



Centre for  
Tropical Livestock  
Genetics and Health

# Conceptual model of phenotypic values

**Chrissy Rochus**

Roslin Institute, University of Edinburgh

Slides adapted from Gregor Gorjanc and Gabriela Mafra Fortuna





Centre for  
Tropical Livestock  
Genetics and Health



THE UNIVERSITY of EDINBURGH  
The Royal (Dick) School  
of Veterinary Studies



Biotechnology and  
Biological Sciences  
Research Council

<https://vet.ed.ac.uk/roslin>



# The Roslin Institute



## Sustainable Agriculture

Improving the health, productivity and welfare of farmed animals while mitigating impacts on the climate and environment.



## Infectious Diseases

Improving the detection, prevention and treatment of infectious diseases of animals, and those that pass from animals to people.



## Enhancing Health

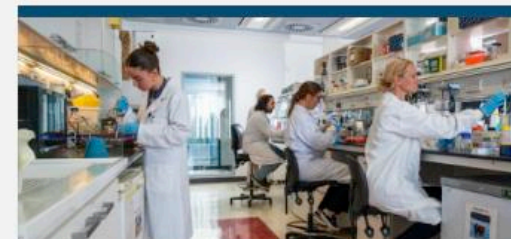
Leveraging our unique expertise, infrastructure and resources to understand animal and human health, ageing and disorders.



## Research

Animal research

INSTITUTE



## Study with us

Studentships



## Work with us

Vacancies





# Conceptual model of phenotypic values

## Learning objectives

- Understanding how some types of variation in DNA looks like and how we work with that variation in computers (allele dosages)
- Connect variation in DNA and environment with the variation in phenotypic values
- Familiarise with the processes of mutation, recombination, segregation of DNA within pedigrees that give rise to DNA lottery
- Train common vector & matrix operations in R



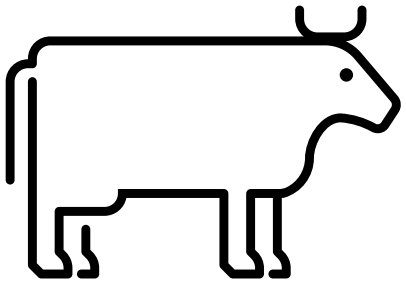
# Conceptual model of phenotypic values

## Materials

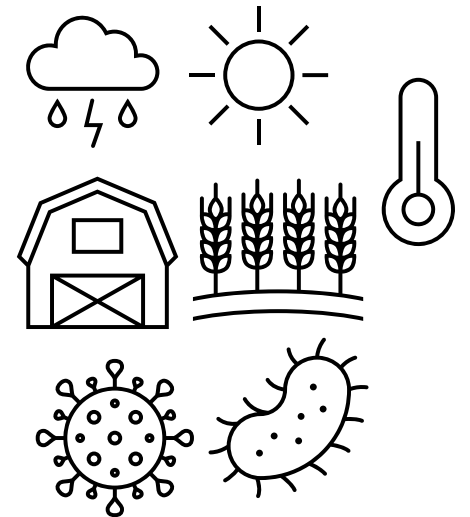
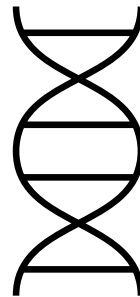
- These slides (shared PDF) & lecture
- Exercises
- Reading material (shared PDF)
- AlphaSimR on-line course (for later)



# Phenotypic variation



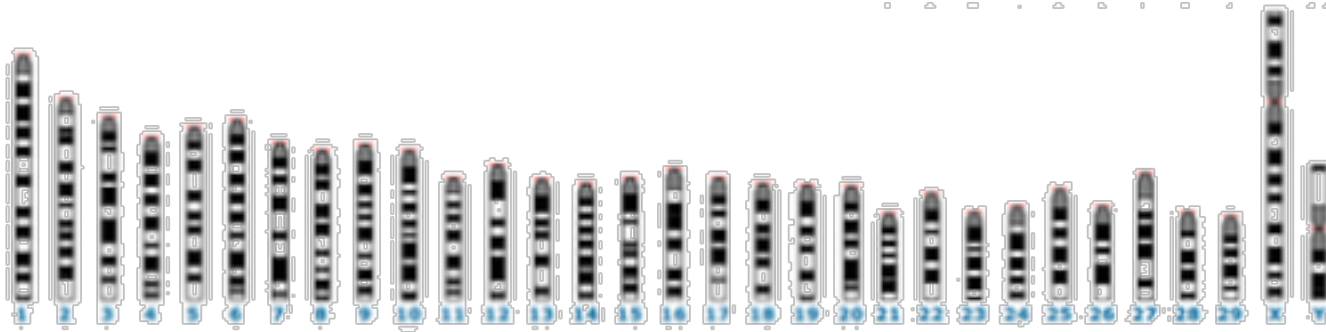
Phenotype = Genetics + Environment





# Genome (cattle example)

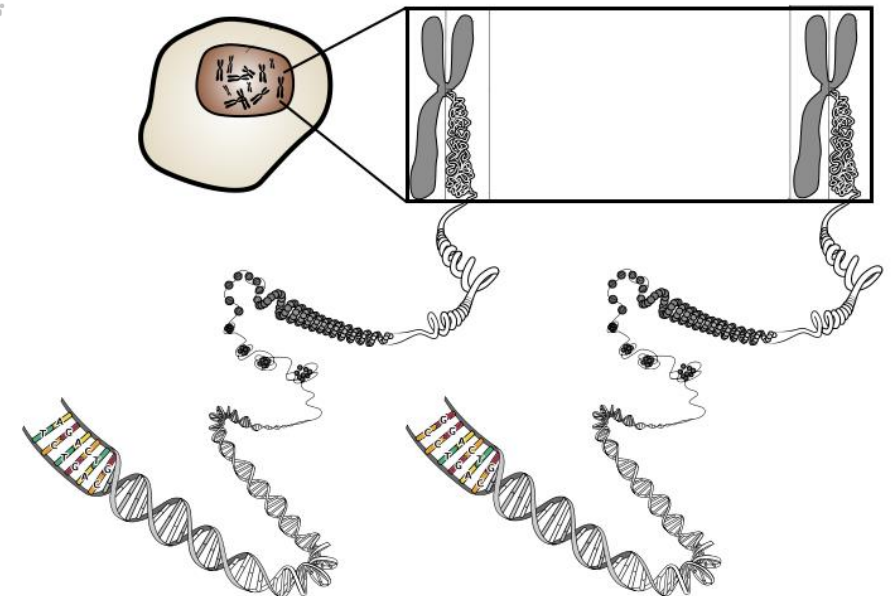
2 x 30 chromosomes



DNA, 2 x 3 billion (3,000,000,000) base pairs

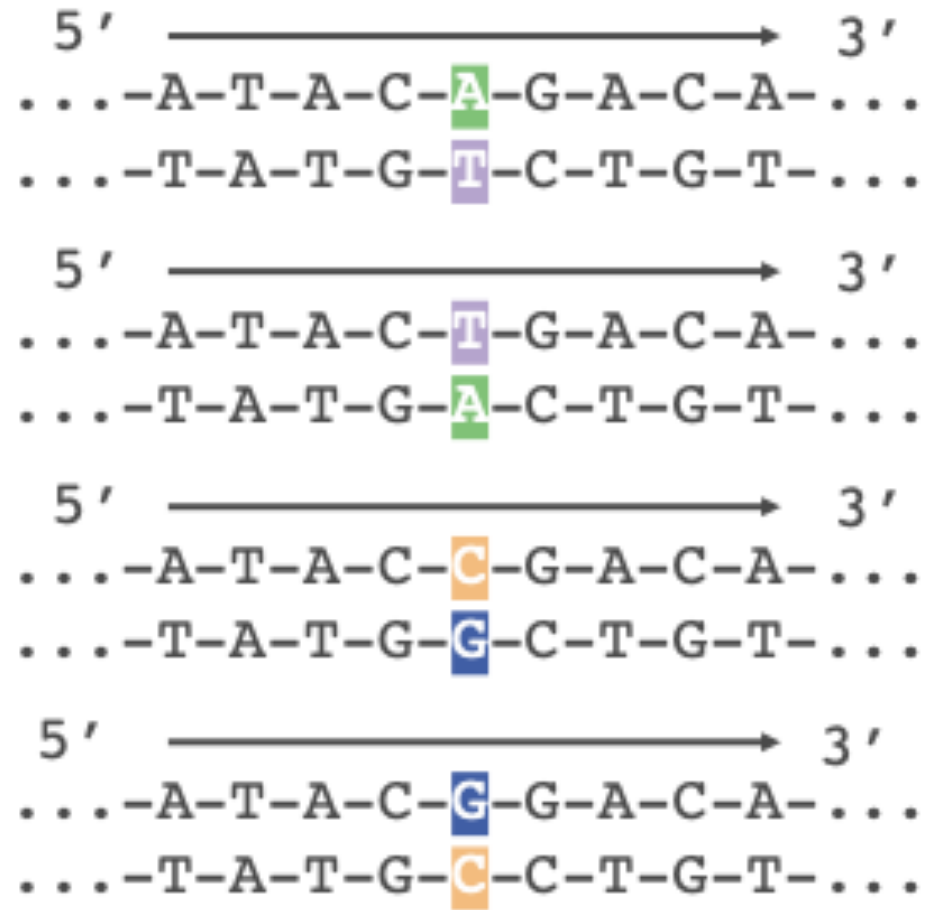
Adenine  
Thymine  
A  
T

Cytosine  
Guanine  
C  
G





# Single Nucleotide Polymorphism





Centre for  
Tropical Livestock  
Genetics and Health

# How many SNPs and other variants

The sequences of >150,119 genomes in the UK biobank

<https://doi.org/10.1038/s41586-022-04965-x>

~600M SNPs (representing 7% of all possible human SNPs)

~60M indels

~1M structural variants

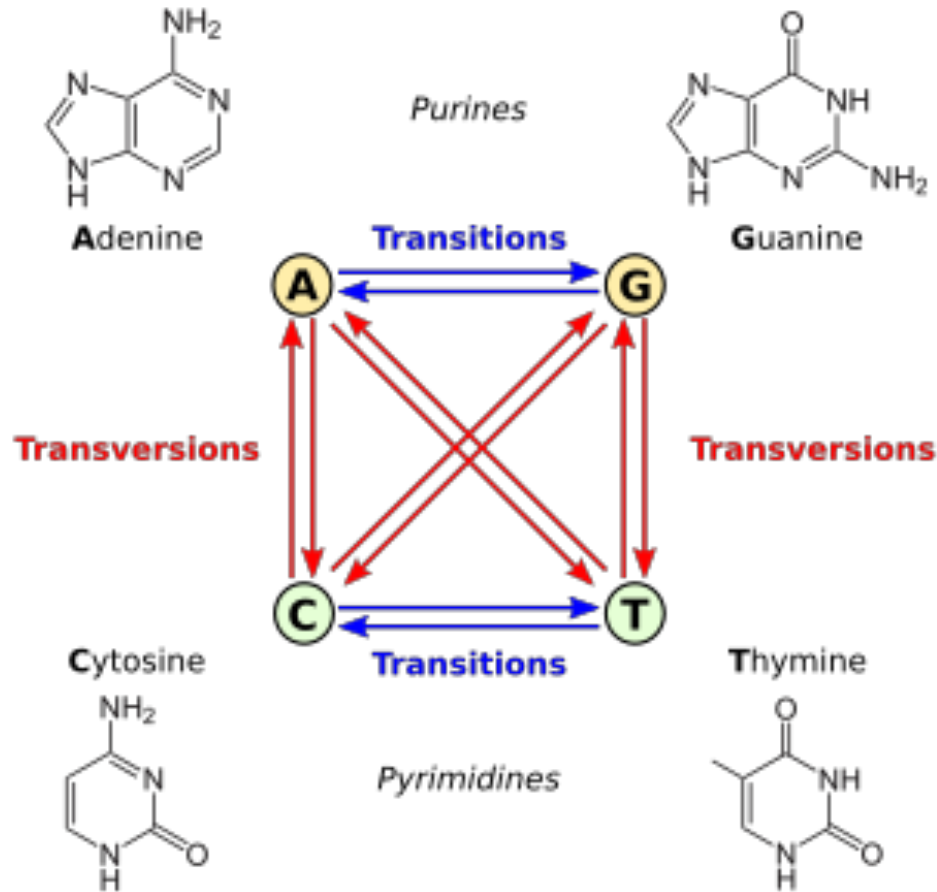
~3M microsatellites

MEGA-SCALE DATA





# Single Nucleotide Polymorphism (SNP)



5' → 3'

...-A-T-A-C-**A**-G-A-C-A-...

...-T-A-T-G-**T**-C-T-G-T-...

5' → 3'

...-A-T-A-C-**T**-G-A-C-A-...

...-T-A-T-G-**A**-C-T-G-T-...

5' → 3'

...-A-T-A-C-**C**-G-A-C-A-...

...-T-A-T-G-**G**-C-T-G-T-...

5' → 3'

...-A-T-A-C-**G**-G-A-C-A-...

...-T-A-T-G-**C**-C-T-G-T-...



# Bi-allelic SNP alleles, genotypes & dosages

5' → 3'  
...-A-T-A-C-**A**-G-A-C-A-...  
...-T-A-T-G-**T**-C-T-G-T-... Ref. allele --> 0  
5' → 3' Ref. allele --> 0 --> 0  
...-A-T-A-C-**A**-G-A-C-A-...  
...-T-A-T-G-**T**-C-T-G-T-...



# Bi-allelic SNP alleles, genotypes & dosages

5' → 3'  
...-A-T-A-C-**A**-G-A-C-A-...  
...-T-A-T-G-**T**-C-T-G-T-... Ref. allele --> 0  
Ref. allele --> 0 --> 0

5' → 3'  
...-A-T-A-C-**A**-G-A-C-A-...  
...-T-A-T-G-**T**-C-T-G-T-... Ref. allele --> 0  
Alt. allele --> 1 --> 1

5' → 3'  
...-A-T-A-C-**G**-G-A-C-A-...  
...-T-A-T-G-**C**-C-T-G-T-... Alt. allele --> 1  
Alt. allele --> 1 --> 2





# Genome-wide haplotypes & genotype

Haplotype 1

0	1	1	0	0	1
---	---	---	---	---	---

Haplotype 2

1	1	1	1	0	0
---	---	---	---	---	---

Genotype

1	2	2	1	0	1
---	---	---	---	---	---



Centre for  
Tropical Livestock  
Genetics and Health

## Take home message 1:

Encoding haplotypes as a series of 0 & 1

Encoding genotypes as a series of 0, 1, & 2





# SNP allele dosages in haplotypes and genotypes

Haplotype 3

0	0	1	1	0	1
---	---	---	---	---	---

Haplotype 4

0	0	0	0	1	1
---	---	---	---	---	---

Genotype

--	--	--	--	--	--



# SNP allele dosages in haplotypes and genotypes

Haplotype 3

0	0	1	1	0	1
---	---	---	---	---	---

Haplotype 4

0	0	0	0	1	1
---	---	---	---	---	---

Genotype

0	0	1	1	1	2
---	---	---	---	---	---



# SNP allele dosages in haplotypes and genotypes

Haplotype 5

1	1	1	0	1	1
---	---	---	---	---	---

Haplotype 6

1	1	0	0	1	1
---	---	---	---	---	---

Genotype

--	--	--	--	--	--



# SNP allele dosages in haplotypes and genotypes

Haplotype 5

1	1	1	0	1	1
---	---	---	---	---	---

Haplotype 6

1	1	0	0	1	1
---	---	---	---	---	---

Genotype

2	2	1	0	2	2
---	---	---	---	---	---



# Summarising SNP genotype data

Table 1: Genotype allele dosages at two loci in five animals and corresponding summary

Animal	SNP1	SNP2
1	0	0
2	2	1
3	2	0
4	1	1
5	0	0
Mean	1.0	0.40
Standard deviation	1.0	0.55
Variance	1.0	0.30
Allele frequency	0.50	0.20
Genic variance	0.50	0.32
Correlation	0.46	





# Summarising SNP genotype data

Table 1: Genotype allele dosages at two loci in five animals and corresponding summary

Animal	SNP1				
1	0	$\bar{x} = \frac{\sum x_i}{N}$	$\sigma^2 = \frac{\sum (x_i - \bar{x})^2}{N - 1}$		
2	2				
3	2				
4	1				
5	0				
Mean	1.0				
Standard deviation	1.0	$\sigma = \sqrt{\sigma^2}$			
Variance	1.0				
Allele frequency	0.50	$\frac{\sum x_i}{2N}$	$\bar{x}(1 - \frac{\bar{x}}{2})$		
Genic variance	0.50				
Correlation					



# Summarising SNP genotype data

Table 1: Genotype allele dosages at two loci in five animals and corresponding summary

Animal	SNP1
1	0
2	2
3	2
4	1
5	0
Mean	1.0
Standard deviation	1.0
Variance	1.0
Allele frequency	0.50
Genic variance	0.50
Correlation	

$$\bar{x} = \frac{\sum x_i}{N} = \frac{0 + 2 + 2 + 1 + 0}{5}$$

$$\begin{aligned}\sigma^2 &= \frac{\sum (x_i - \bar{x})^2}{N - 1} \\ &= \frac{(0 - 1)^2 + (2 - 1)^2 + (2 - 1)^2 + (1 - 1)^2 + (0 - 1)^2}{5 - 1}\end{aligned}$$

$$\sigma = \sqrt{\sigma^2} = \sqrt{1}$$

$$\frac{\sum x_i}{2N} = \frac{0 + 2 + 2 + 1 + 0}{2(5)}$$

$$\bar{x}(1 - \frac{\bar{x}}{2}) = 1(1 - \frac{1}{2})$$

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}$$



## Genotype summary (2)

Animal	SNP3	SNP4
1	1	0
2	0	1
3	2	2
4	0	1
5	0	0
Mean		
Standard deviation		
Variance		
Allele frequency		
Genic variance		
Correlation		



## Genotype summary (2)

Animal	SNP3	SNP4
1	1	0
2	0	1
3	2	2
4	0	1
5	0	0
Mean	0.6	0.8
Standard deviation	0.89	0.84
Variance	0.8	0.7
Allele frequency	0.3	0.4
Genic variance	0.42	0.48
Correlation	0.55	



## Genotype summary (3)

Animal	SNP5	SNP6
1	1	0
2	0	1
3	1	0
4	0	1
5	0	0

Mean

Standard deviation

Variance

Allele frequency

Genic variance

Correlation





## Genotype summary (3)

Animal	SNP5	SNP6
1	1	0
2	0	1
3	1	0
4	0	1
5	0	0
Mean	0.6	0.4
Standard deviation	0.55	0.55
Variance	0.3	0.3
Allele frequency	0.3	0.2
Genic variance	0.42	0.32
Correlation	-1	



# Variation in phenotypic values

$$y_i = \mu + g_i + e_i + g_i \times e_i$$

$y_i$  - Phenotypic value

$\mu$  - Population mean(s)

$g_i$  - Genetic value (deviation from mean)

$e_i$  - Environmental value (deviation from mean)

$g_i \times e_i$  - Genotype-by-environment interaction (deviation from mean)



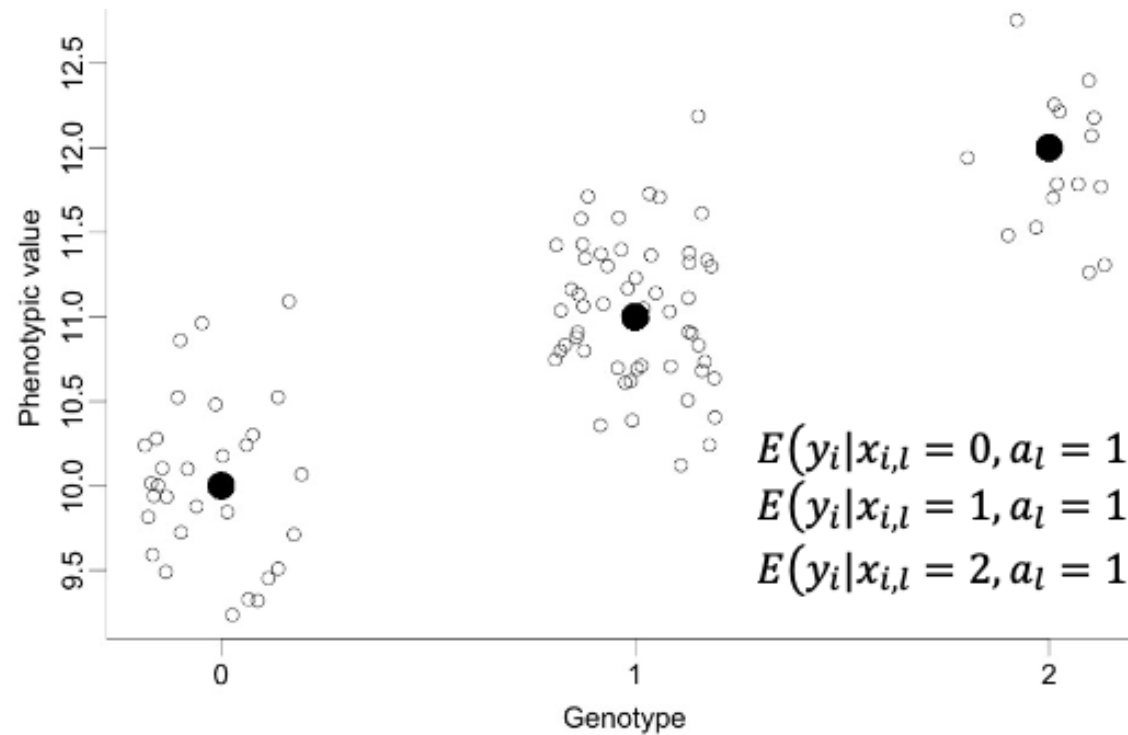
Centre for  
Tropical Livestock  
Genetics and Health

Genetic values comprise individual DNA  
locus effects

$$g_i = g_{i,1} + g_{i,2} + \dots + g_{i,k} + g_{i,1} \times g_{i,2} + \dots$$



$$y_i = \mu + g_i + e_i$$



$$E(y_i | x_{i,l} = 0, a_l = 1) = \mu + x_{i,l}a_l = 10 + 0 \times 1 = 10,$$

$$E(y_i | x_{i,l} = 1, a_l = 1) = \mu + x_{i,l}a_l = 10 + 1 \times 1 = 11,$$

$$E(y_i | x_{i,l} = 2, a_l = 1) = \mu + x_{i,l}a_l = 10 + 2 \times 1 = 12.$$



## Exercise 4

3 genotypes (0, 1, 2)

Allele substitution effect (-2)

Population mean (15)

$$y = 15 + 0 \times (-2)$$

$$y = 15 + 1 \times (-2)$$

$$y = 15 + 2 \times (-2)$$

Produce expected phenotype

Sample environment effect ( $\sigma_e=1$ )

Produce realised phenotype





## Multi-locus version

Haplotype 1

0	1	1	0	0	1
---	---	---	---	---	---

Haplotype 2

1	1	1	1	0	0
---	---	---	---	---	---

Genotype

1	2	2	1	0	1
---	---	---	---	---	---

X

+1	+2	-1	+1	+1	-2
----	----	----	----	----	----

Allele  
substitution  
effect



Haplotype 1

0	+2	-1	0	0	-2
---	----	----	---	---	----

Haplotype 2

+1	+2	-1	+1	0	0
----	----	----	----	---	---

Genotype

+1	+4	-2	+1	0	-2
----	----	----	----	---	----

-1

+3

+2

Values



## Allele dosage to genetic value

Haplotype 3

0	0	1	1	0	1
---	---	---	---	---	---

Haplotype 4

0	0	0	0	1	1
---	---	---	---	---	---

Genotype

0	0	1	1	1	2
---	---	---	---	---	---



## Allele dosage to genetic value

Haplotype 3

0	0	1	1	0	1
---	---	---	---	---	---

Haplotype 4

0	0	0	0	1	1
---	---	---	---	---	---

Genotype

0	0	1	1	1	2
---	---	---	---	---	---

X

+1	+1	-3	+2	-2	-1
----	----	----	----	----	----



Haplotype 3

--	--	--	--	--	--

Haplotype 4

--	--	--	--	--	--

Genotype

--	--	--	--	--	--

Values



## Allele dosage to genetic value

Haplotype 3

0	0	1	1	0	1
---	---	---	---	---	---

Haplotype 4

0	0	0	0	1	1
---	---	---	---	---	---

Genotype

0	0	1	1	1	2
---	---	---	---	---	---

Allele dosage to genetic

$\times$

+1	+1	-3	+2	-2	-1
----	----	----	----	----	----



Haplotype 3

0	0	-3	+2	0	-1
---	---	----	----	---	----

-2

Haplotype 4

0	0	0	0	-2	-1
---	---	---	---	----	----

-3

Values

Genotype

0	0	-3	+2	-2	-2
---	---	----	----	----	----

-5



## Allele dosage to genetic value

Haplotype 5

1	1	1	0	1	1
---	---	---	---	---	---

Haplotype 6

1	1	0	0	1	1
---	---	---	---	---	---

Genotype

--	--	--	--	--	--

X

+1	+1	-3	+2	-2	-1
----	----	----	----	----	----



Haplotype 5

--	--	--	--	--	--

Haplotype 6

--	--	--	--	--	--

Genotype

--	--	--	--	--	--

Values



# Allele dosage to genetic value

Haplotype 5

1	1	1	0	1	1
---	---	---	---	---	---

Haplotype 6

1	1	0	0	1	1
---	---	---	---	---	---

Genotype

2	2	1	0	2	2
---	---	---	---	---	---

X

+1	+1	-3	+2	-2	-1
----	----	----	----	----	----



Haplotype 5

+1	+1	-3	0	-2	-1
----	----	----	---	----	----

-4

Haplotype 6

+1	+1	0	0	-2	-1
----	----	---	---	----	----

-1

Values

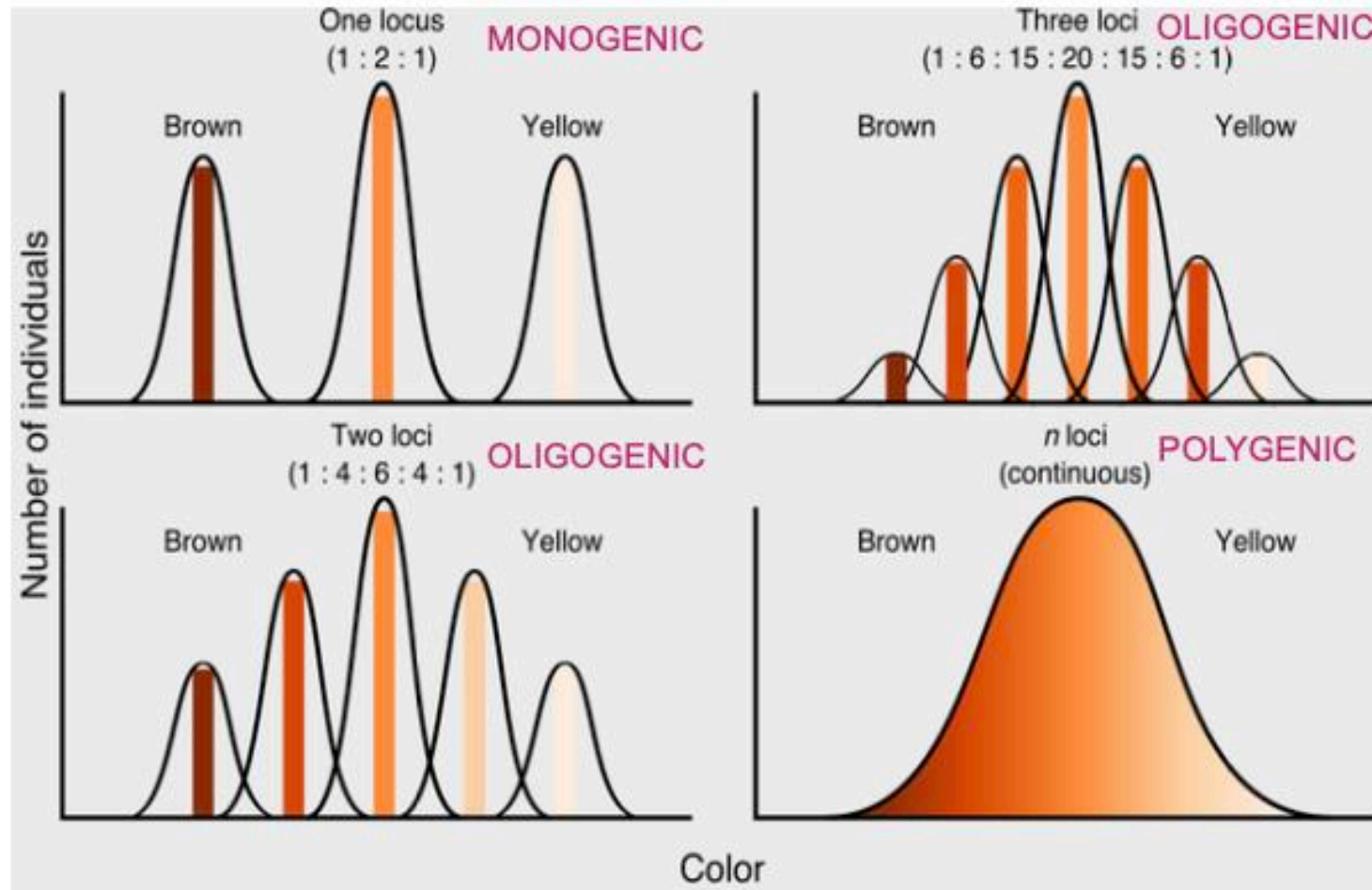
Genotype

+2	+2	-3	0	-4	-2
----	----	----	---	----	----

-5



# Hypothetical architecture for cattle coat colour





Centre for  
Tropical Livestock  
Genetics and Health

# Sources of variation between individuals

Discuss!





# Sources of variation between individuals

## Different DNA

- Due to different mutations
- These were inherited from parents following recombination and segregation of parental genomes
- Individual mutations can have effects
- Combinations of mutations can have effects (dominance and epistasis)

## Different environment

- Nutrition, water, health, ...

## Different combination of DNA and environment



Centre for  
Tropical Livestock  
Genetics and Health

# Sources of quantitative genetic variation within family





Centre for  
Tropical Livestock  
Genetics and Health

# Sources of quantitative genetic variation within family



F – frizzled  
f – normal

<https://www.poultryclub.org/breeds/>

Frizzle

Probability of each  
offspring being  
frizzled:



50%



# Sources of quantitative genetic variation within family



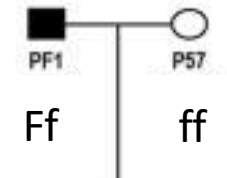
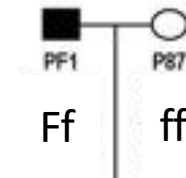
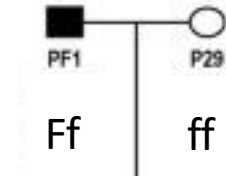
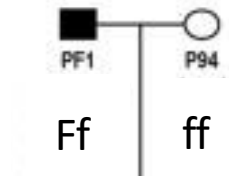
F – frizzled  
f – normal

<https://www.poultryclub.org/br>

Frizzle

Probability of each  
offspring being  
frizzled:

50%





# Sources of quantitative genetic variation within family



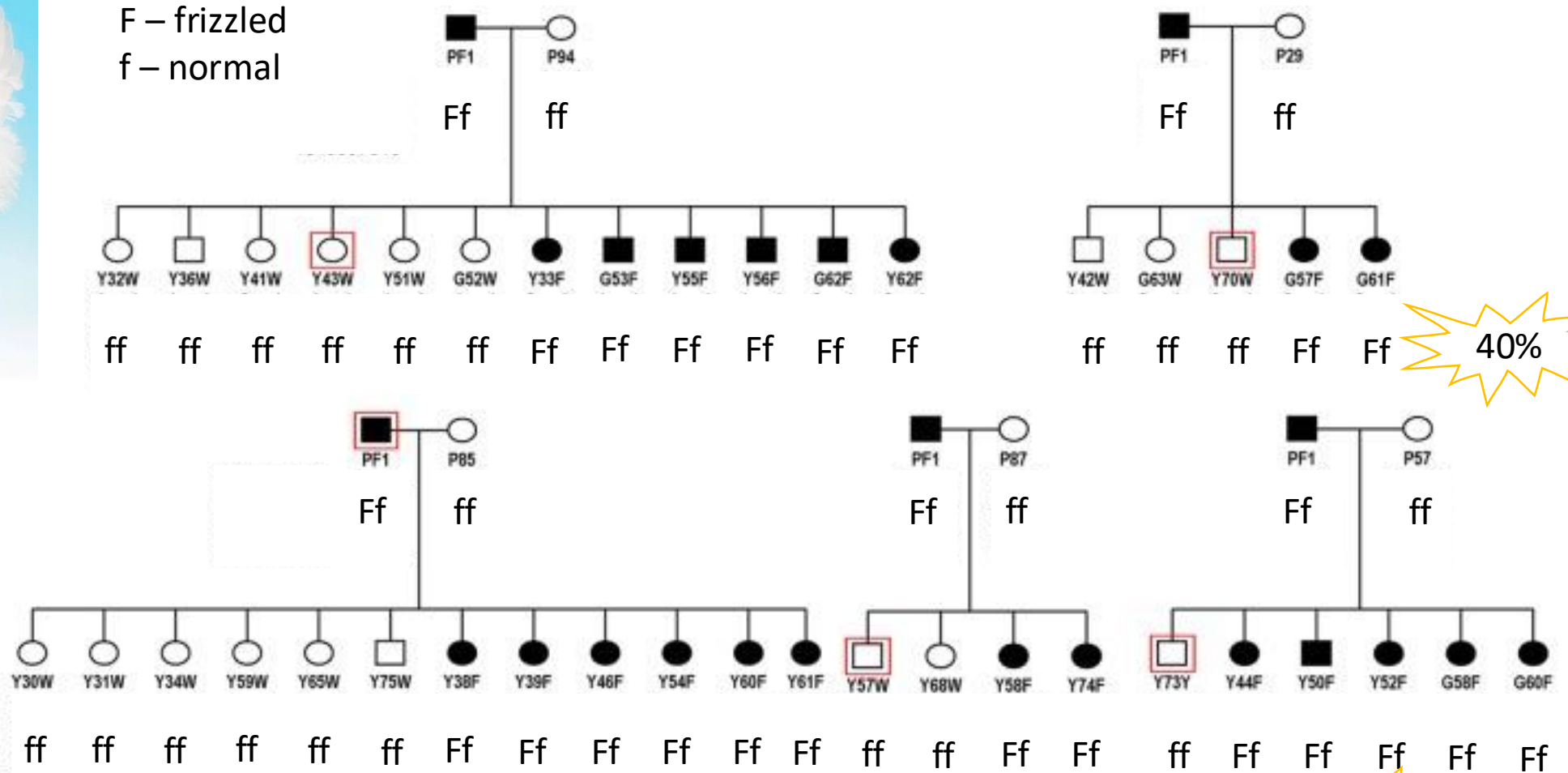
<https://www.poultryclub.org/br>

Frizzle

Probability of each  
offspring being  
frizzled:

50%

F – frizzled  
f – normal



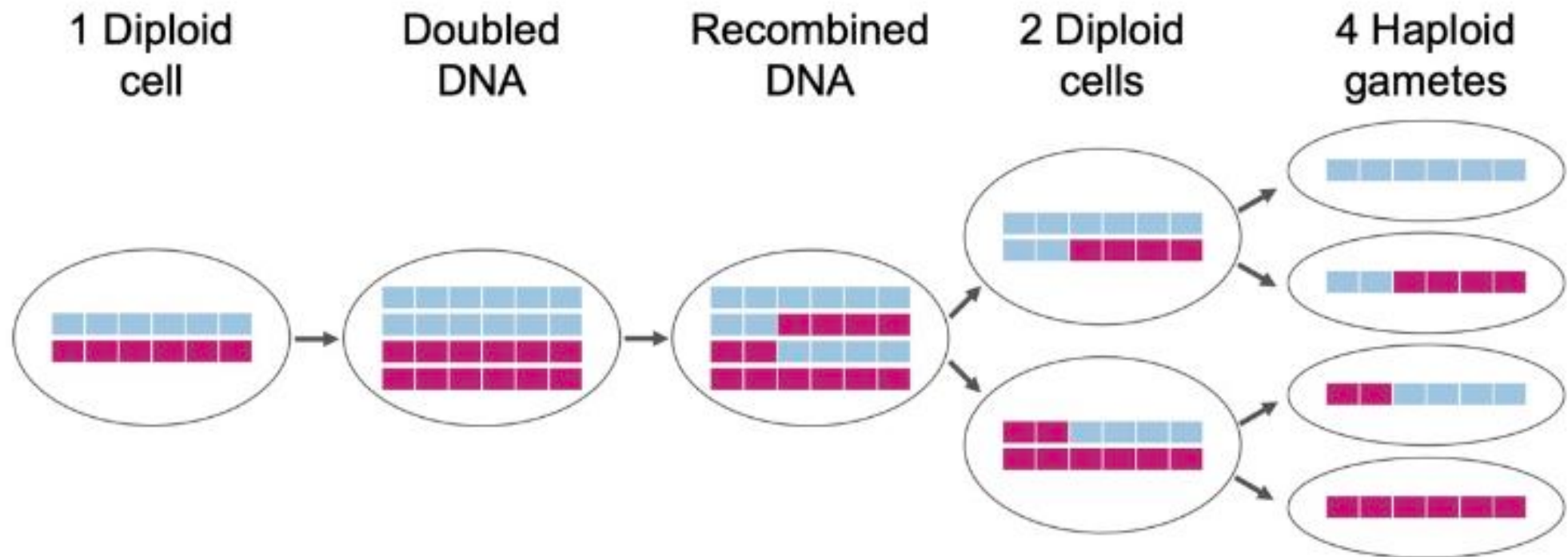


Generate combinations of father's chromosomes in his gametes

- 1 chromosome pair [1L 1D]
- 2 chromosome pairs [1L 1D, 2L 2D]
- 3 chromosome pairs [1L 1D, 2L 2D, 3L 3D]

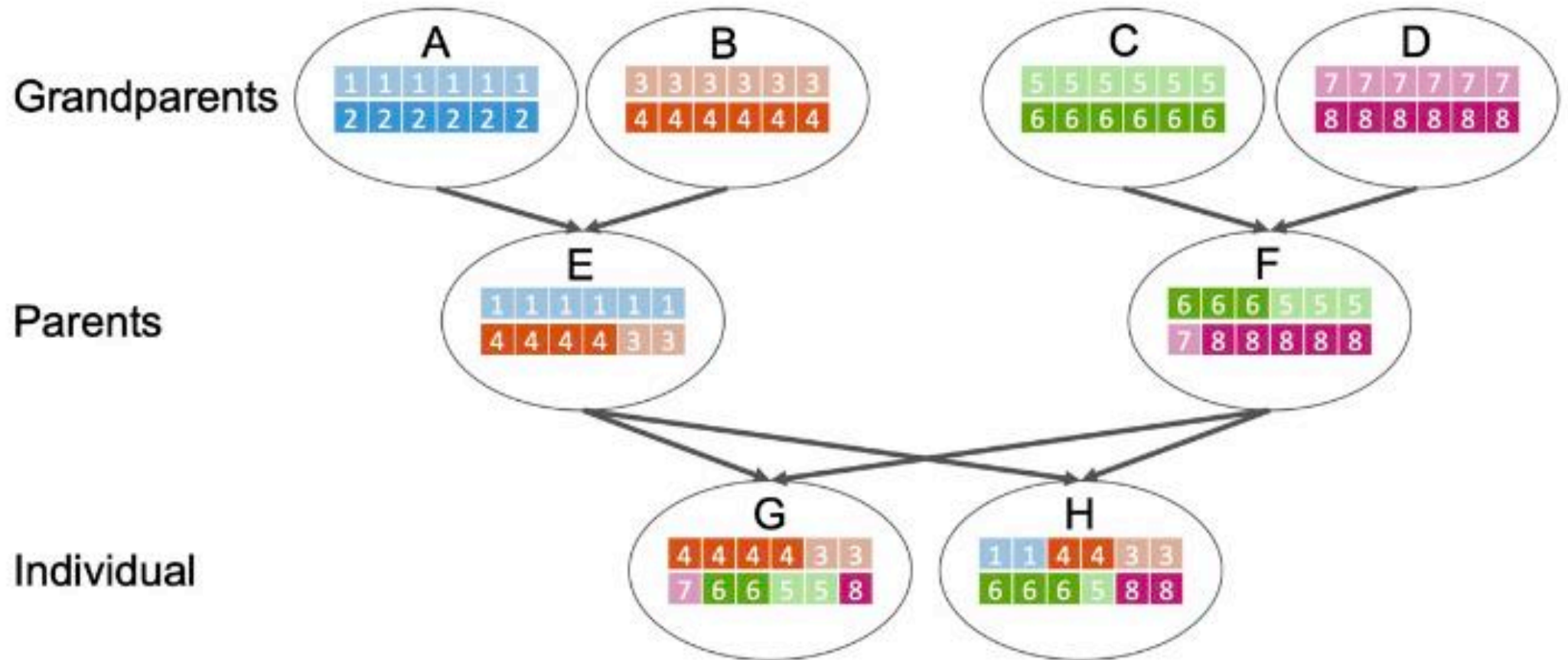


# Meiosis





# Meiosis in a pedigree (IBS and IBD)







Generate more progeny with parents E & F

1	1	1	1	1	1
4	4	4	4	3	3

6	6	6	5	5	5
7	8	8	8	8	8



# Parent average and Mendelian sampling terms

$$g_i = \frac{1}{2} g_{f(i)} + \frac{1}{2} g_{m(i)} + r_i$$

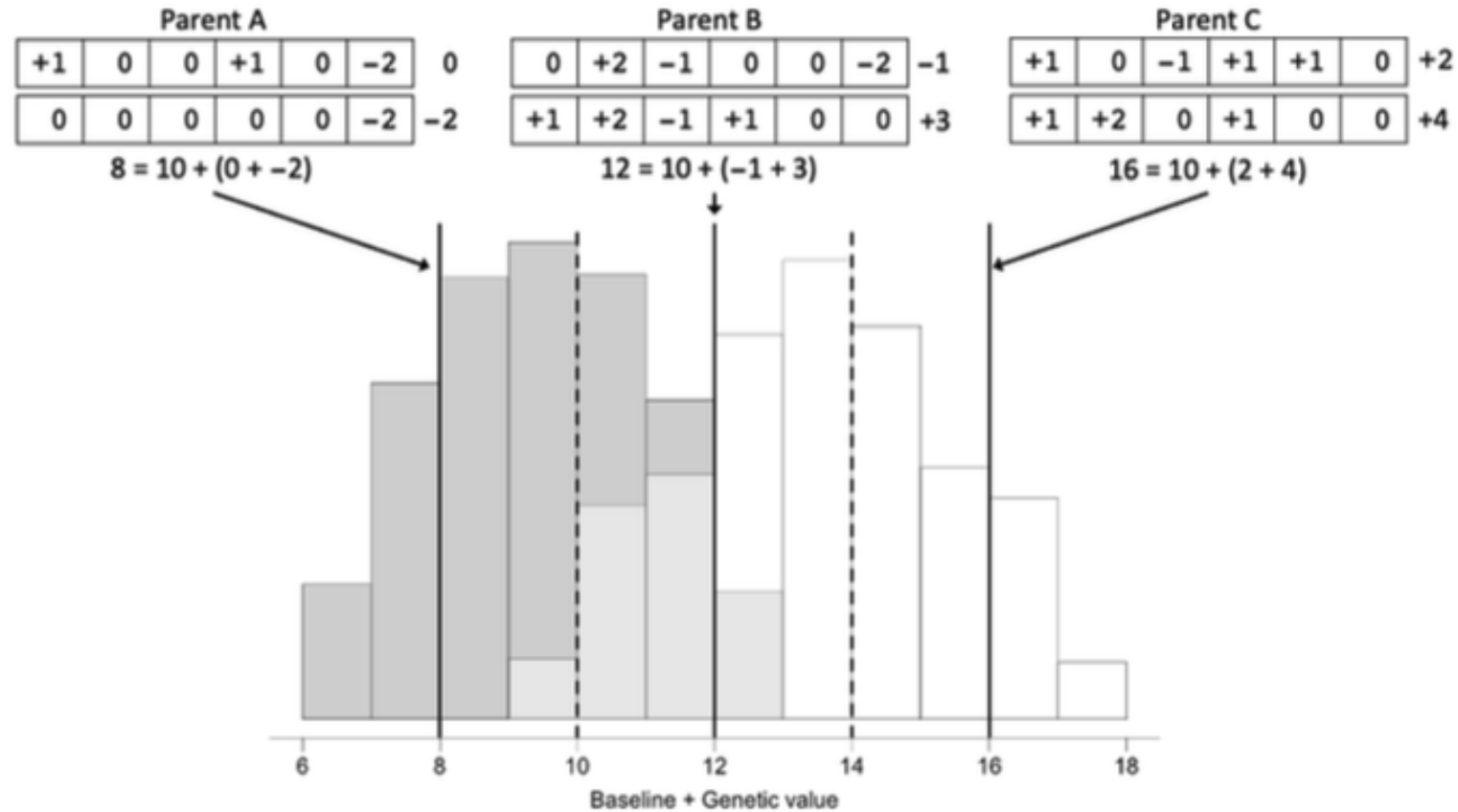
$$g_i = g_{i,1} + g_{i,2},$$

$$g_{i,1} = \frac{1}{2} g_{f(i),1} + \frac{1}{2} g_{f(i),2} + r_{i,1},$$

$$g_{i,2} = \frac{1}{2} g_{m(i),1} + \frac{1}{2} g_{m(i),2} + r_{i,2}.$$

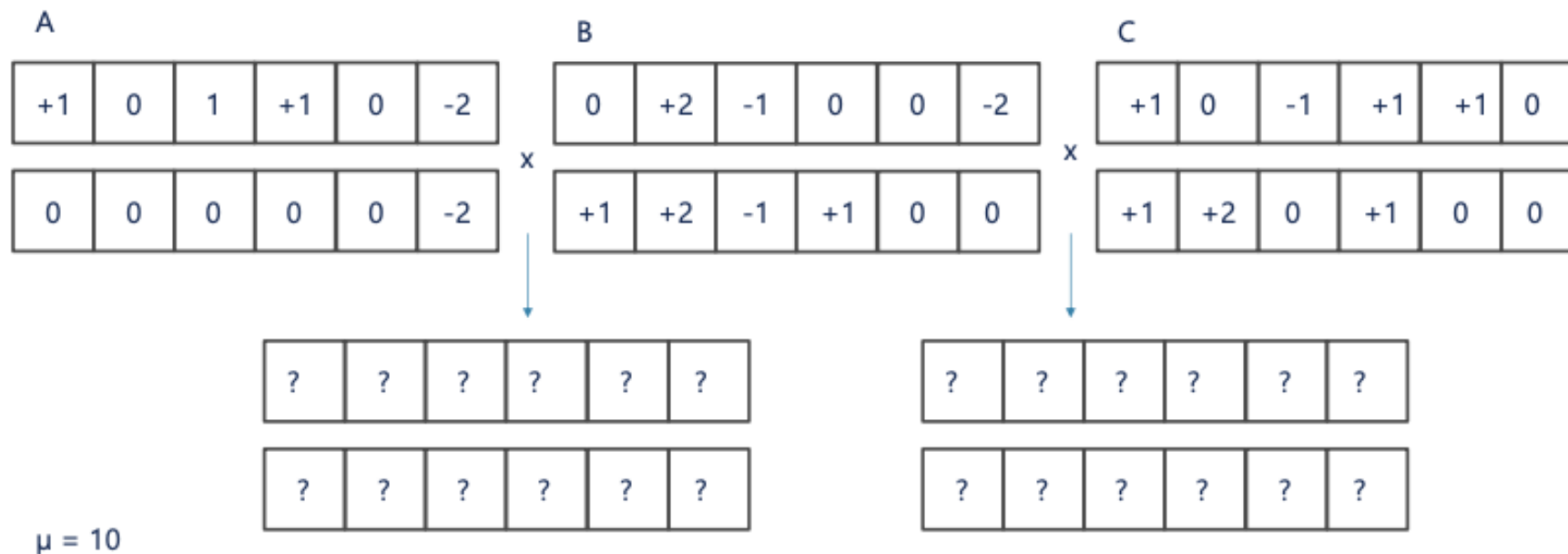


# Between and within family genetic variation



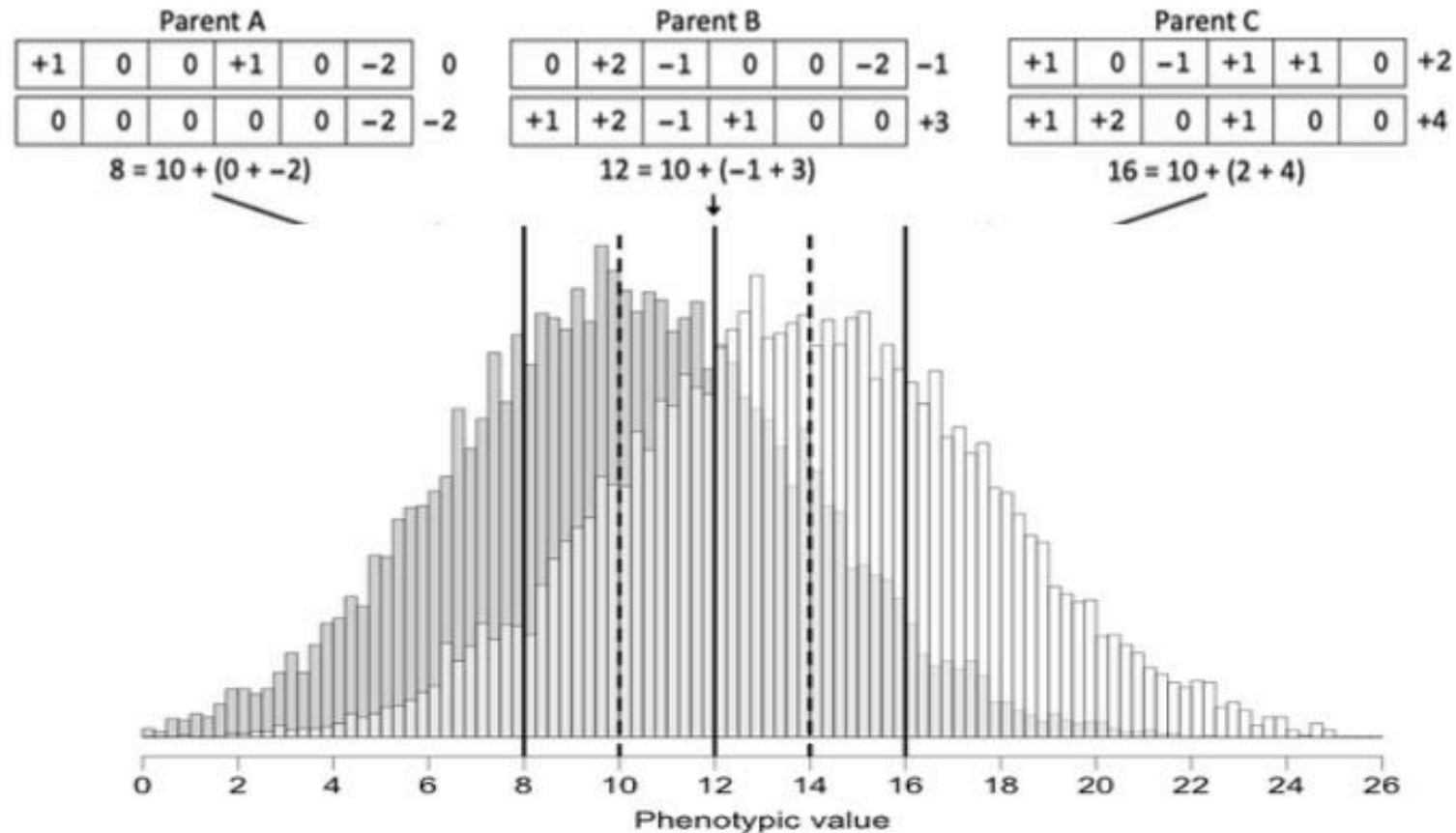


Generate progeny from the pairs and calculate their genetic value





# Between and within family phenotypic variation





Take the 4 genetic values from the previous example and add environment effect  
(can use the `rnorm()`)

```
for (Year in 1:10) {  
  Pop = randCross2(males = Sires,  
    females = Dams,  
    nCrosses = 750,  
    nProgeny = 100)  
  
  Dams = selectInd(Pop,  
    nInd = 750,  
    sex = "F")  
  
  Sires = selectInd(Pop,  
    nInd = 25,  
    sex = "M")  
}
```

```
for (Year in 1:10) {  
  Variety = selectInd(EYT, nInd = 1)  
  EYT = selectInd(AYT, nInd = 10)  
  AYT = selectInd(PYT, nInd = 50)  
  PYT = selectInd(HDRW, nInd = 500)  
  HDRW = makeDH(F1, nDH = 100)  
  Parents = c(EYT, AYT)  
  F1 = randCross(Parents, nCrosses = 100)  
}
```

Free short online course

# Breeding Programme Modelling with AlphaSimR



THE UNIVERSITY  
of EDINBURGH



THE  
**DATA LAB**  
value from data



Data-Driven  
Innovation

Part of the Edinburgh & South East Scotland City Region Deal





Centre for  
Tropical Livestock  
Genetics and Health

# Free online course on AlphaSimR

edX platform

<https://www.edx.org/learn/animal-breeding/the-university-of-edinburgh-breeding-programme-modelling-with-alphasimr>

5 weeks (~20+ hours)

1. The big question
2. Simulation of DNA and phenotypes
3. DNA lottery
4. Selection
5. Breeding programme





# Conceptual model of phenotypic values

## Learning objectives

- Understanding how some types of variation in DNA looks like and how we work with that variation in computers (allele dosages)
- Connect variation in DNA and environment with the variation in phenotypic values
- Familiarise with the processes of mutation, recombination, segregation of DNA within pedigrees that give rise to DNA lottery
- Train common vector & matrix operations in R