



# Stochastic models help understanding the evolution of partially clonal species and inferring rate of clonality from genetic diversity

Solenn Stoeckel, INRAE, France

INRAE – Institute of Genetics, Environment and Plant Protection  
Team DEMEcology: Dynamics, Evolution, Modelling & Ecology

## Clonix consortium:

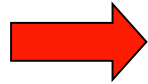
CNRS, Ifremer, INRAE, IRD, **UACH**, **UC Chile**, University of Kansas (USA),  
UAB (USA), UAY (Mexico), RAS (Russia), CSIRO (Australia)



# Asexuality / Clonality

Reproductive mode in which descendents are genetically identical to their only ancestor at all loci in their genomes, except at sites that have experienced somatic mutations and recombinations.

Richards 2003; DeMeeüs *et al.* 2007



**No segregation, No fusion**

→  $n$  alleles at each gene are transmitted together to the descendent.

**No meiosis** (thus no massive recombination)

→ alleles at different genes on a chromatid are transmitted together.

# Why studying clonality?

... most Eukaryote clades reproduce using clonality!



# Species using clonal reproduction: The mainstay of ecosystems and human activities



**Engineers**



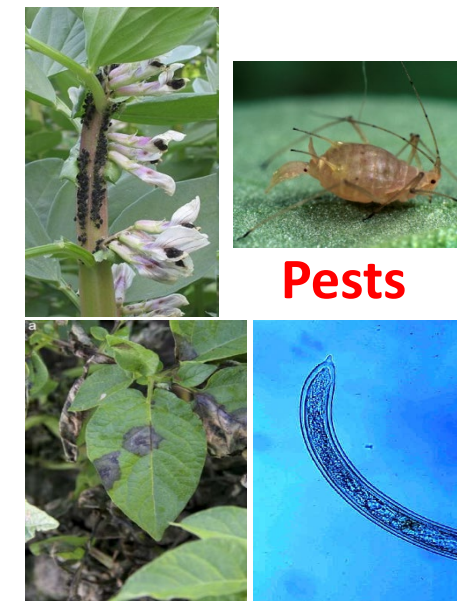
**Pathogens**



**Resources**



**Invasives**



**Pests**

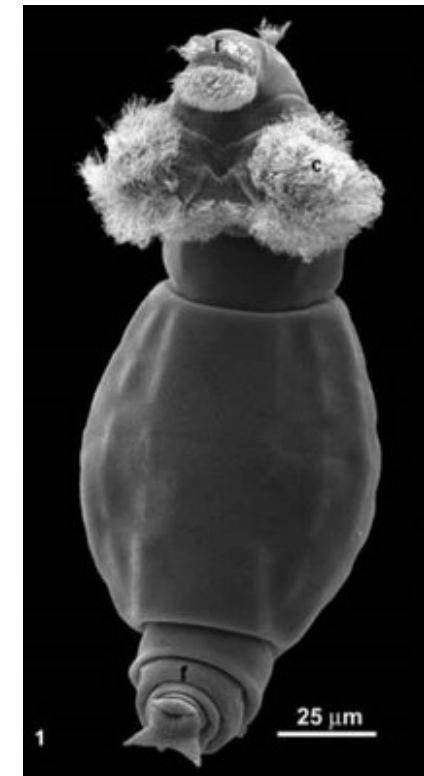
# In Eukaryotes: Clonality = Partial Clonality



$$\text{Rate of clonality: } c = \frac{\text{nb clonal descendents}}{\text{total nb descendents}}$$

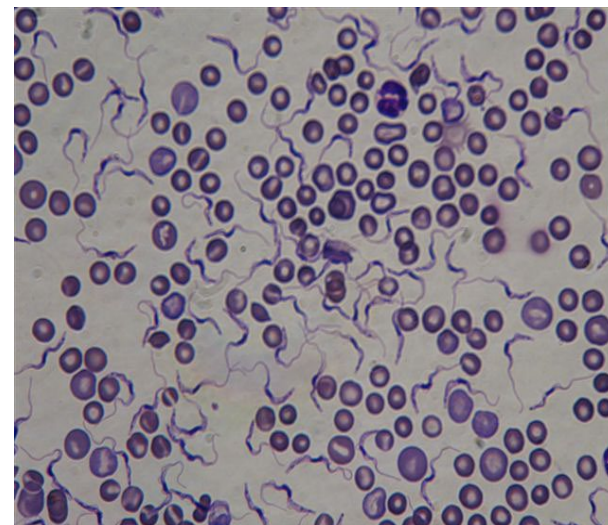
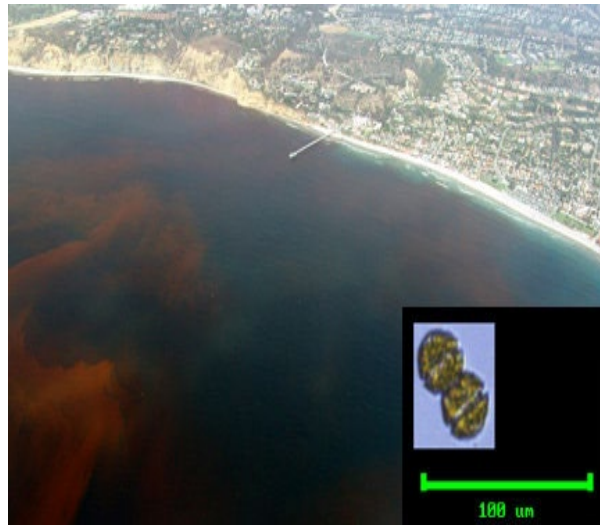
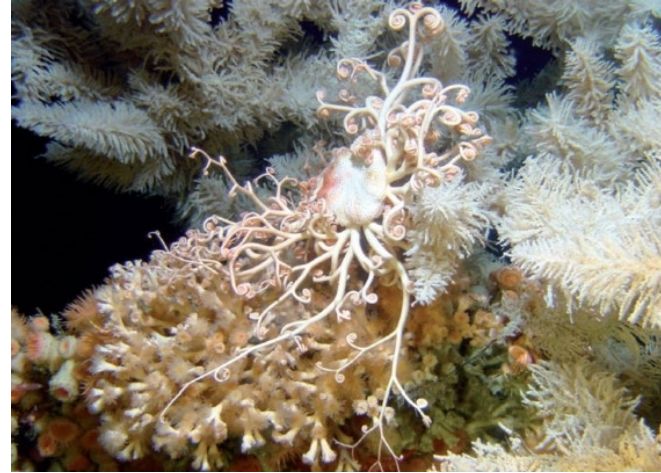
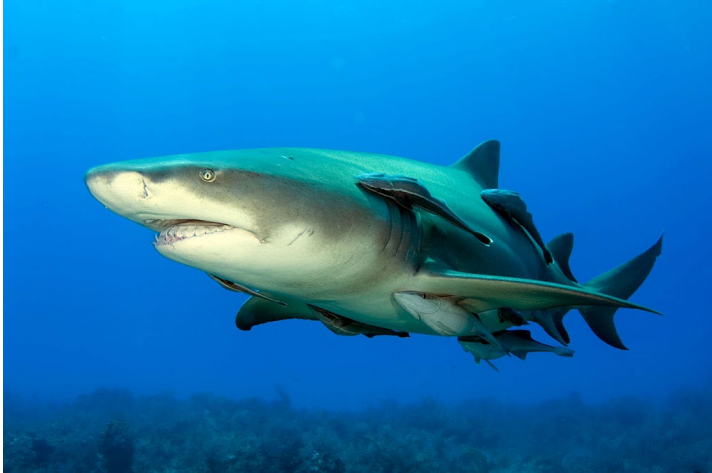
→ How estimating those rates in populations?

1) Inference



# 1) Identify clonal species, assess their level of clonality

→ troubles: cost of observations, how to deal with cryptic organisms, *etc.*



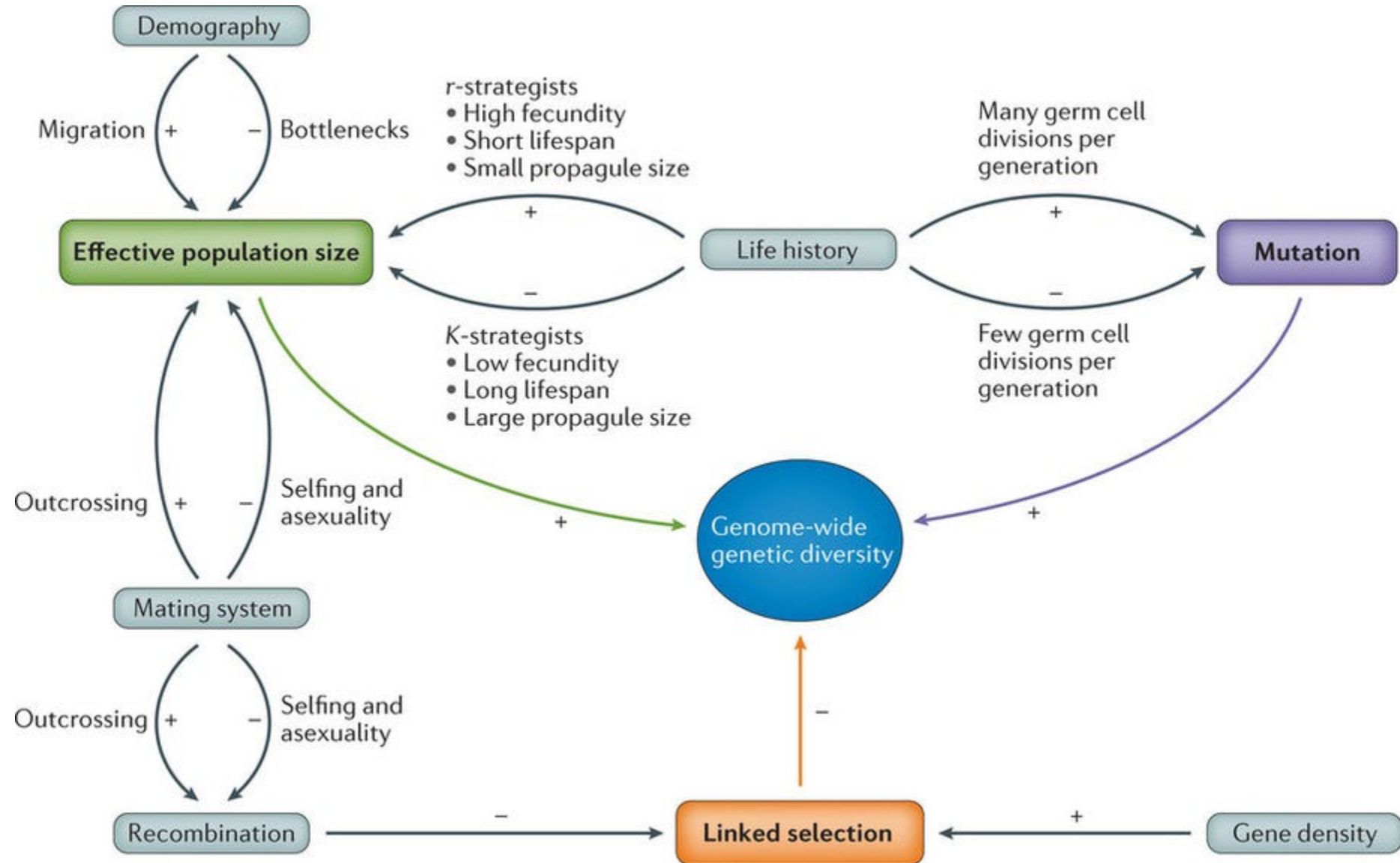
# Understand the effects of partial clonality on ecological and evolutionary dynamics of focal species

- Biotech John C. Avise (2015). Clonal reproduction in vertebrate animals. PNAS, 112 (29), p.8867-8873
- Invasive species under global changes Baker's conjecture (1955)
- Breeding and mass selection



→ How evolve genetic diversity and genomes in such species?  
II) Prediction

# DNA = a barcode to record all those effects:





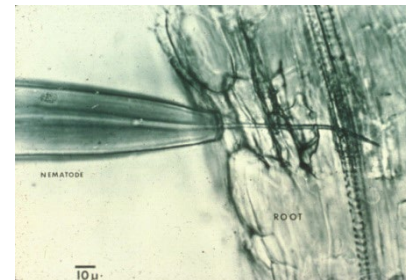
# → What the early models say?

Genetic diversity in partial clonality = sexuality as long as clonality rate is under 90%

$$E[T]_f = 1 + \frac{1}{2\Omega} - \Phi, \text{ where } \Phi = \frac{f(\Omega + f + 1)}{2(2\Omega + 1)}.$$

Marshall & Weir 1979, Orive 1993, Berg & Lascoux 2000, Bengtson 2003, Ceplitis 2003, Balloux et al. 2003, Yonezawa et al. 2004

# → What the data say?



Three main *kuhnian*\* anomalies

- 1- over-representation of heterozygotes in partially clonal populations ( $0 < c < 0.9$ )
- 2- Observed genetic variances  $\gg$  predicted variances in highly clonals ( $c > 0.9$ )
- 3- high volatility of genetic indices over time and space  $\neq$  stabilité prédictions

\*Thomas S. Kuhn. (1962). The Structure of Scientific Revolutions. University of Chicago Press.

# Wright-Fisher + clonal rate + genetic drift + reciprocal mutation

Genetic state of a population

State  $t$

11 AA, 9 Aa, 10 aa

Transition probability



State  $t+1$

7 AA, 12 Aa, 9 aa

$$\begin{cases} \text{rate of clonality} = c \\ \text{rate of sexuality} = 1 - c \end{cases}$$

Ratio Clonality - Sexuality

$$\begin{cases} P(aa|\text{clonality}) = \frac{r_{aa}}{N} = p_{aa} \\ P(aA|\text{clonality}) = \frac{r_{aA}}{N} = p_{aA} \\ P(AA|\text{clonality}) = \frac{r_{AA}}{N} = p_{AA} \end{cases}$$

Clonal reproduction

$$\begin{cases} P(aa|\text{sexuality}) = q_{aa} = \left(\frac{2r_{aa} + r_{aA}}{2N}\right)^2 = \left(p_{aa} + \frac{1}{2}p_{aA}\right)^2 \\ P(aA|\text{sexuality}) = q_{aA} = 2 \cdot \frac{2r_{aa} + r_{aA}}{2N} \cdot \frac{2r_{AA} + r_{aA}}{2N} = 2 \left(p_{aa} + \frac{1}{2}p_{aA}\right) \left(p_{AA} + \frac{1}{2}p_{aA}\right) \\ P(AA|\text{sexuality}) = q_{AA} = \left(\frac{2r_{AA} + r_{aA}}{2N}\right)^2 = \left(p_{AA} + \frac{1}{2}p_{aA}\right)^2 \end{cases}$$

Sexual reproduction

$$\begin{cases} \pi_{aa}^{n+1} = p_{aa} \cdot c + q_{aa} \cdot (1 - c) \\ \pi_{Aa}^{n+1} = p_{aA} \cdot c + q_{aA} \cdot (1 - c) \\ \pi_{AA}^{n+1} = p_{AA} \cdot c + q_{AA} \cdot (1 - c) \end{cases}$$

under infinite pop size

---


$$P(s_{aa}, s_{Aa}, s_{AA} / r_{aa}, r_{aA}, r_{AA}) = \frac{N!}{s_{aa}! s_{Aa}! s_{AA}!} (\pi_{aa}^{n+1})^{s_{aa}} (\pi_{Aa}^{n+1})^{s_{Aa}} (\pi_{AA}^{n+1})^{s_{AA}}$$

Transition probabilities under finite size  $N$

# Wright-Fisher + clonal rate + genetic drift + reciprocal mutation

Let  $u$  being a symmetrical mutation rate

$$\begin{cases} \text{Taux de reproduction asexu e} = c \\ \text{Taux de reproduction sexu e} = 1 - c \end{cases}$$

$$\begin{cases} P(\text{observer } aa | \text{repro. asexu e}) = \frac{r_{aa}}{N} = p_{aa} \\ P(\text{observer } aA | \text{repro. asexu e}) = \frac{r_{aA}}{N} = p_{aA} \\ P(\text{observer } AA | \text{repro. asexu e}) = \frac{r_{AA}}{N} = p_{AA} \end{cases}$$

$$\begin{cases} p_{aa}^{(n+1)} = (1-u)^2 \cdot p_{aa}^{(n)} + u \cdot (1-u) \cdot p_{aA}^{(n)} + u^2 \cdot p_{AA}^{(n)} \\ p_{aA}^{(n+1)} = 2 \cdot u \cdot (1-u) \cdot p_{aa}^{(n)} + [(1-u)^2 + u^2] \cdot p_{aA}^{(n)} + 2 \cdot u \cdot (1-u) \cdot p_{AA}^{(n)} \\ p_{AA}^{(n+1)} = u^2 \cdot p_{aa}^{(n)} + u \cdot (1-u) \cdot p_{aA}^{(n)} + (1-u)^2 \cdot p_{AA}^{(n)} \end{cases}$$

$$\begin{cases} P(\text{observer } aa | \text{repro. sexu e}) = q_{aa} = \left( \frac{2r_{aa} + r_{aA}}{2N} \right)^2 \\ P(\text{observer } aA | \text{repro. sexu e}) = q_{aA} = 2 \cdot \left( \frac{r_{aA}}{2N} \right) \cdot \left( \frac{r_{aa} + r_{AA}}{2N} \right) \\ P(\text{observer } AA | \text{repro. sexu e}) = q_{AA} = \left( \frac{r_{aA} + 2r_{AA}}{2N} \right)^2 \end{cases}$$

$$\begin{cases} q_{aa}^{n+1} = \left[ (1-\mu) \cdot p_{aa}^n + \mu \cdot p_{AA}^n + \frac{1}{2} p_{aA}^n \right]^2 \\ q_{aA}^{n+1} = 2 \cdot \left[ (1-\mu) \cdot p_{aa}^n + \frac{1}{2} \cdot p_{AA}^n + \mu \cdot p_{aA}^n \right] \cdot \left[ (1-\mu) \cdot p_{AA}^n + \frac{1}{2} \cdot p_{aA}^n + \mu \cdot p_{aa}^n \right] \\ q_{AA}^{n+1} = \left[ (1-\mu) \cdot p_{AA}^n + \mu \cdot p_{aa}^n + \frac{1}{2} p_{aA}^n \right]^2 \end{cases}$$

$$\begin{cases} \pi_{aa}^{n+1} = p_{aa} \cdot c + q_{aa} \cdot (1-c) \\ \pi_{aA}^{n+1} = p_{aA} \cdot c + q_{aA} \cdot (1-c) \\ \pi_{AA}^{n+1} = p_{AA} \cdot c + q_{AA} \cdot (1-c) \end{cases}$$

$$P(s_{aa}, s_{aA}, s_{AA} / r_{aa}, r_{aA}, r_{AA}) = \frac{N!}{s_{aa}! s_{aA}! s_{AA}!} (\pi_{aa}^{n+1})^{s_{aa}} (\pi_{aA}^{n+1})^{s_{aA}} (\pi_{AA}^{n+1})^{s_{AA}}$$

Computational/calculus complexity:

$$\Upsilon = \frac{(N + \frac{(n_a + pl - 1)!}{pl! \cdot (n_a - 1)!} - 1)!}{(N! \cdot \frac{(n_a + pl - 1)!}{pl! \cdot (n_a - 1)!})}$$



$N$	$\mathcal{P}$	$\mathcal{L}$	$\mathcal{A}$	$g$	$ S $	memory uptake
20	2	1	2	3	231	420 KB
100	2	1	2	3	5 151	205 MB
500	2	1	2	3	125 751	120 GB
1000	2	1	2	3	501 501	2 TB
20	4	1	2	5	10 626	865 MB
20	2	2	2	9	3 108 105	75 TB
20	2	1	4	10	10 015 005	730 TB
20	2	2	4	100	$9.8 \times 10^{20}$	$6.5 \times 10^{21}$ YB

Table 1: Examples of matrix size based on the Stoeckel-Masson model. Memory sizes are approximate and assume 64-bit accuracy.

## Forward Itô diffusive SDE approximation

Assuming that  $1-c$  and  $u$  converge to 0 at the speed of  $1/N$

$$dY_t = b(Y_t)dt + \sigma(Y_t)dW_t$$

$$u = \frac{\alpha}{N} \quad c = 1 - \frac{\beta}{N}$$

$$Y = (\pi_{aa}, \pi_{Aa}, \pi_{AA}) = (p, q, r)$$

Homoz

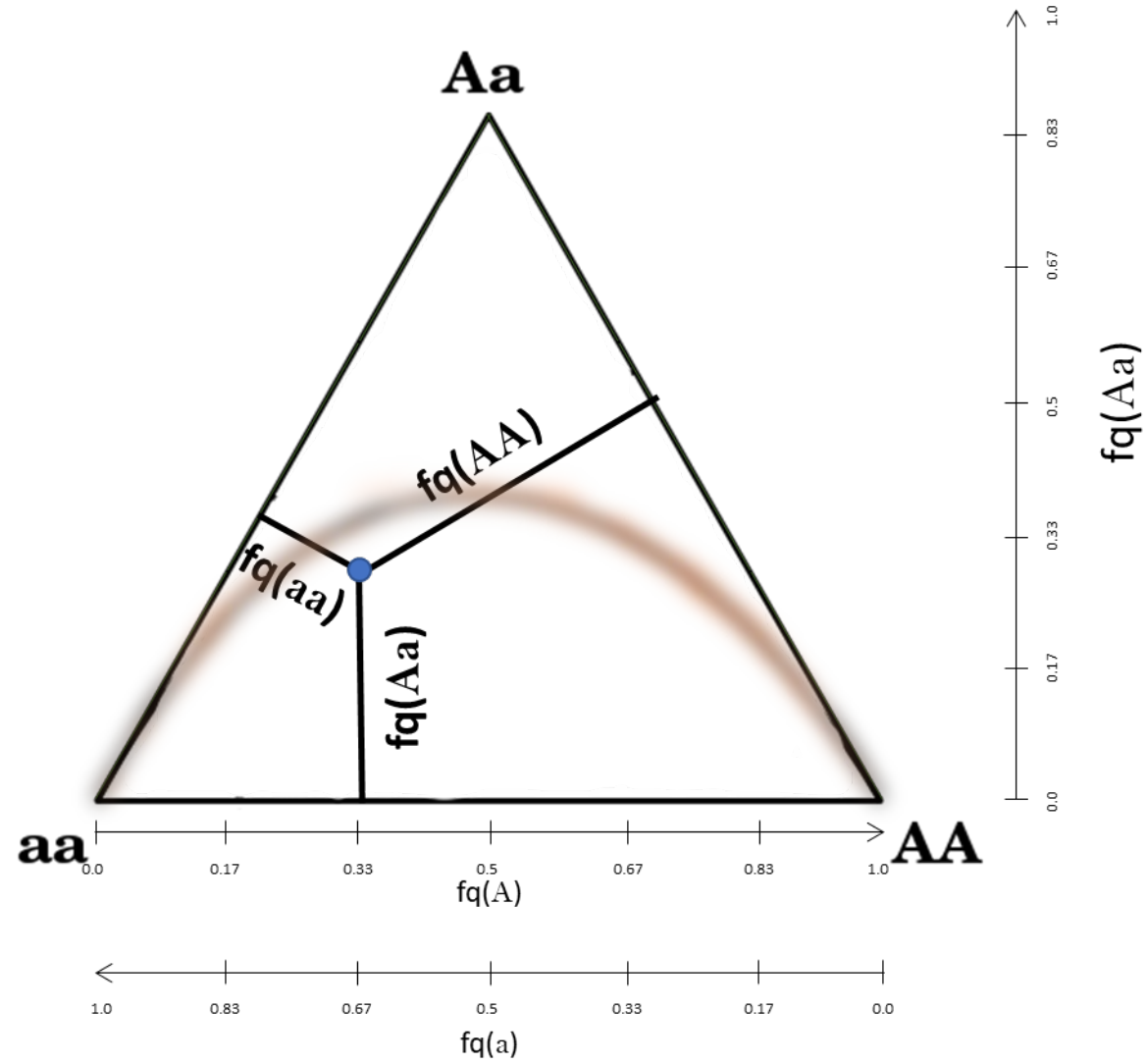
$$dp = \left[ (-2p + q)\alpha + \left( (p + q/2)^2 - p \right)\beta \right] dt + (1-p)\sqrt{p}dB^p - p\sqrt{q}dB^q - p\sqrt{r}dB^r$$

Heteroz

$$dq = \left[ 2(p - q + r)\alpha + \left( 2(p + q/2)(r + q/2) - q \right)\beta \right] dt - q\sqrt{p}dB^p + (1-q)\sqrt{q}dB^q - q\sqrt{r}dB^r$$

# Parenthesis: Visualizing equation results using DeFinetti Diagram

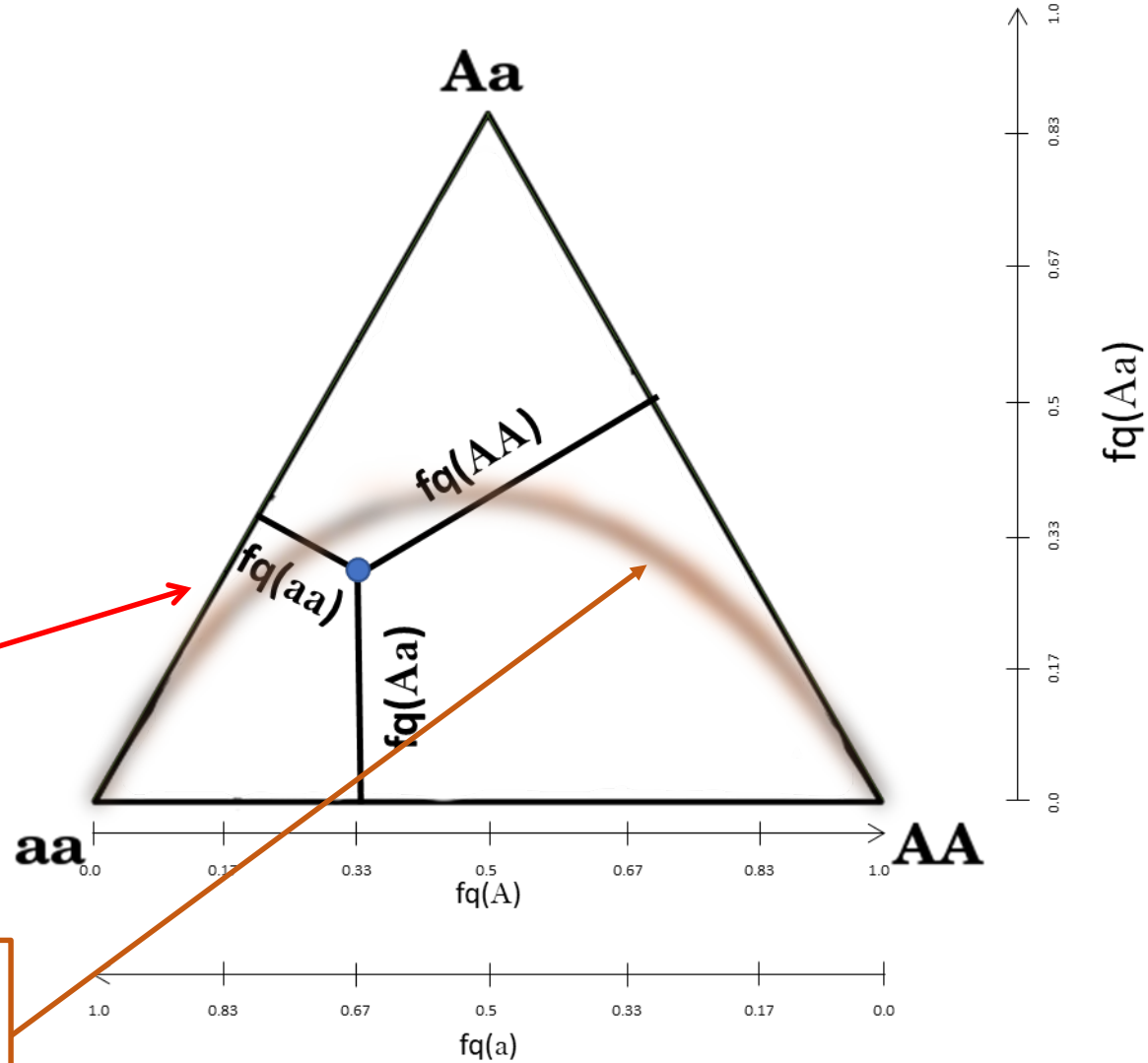
1 locus, 2 alleles (A & a)



# DeFinetti Diagram

1 locus, 2 alleles (A&a)  
N=6

**(aa=1,Aa=2,AA=3)**



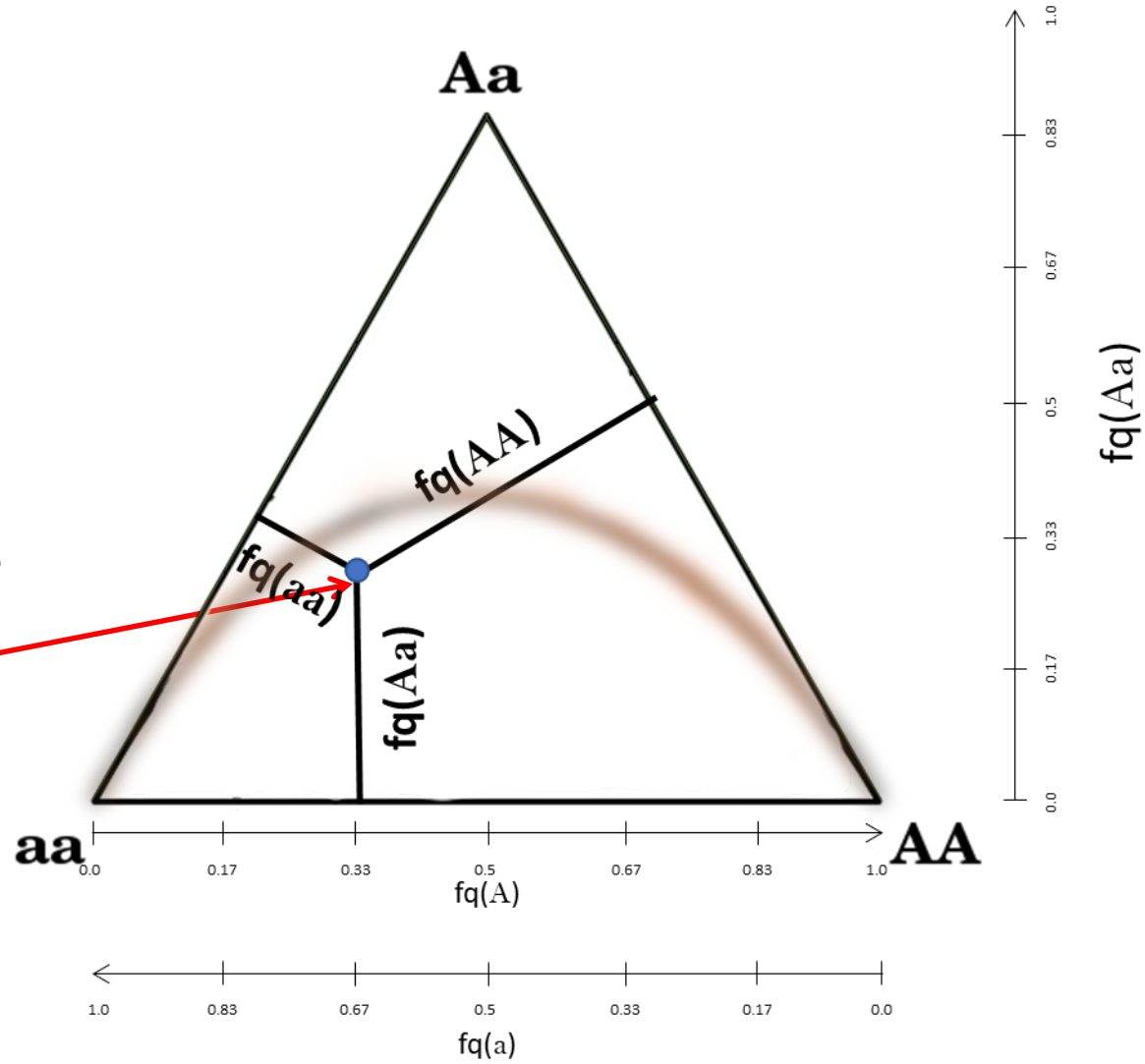
All genetic states that respect genotype frequencies expected under Hardy-Weinberg hypotheses

# DeFinetti Diagram

1 locus, 2 alleles (A&a)  
N=6

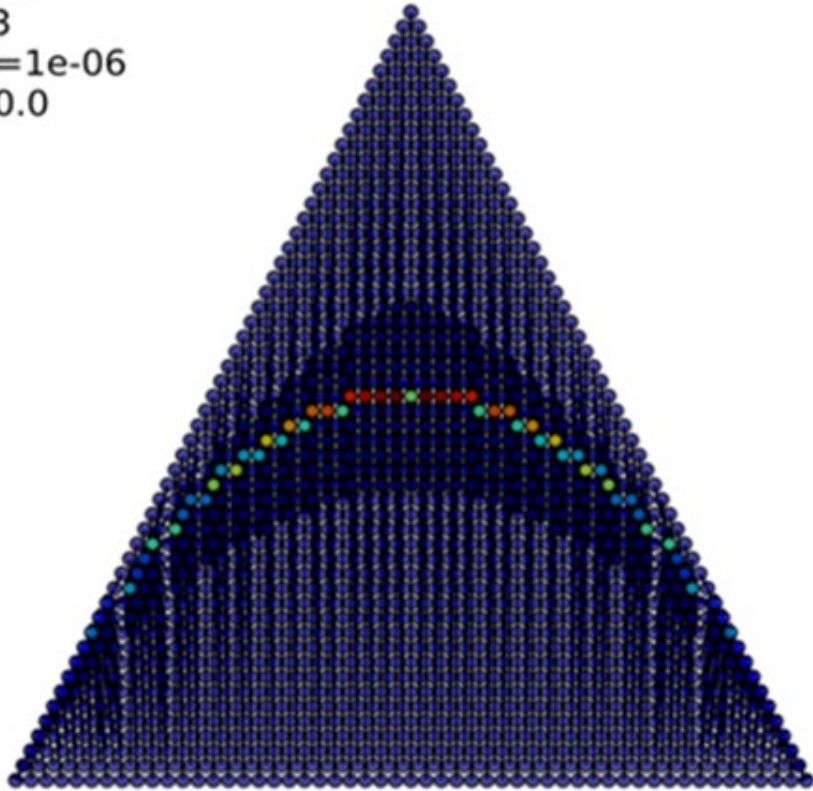
One population genetic state

**(aa=1,Aa=2,AA=3)**

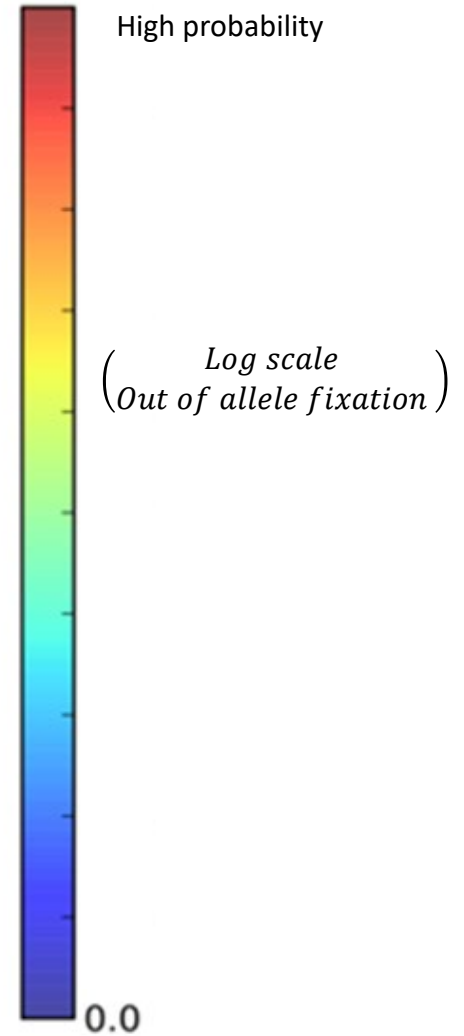
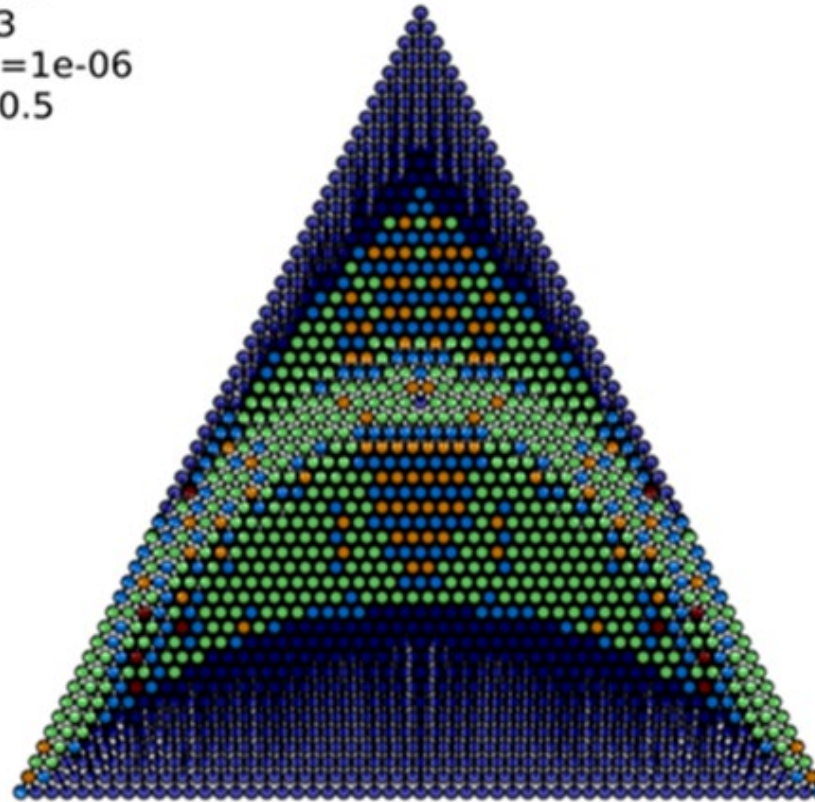


# results

N=52  
n=3  
 $\mu=1e-06$   
C=0.0



N=52  
n=3  
 $\mu=1e-06$   
C=0.5



1/ a few clonality change the probabilities of population genetic states at stationarity (equilibrium)

2/ Transition probabilities are identifiable considering rates of clonality



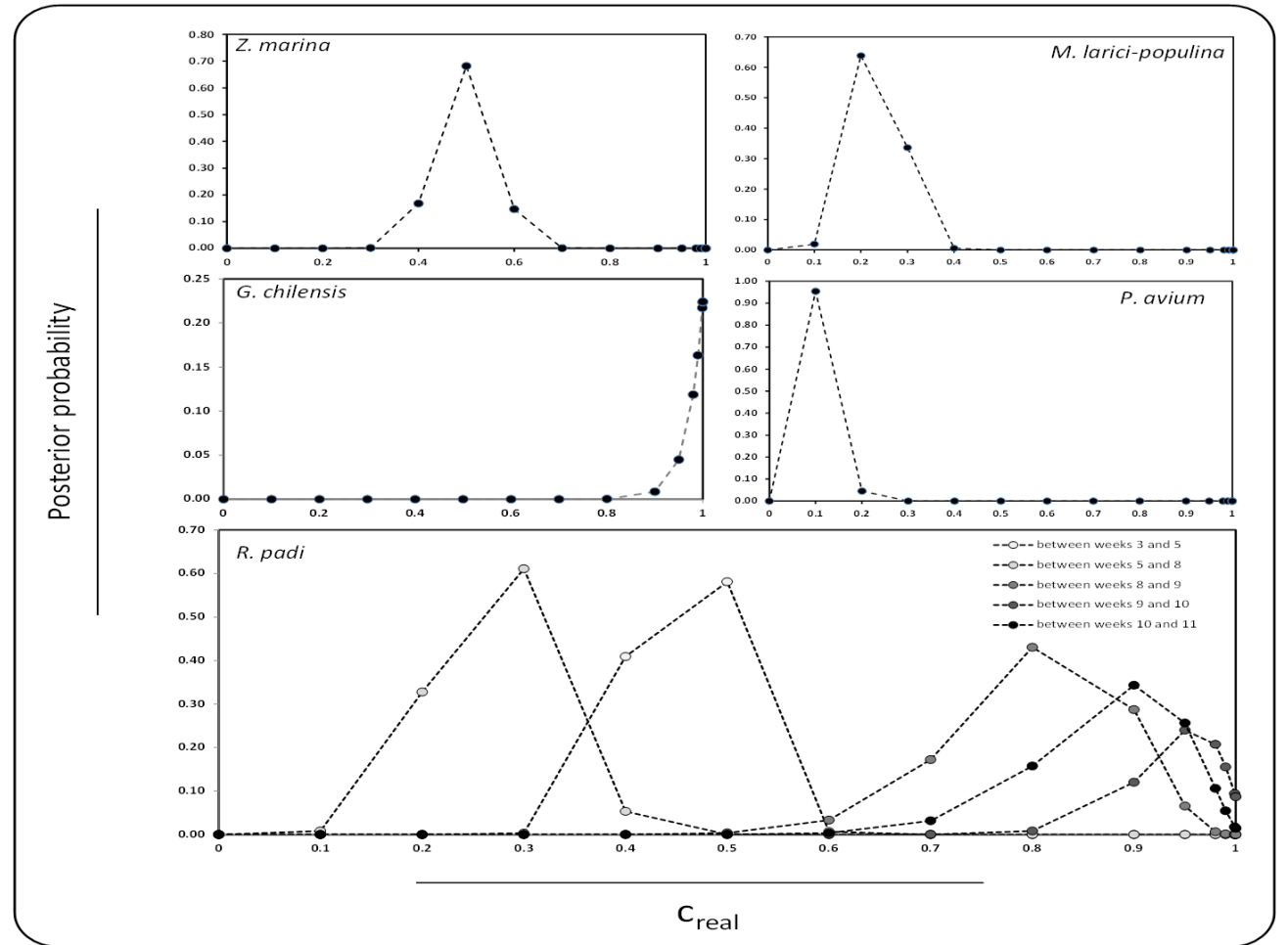
# ClonEstiMate: a Bayesian method for quantifying rates of clonality of populations genotyped at two-time steps

$$L(\tilde{X}_t, \tilde{X}_{t+1}) = \frac{N_\sigma!}{\prod_{i=1}^{i=n} (q_{A_i A_i})! \prod_{i=1, j \neq i}^{i=n} (q_{A_i A_j})!} \cdot \prod_{i=1}^{i=n} (p_{A_i A_i}^*)^{q_{A_i A_i}} \cdot \prod_{i=1, j \neq i}^{i=n} (p_{A_i A_j}^*)^{q_{A_i A_j}}$$

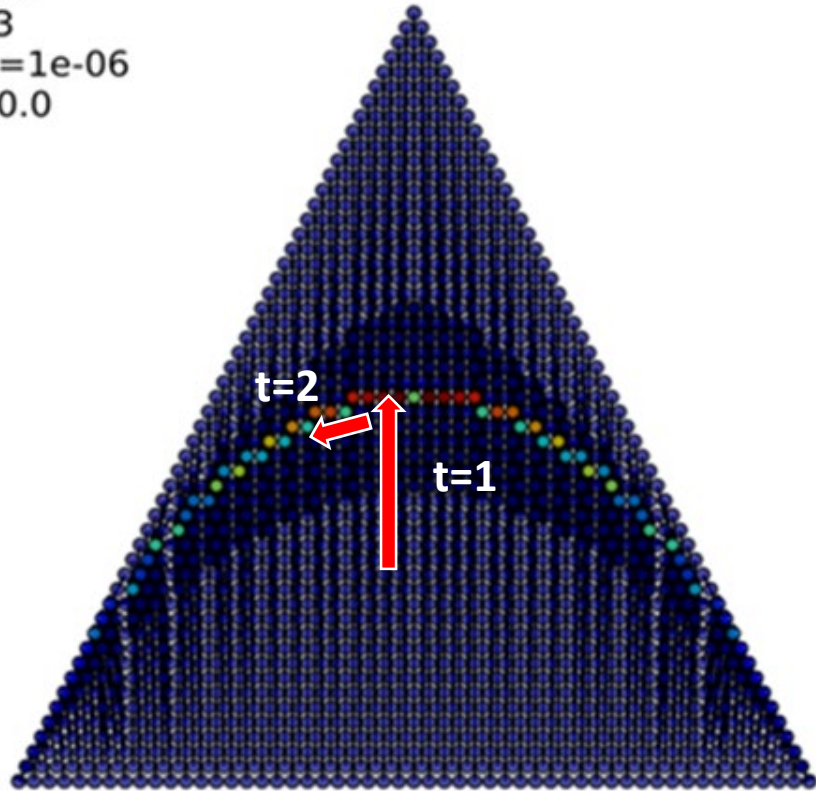
with  $N_\sigma = \sum_i q_{A_i A_i} + \sum_{i,j} q_{A_i A_j}$  and  $\forall i, j \in \{1, \dots, n\}$

## Efficient inference of clonal rate recipe

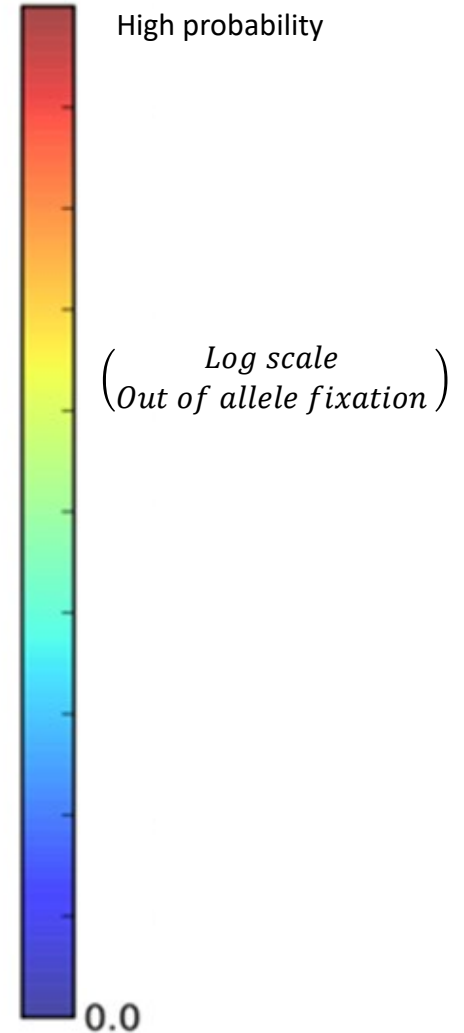
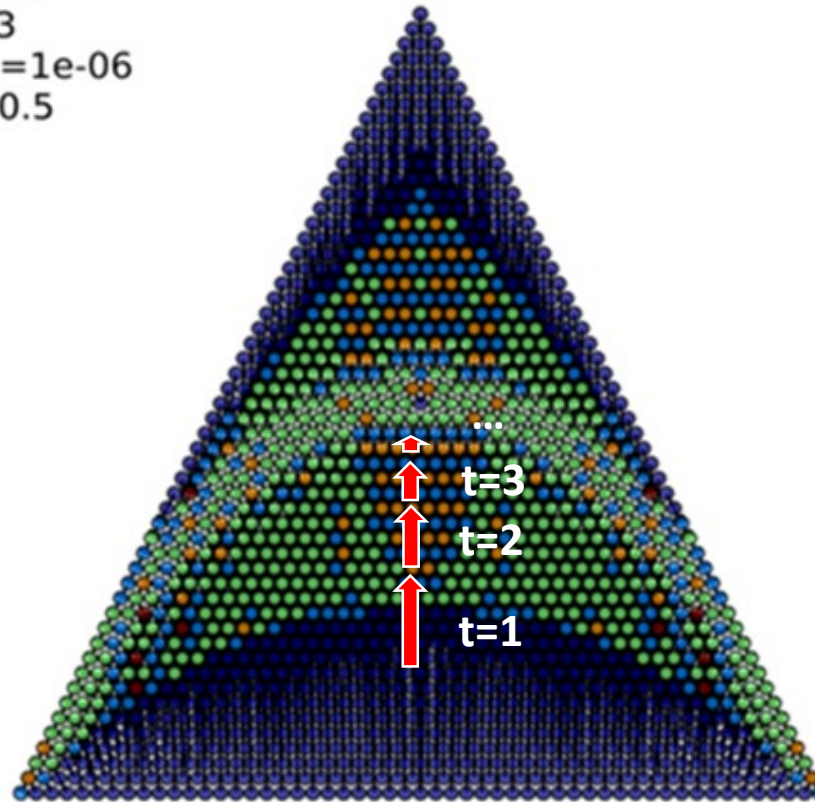
- 30 genotyped individuals
- 10 loci with 2 to 4 alleles (SNP)
- Two-time step sampling



$N=52$   
 $n=3$   
 $\mu=1e-06$   
 $C=0.0$

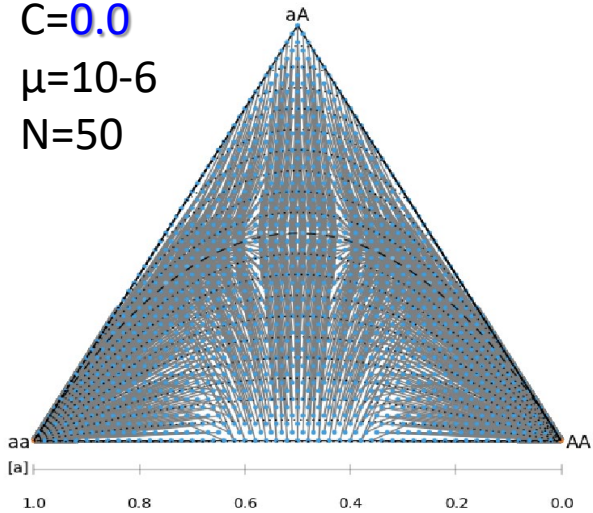


$N=52$   
 $n=3$   
 $\mu=1e-06$   
 $C=0.5$

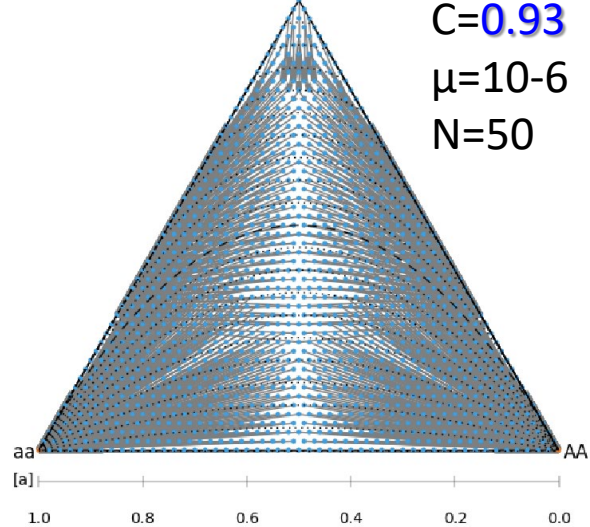


**From low rate of clonality, getting back to genotype frequencies expected under Hardy-Weinberg hypotheses takes more and more generations as we are approaching them!**

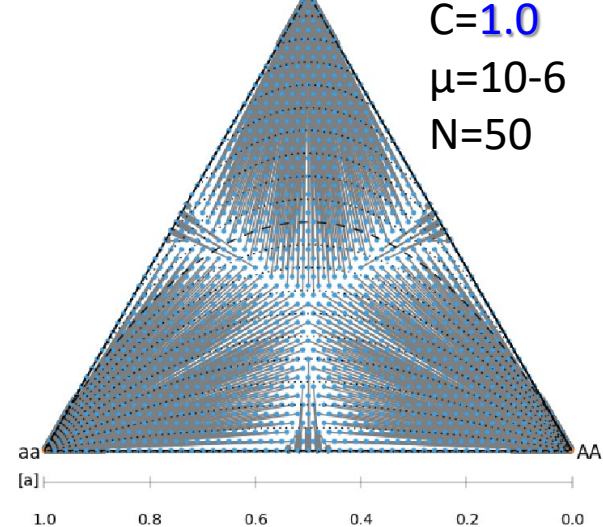
$C=0.0$   
 $\mu=10^{-6}$   
 $N=50$



$C=0.93$   
 $\mu=10^{-6}$   
 $N=50$

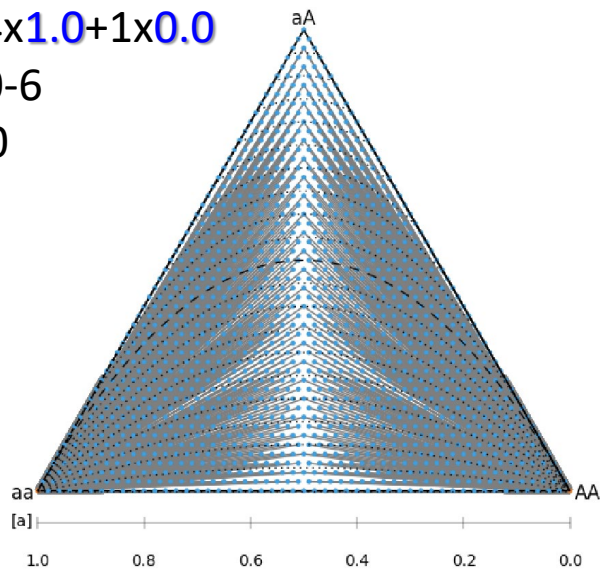


$C=1.0$   
 $\mu=10^{-6}$   
 $N=50$

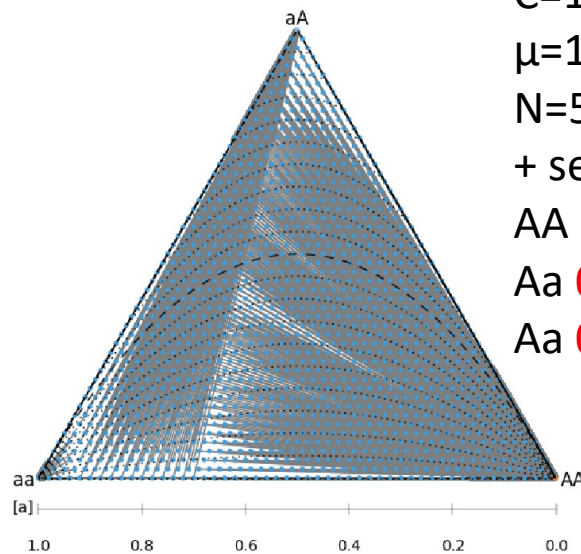


Non cyclical partial clonality

$C=14 \times 1.0 + 1 \times 0.0$   
 $\mu=10^{-6}$   
 $N=50$



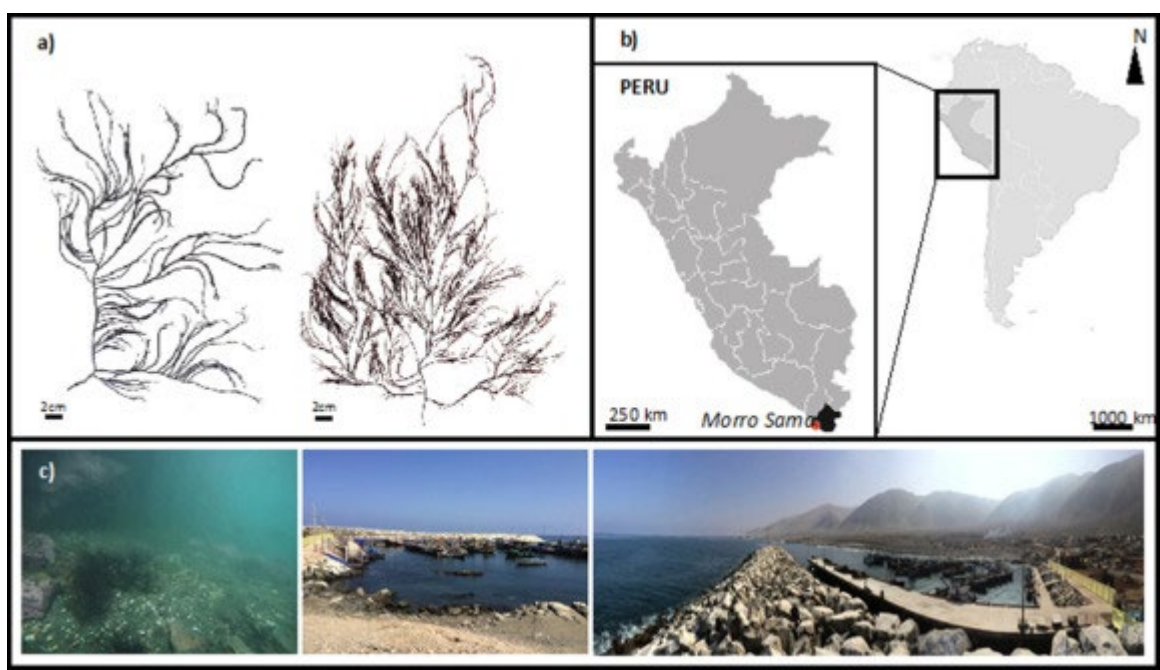
$C=14 \times 1.0 + 1 \times 0.0$   
 $\mu=10^{-6}$   
 $N=50$   
+ selection  
AA **1.0**  
Aa **0.95**  
Aa **0.9**



Cyclical parthenogenesis

(Fungi, Insects, Crustaceans, Rotifers,...)





## Collaborations with

Marie-Laure Guillemin,

Ciencias Ambientales y Evolutivas, Universidad Austral de Chile

Sylvain Faugeron,

Ecología, Pontificia Universidad Católica de Chile

# Simplified biologic system & environment

Warmer waters

introduction, 1930

First recorded sexual reproduction 1993

Population adventice isolée pendant 30 ans

$$\begin{cases} dp = [(-2p+q)\alpha + ((p+q/2)^2 - p)\beta] dt + (1-p)\sqrt{pd}B^p - p\sqrt{qd}B^q - p\sqrt{r}dB^r \\ dq = [2(p-q+r)\alpha + (2(p+q/2)(r+q/2) - q)\beta] dt - q\sqrt{pd}B^p + (1-q)\sqrt{qd}B^q - q\sqrt{r}dB^r \\ dr = [(-2r+q)\alpha + ((r+q/2)^2 - r)\beta] dt - r\sqrt{pd}B^p - r\sqrt{qd}B^q + (1-r)\sqrt{r}dB^r \end{cases}$$

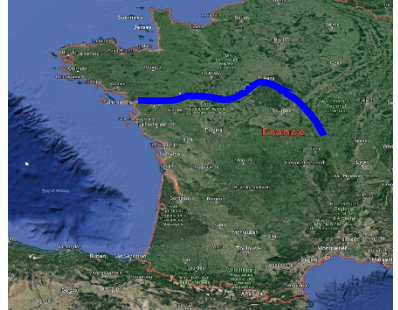
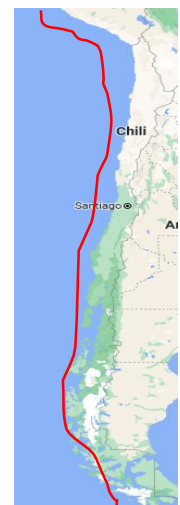
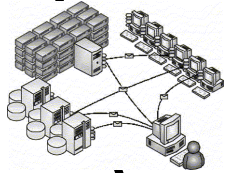
## Stochastic mathematical models

## Signal identifiability Model analyses

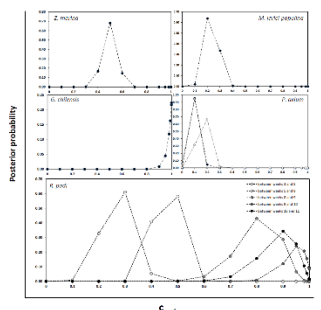
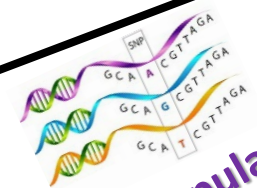
## Simulations

```

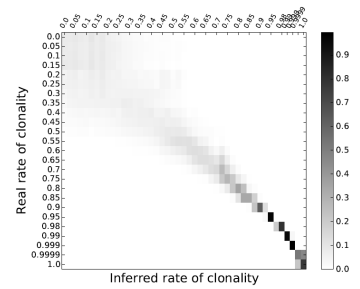
## reproduction
for p in xrange(nbpop):
    #####
    # S E X U A L #
    #####
    NbDeaX = (1-q)*(nbind+0.0000000000000001)*restesex
    restesex = NbDeaX - int(NbDeaX)
    NbDeaX = int(NbDeaX)
    #choix des parents
    Lip = numpy.random.randint(nbind, size=NbDeaX) # liste des indices
    Lm = numpy.random.randint(nbind, size=NbDeaX) # liste des indices
    for i,j in zip(Lip,Lm):
        temp = numpy.empty(nblocus*2, dtype=DATA.dtype)
        o = numpy.random.randint(0, 2, nblocus) + numpy.arange(0, 2 * nblocus)
        temp[o] = DATA[p,i,o]
        temp[o+nblocus] = DATA[p,j,o]
        DATA[p,ind,i] = temp
        ind+=1
    #####
    # A S S E M B L E #
    #####
    NbDeaA = nbind - NbDeaX
    pl=numpy.random.randint(nbind, size=NbDeaA)
    
```



Hiplex  
Genotyping populations



## Quantitative statistical inferences



## Clonix2D

The genetic consequences of partial clonality in populations colonizing new areas

### Task1:

Theoretical dynamics of genetic diversity in populations colonizing new areas at increasing rates of clonality

Resp. S.Stoeckel

### Task2:

Inference of clonal dispersal from various datasets

Resp. S.Arnaud-Haond

### Task3:

Inference of selective events when partially clonal populations colonize new areas

Resp. F.Halkett

# Clonix2D

The genetic consequences of partial clonality in populations colonizing new areas

**Task1:**  
Theoretical dynamics  
of genetic diversity in  
populations colonizing  
new areas at  
increasing rates of  
clonality

**Task2:**  
Inference of clonal  
dispersal from various

**Task3:**  
Inference of selective  
events when partially  
clonal populations  
colonize new areas

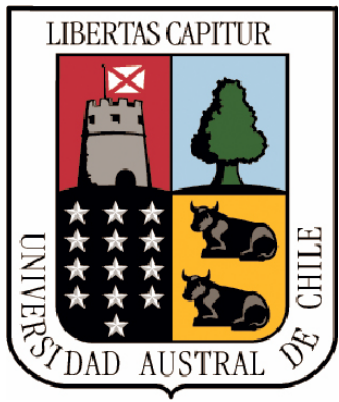
**WG1: Data acquisitions  
and managements**  
SM & MLG

**WG5: Sharing of  
progresses**  
SAH & SKH

**WG2: Handling and  
production of genomic  
data**  
MV & DB

**WG4: Statistical  
inferences**  
SdM & NP

**WG3: Mathematical  
models, numerical  
computations**  
MEO, FM & SS



INRAE



UQÀM





# ClonX consortium

A comprehensive theoretical framework to interpret molecular data  
in the framework of population genetics

## ○ WHO?

- INRAE: aphids, fungi, trees, plants, daphnia
- Ifremer: Seagrasses and corals
- CNRS: Algae
- IRD: human pathogens

## ○ OBJECTIVES:

through mathematics and simulations:

- assess the effect of increasing rate of clonality on the genotypic and genetic composition of populations
- In turn develop estimates of rates of clonality
- Improve dispersal estimates

How does clonality affect the genomic and genetic architecture of populations?

How do genetic and genotypic parameters relate to the extent of clonality and can they lead to its estimation?





Sophie Arnaud-Haond

# Clon*x*



Jean-Pierre Masson



Barbara Porro



Katja Reichel



Ronan Becheler

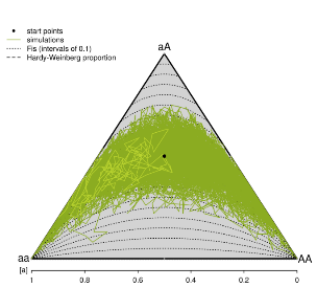
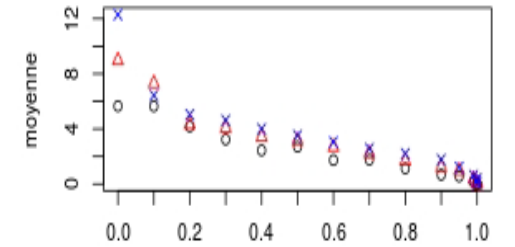


Diane Bailleul

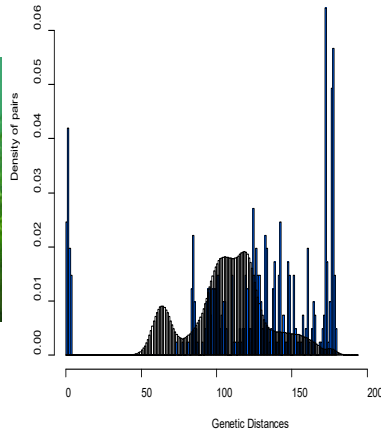


Romuald Rouger

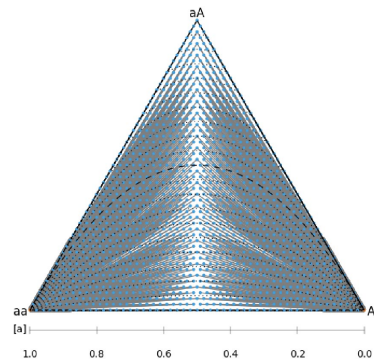
$$\begin{cases} P(\text{observer } aa | \text{repro. asexuée}) = \frac{r_{aa}}{N} = p_{aa} \\ P(\text{observer } aA | \text{repro. asexuée}) = \frac{r_{aA}}{N} = p_{aA} \\ P(\text{observer } AA | \text{repro. asexuée}) = \frac{r_{AA}}{N} = p_{AA} \end{cases}$$



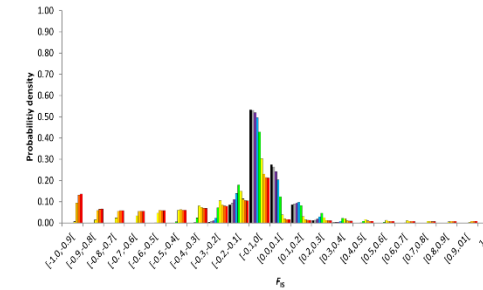
ClonEstiMate



RClone



Cyclical partheno



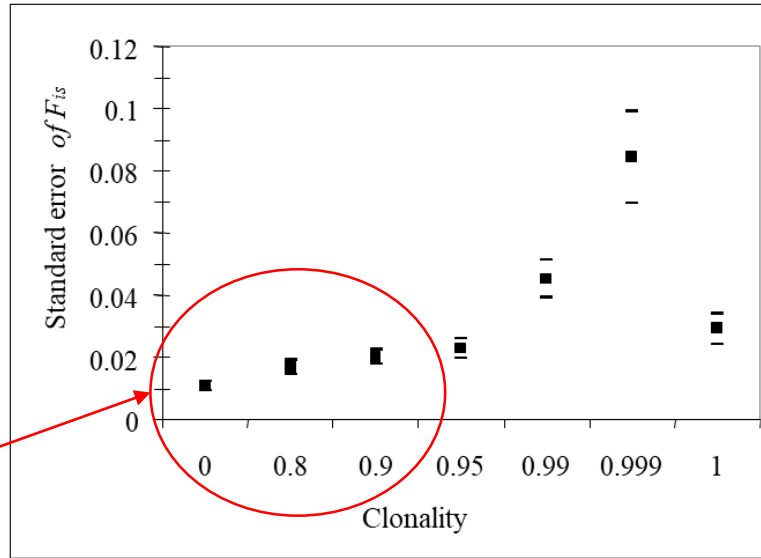
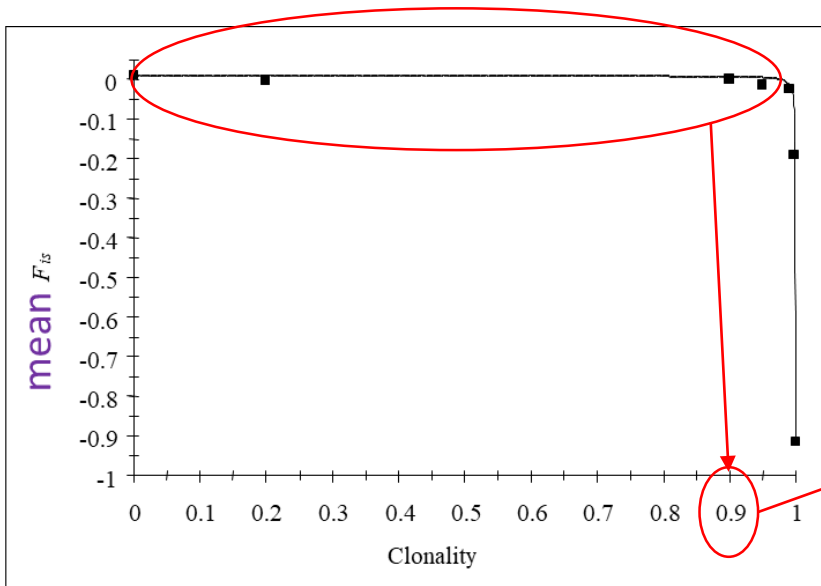
Fabien Halkett

# What the early models say?

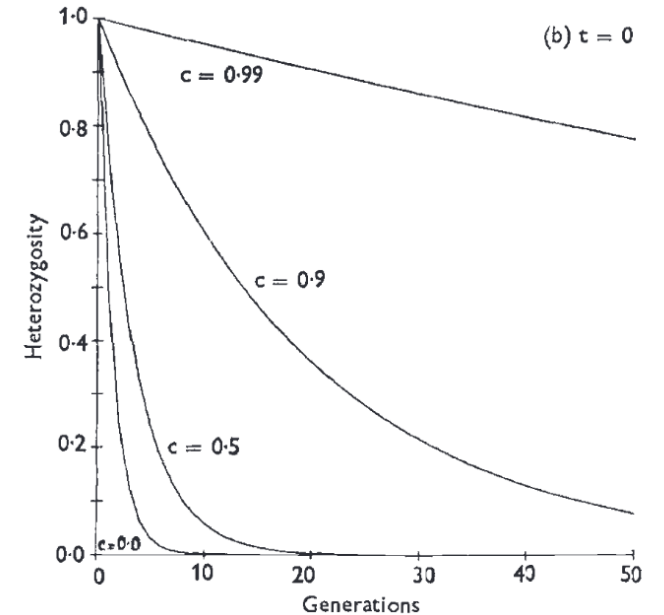
Hypothesis: Clonality impacts genomes as its rate increases (Lewis 1987)?

Marshall & Weir 1979, Orive 1993, Berg & Lascoux 2000, Bengtson 2003, Ceplitis 2003, Balloux et al. 2003, Yonezawa et al. 2004

$$E[T]_f = 1 + \frac{1}{2\Omega} - \Phi, \text{ where } \Phi = \frac{f(\Omega + f + 1)}{2(2\Omega + 1)}.$$



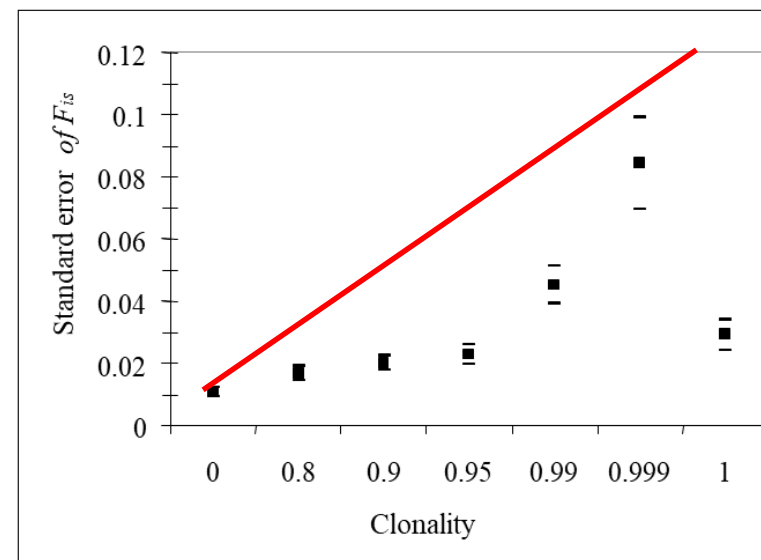
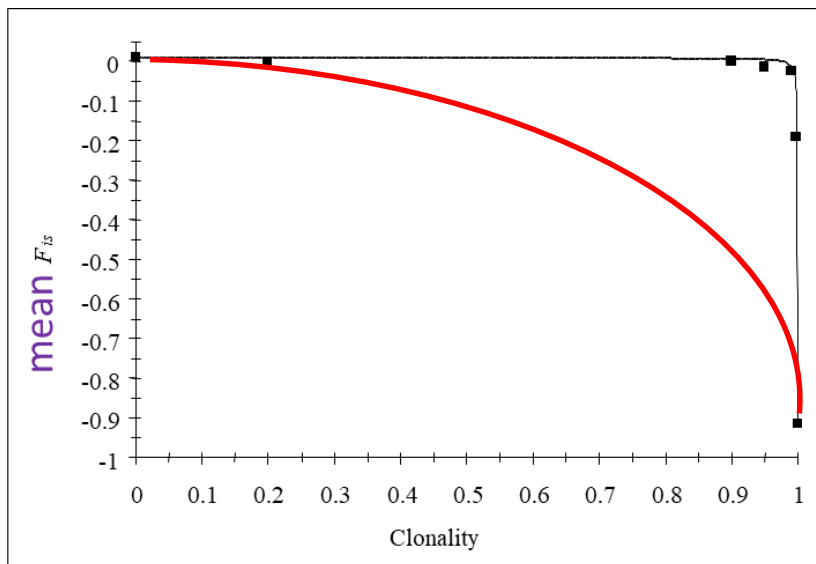
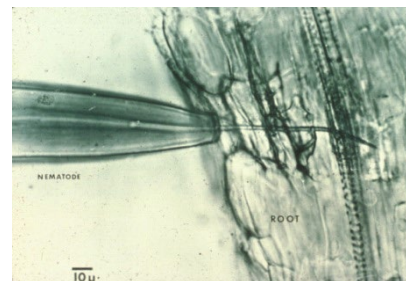
Balloux et al. 2003 Genetics



Marshall & Weir 1979

→ No, as long as rates of clonality are not very high ( $c > 0.9$ ), genetic diversity in partial clonality = sexuality.

# What the data say?



Balloux et al. 2003 Genetics

Three main *kuhnian*\* anomalies

- 1- over-representation of heterozygotes in partially clonal populations ( $0 < c < 0.9$ )
- 2- Observed genetic variances  $\gg$  predicted variances in highly clonals ( $c > 0.9$ )
- 3- high volatility of genetic indices over time and space  $\neq$  stabilité prédictions

\*Thomas S. Kuhn. (1962). The Structure of Scientific Revolutions. University of Chicago Press.

# Reconciling theory with observations

I) Inference: How estimating those rates in field populations?

II) Prediction: How evolve genetic diversity and genomes in such species?

→ Models allowing tracking trajectories, putting probabilities on observations rather than average values of indices (empirically-inspired models rather than theoretically-driven models or data-driven models)

→ Adapted methodologies

→ Adapted datasets



**consortium**

<https://wwz.ifremer.fr/clonix/>

[https://www6.rennes.inra.fr/igepp\\_eng/Productions/Software](https://www6.rennes.inra.fr/igepp_eng/Productions/Software)