Inférence en génétique des populations

François Rousset & Raphaël Leblois

M2 Biostatistiques 2015–2016

Outline of course

Buts: présenter des thématiques de recherche méthodologiques actuelles, et faciliter la compréhension de la littérature

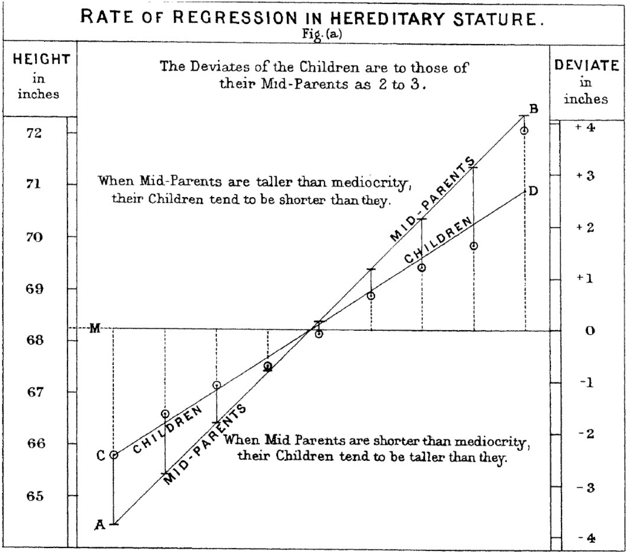
- Rappels de génétique (FR)
- Likelihood inference under simple models; the coalescent (FR)
  Molecular markers (RL)
- TD Coalescence (RL)
- Moment methods (FR)
- Algorithms for likelihood inference under neutral models (RL)
- Simulation-based inference: ABC (Jean-Michel Marin)
- Analyse d'articles

Why is (statistical) regression called regression?

Why is (statistical) regression called regression?

Regression *towards* Mediocrity *in* Hereditary Stature.

By Francis Galton, F.R.S., &c.



RATE OF REGRESSION IN HEREDITARY STATURE.

Fig. (a)

Today's outline

Population genetics = analysis of the processes controlling genetic polymorphisms in populations

- Developed to understand evolution
- From Mendel's rules to population processes
- Population genetics

A familiar example: our mosquitoes

In the '60s: development of tourism.

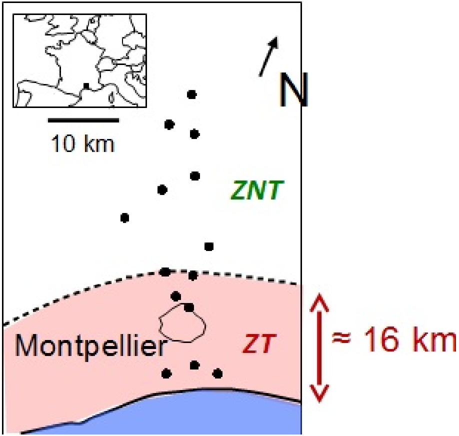
Insecticide treatments 1969- First resistance in 1972

A familiar example: our mosquitoes

In the '60s: development of tourism.

Insecticide treatments 1969-

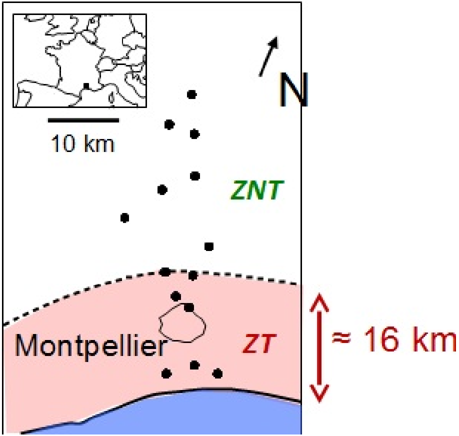
First resistance in 1972



A familiar example: our mosquitoes

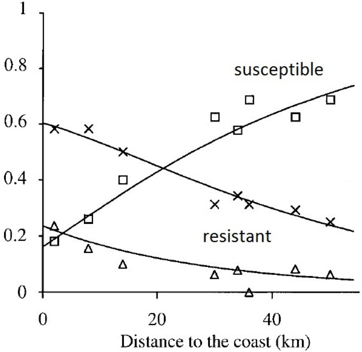
In the '60s: development of tourism.

Insecticide treatments 1969-



First resistance in 1972

October 1996

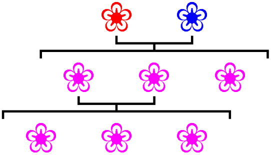


How does natural selection work?

- artificial breeding: we know that selection works even if we do not know the mechanisms of heredity
    - Variation
    - Differential reproductive success (*fitness*)
    - Heredity
- Was not compatible with some early ideas about heredity

Heredity matters

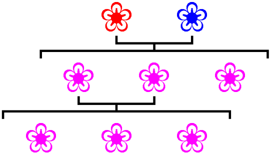
The misconception of blending inheritance



- Assuming $X_{\mathrm{descendant}} = \bar{X}_{\mathrm{parents}}$
- How does the variance of trait evolve?

Heredity matters

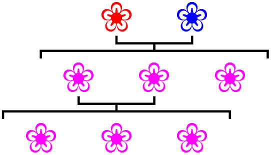
The misconception of blending inheritance



- Assuming $X_{\text{descendant}} = \bar{X}_{\text{parents}}$
- Variance of trait quickly vanishes
  $\mathrm{Var}(X)_{\text{among descendants}} =$
  $\mathrm{Var}[(X_{\text{mother}} + X_{\text{father}})/2]_{\text{among descendants}}$
  $\Rightarrow$ No variation to select from!

Heredity matters

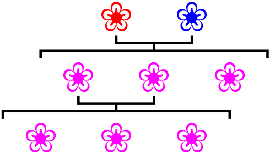
The misconception of blending inheritance



- Assuming $X_{\text{descendant}} = \bar{X}_{\text{parents}}$
- Variance of trait quickly vanishes
  $\text{Var}(X)_{\text{among descendants}} = \text{Var}[(X_{\text{mother}} + X_{\text{father}})/2]_{\text{among descendants}}$
  $\Rightarrow$ No variation to select from!
- But of course, $X_{\text{descendant}} \neq \bar{X}_{\text{parents}}$

Heredity matters

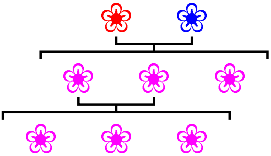
The misconception of blending inheritance



- Assuming $X_{\mathrm{descendant}} = \bar{X}_{\mathrm{parents}}$
- Variance of trait quickly vanishes $\mathrm{Var}(X)_{\mathrm{among\ descendants}} = \mathrm{Var}[(X_{\mathrm{mother}} + X_{\mathrm{father}})/2]_{\mathrm{among\ descendants}}$ $\Rightarrow$ No variation to select from!
- But of course, $X_{\mathrm{descendant}} \neq \bar{X}_{\mathrm{parents}}$
- Elaborations, e.g. regression on ancestral values (Galton) $X_{t+1} = \frac{2\bar{X}_t}{3} + \frac{4\bar{X}_{t-1}}{9} + \frac{8\bar{X}_{t-2}}{27} + \cdots$

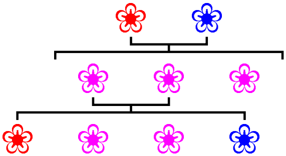
Heredity matters

The misconception of blending inheritance

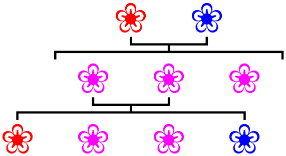


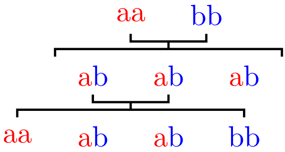
- Assuming $X_{\text{descendant}} = \bar{X}_{\text{parents}}$
- Variance of trait quickly vanishes
  $\text{Var}(X)_{\text{among descendants}} =$
  $\text{Var}[(X_{\text{mother}} + X_{\text{father}})/2]_{\text{among descendants}}$
  $\Rightarrow$ No variation to select from!
- But of course, $X_{\text{descendant}} \neq \bar{X}_{\text{parents}}$
- Elaborations, e.g. regression on ancestral values (Galton)
  $X_{t+1} = \frac{\bar{X}_t}{2} + \frac{\bar{X}_{t-1}}{4} + \frac{\bar{X}_{t-2}}{8} + \cdots$

Mendelian segregation



Mendelian segregation



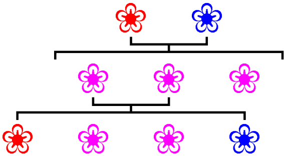


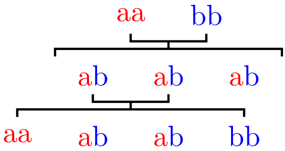
aa bb

ab ab ab

aa ab ab bb

Mendelian segregation





aa bb

ab ab ab

aa ab ab bb

Allows continued selection of initial variation over many generations

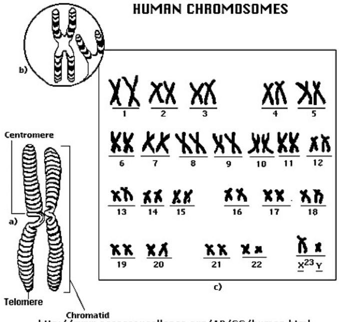
Two developments

Concepts of particulate inheritance and its physical basis

Two developments

Concepts of particulate inheritance and its physical basis

chromosomes

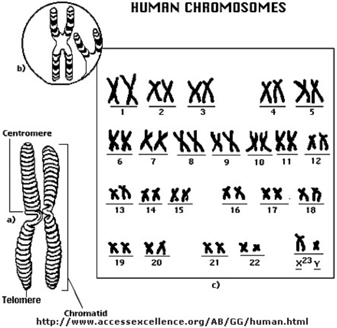


http://www.accessexcellence.org/AB/GG/human.html

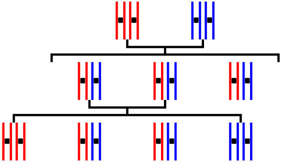
Two developments

Concepts of particulate inheritance and its physical basis

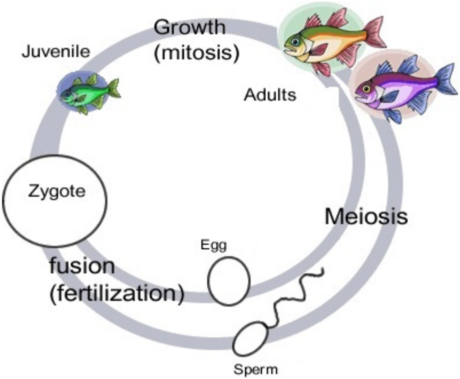
chromosomes



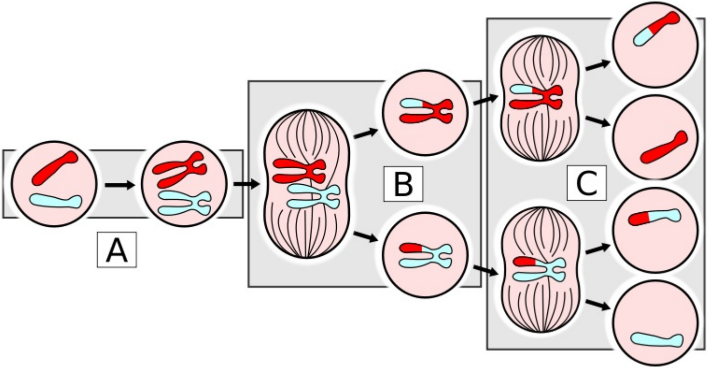
http://www.accessexcellence.org/AB/GG/human.html



Meiosis


Growth
(mitosis)
Juvenile
Adults
Zygote
Meiosis
Egg
fusion
(fertilization)
Sperm

Meiosis

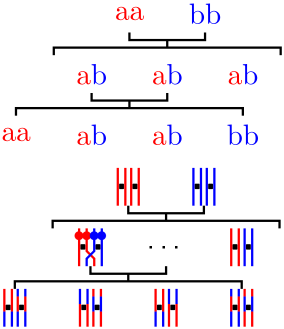


Two developments

Concept of particulate inheritance and its physical basis

chromosomes

Linkage maps



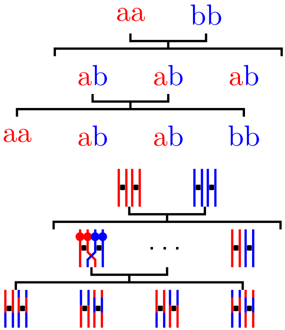
Two developments

Concept of particulate inheritance and its physical basis

chromosomes

Linkage maps

Quantitative theory of evolution



The language of Mendelian and population genetics

At an (autosomal) locus you have two genes (one from each parent) but maybe a single allele.

The language of Mendelian and population genetics

At an (autosomal) locus you have two genes (one from each parent) but maybe a single allele.

- Phenotype[1] := anything (❀)
- Genotype[1] := set of transmitted determinants of the phenotype, each of which is transmitted independently of the environment. (aa/ab/bb)
- Gene[1] := an element of the genotype.
  - May or may not be DNA
  - May or may not code for a protein
- Allele[1] := a form of the gene (a as opposed to b)

[1] After Johannsen, 1911

The language of Mendelian and population genetics

At an (autosomal) locus you have two genes (one from each parent) but maybe a single allele.

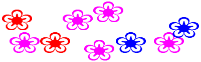
- Phenotype[1] := anything (❀)
- Genotype[1] := set of transmitted determinants of the phenotype, each of which is transmitted independently of the environment. ($\mathrm{aa}/\mathrm{ab}/\mathrm{bb}$)
- Gene[1] := an element of the genotype.
  - May or may not be DNA
  - May or may not code for a protein
- Allele[1] := a form of the gene (a as opposed to b)
- Locus := position of a gene on a genetic (or physical) map

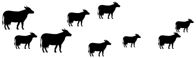
The language of Mendelian and population genetics

At an (autosomal) locus you have two gene copies (one from each parent) but maybe a single allele.

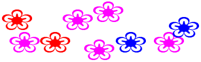
- Phenotype[1] := anything (❀)
- Genotype[1] := set of transmitted determinants of the phenotype, each of which is transmitted independently of the environment. (aa/ab/bb)
- Gene[1] := an element of the genotype.
  - May or may not be DNA
  - May or may not code for a protein
- Allele[1] := a form of the gene (a as opposed to b)
- Locus := position of a gene on a genetic (or physical) map

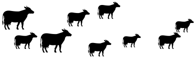
From crosses to populations

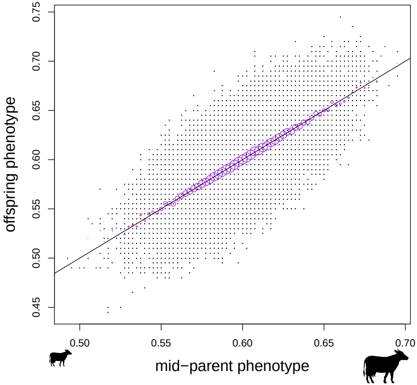




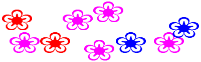
From crosses to populations

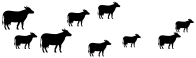


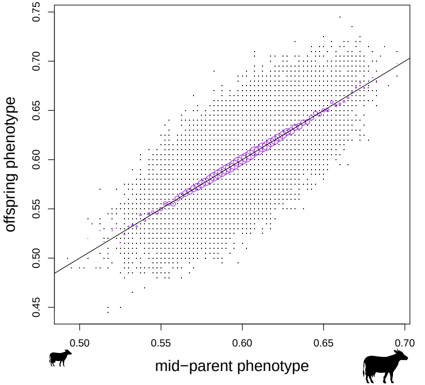




From crosses to populations



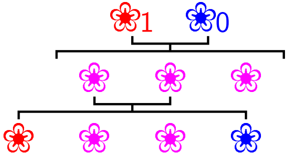




Regression coefficient=heritability; quantifies response to selection

Parent-offspring regressions under Mendelian inheritance

One locus with semi-dominance, i.e.

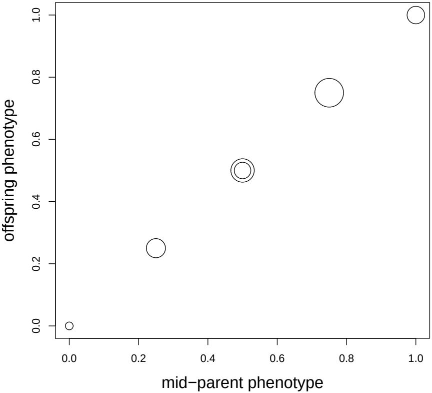


Further assume $p_{\mathrm{b}} = 0.4$

Parent-offspring regressions under Mendelian inheritance

One locus with semi-dominance

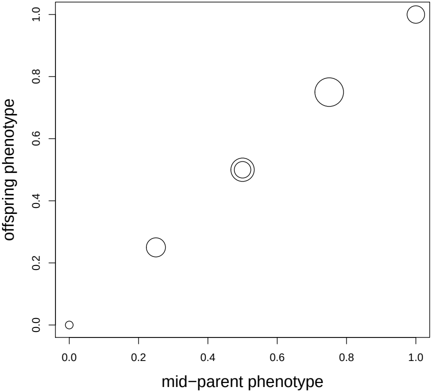
Further assume $p_{\mathrm{b}} = 0.4$



Parent-offspring regressions under Mendelian inheritance

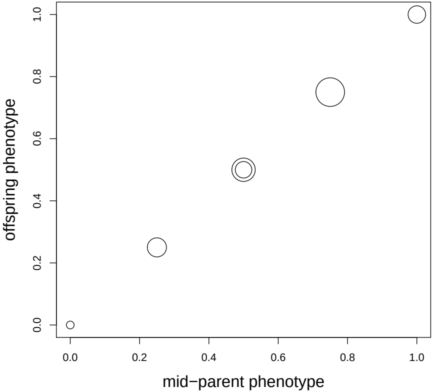
One locus with semi-dominance

Further assume $p_{\mathrm{b}} = 0.4$

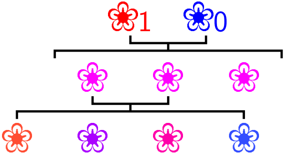


Parent-offspring regressions under Mendelian inheritance

One locus with semi-dominance
Further assume $p_{\mathrm{b}} = 0.4$

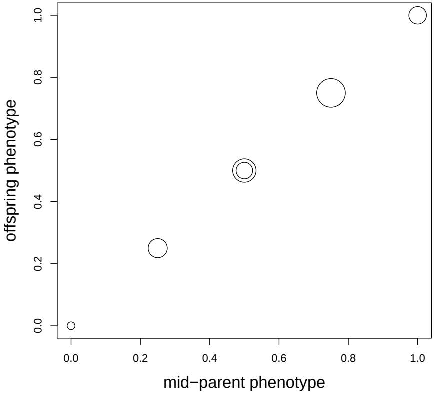


100 loci, additive effect among loci, semi-dominance within loci, all $p_{\mathrm{b}} = 0.4$

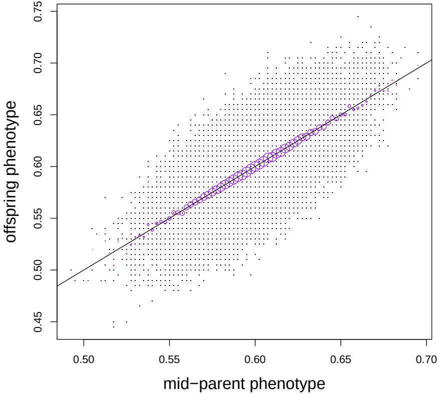


Parent-offspring regressions under Mendelian inheritance

One locus with semi-dominance
Further assume $p_{\mathrm{b}} = 0.4$



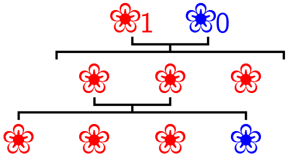
100 loci, additive effect among loci, semi-dominance within loci, all $p_{\mathrm{b}} = 0.4$



Parent-offspring regressions under Mendelian inheritance

One locus with dominance,

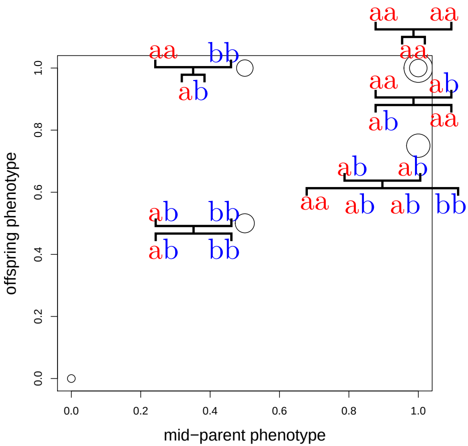
$p_{\mathrm{b}} = 0.4$



Parent-offspring regressions under Mendelian inheritance

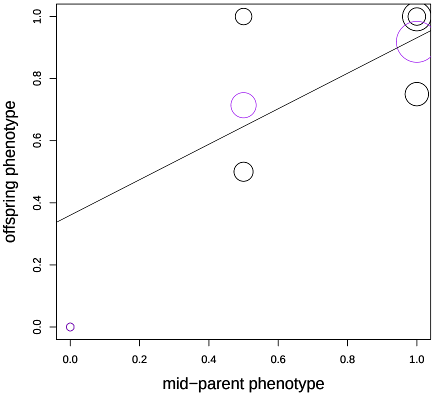
One locus with dominance,

$p_{\mathrm{b}} = 0.4$



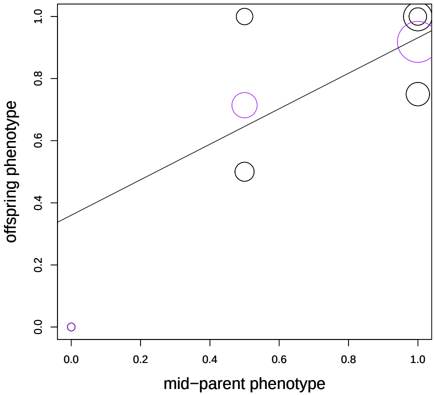
Parent-offspring regressions under Mendelian inheritance

One locus with dominance,
$p_{\mathrm{b}} = 0.4$



Parent-offspring regressions under Mendelian inheritance

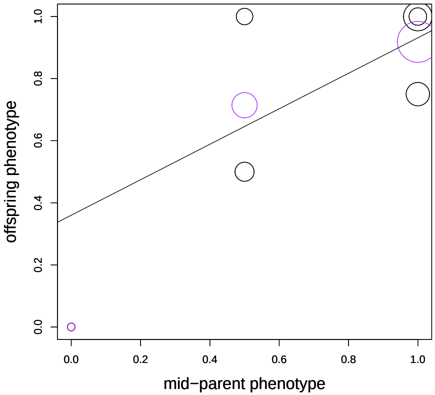
One locus with dominance, $p_{\mathrm{b}} = 0.4$



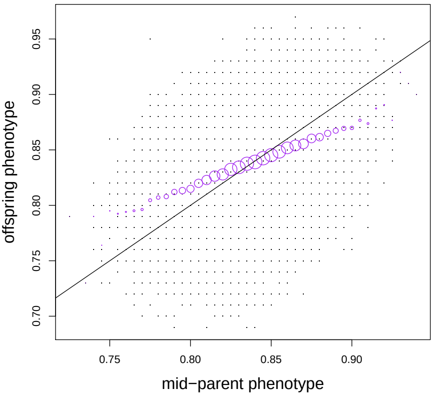
100 loci, additive effect among loci, dominance within loci, all $p_{\mathrm{b}} = 0.4$

Parent-offspring regressions under Mendelian inheritance

One locus with dominance, $p_{\mathrm{b}} = 0.4$



100 loci, additive effect among loci, dominance within loci, all $p_{\mathrm{b}} = 0.4$



Parent-offspring regressions under Mendelian inheritance

Many complications ignored in previous examples: environmental effects, non-additive effects of different loci (epistasis)

Changes in allele frequencies: classification of causes

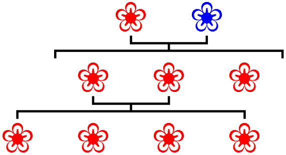
Analysis of changes in genotype frequencies in terms of

- Selection
- Mutation
- Immigration (“gene flow”)
- Drift

Additional effects of the mating system on the diploid genotype frequencies
Additional effects of recombination on multilocus genotype frequencies

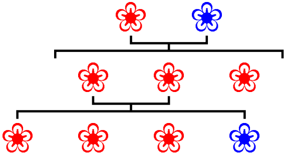
When nothing happens: Hardy-Weinberg (HW) equilibrium

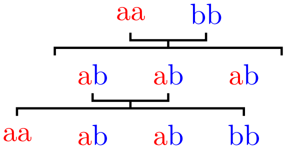
Initially addressed an early misconception about the transmission of dominant characters:



When nothing happens: Hardy-Weinberg (HW) equilibrium

Initially addressed an early misconception about the transmission of dominant characters:





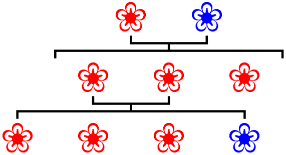
aa bb

ab ab ab

aa ab ab bb

When nothing happens: Hardy-Weinberg (HW) equilibrium

Initially addressed an early misconception about the transmission of dominant characters:



HW equilibrium: allele frequencies do not change over generations (in the absence of selection, mutation and drift)

Random mating (panmixia) ⇒ HW genotype frequencies $p^2 : 2pq : q^2$ (using traditional notation p for the frequency of an allele in a population, and $q := 1 - p$)

Genotype frequencies also constant over generations

Non-random mating

E.g. partial selfing with probability s

$$\mathbb{P}(\mathrm{ab})' = (1-s)2pq + s\mathbb{P}(\mathrm{ab})/2$$

Still equilibrium: allele frequencies do not change over generations (in the absence of selection and drift)
$\Rightarrow$ Asymptotic equilibrium,

$$\mathbb{P}(\mathrm{ab}) = 2pq\frac{1-s}{1-s/2} = 2pq(1-F_{\mathrm{IS}}) \text{ for } F_{\mathrm{IS}} = \frac{s}{2-s}.$$

Genotype frequencies $p^2 + pqF_{\mathrm{IS}} : 2pq(1-F_{\mathrm{IS}}) : q^2 + pqF_{\mathrm{IS}}$

Mutation

Example: insecticide resistance

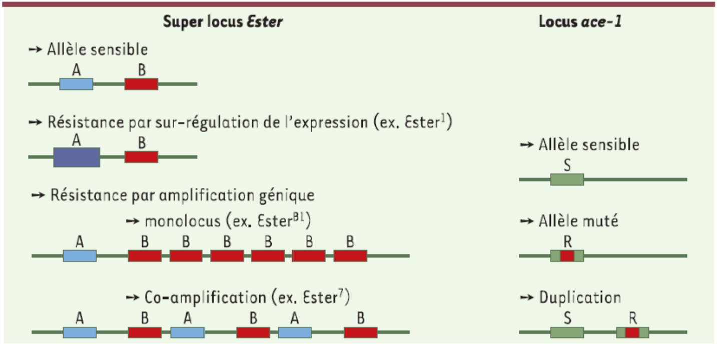
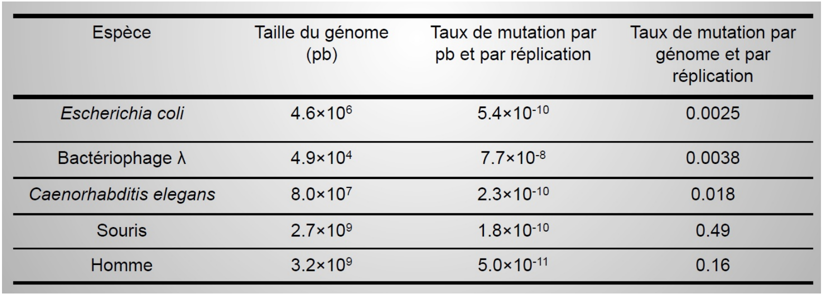

Super locus *Ester*
→ Allèle sensible
A
B
→ Résistance par sur-régulation de l'expression (ex. Ester[1])
A
B
→ Résistance par amplification génique
→ monolocus (ex. Ester[B1])
A B B B B B B
→ Co-amplification (ex. Ester[7])
A B A B A B
Locus *ace-1*
→ Allèle sensible
S
→ Allèle muté
R
→ Duplication
S R

Figure 2. ***Gènes impliqués dans la résistance aux OP chez* Culex pipiens.** *Est-2* et *Est-3*, super locus *Ester*, codent pour les estérases A et B qui piègent les insecticides. Dans les cas de résistance, ces estérases sont produites en excès grâce à un processus d'amplification du nombre de copies des gènes qui les codent dans le génome ou de sur-régulation de leur expression. Le gène *ace-1* code pour la cible des insecticides, l'acétylcholinestérase1 (AChE1). Dans les cas de résistance, cette cible est mutée, ce qui réduit son affinité pour les OP.

Mutation

Anything that changes the allelic state: single nucleotide, deletions, insertions, chromosomal inversions and translocations....

Rates of point mutation per gene copy per generation:



Espèce	Taille du génome (pb)	Taux de mutation par pb et par réplication	Taux de mutation par génome et par réplication
Escherichia coli	4.6×10^{6}	5.4×10^{-10}	0.0025
Bactériophage λ	4.9×10^{4}	7.7×10^{-8}	0.0038
Caenorhabditis elegans	8.0×10^{7}	2.3×10^{-10}	0.018
Souris	2.7×10^{9}	1.8×10^{-10}	0.49
Homme	3.2×10^{9}	5.0×10^{-11}	0.16

After Drake et al. (1998) *Genetics*

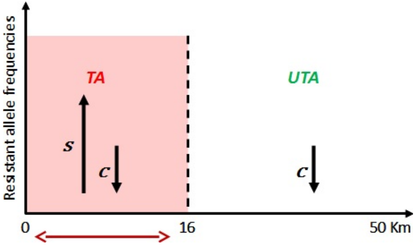
Selection

Selection: causal link between parent i's alleles and their reproductive success.

Selection

Selection: causal link between parent i's alleles and their reproductive success.

Example: insecticide resistance



$$\mathrm{E}(\text{survival}) = 1 - \mathbf{1}_{\text{Treated}}(x)\left[\frac{s_\text{a}}{2}(2-\#_\text{A}) + \frac{s_\text{e}}{2}(2-\#_\text{E})\right] - c_\text{a}\frac{\#_\text{A}}{2} - c_\text{e}\frac{\#_\text{E}}{2}$$

Selection

Selection: causal link between parent i's alleles and their reproductive success.

$$\begin{aligned}\mathsf{E}[p'_{\mathrm{a}}] &= \sum_{\text{parents } i} \mathbb{P}(\text{parent is } i)\mathbf{1}_{\mathrm{a}}(i)\\ &= \sum_{\text{parents } i} \frac{\mathbb{P}(\text{survival of } i)}{\sum_{\text{parents } k} \mathbb{P}(\text{survival of } k)}\mathbf{1}_{\mathrm{a}}(i).\end{aligned}$$

Selection

Selection: causal link between parent i's alleles and their reproductive success.

General:

$$\mathsf{E}[p'_{\mathrm{a}}] = \sum_{\text{parents } i} \mathbb{P}(\text{parent is } i)\mathbf{1}_{\mathrm{a}}(i) = \frac{1}{N}\sum N\mathbb{P}(\text{parent is } i)\mathbf{1}_{\mathrm{a}}(i)$$

$N\mathbb{P}(\text{parent is } i)$ is the expected number of descendants from parent i.

Selection

Selection: causal link between parent i's alleles and their reproductive success.

General:

$$\mathrm{E}[p'_{\mathrm{a}}] = \sum_{\text{parents } i} \mathbb{P}(\text{parent is } i)\mathbf{1}_{\mathrm{a}}(i) = \frac{1}{N}\sum N\mathbb{P}(\text{parent is } i)\mathbf{1}_{\mathrm{a}}(i)$$

$N\mathbb{P}(\text{parent is } i)$ is the expected number of descendants from parent i.
It may be taken as a definition of the *fitness* w_i of individual i, such that

$$\mathrm{E}[p'_{\mathrm{a}}] - p_{\mathrm{a}} = \mathrm{Cov}[w_i, \mathbf{1}_{\mathrm{a}}(i)].$$

Some traditional or memorable formulas

For deterministic models, in terms of allelic fitnesses w_{a} and w_{b}

Some traditional or memorable formulas

For deterministic models, in terms of allelic fitnesses w_{a} and w_{b}

$$\left(\frac{p_{\mathrm{a}}}{p_{\mathrm{b}}}\right)' = \frac{w_{\mathrm{a}}}{w_{\mathrm{b}}}\frac{p_{\mathrm{a}}}{p_{\mathrm{b}}}$$

$$\begin{aligned} p'_{\mathrm{a}} - p_{\mathrm{a}} =& (w_{\mathrm{a}} - w_{\mathrm{b}})p_{\mathrm{a}}(1 - p_{\mathrm{a}}) \\ =& \beta_{w,\mathbf{1}_{\mathrm{a}}} \operatorname{Var}(\mathbf{1}_{\mathrm{a}}) = \operatorname{Cov}[w_i, \mathbf{1}_{\mathrm{a}}(i))] \end{aligned}$$

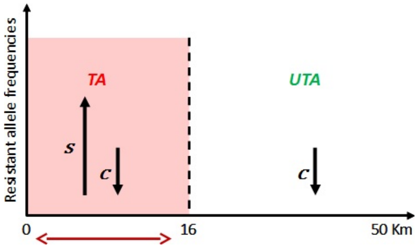
Fitness is often more vaguely defined, up to a constant $\bar{w}$, such that

$$p'_{\mathrm{a}} - p_{\mathrm{a}} = \frac{(w_{\mathrm{a}} - w_{\mathrm{b}})}{\bar{w}} p_{\mathrm{a}}(1 - p_{\mathrm{a}})$$

E.g., “fitness” defined as survival in previous example.

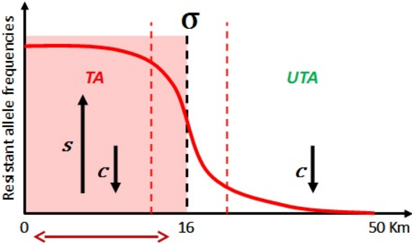
Migration

Example: insecticide resistance



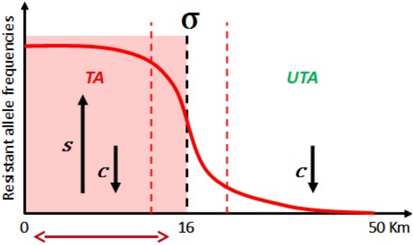
Migration

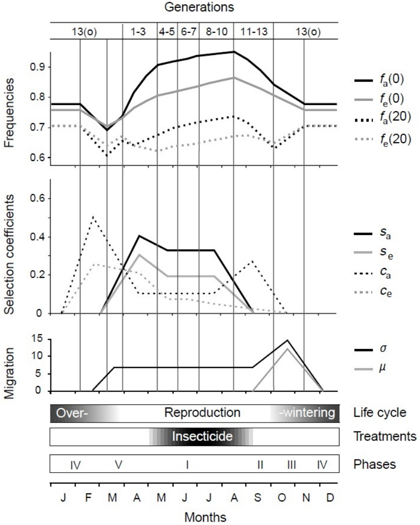
Example: insecticide resistance



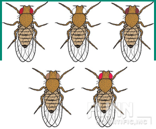
Components of fitness can be estimated

Example: insecticide resistance

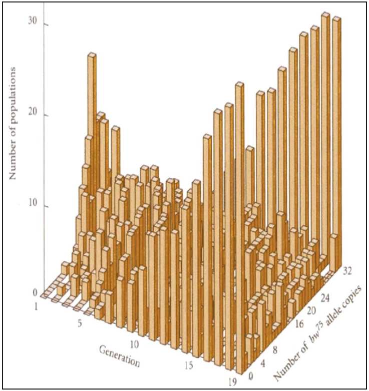




Genetic drift

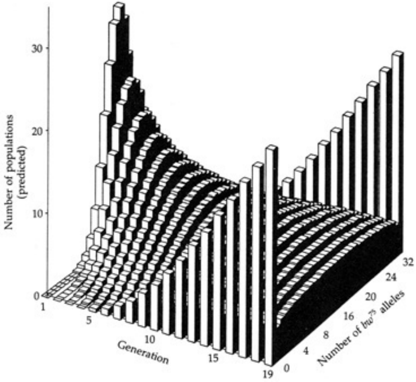


107 lines founded each by 16 heterozygous flies



Buri (1956)

Wright-Fisher model



Wright-Fisher model

Assumptions

N parents each producing a Poisson-distributed number (with mean $\gg N$) of juveniles.

N descendants are drawn from all juveniles.

Elementary questions

Distribution of number of drawn offspring of each parent?

Two alleles a and b: Distribution of number of drawn offspring of type a?

Simplest version: no mutation nor selection

Markov chain on n_{a} with transition probabilities $\mathbb{P}(n'_{\mathrm{a}}|n_{\mathrm{a}})$:

Wright-Fisher model

Assumptions

N parents each producing a Poisson-distributed number (with mean $\gg N$) of juveniles.

N descendants are drawn from all juveniles.

Elementary questions

Distribution of number of drawn offspring of each parent?

Two alleles a and b: Distribution of number of drawn offspring of type a?

Simplest version: no mutation nor selection

Markov chain on n_{a} with transition probabilities $\mathbb{P}(n'_{\mathrm{a}}|n_{\mathrm{a}})$:

$$\binom{N}{n'_{\mathrm{a}}}(n_{\mathrm{a}}/N)^{n'_{\mathrm{a}}}(1-n_{\mathrm{a}}/N)^{N-n'_{\mathrm{a}}} = \binom{N}{n'_{\mathrm{a}}}p_{\mathrm{a}}^{n'_{\mathrm{a}}}(1-p_{\mathrm{a}})^{N-n'_{\mathrm{a}}}$$

Wright-Fisher model

Assumptions

N parents each producing a Poisson-distributed number (with mean $\gg N$) of juveniles.

N descendants are drawn from all juveniles.

Elementary questions

Distribution of number of drawn offspring of each parent?

Two alleles a and b: Distribution of number of drawn offspring of type a?

Simplest version: no mutation nor selection

Markov chain on n_a with transition probabilities $\mathbb{P}(n'_\mathrm{a}|n_\mathrm{a})$:

$$\binom{N}{n'_\mathrm{a}}(n_\mathrm{a}/N)^{n'_\mathrm{a}}(1-n_\mathrm{a}/N)^{N-n'_\mathrm{a}}=\binom{N}{n'_\mathrm{a}}p_\mathrm{a}^{n'_\mathrm{a}}(1-p_\mathrm{a})^{N-n'_\mathrm{a}}$$

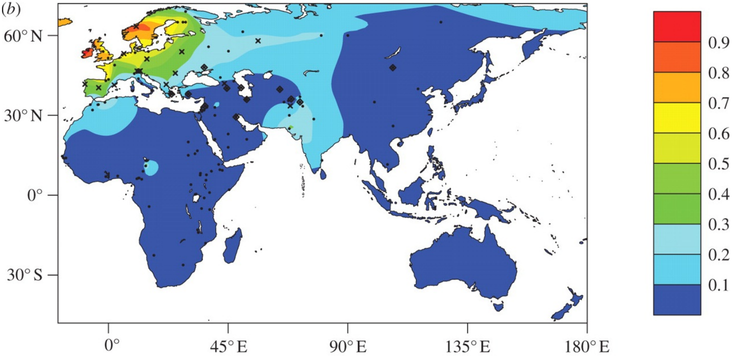
(Symmetric) mutation:

$$\binom{N}{n'_\mathrm{a}}\wp^{n'_\mathrm{a}}(1-\wp)^{N-n'_\mathrm{a}}$$

with $\wp=p_\mathrm{a}+\mu(1-2p_\mathrm{a})$

Complex patterns can result from interactions between the different processes

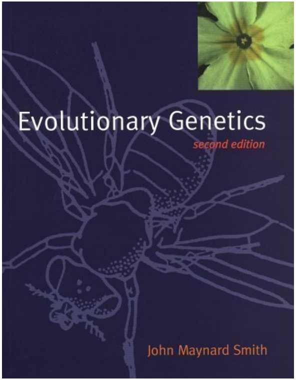
Frequency of a mutant controlling expression of lactase in human populations



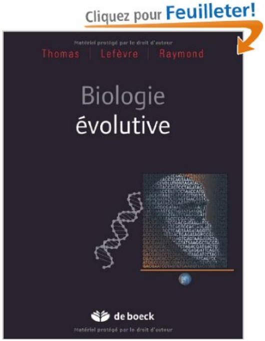
Need for formal model-based inferences

References

Maynard Smith



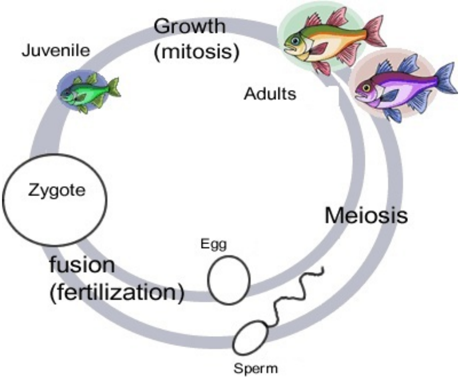
Chapitre 1 Biologie Evolutive


Cliquez pour Feuilleter!
Matériel protégé par le droit d'auteur
Thomas | Lefèvre | Raymond
Biologie
évolutive
de boeck
Matériel protégé par le droit d'auteur

http://kimura.univ-montp2.fr/~rousset/courses.html

Sexual life cycles

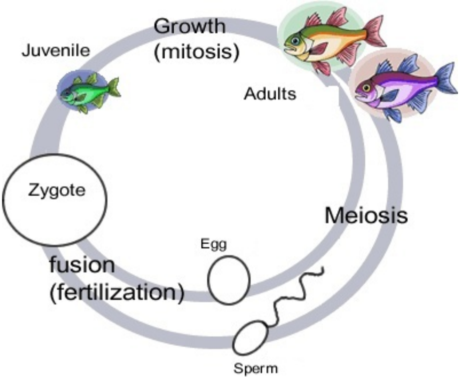
"Diploid" organism



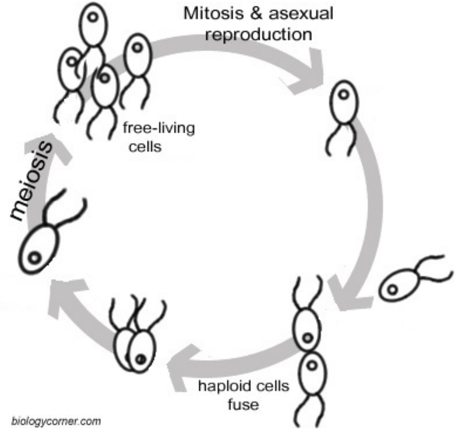
"Haploid" organism

Sexual life cycles

“Diploid” organism

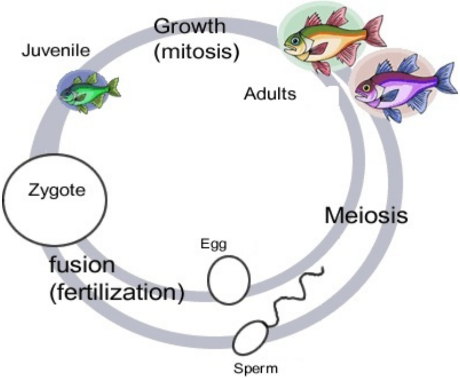


“Haploid” organism

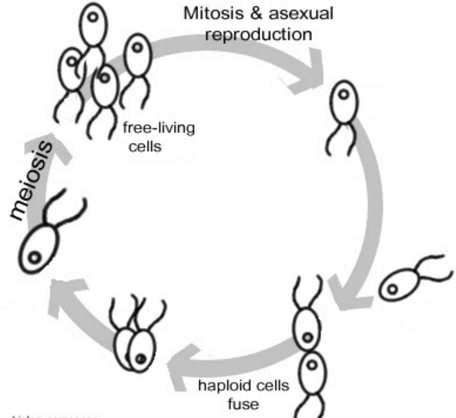


Sexual life cycles

“Diploid” organism



“Haploid” organism



biologycorner.com

A single haplo-diploid cycle with a unique transmission rule