

# Vector Competence: Population Genomics, Price Decomposition, and Causal Graphs

Claudio José Struchiner<sup>1</sup>   José M.C. Ribeiro<sup>2</sup>  
Bruno Arca<sup>3</sup>

<sup>1</sup>FGV and IMS/UERJ  
claustru@gmail.com

<sup>2</sup>National Institutes of Health

<sup>3</sup>Sapienza University of Rome

COLMEA 2019

# Evolution (Dennett)

“Noise”

Evolution (Dennett)

Noise into signal

Motivation /  
Implications

Hands-on  
analytical tools

Wrap-up

Causal Inference

“Evolution is all about turning “bugs” into “features”, turning “noise” into “signal”, and the fuzzy boundaries between these categories are not optional; the opportunistic open-endedness of natural selection depends on them.

## Interspecies Vectorial Capacity

"Noise"

Noise into signal

Vector trait recurrence

Results

Competence

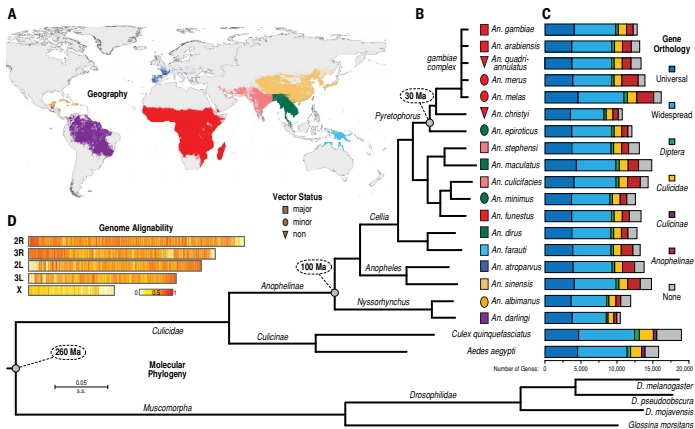
Vectorial Capacity

Mosquitoes Susceptibility

Motivation /  
ImplicationsHands-on  
analytical tools

Wrap-up

Causal Inference



Geography, vector status, molecular phylogeny, gene orthology, and genome alignability of the 16 newly sequenced anopheline mosquitoes and selected other dipterans. (Neafsey et al., 2015).

“Noise”

Noise into signal

Vector trait recurrence

Results

Competence

Vectorial Capacity

Mosquitoes Susceptibility

Motivation /  
Implications

Hands-on  
analytical tools

Wrap-up

Causal Inference

# LETTER

doi:10.1038/nature24995

## Genetic diversity of the African malaria vector *Anopheles gambiae*

The *Anopheles gambiae* 1000 Genomes Consortium\*

"Noise"

Noise into signal

Vector trait recurrence

Results

Competence

Vectorial Capacity

Mosquitoes Susceptibility

Motivation /  
ImplicationsHands-on  
analytical tools

Wrap-up

Causal Inference

**Science**express**Research Article**

## **Extensive introgression in a malaria vector species complex revealed by phylogenomics**

partly due to the challenges involved in accurately identifying introgression in the very groups where it is most likely to occur. Recently diverged species often have incomplete reproductive barriers, hence may hybridize in sympatry. However, another fea-

## RESEARCH ARTICLE

## MOSQUITO GENOMICS

**Highly evolvable malaria vectors: The genomes of 16 *Anopheles* mosquitoes**

Comparative analyses show:

- ▶ faster rates of gene gain and loss
- ▶ elevated gene shuffling on the X chromosome
- ▶ more intron losses, relative to *Drosophila*
- ▶ chemosensory genes, do not show elevated turnover but instead diversify through protein-sequence changes.

This dynamism of anopheline genes and genomes may contribute to their flexible capacity to take advantage of new ecological niches, including adapting to humans as primary hosts.

# Competence is a Variable Trait

“Noise”

Noise into signal

Vector trait recurrence

Results

Competence

Vectorial Capacity

Mosquitoes Susceptibility

Motivation /  
Implications

Hands-on  
analytical tools

Wrap-up

Causal Inference

## Why is this trait so variable?

- ▶ differences in mosquito physiology, molecular biology, and/or behavior
- ▶ differences in the environment and the climate
- ▶ differences in host and parasite physiology, molecular biology, and/or behavior

# Vector Competence as a Phenotype

“Noise”

Noise into signal

Vector trait recurrence

Results

Competence

Vectorial Capacity

Mosquitoes Susceptibility

Motivation /  
Implications

Hands-on  
analytical tools

Wrap-up

Causal Inference

Vector  
Competence

hemophagy  
anthropophily  
immune diversity  
exflagelation  
midgut penetration  
salivary gland

**quadrifasciatus** close to gambiae but shows no  
anthropophilic behavior; competent if  
artificially fed; long lasting

**farauti** everything ok but does not live long enough

**darlingi** new world; distance to gambiae = man to  
rat; reasonable vector; contact with parasite  
500 yrs ago



# Vectorial Capacity B.G.

- ▶ Garrett-Jones, 1964: Vectorial capacity is the expected number of humans infected per infected human, per day; parameters that have to do with mosquitoes

$$C = \frac{ma^2be^{-\mu n}}{\mu}$$

- ▶ Ross-MacDonald

$$\frac{dx(t)}{dt} = maby(t)(1 - x(t)) - rx(t)$$

$$\frac{dy(t)}{dt} = ax(t)(1 - y(t)) - \mu y(t)$$

- ▶ Basic reproductive number (threshold condition)

$$R_0 = \frac{ma^2bce^{-\mu n}}{\mu r} = \frac{C \times c}{r}$$

"Noise"

Noise into signal

Vector trait recurrence

Results

Competence

Vectorial Capacity

Mosquitoes Susceptibility

Motivation /  
ImplicationsHands-on  
analytical tools

Wrap-up

Causal Inference

# One Gene, One Parameter

“Noise”

Noise into signal

Vector trait recurrence

Results

Competence

Vectorial Capacity

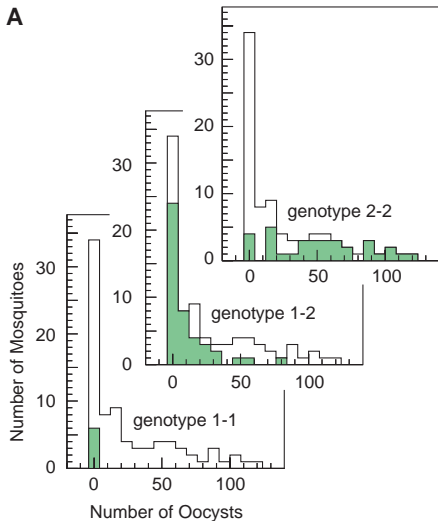
Mosquitoes Susceptibility

Motivation /  
Implications

Hands-on  
analytical tools

Wrap-up

Causal Inference



Infection phenotype of wild *A. gambiae* pedigrees with natural *P. falciparum*. (Niaré et al., 2002)

# Genetically Modified Mosquitoes

"Noise"

Noise into signal

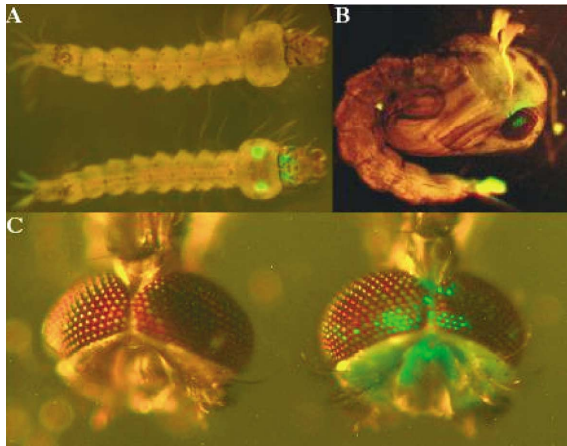
Motivation /  
Implications

GMM

Mosquito "Vaccines":  
WolbachiaHands-on  
analytical tools

Wrap-up

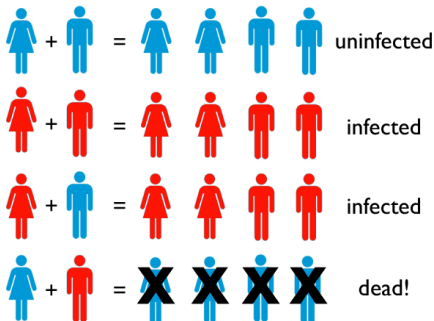
Causal Inference



Pattern of green fluorescent protein (GFP) expression in transgenic *Anopheles stephensi* mosquitoes transformed with a piggyBac vector (Horn et al., 2000). The GFP gene was under the control of the eye-specific 3XP3 promoter. (A) Two larvae: transgenic (bottom) and non-transgenic (top). GFP is visible in the ocelli and salivary glands of the transgenic larva. (B) Transgenic pupa. Note GFP fluorescence in some of the eye ommatidia. (C) Eyes of a non-transgenic (left) and transgenic (right) mosquito. Note that while all eye ommatidia of the transgenic mosquito express GFP, the pattern of fluorescence depends on the angle of incident light (Moreira et al., 2002)

# Cytoplasmic Incompatibility

## Cytoplasmic Incompatibility



Infected males are incompatible with uninfected females

Cytoplasmic incompatibility caused by maternally inherited *Wolbachia* bacteria in arthropods. In diploids, the offspring of crosses between uninfected mothers and infected fathers die early in development. All other crosses are compatible (Werren Lab).

## Molecular evolution

"Noise"

Noise into signal

Motivation /  
ImplicationsHands-on  
analytical tools

Molecular evolution

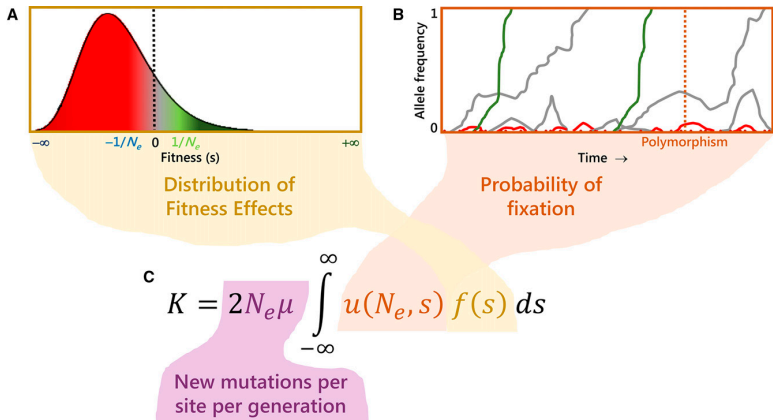
Natural Selection Footprints

MK tables

Price

Wrap-up

Causal Inference



Molecular evolutionary rate ( $K$ ) as a function of (A) the DFE, (B) the probability of fixation of new mutations entering the population, and (C) the rate at which new mutations enter the population per site per generation. Different selection coefficients of mutations are colored in a gradient from maroon (strongly deleterious), red (slightly deleterious), gray (neutral), light green (slightly advantageous), and dark green (advantageous). (Casillas and Barbadilla, 2017)

# Natural Selection Footprints

“Noise”

Noise into signal

Motivation /  
Implications

Hands-on  
analytical tools

Molecular evolution

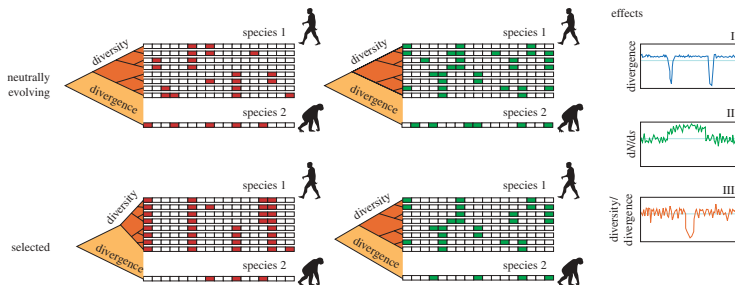
Natural Selection Footprints

MK tables

Price

Wrap-up

Causal Inference



Strategies for detection of the genome-wide selection signatures. SNP variation: white and green alleles (synonymous variants) and also a codon-altering non-synonymous allele (red and white) (Oleksyk, 2009).

# Natural Selection Footprints

"Noise"

Noise into signal

Motivation /  
ImplicationsHands-on  
analytical tools

Molecular evolution

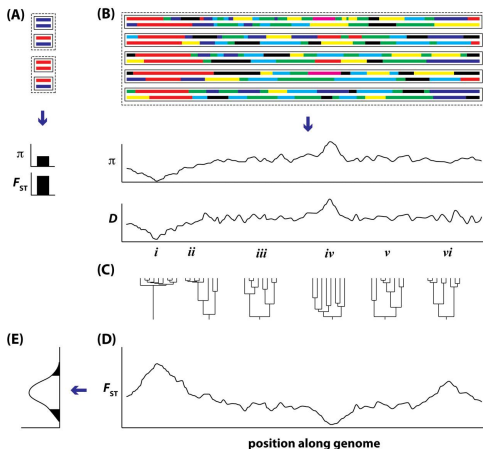
Natural Selection Footprints

MK tables

Price

Wrap-up

Causal Inference



A, Traditional population genetics (individuals within populations); B, summary statistics as continuous variables along the length of the genome, such as  $\pi$  (nucleotide diversity) and allele frequency spectrum (Tajima's D); Signature of evolutionary processes: **i**, hard selective sweep; **ii**, region linked to hard sweep; **iii**, neutral expectation; **iv**, balancing selection; **v**, neutral expectation; **vi**, soft sweep. C, The coalescent structure. D, statistics comparing genetic variation across populations, such as  $F_{ST}$ . E, Collapsing the genomic distribution of a statistic (Hohenlohe, 2010).

# MK table

"Noise"

Noise into signal

Motivation /  
ImplicationsHands-on  
analytical tools

Molecular evolution

Natural Selection Footprints

MK tables

Price

Wrap-up

Causal Inference

**Table 1.** MK table.

	MK		SniPRE		
	Polymorphic	Divergent	Polymorphic	Divergent	
Synonymous	$P_S$	$D_S$	$y_{00}$	$y_{01}$	$n_1$
Non-Synonymous	$P_N$	$D_N$	$y_{10}$	$y_{11}$	$n_2$
		$d$			

Notation used for the MK statistic and SniPRE.  $y_{ij}$  = the number of mutations a gene has in category  $ij$ ;  $i = 1$  if the mutations are non-synonymous, 0 otherwise;  $j = 1$  if the mutations are divergent, 0 otherwise.

McDonald and Kreitman table (Eilertson et al., 2012 Plos CB)



"Noise"

Noise into signal

Motivation /  
ImplicationsHands-on  
analytical tools

Molecular evolution

Natural Selection Footprints

MK tables

Price

Wrap-up

Causal Inference

$$\log(\mu_{ijk}) = \log(Tsites_i) + \beta + \beta^{N_i} + \beta^{D_j} + \beta^{ND_{ij}} + \beta_k^G + \beta_k^{NG_i} + \beta_k^{DG_j} + \beta_k^{NDG_{ij}}$$

 $N_i$  synonymous vs non-synonymous $D_j$  divergent vs polymorphic $\beta_k$  random effects

# Selection vs Constraint (drosophila)

“Noise”

Noise into signal

Motivation /  
Implications

Hands-on  
analytical tools

Molecular evolution

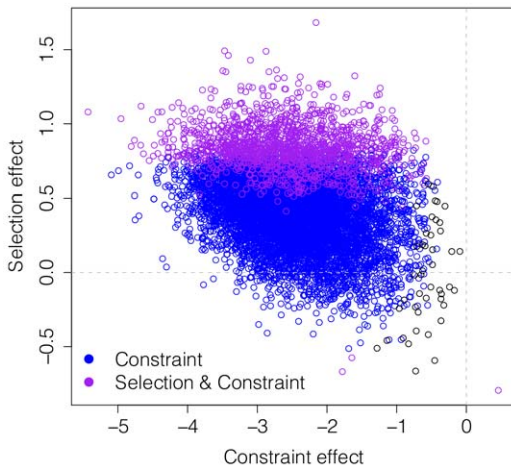
Natural Selection Footprints

MK tables

Price

Wrap-up

Causal Inference



*D. simulans* estimated selection effects and non-synonymous effects for 8,887 genes (Eilertson et al., 2012).

# Selection vs Constraint (humans)

"Noise"

Noise into signal

Motivation /  
Implications

Hands-on  
analytical tools

Molecular evolution

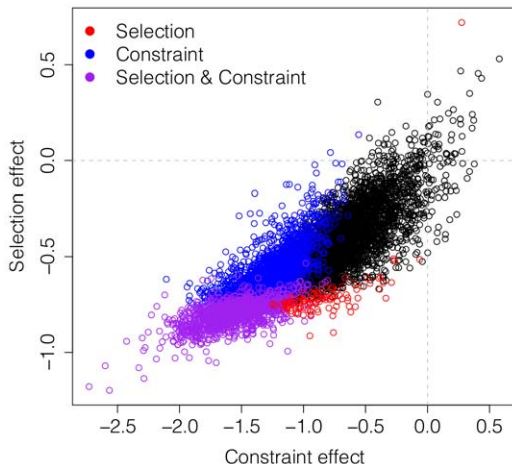
Natural Selection Footprints

MK tables

Price

Wrap-up

Causal Inference



Human estimated selection effects and non-synonymous effects for 11,624 genes. (Eilertson et al., 2012).

# Selection vs Constraint (anopheles)

“Noise”

Noise into signal

Motivation /  
Implications

Hands-on  
analytical tools

Molecular evolution

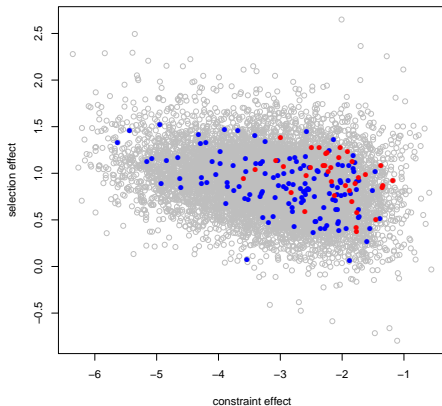
Natural Selection Footprints

MK tables

Price

Wrap-up

Causal Inference



Background, saliva (red) and immune (blue) genes selection and constraint effects.

# Parameters for each gene

“Noise”

Noise into signal

Motivation /  
Implications

Hands-on  
analytical tools

Molecular evolution

Natural Selection Footprints

MK tables

Price

Wrap-up

Causal Inference

LME  $\iff$  PRF

$\gamma$  selection coefficient;  $\gamma = 2 \times N_e \times s$

$1 + s$  fitness of mutants

$\tau$  divergence time

$\mu$  nucleotide mutation rate

$\theta$  per-locus mutation rate;  $\theta = 2 \times N_e \times \mu$

$f$  proportion of nonsynonymous mutations that are non-lethal;  $1 - f$  mutation constraint

$N_e$  effective population size

# Phenotype changes across generations

“Noise”

Noise into signal

Motivation /  
Implications

Hands-on  
analytical tools

Molecular evolution

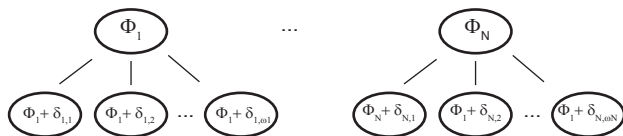
Natural Selection Footprints

MK tables

Price

Wrap-up

Causal Inference



Phenotype of ancestors and descendants (Rice, 2004)

# Phenotype changes across generations

“Noise”

Noise into signal

Motivation /  
Implications

Hands-on  
analytical tools

Molecular evolution

Natural Selection Footprints

MK tables

Price

Wrap-up

Causal Inference

$$\bar{\Phi}' - \bar{\Phi} = \Delta\bar{\Phi} = \frac{1}{\bar{W}}[\text{cov}(W, \Phi) + E(W\bar{\delta})]$$

$N$  Population size

$\Phi_i$  Phenotype of individual  $i$  ( $0 < i \leq N$ )

$\bar{\Phi}$  Mean phenotype in the population

$\delta_{i,j}$  Difference between phenotype of the  $j^{\text{th}}$  descendant of individual  $i$  and  $i$ 's phenotype

$\bar{\delta}_i$  Difference between the mean value of  $\Phi$  among  $i$ 's descendants and  $\Phi_i$

$\bar{\Phi}'$  Mean phenotype of the descendants

$W_i$  Number of descendants of individual  $i$

$\bar{W}$  Mean number of descendants per individual

# Fundamental Theorems of Evolution

David C. Queller\*

Fundamental theorem of evolution  
(Price equation)

$$\Delta \bar{w} = \frac{1}{\bar{w}} (\text{cov}(w, \phi) + E(w\delta))$$

Partitions selection and transmission bias

$E(w\delta) = 0$  (no transmission bias)  
 $\phi = p$  (measured genotypes)

$E(w\delta) = 0$  (no transmission bias)  
 $\phi = g$  (unmeasured genotypes)

Fundamental theorem  
of gene selection  
(Average excess)

$$\Delta \bar{p} = \frac{\bar{p}a_A}{\bar{w}}$$

Selective change in gene frequencies

Fundamental theorem  
of phenotype selection  
(Robertson equation)

$$\Delta \bar{z} = \frac{1}{\bar{w}} \text{cov}(w, g)$$

Selective change in traits

Fundamental theorem  
of selection and heritability  
(Breeder's equation)

$$\Delta \bar{z} = sh^2$$

Partitions selection and heritability

Fundamental theorem  
of adaptation  
(Fisher's fundamental theorem)

$$\Delta \bar{w} = \frac{1}{\bar{w}} \text{var}(g_{(w)})$$

Selective change in fitness

$$r_{w\phi} = 0$$

$$g = g_{(w)}$$

( $\phi$  = any trait value;  $\delta$  = the change in  $\phi$  from parent to offspring;  $w$  = fitness;  $p$  = allele frequency;  $a_A$  = average excess;  $g$  = breeding value;  $z$  = phenotype value;  $r$  = partial correlation;  $s$  = selection differential,  $h^2$  = heritability.)

"Noise"

Noise into signal

Motivation /  
Implications

Hands-on  
analytical tools

Molecular evolution

Natural Selection Footprints

MK tables

Price

Wrap-up

Causal Inference



# Directional selection

“Noise”

Noise into signal

Motivation /  
Implications

Hands-on  
analytical tools

Molecular evolution

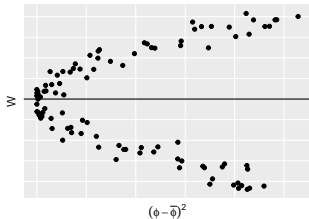
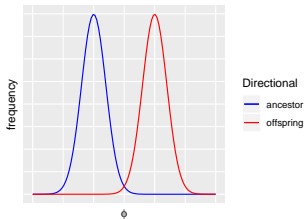
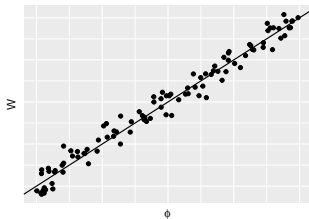
Natural Selection Footprints

MK tables

Price

Wrap-up

Causal Inference



Directional selection (Rice 2004)

# Stabilizing and disruptive selection

“Noise”

Noise into signal

Motivation /  
Implications

Hands-on  
analytical tools

Molecular evolution

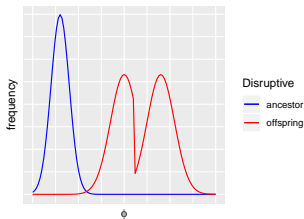
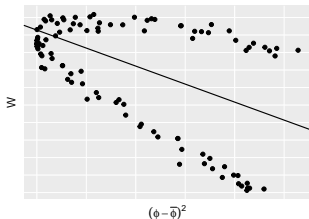
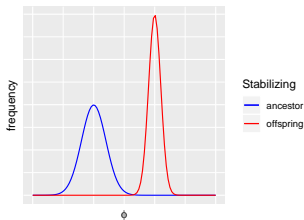
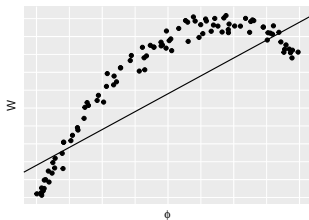
Natural Selection Footprints

MK tables

Price

Wrap-up

Causal Inference



Stabilizing and disruptive selection (Rice 2004)

## Vectorial Capacity B.G.

- ▶ Garrett-Jones, 1964: Vectorial capacity is the expected number of humans infected per infected human, per day; parameters that have to do with mosquitoes

$$C = \frac{ma^2be^{-\mu n}}{\mu}$$

- ▶ Ross-MacDonald

$$\frac{dx(t)}{dt} = maby(t)(1 - x(t)) - rx(t)$$

$$\frac{dy(t)}{dt} = ax(t)(1 - y(t)) - \mu y(t)$$

- ▶ Basic reproductive number (threshold condition)

$$R_0 = \frac{ma^2bce^{-\mu n}}{\mu r} = \frac{C \times c}{r}$$

# Evolutionary dynamics and epidemiological dynamics

“Noise”

Noise into signal

Motivation /  
Implications

Hands-on  
analytical tools

Wrap-up

B.G.

A.G.

Price Like

Causal Inference

estimable from  
sequences

$$\frac{dq_i}{dt} = q_i(r_i - \bar{r}) - \eta q_i + \eta \sum_j \delta_{ji} q_j$$

field data

$$\frac{dx(t)}{dt} = m\bar{a}b y_T(t)(1 - x(t)) - rx(t)$$

$$\frac{dy_T(t)}{dt} = \bar{a}x(t)(1 - y_T(t)) - \bar{\mu}y_T(t)$$

latent class

## Going “genomics”

“Noise”

Noise into signal

Motivation /  
ImplicationsHands-on  
analytical tools

Wrap-up

B.G.

A.G.

Price Like

Causal Inference

$y_T = \sum_i y_i$  total number of infected mosquitoes

$q_i = \frac{y_i}{y_t}$  frequency of mosquito of strain  $i$

$\eta$  mutation rate

$r_i$  fitness of strain  $i$

probability that, given a mutation, a

$\delta_{ji}$  strain of genotype  $j$  changes to  
one of genotype  $i$

$\bar{a} = \sum_i a_i q_i$

$\bar{\mu} = \sum_i \mu_i q_i$

$m\bar{a}\bar{b} = m \sum_i a_i q_i \sum_i b_i q_i$

# Quantities of interest (Price 1970)

“Noise”

Noise into signal

Motivation /  
Implications

Hands-on  
analytical tools

Wrap-up

B.G.

A.G.

Price Like

Causal Inference

- ▶ trait dynamics

$$\frac{d\bar{a}}{dt} = \text{cov}(\mathbf{a}_i, r_i) - \eta(\bar{a} - \bar{a}_\delta)$$

- ▶ average value of trait among all new mutations

$$\bar{a}_\delta = \sum_{ij} a_i \delta_{ji} q_j$$

# Environment and Niche Construction

“Noise”

Noise into signal

Motivation /  
Implications

Hands-on  
analytical tools

Wrap-up

Causal Inference  
Research agenda

Temperature {  
vector mortality  
oviposition  
adult lifespan  
egg hatching and development  
contact and transmission dynamics  
parasite dynamics

Urbanization {  
Increased Global Travel { expansion  
reestablishment

## Representation as a DAG

"Noise"

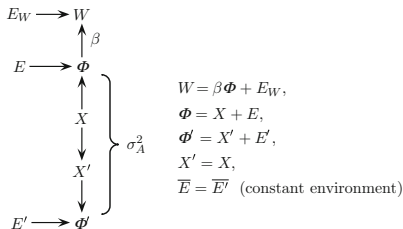
Noise into signal

Motivation /  
ImplicationsHands-on  
analytical tools

Wrap-up

Causal Inference

Research agenda

 $\sigma_A^2$  additive genetic effect (Cov( $\Phi$ ,  $\Phi'$ )) $\Delta\bar{\Phi}$  evolutionary response

$$\Delta\bar{\Phi} = \frac{1}{\beta\bar{\Phi} + \bar{E}_W} \beta\sigma_A^2$$

assumptions: faithful gene transmission ( $\bar{X} = \bar{X}'$ ); homogeneous environments ( $\bar{E} = \bar{E}'$ ); no reproductive bias ( $\bar{\Phi}' - \bar{\Phi} = 0$ )



## Niche construction

"Noise"

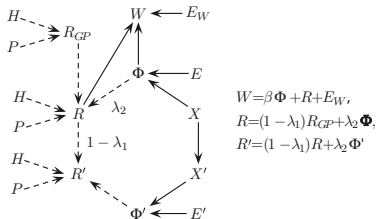
Noise into signal

Motivation /  
ImplicationsHands-on  
analytical tools

Wrap-up

Causal Inference

Research agenda



Organisms as causal agents of environmental changes and evolutionary fate. Modified from Otsuka, 2014, Laland et al. 1999, and Lewontin, 1983

$R_{GP}, R, R'$  environmental resources (niches) of the grandparental, parental, and offspring generations, respectively

$H, P$  human (resources) and parasites (predators) trophic levels

$\lambda_2$  linear coefficient measuring the influence of  $\Phi$  on  $R$

$\lambda_1$  persistence/inheritance of the environmental resource between generations

$$\Delta \bar{\Phi} = \frac{1}{\beta \bar{\Phi} + R} (\beta + \lambda_2) \sigma_A^2$$

assumptions:  $R$  is a group or contextual variable (selection at the group level); it keeps changing across generations and may counteract selection measured by  $\beta$

# Research agenda

“Noise”

Noise into signal

Motivation /  
Implications

Hands-on  
analytical tools

Wrap-up

Causal Inference  
Research agenda

- ▶ study designs (population genomics, climate, behavior):
  - ▶ causal measures,
  - ▶ confounding: backdoor, colliders
  - ▶ mendelian randomization
- ▶ experimental evolution:
  - ▶ sequencing following experimental challenge;
  - ▶ niche manipulation
- ▶ software development
- ▶ what drives us into the future?