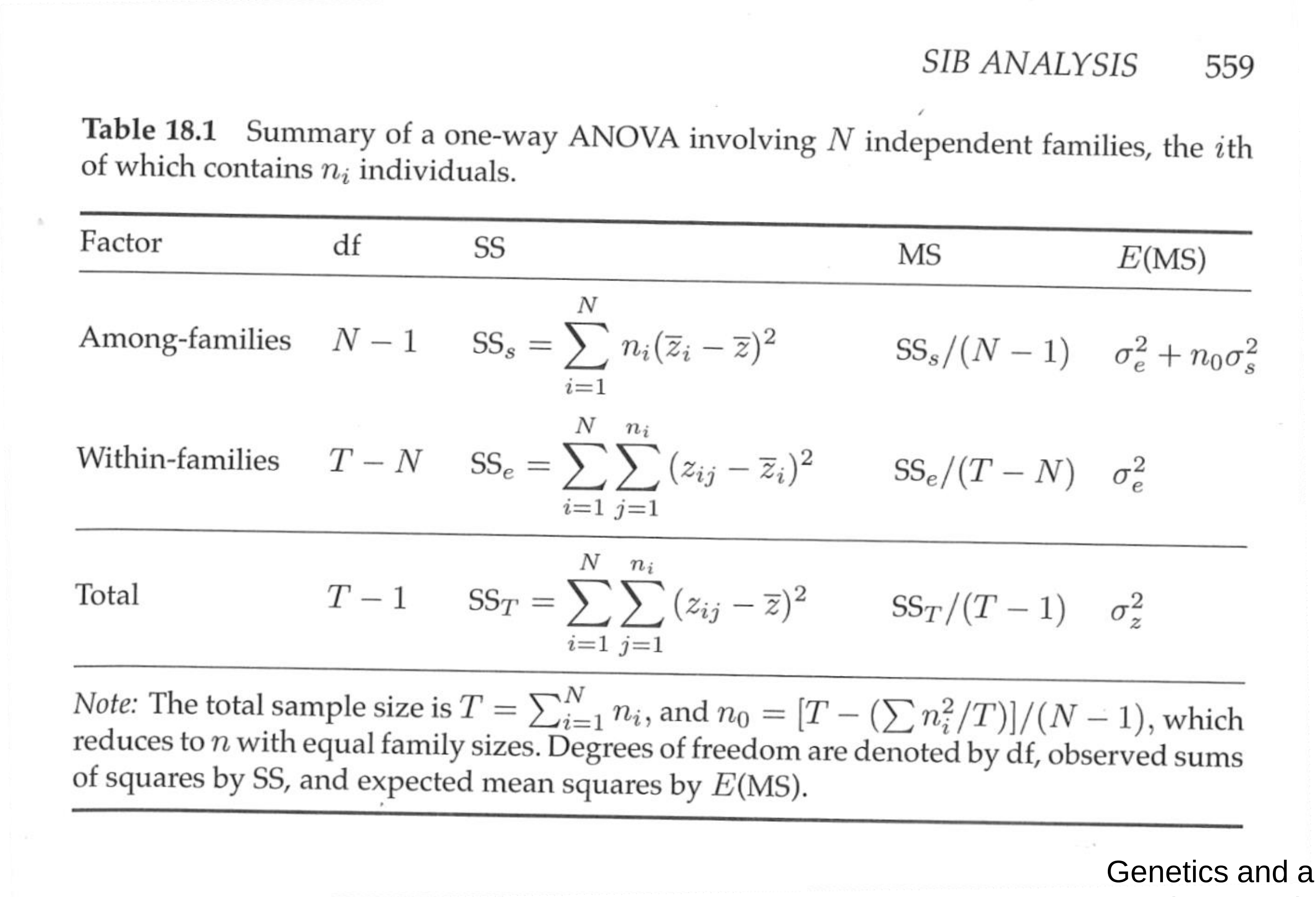


SIB ANALYSIS 555

Figure 18.1 A paternal half-sib mating design. Each male is mated to several unique (unrelated) females, and a single offspring from each female is assayed.

Genetics and analysis of quantitative traits, Lynch & Walsh

Estimation of heritability

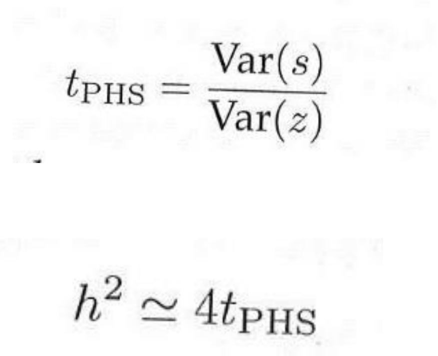


SIB ANALYSIS 559

Table 18.1 Summary of a one-way ANOVA involving N independent families, the ith of which contains n_i individuals.

Factor	df	SS	MS	E(MS)
Among-families	$N-1$	$SS_s = \sum_{i=1}^{N} n_i(\overline{z}_i - \overline{z})^2$	$SS_s/(N-1)$	$\sigma_e^2 + n_0\sigma_s^2$
Within-families	$T-N$	$SS_e = \sum_{i=1}^{N}\sum_{j=1}^{n_i}(z_{ij} - \overline{z}_i)^2$	$SS_e/(T-N)$	σ_e^2
Total	$T-1$	$SS_T = \sum_{i=1}^{N}\sum_{j=1}^{n_i}(z_{ij} - \overline{z})^2$	$SS_T/(T-1)$	σ_z^2

Note: The total sample size is $T = \sum_{i=1}^{N} n_i$, and $n_0 = [T - (\sum n_i^2/T)]/(N-1)$, which reduces to n with equal family sizes. Degrees of freedom are denoted by df, observed sums of squares by SS, and expected mean squares by E(MS).



$$t_{\text{PHS}} = \frac{\text{Var}(s)}{\text{Var}(z)}$$

$$h^2 \simeq 4t_{\text{PHS}}$$

Genetics and analysis of quantitative traits,
Lynch & Walsh

Estimation of heritability

- Usually we do not have balanced design
- Also fixed effects need to be counted
- For more general designs or pedigrees ANOVA-type of methods cannot be used
- REML (restricted maximum likelihood method)
    - Under strong development in 80’s and 90’s
    - Computational aspects needs lot of attention
    - Iterative solutions: average information and/or EM algorithm

Estimation of breeding values

- Information sources
  - Pedigree information → relationship matrix A
  - Phenotypic records (y-values)
  - Systematic effects (gender, year) Fixed effects
  - Random effects (animal, permanent environment etc)
  - Variance components related to random effect → h^2
- Form a mixed linear model
- Form mix model equations
- Solve the equations
  - Iterative methods → no need to invert large matrices

Estimation of breeding values

Matrix notation for BLUP

$$\begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z+A^{-1}\,\sigma^2_e/\sigma^2_a \end{bmatrix} \begin{bmatrix} b \\ \hat{a} \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \end{bmatrix}$$

Estimation of breeding values

$$\begin{bmatrix} b \\ \hat{a} \end{bmatrix} = \begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z + A^{-1}\,\sigma^2_e/\sigma^2_a \end{bmatrix}^{-1} \begin{bmatrix} X'y \\ Z'y \end{bmatrix}$$

Estimation of breeding values

- The linear model needs to take into accout all sources of factors that can have an effect on the phenotype
- Here is an example of the direct genetic, maternal and permanent environmental effects, all are treated as random effects

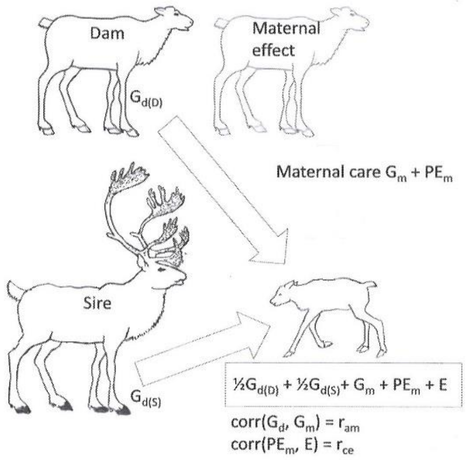


Figure 1 Direct and maternal effects on offspring phenotype. The direct effect (G_d) is split into the sire's (S) and dam's (D) additive genetic effect. The maternal effect consists of the maternal genetic effect (G_m), maternal permanent environmental effect (PE_m) and maternal temporary environmental effect (E_m). The direct-maternal genetic correlation is denoted as r_{am} and the respective direct-maternal environmental correlation as r_{ce} (the reindeer figures are modifications of those by Mauri Nieminen).

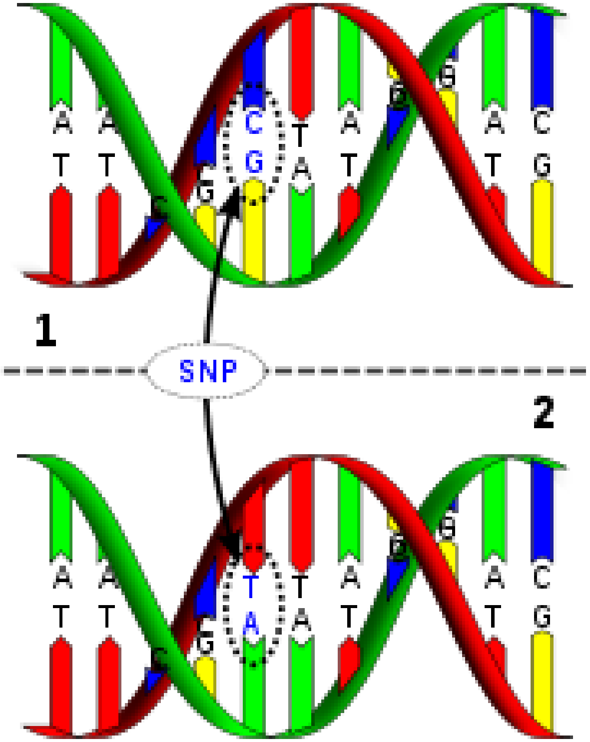
Kirsi Muuttoranta, PhD thesis

SNP: single nucleotide polymorphism

- DNA sequence variations that occur when a single nucleotide

  (A, T, C, or G)

  in the genome sequence is altered



Commercial chips

- Illumina:
  - BovineLD
  - BovineSNP50
  - BovineHD
  - PorcineSNP60
  - OvineSNP50
- Affymetrix
  - Axiom Genome-Wide BOS 1 Array
- Price 20-100€/sample

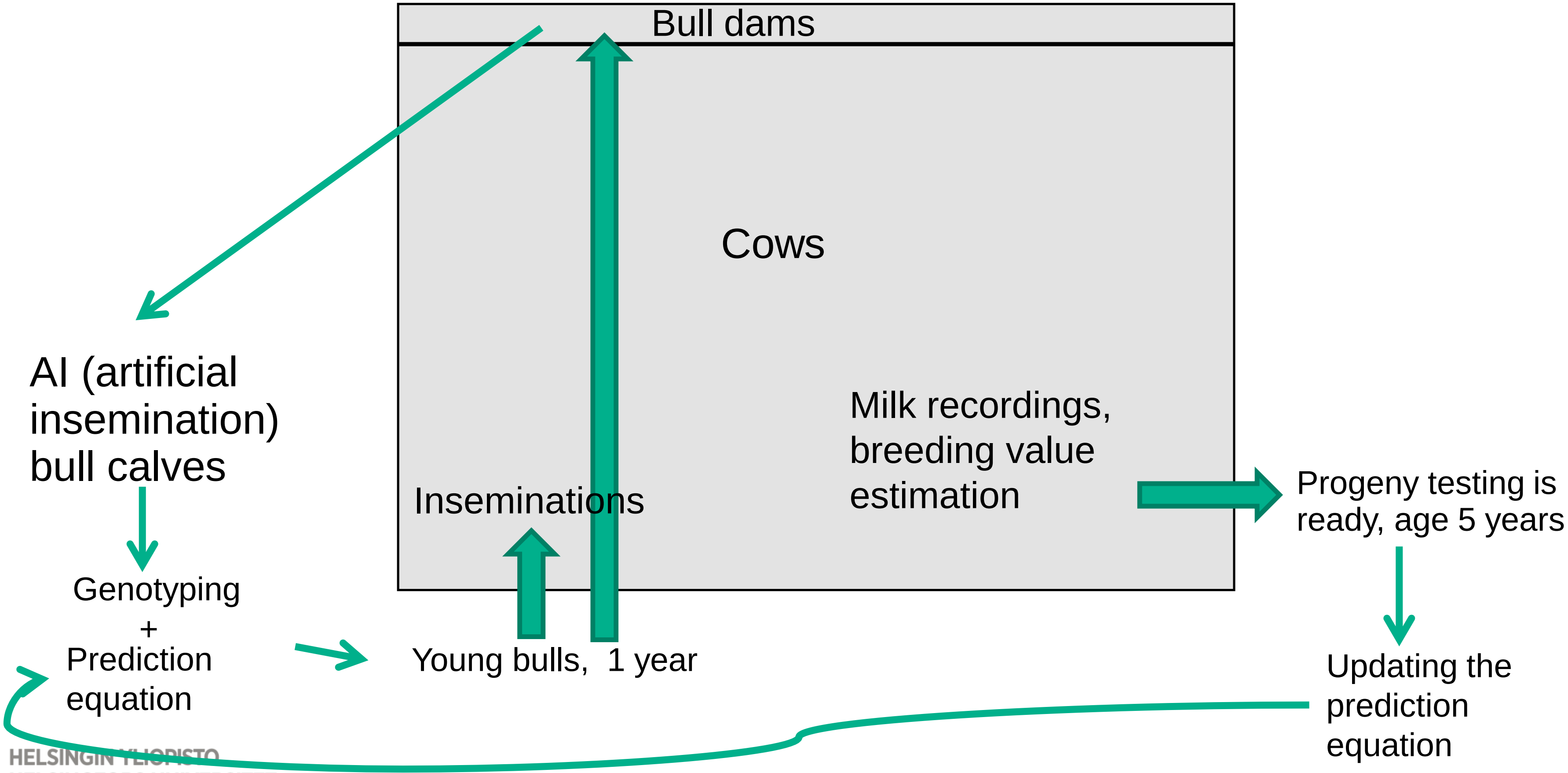
Applications of the SNP chips

- Genomic selection
  - Applied mainly in dairy cattle breeding
- Gene mapping
  - Single gene Mendelian traits
  - Polygenic traits
  - More powerful than microsatellites because the high density of SNPs allow linkage disequilibrium based gene mapping methods → the causal mutation can be mapped with more precise position

Genomic selection

- Genomic selection is the biggest biotechnological revolution in animal breeding since artificial insemination
- Moderate price of chips has enabled genotyping of thousands (millions) of production animals
- New statistical methods and programs have been developed to combine genotypic and phenotypic information in breeding value evaluation
- Genetic merit == breeding value of a animal can be predicted from it’s DNA → generation interval is short → important especially in dairy and beef cattle

An example of genomic selection in dairy cattle



Bull dams

Cows

AI (artificial insemination) bull calves

Genotyping
+
Prediction equation

Inseminations

Young bulls, 1 year

Milk recordings, breeding value estimation

Progeny testing is ready, age 5 years

Updating the prediction equation

Prediction equation

- Reference population is needed
- Animals with reliable estimates of breeding values: AI-bulls (boars) with tens to hundreds of progeny
- These animas will be genotyped (sperm, blood, hair samples)
- The number of animals in reference population varies from hundreds to thousands, Eurogenomics >25,000 AI-bulls + tens of thousands of cows

Prediction equation

- Statistical methods to estimate effect of each SNP on production
- A separate prediction model must be estimated for each trait and population
- SNPs that have the biggest effect on the trait will have the biggest weight in the equation
- Accuracy of breeding value estimates based on genomic selection varies from 50 to 80% depending on the size of the reference population, trait and homogeneity of the population

SNP-BLUP

- SNP-BLUP, mixed model:
- $\mathbf{y} = \mathbf{1}\mu + \mathbf{Zg} + \mathbf{e}$;
  - **y** is vector of observations (reference animals)
  - μ is the general mean,
  - **Z** is a n x m matrix containing a column for each marker locus,
  - **g** is a vector of length m of random SNP marker effects
  - **e** random residual effect

- An assumption: $\mathbf{g} \sim N(0, \mathbf{I}\, \sigma^2_g)$

SNP-BLUP

Matrix notation for SNP-BLUP

$$\begin{bmatrix} \mathbf{1'1} & \mathbf{1'Z} \\ \mathbf{Z'1} & \mathbf{Z'Z+I}\sigma^2_e/\sigma^2_g \end{bmatrix} \begin{bmatrix} \boldsymbol{\mu} \\ \hat{\mathbf{g}} \end{bmatrix} = \begin{bmatrix} \mathbf{1'y} \\ \mathbf{Z'y} \end{bmatrix}$$

Scalar notation for 5 SNPs:

$$y_i = \mu + z_i g_1 + z_i g_2 + z_i g_3 + z_i g_4 + z_i g_5 + e_i$$

Genomic breeding values (GEBV) or direct genomic values (DGV)

GEBV = $\mu + \mathbf{Z\hat{g}}$

For 5 SNPs and animal i:

$\text{GEBV}_i = \mu + z_i\hat{g}_1 + z_i\hat{g}_2 + z_i\hat{g}_3 + z_i\hat{g}_4 + z_i\hat{g}_5$

Hypothetical example:

$\text{GEBV}_i = 2.3 + 1*0.02 + 0*0.01 + 2*0.02 + 2*(-0.01) + 0*0.01$

$\text{GEBV}_i = 2.34$

GBLUP

Matrix notation for GBLUP

$$\begin{bmatrix} \mathbf{1'1} & \mathbf{1'} \\ \mathbf{1} & \mathbf{I} + \lambda \mathbf{G}^{-1} \end{bmatrix} \begin{bmatrix} \boldsymbol{\mu} \\ \hat{\mathbf{a}} \end{bmatrix} = \begin{bmatrix} \mathbf{1'y} \\ \mathbf{Z'y} \end{bmatrix}$$

$\mathbf{G}^{-1}$ is the genomic relatioship matrix

$\lambda = \sigma^2_e/\sigma^2_g$ and GEBV = $\mu + \hat{\mathbf{a}}$