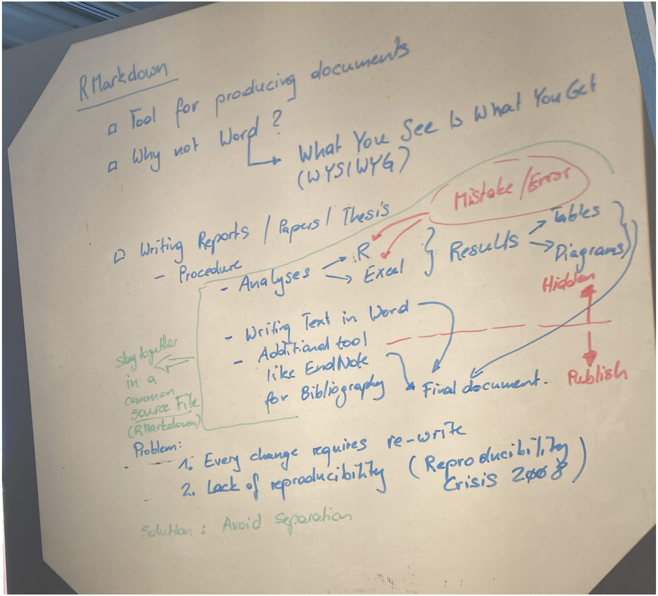
OHP Page 1



R Markdown

- ☐ Tool for producing documents
- ☐ Why not Word? → What You See Is What You Get (WYSIWYG)
- ☐ Writing Reports / Papers / Thesis
  - Procedure
    - Analyses → R, Excel } Results → Tables, Diagrams
    - Writing Text in Word
    - Additional tool like EndNote for Bibliography

→ Final document. Publish

Mistake / Error

Hidden

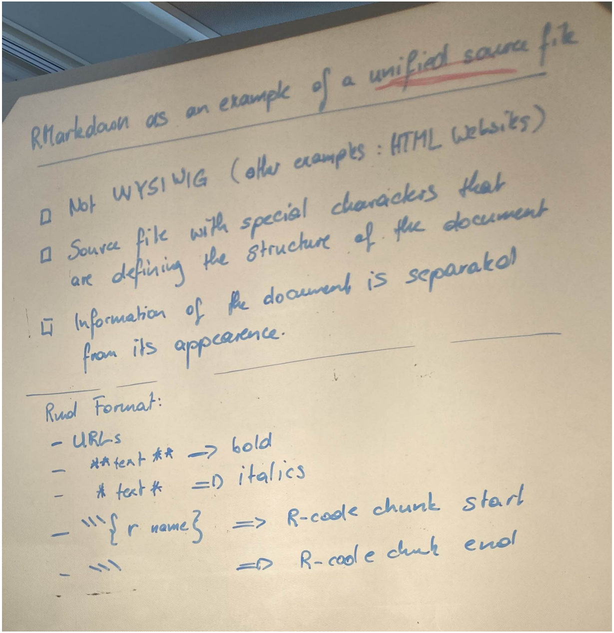
stay together in a common source File (RMarkdown)

Problem:

1. Every change requires re-write
2. Lack of reproducibility (Reproducibility Crisis 2008)

Solution: Avoid Separation

OHP Page 2



RMarkdown as an example of a unified source file

☐ Not WYSIWIG (other examples: HTML websites)

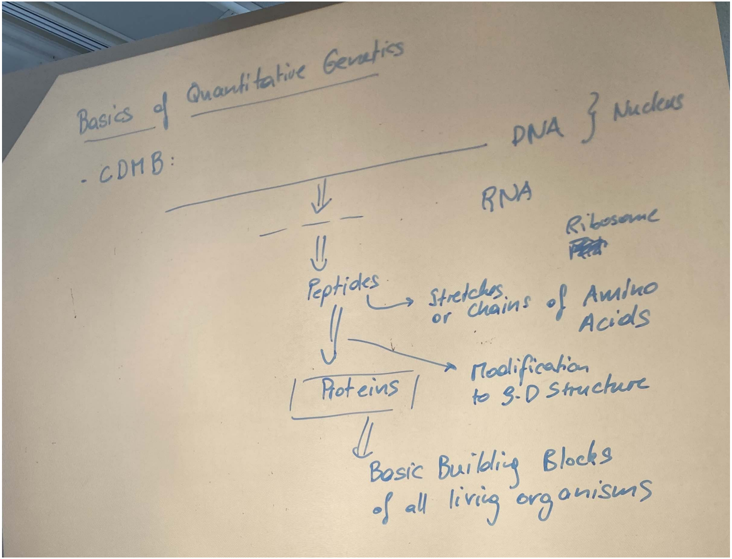
☐ Source file with special characters that are defining the structure of the document

☐ Information of the document is separated from its appearence.

Rmd Format:

- URLs
- **text** -> bold
- *text* => italics
- `` ```{r name} `` => R-code chunk start
- `` ``` `` => R-code chunk end

OHP Page 3



Basics of Quantitative Genetics

- CDMB:

DNA } Nucleus

⇓

RNA

⇓

Ribosome

[illegible]

Peptides → Stretches or chains of Amino Acids

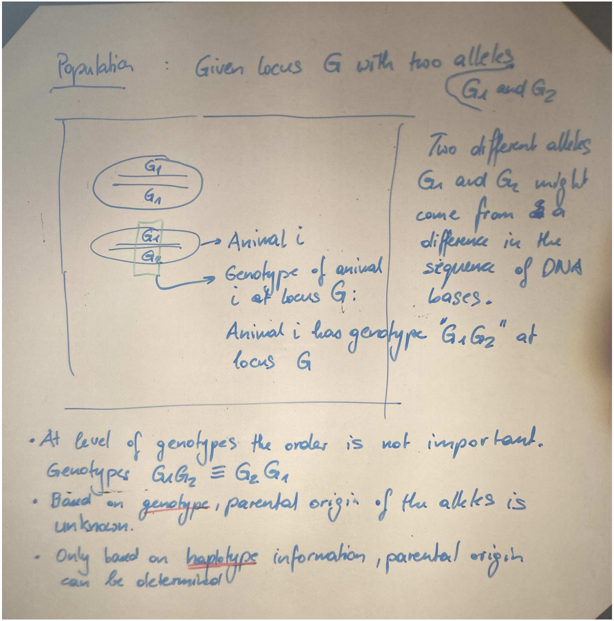
⇓ → Modification to 3-D Structure

Proteins

⇓

Basic Building Blocks of all living organisms

OHP Page 4



Population : Given locus G with two alleles (G_1 and G_2)

Two different alleles G_1 and G_2 might come from a difference in the sequence of DNA bases.

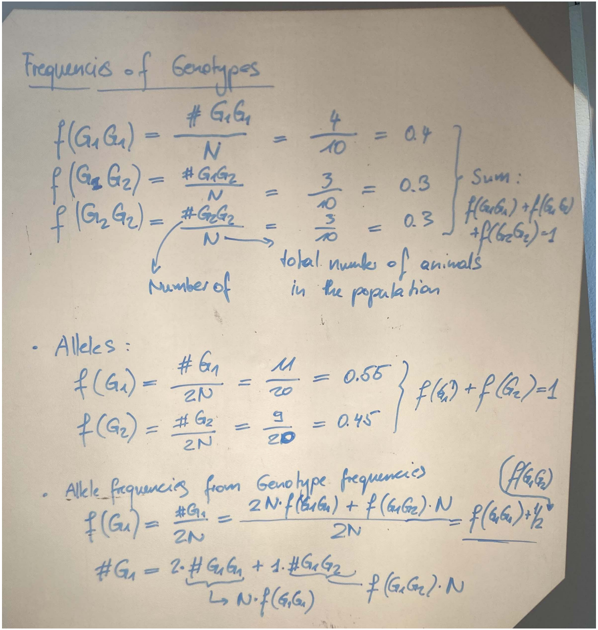
Animal i

Genotype of animal i at locus G:

Animal i has genotype "G_1G_2" at locus G

- At level of genotypes the order is not important. Genotypes $G_1G_2 \equiv G_2G_1$
- Based on genotype, parental origin of the alleles is unknown.
- Only based on haplotype information, parental origin can be determined

OHP Page 5



Frequencis of Genotypes

$$f(G_1G_1) = \frac{\# G_1G_1}{N} = \frac{4}{10} = 0.4$$

$$f(G_1G_2) = \frac{\# G_1G_2}{N} = \frac{3}{10} = 0.3$$

$$f(G_2G_2) = \frac{\# G_2G_2}{N} = \frac{3}{10} = 0.3$$

Sum:
$f(G_1G_1) + f(G_1G_2) + f(G_2G_2) = 1$

Number of

N → total number of animals in the population

- Alleles:

$$f(G_1) = \frac{\# G_1}{2N} = \frac{11}{20} = 0.55$$

$$f(G_2) = \frac{\# G_2}{2N} = \frac{9}{20} = 0.45$$

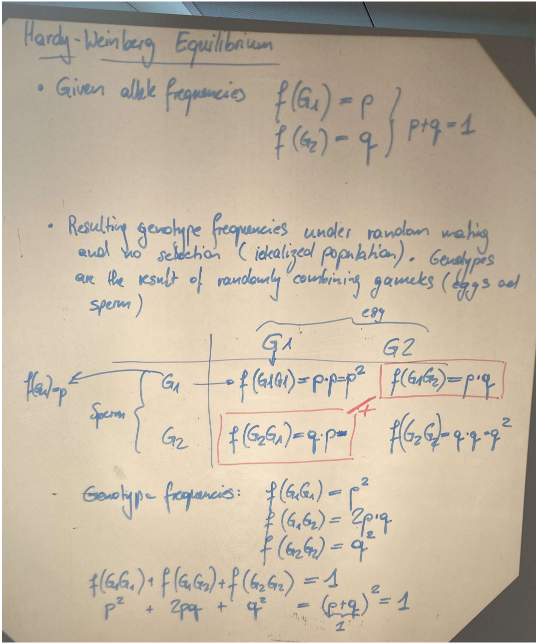
$f(G_1) + f(G_2) = 1$

- Allele frequencies from Genotype frequencies

$$f(G_1) = \frac{\# G_1}{2N} = \frac{2N \cdot f(G_1G_1) + f(G_1G_2) \cdot N}{2N} = \underline{f(G_1G_1) + \tfrac{1}{2} f(G_1G_2)}$$

$$\# G_1 = 2 \cdot \underbrace{\# G_1G_1}_{\hookrightarrow N \cdot f(G_1G_1)} + 1 \cdot \underbrace{\# G_1G_2}_{f(G_1G_2) \cdot N}$$

OHP Page 6



Hardy-Weinberg Equilibrium

- Given allele frequencies $\left.\begin{array}{l} f(G_1) = p \\ f(G_2) = q \end{array}\right\} p+q=1$

- Resulting genotype frequencies under random mating and no selection (idealized population). Genotypes are the result of randomly combining gametes (eggs and sperm)

$f(G_1)=p$

		egg	
		$G1$	$G2$
Sperm	G_1	$f(G_1G_1)=p\cdot p=p^2$	$\boxed{f(G_1G_2)=p\cdot q}$
	G_2	$\boxed{f(G_2G_1)=q\cdot p}$	$f(G_2G_2)=q\cdot q=q^2$

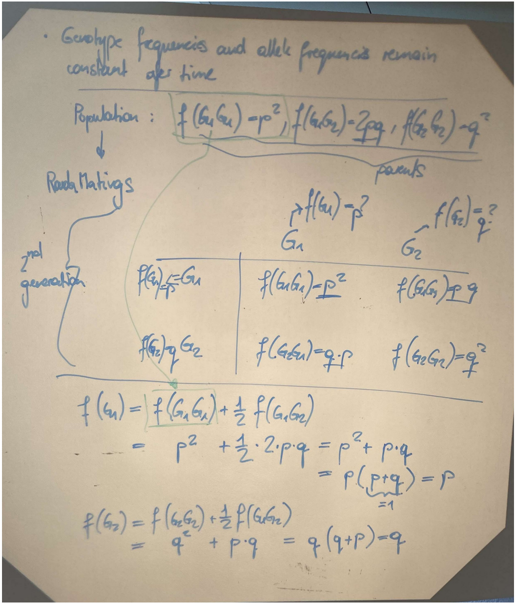
+

- Genotype frequencies: $f(G_1G_1)=p^2$
  $f(G_1G_2)=2p\cdot q$
  $f(G_2G_2)=q^2$

$$f(G_1G_1)+f(G_1G_2)+f(G_2G_2)=1$$

$$p^2 + 2pq + q^2 = \underbrace{(p+q)}_{1}^2 = 1$$

OHP Page 7



- Genotype frequencies and allele frequencies remain constant after time

Population : $f(G_1G_1)=p^2,\ f(G_1G_2)=2pq,\ f(G_2G_2)=q^2$

parents

↓

Random Matings

$G_1 \rightarrow f(G_1)=p$? $\qquad G_2 \rightarrow f(G_2)=q$?

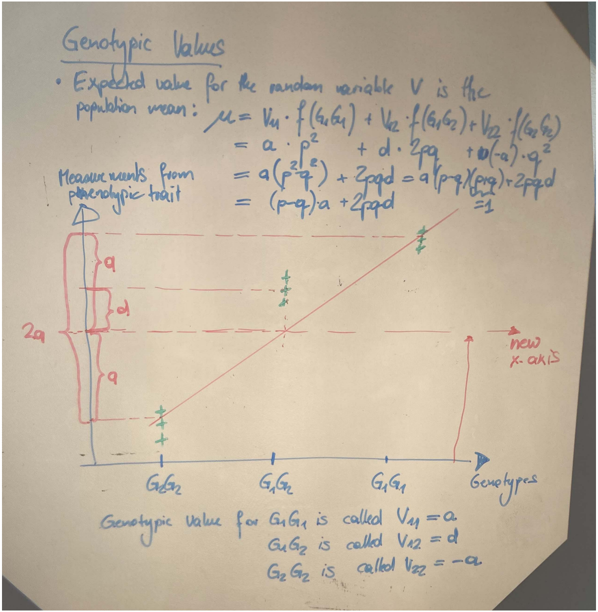
2nd generation

	G_1	G_2
$f(G_1)=p$ G_1	$f(G_1G_1)=p^2$	$f(G_1G_2)=p\,q$
$f(G_2)=q$ G_2	$f(G_2G_1)=q\cdot p$	$f(G_2G_2)=q^2$

$$f(G_1)=\boxed{f(G_1G_1)}+\frac{1}{2}f(G_1G_2)$$
$$=p^2+\frac{1}{2}\cdot 2\cdot p\cdot q=p^2+p\cdot q$$
$$=p\,(\underbrace{p+q}_{=1})=p$$

$$f(G_2)=f(G_2G_2)+\frac{1}{2}f(G_1G_2)$$
$$=q^2+p\cdot q=q\,(q+p)=q$$

OHP Page 8



Genotypic Values

- Expected value for the random variable V is the population mean: $\mu = V_{11} \cdot f(G_1G_1) + V_{12} \cdot f(G_1G_2) + V_{22} \cdot f(G_2G_2)$

$$= a \cdot p^2 + d \cdot 2pq + (-a) \cdot q^2$$

$$= a(p^2 - q^2) + 2pqd = a(p-q)\underbrace{(p+q)}_{=1} + 2pqd$$

$$= (p-q) \cdot a + 2pqd$$

Genotypic value for G_1G_1 is called $V_{11} = a$

G_1G_2 is called $V_{12} = d$

G_2G_2 is called $V_{22} = -a$