






What can molecular markers add to the development of new biocontrol agents?




BINGO







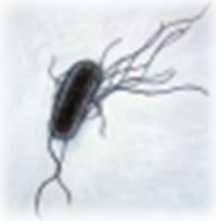

What can genetics add to the development of new biocontrol agents?

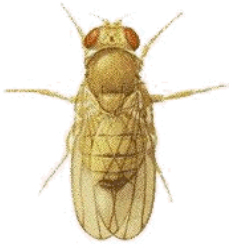



BINGO

Outline:

- Background: matching the enemy to the pest
- Genetic improvement of natural enemies: analysis of relevant factors
- Exploiting genetic variation; a guideline to use what and when
- Outlook






Matching the enemy to the pest

Nature Agriculture

Butterfly **Pest**

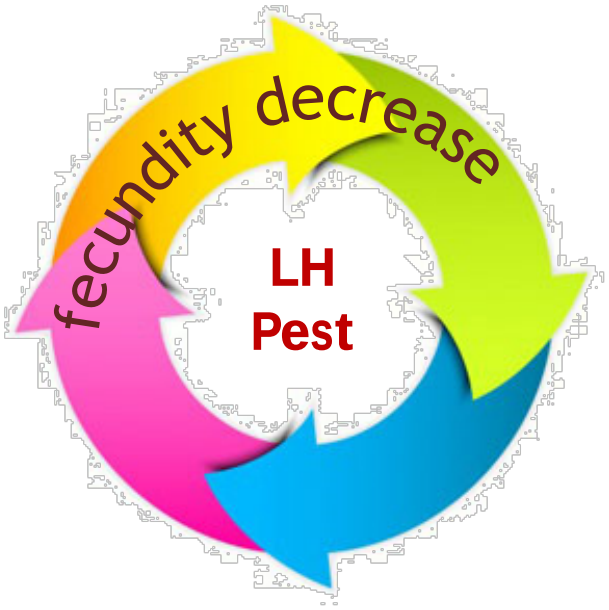
DAMAGE **DAMAGE**

Parasitoid wasp **Natural enemy**

Demands/wishes of the stakeholders

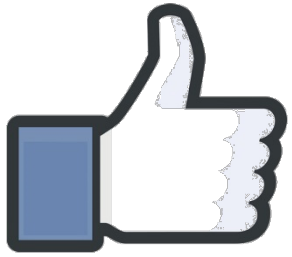
Matching the enemy to the pest

Matching must concern the full life history



DAMAGE

Demands/wishes of the stakeholders



VARIATION

HEREDITY

FITNESS

Adaptive

EVOLUTION

VARIATION

HEREDITY

FITNESS

Desirable Phenotype

Analysis of relevant factors

Exploiting variation in natural enemies

(i) Geographic variation

(ii) "Experimental" approaches

- **Useful to uncover G-P map**
- **Estimating fitness of trait combinations and trade-offs**
- **Evolutionary history**

- **Producing tailor-made strains through genetic improvement**

Analysis of relevant factors

"Desirable phenotype"

Single/simple trait

- E.g. insecticide resistance

Composite/complex traits

- LH adaptations
- Prey handling times
- Stress resistance
- …

Genetic variation and nature of traits

consequences for analysis

discrete classes

alleles individually identified

Population Genetics

continuous variation

partitioning of variances

Quantitative Genetics

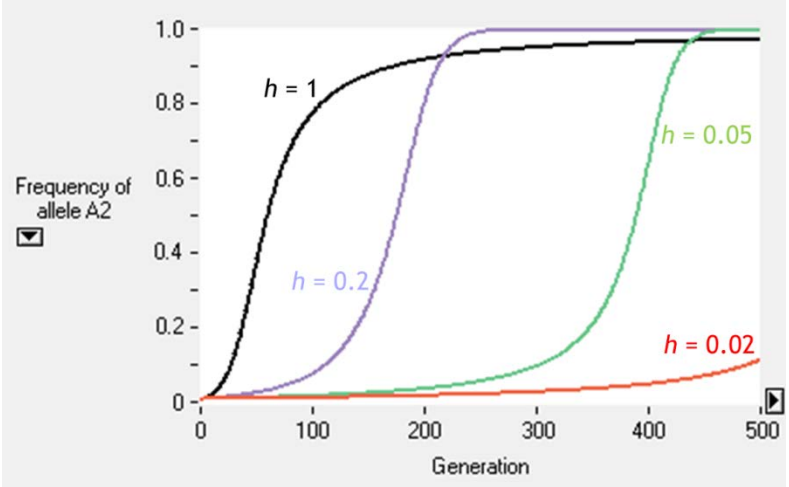
QTLs, GWAS, Genetical Genomics *et cetera*

Analysis of relevant factors

Heredity

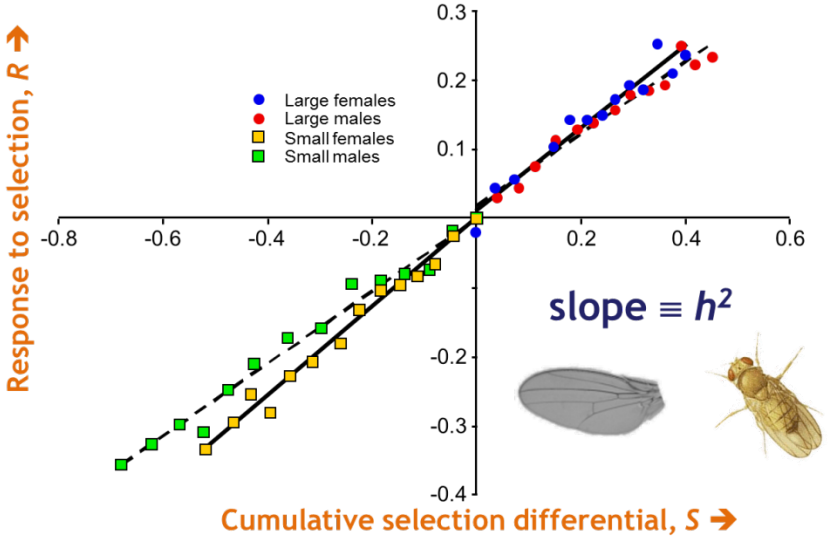
Population Genetics

- *One locus, two alleles;* $q_0 = 0.01$, $s = 0.1$



Quantitative Genetics

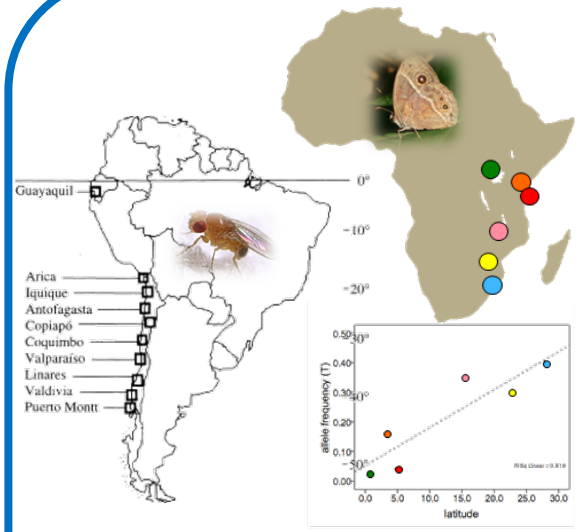
- Heritability (h^2) ~ the proportion of the phenotypic variance (V_P) that is attributable to (additive) genetic variance (V_A)
- $h^2 = {}^{V_A}/_{V_p}$
- Breeders equation, $R = h^2 \cdot S$
- h^2 determines response to selection
- h^2 tends to be low for LH traits



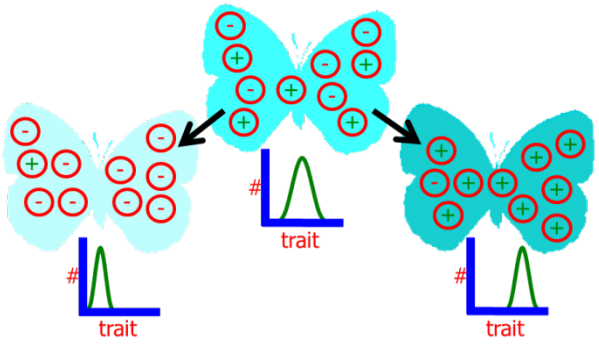
Analysis of relevant factors

Exploiting variation in natural enemies

- (i) *De novo* mutations
- (ii) Standing genetic variation
- Microbes (prokaryotes and simple eukaryotes) usually (i)
- Eukaryotes (small N_e, longer generations times) usually (ii)



Latitudinal clines

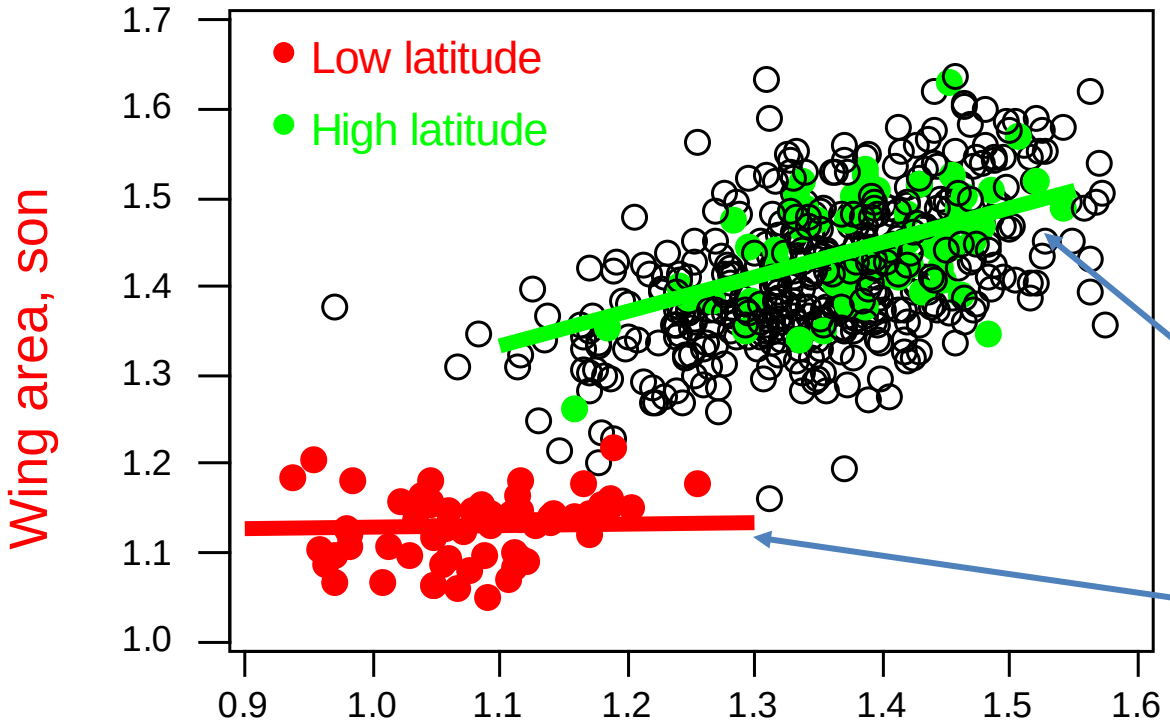


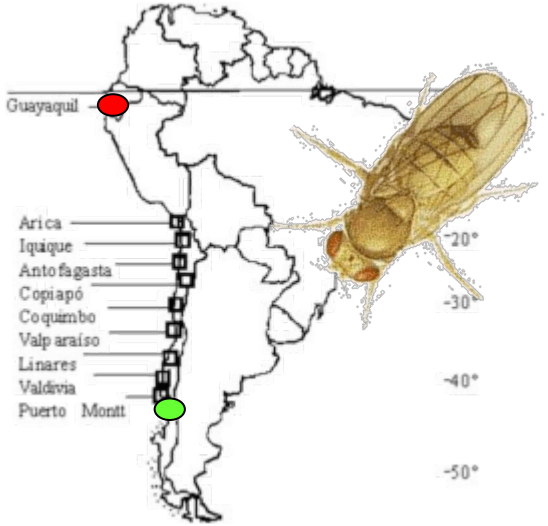
Artificial selection

Experimental evolution

Exploiting variation; what and when?

Heritability of wing area





0.388*2=0.776

0.018*2=0.036

Exploiting variation; what and when?

Adaptation in European *Drosophila* populations

DrosEU

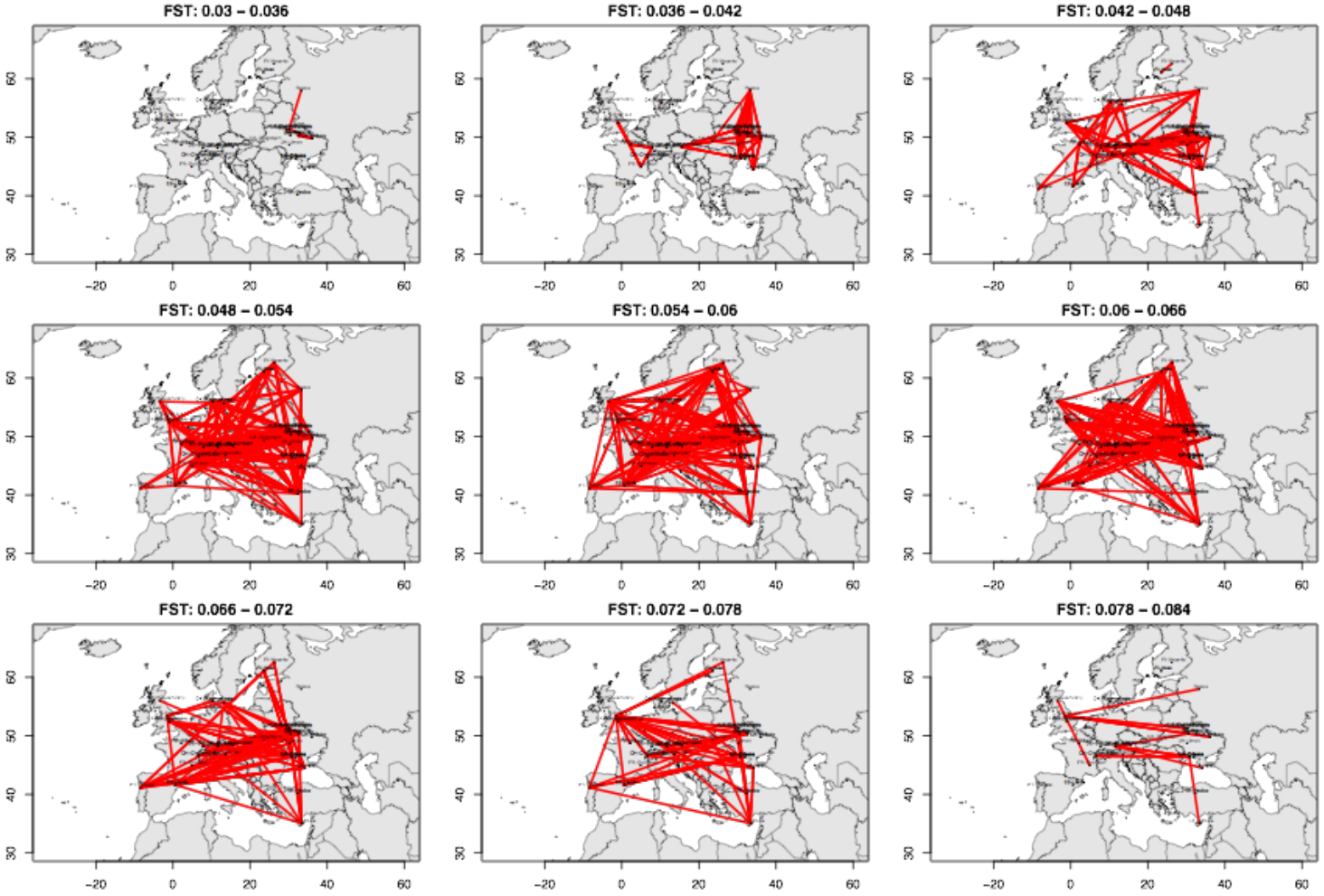
2013

eseb
European Society for Evolutionary Biology

STN - 2016

- Cooperate in collecting, generating and analyzing genomic and environmental data,
- In order to track the eco-evolutionary dynamics for numerous Drosophila populations across Europe (and beyond),
- And to foster the integration and exchange of population genomic information and data

Population genetic differentiation - F_{ST}

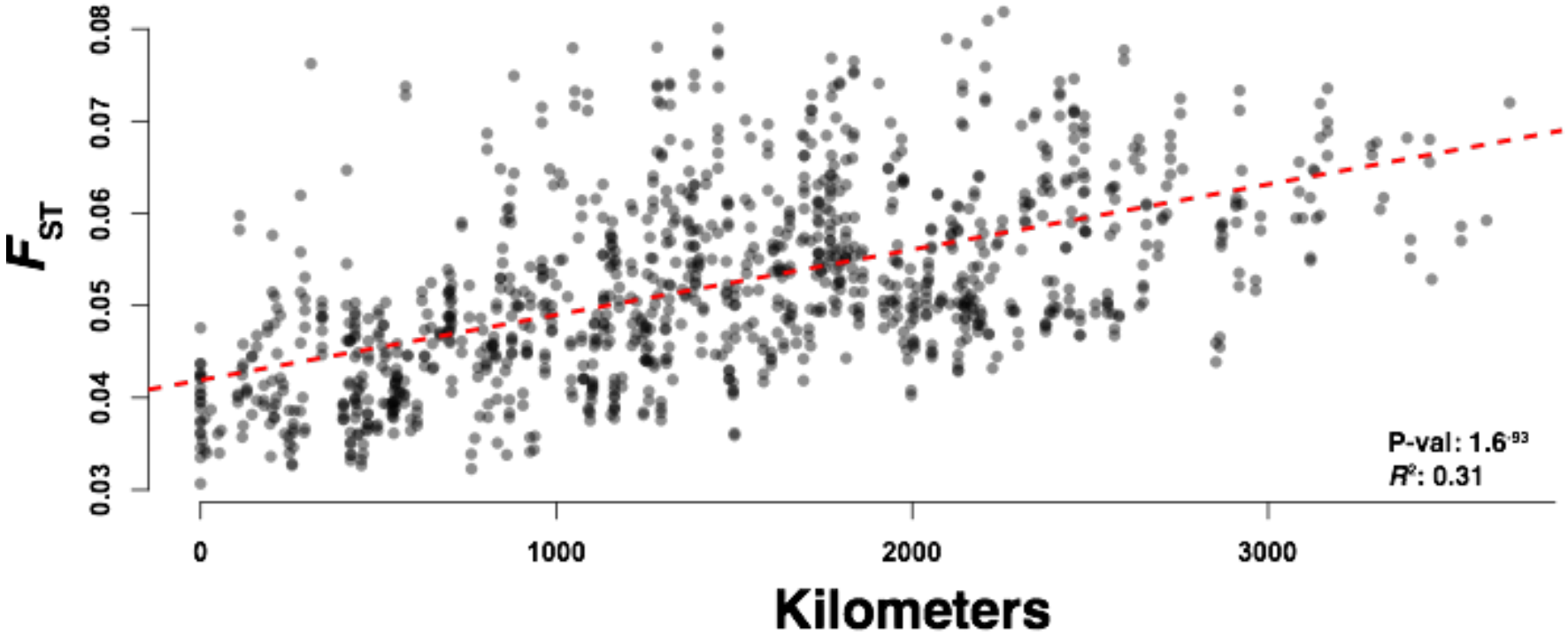


- SNPs located in short introns (<60 bp)
- SNPs in 100kb distance from inversions
- n = 21008

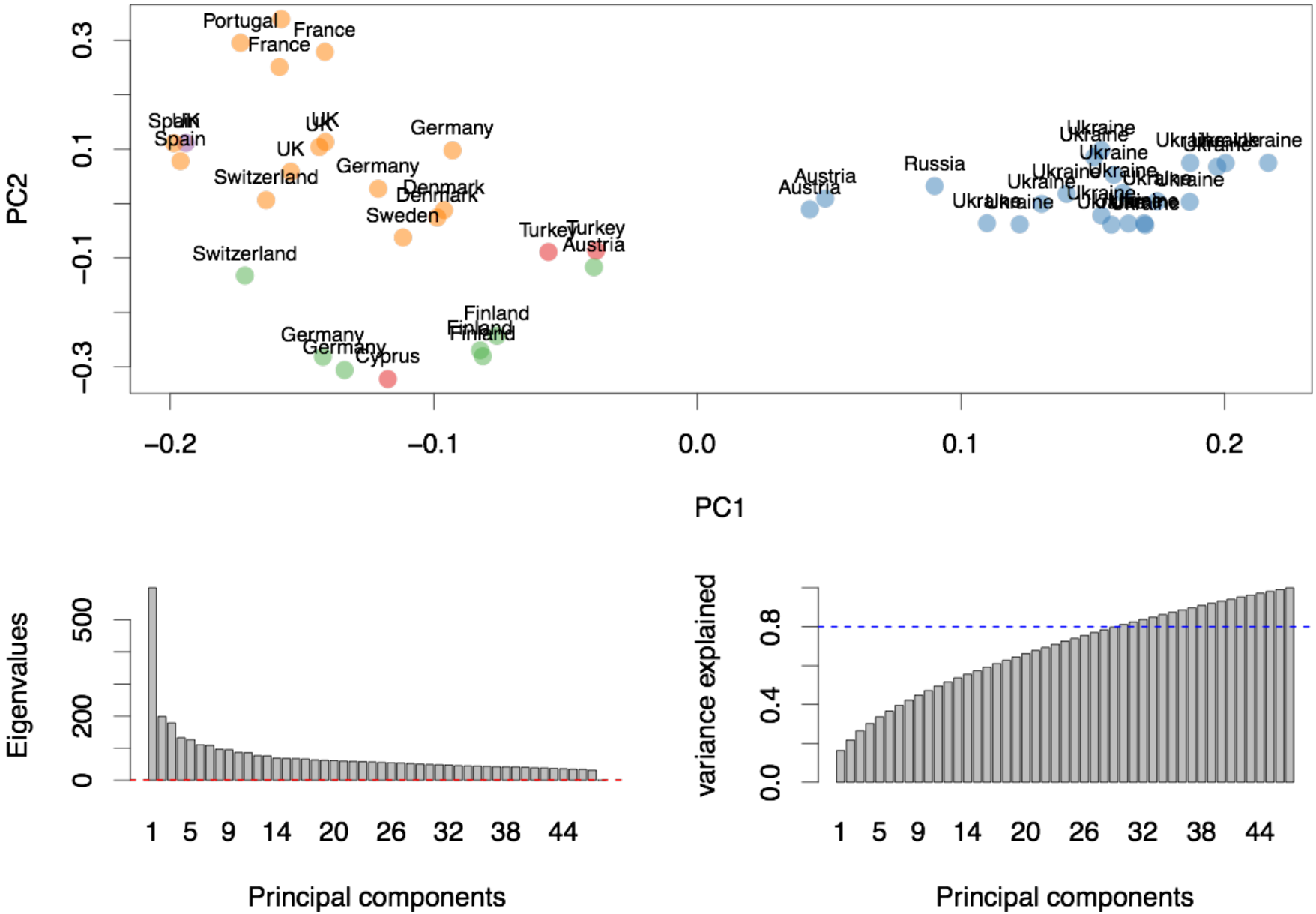
Population genetic differentiation - *IBD*

Correlation: F_{ST} / geographic distance

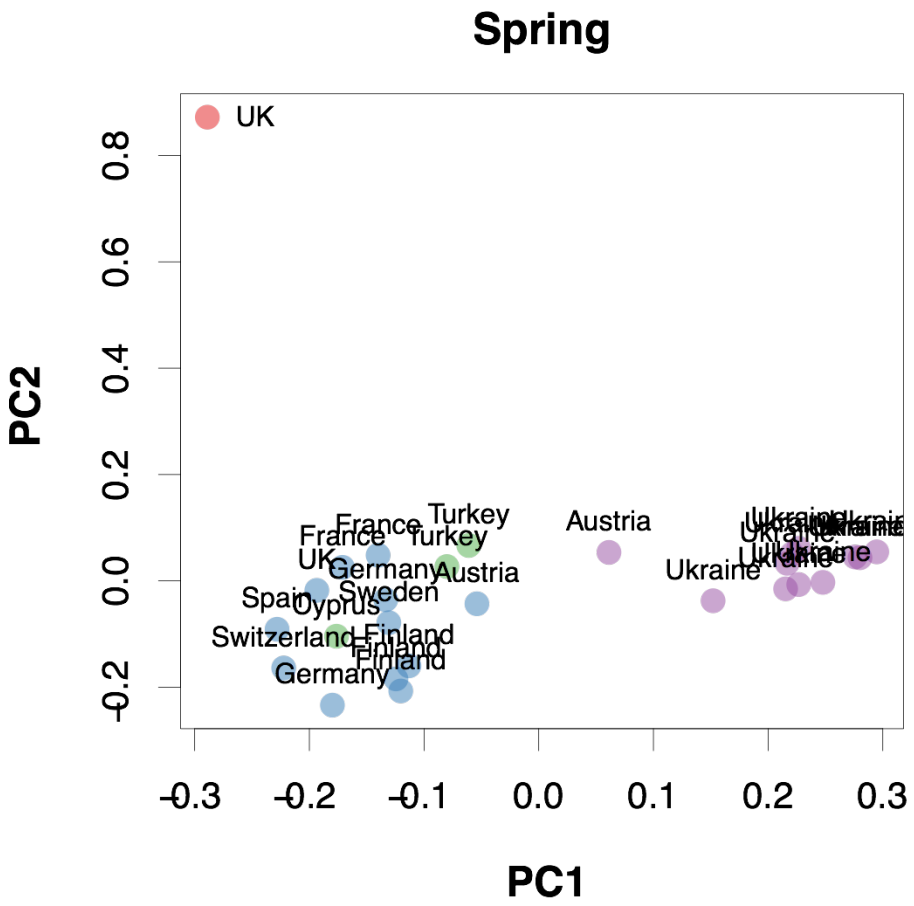
neutral intronic SNPs

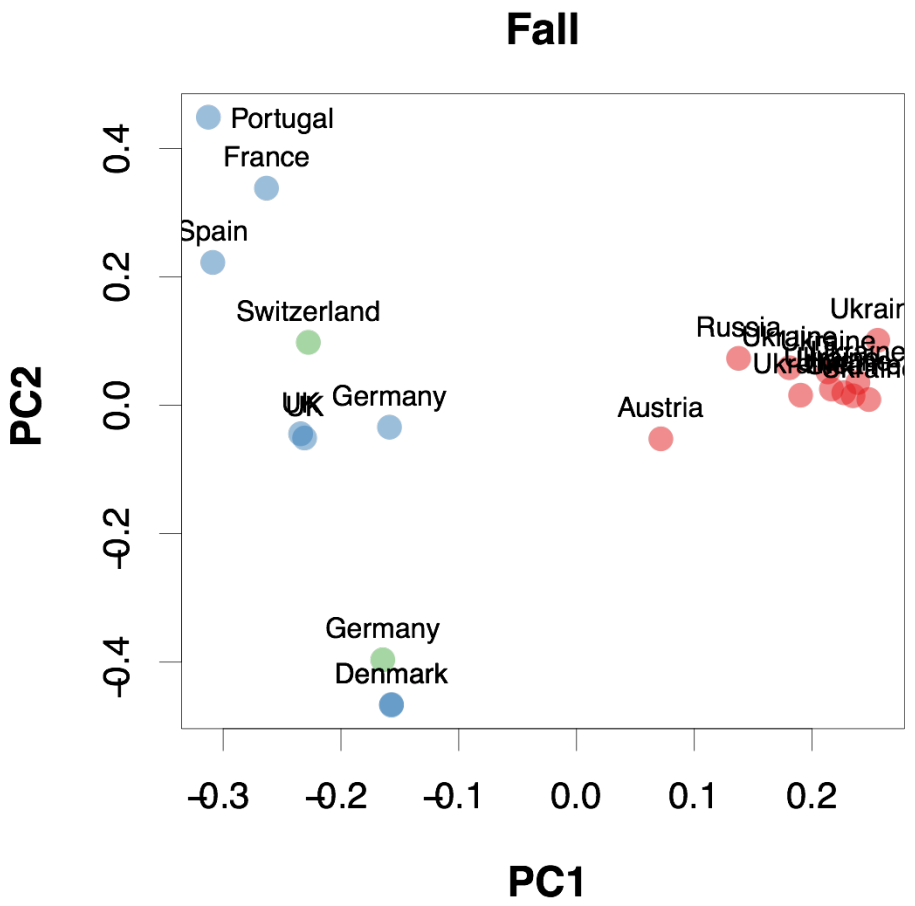


Population genetic differentiation - *IBD*



Population genetic differentiation - *IBD*





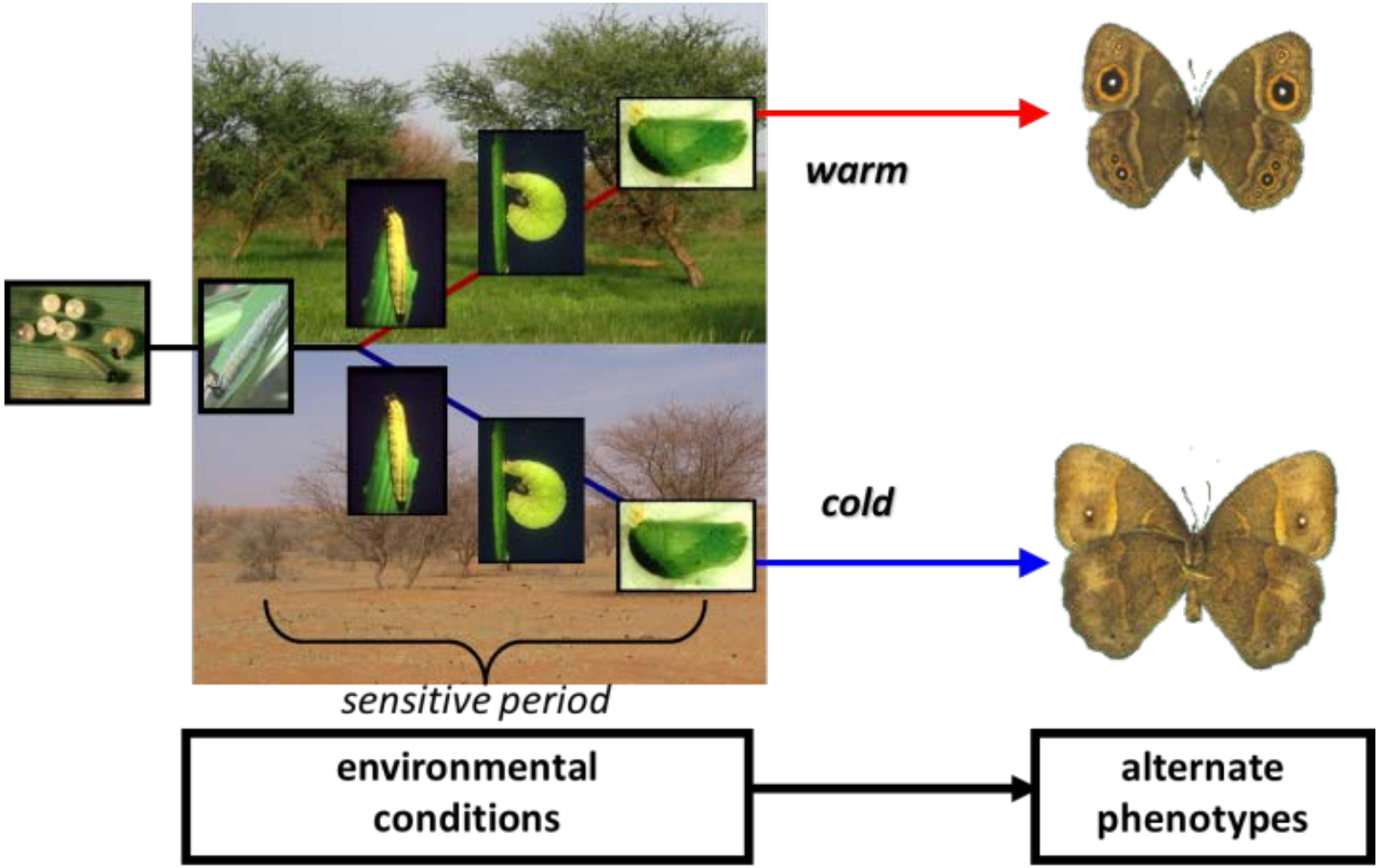
Population genetic differentiation

Conclusions

- X chromosome with highest F_{ST}
- Overall low amounts of genetic differentiation
- Low but highly significant IBD
- Pronounced longitudinal variation/ population structure

Exploiting variation; what and when?

Footprints of selection in wild populations of *Bicyclus anynana*



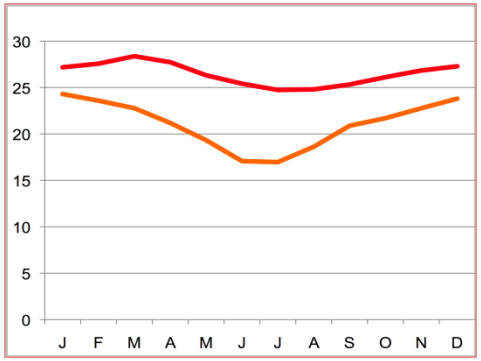
B. anynana: species range and climate

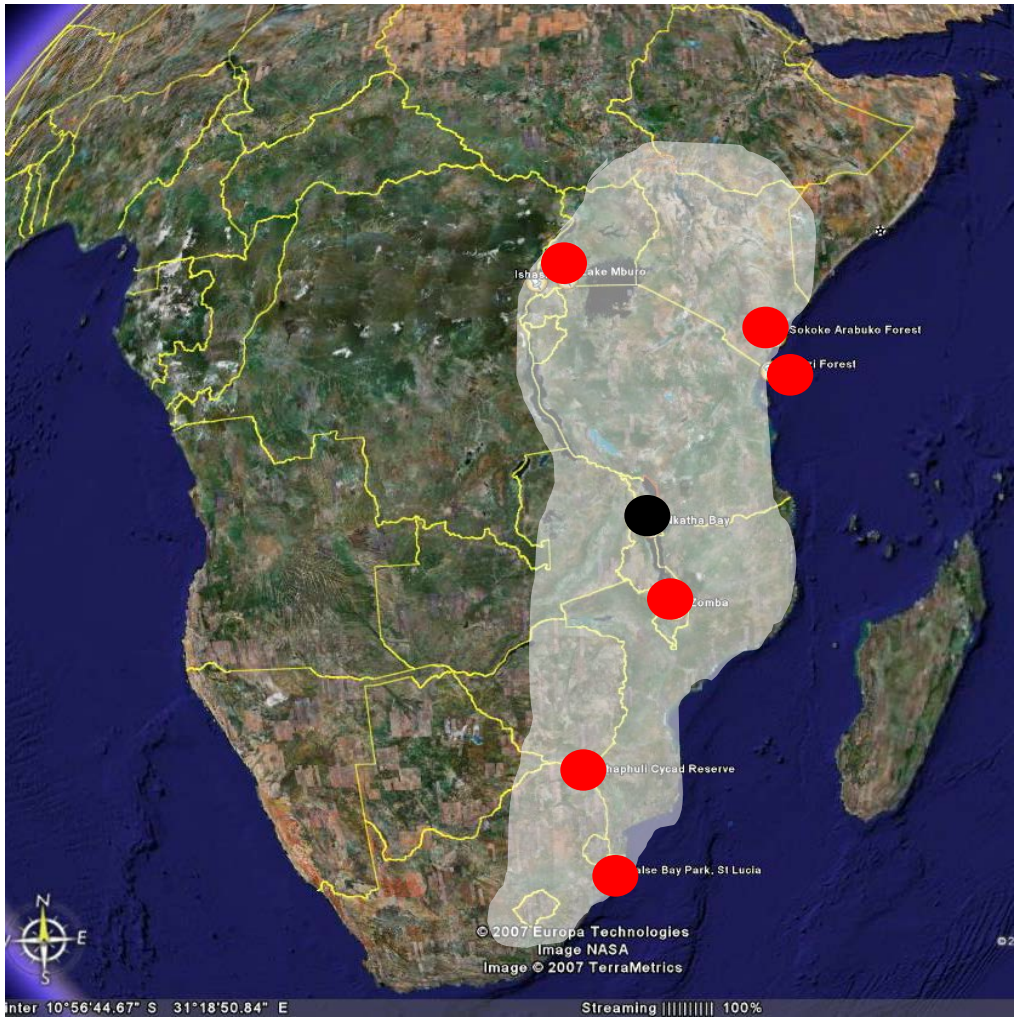
Eastern Africa: equator to subtropics

Towards equator:
- increase average temperatures
- decrease yearly temperature differences

Sampled populations across 3000 km
6 populations: #35 to #50

Kenya
South Africa

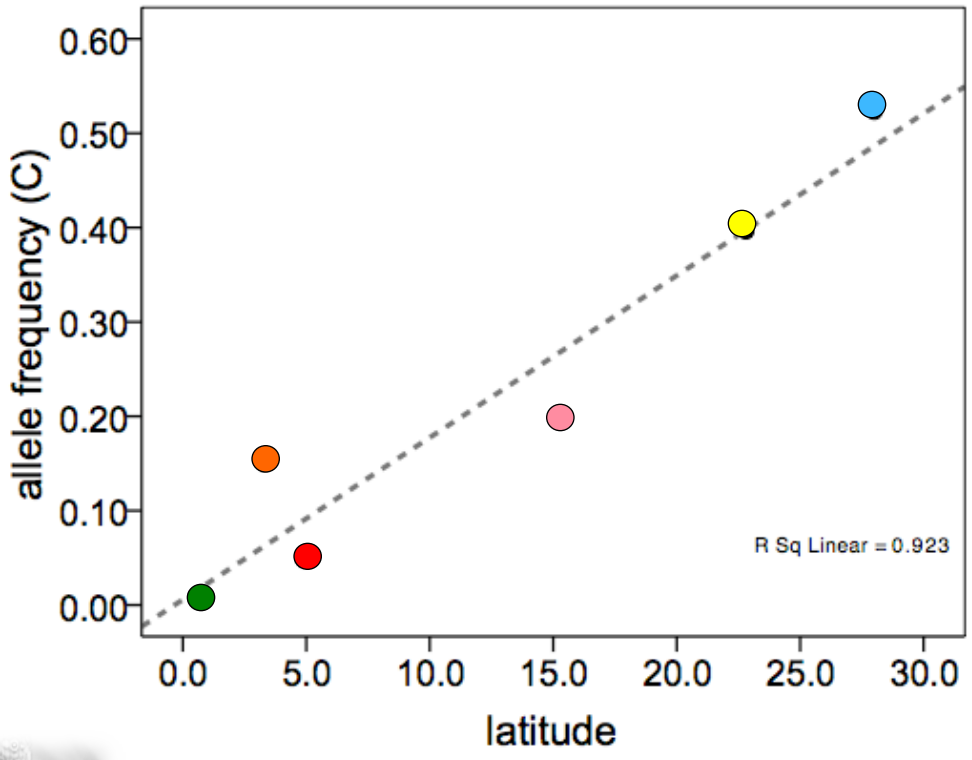


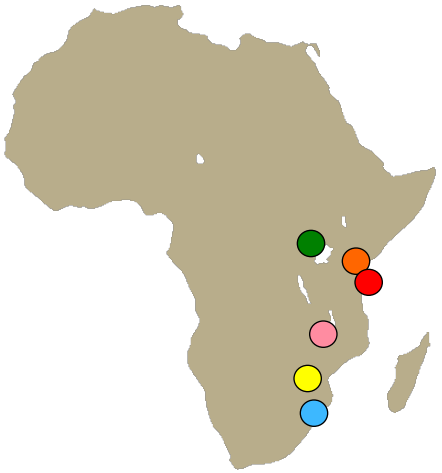


Significant cline:

UGPase

(UDP-glucose-pyrophosphorylase)





Sampled length: 408 bp

1 replacement SNP

Fst = 0.31

R sq = 0.966

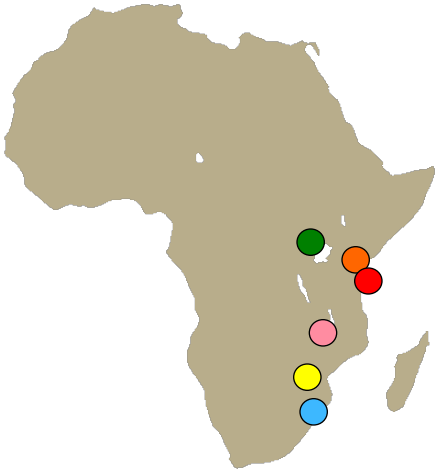
p = 0.002

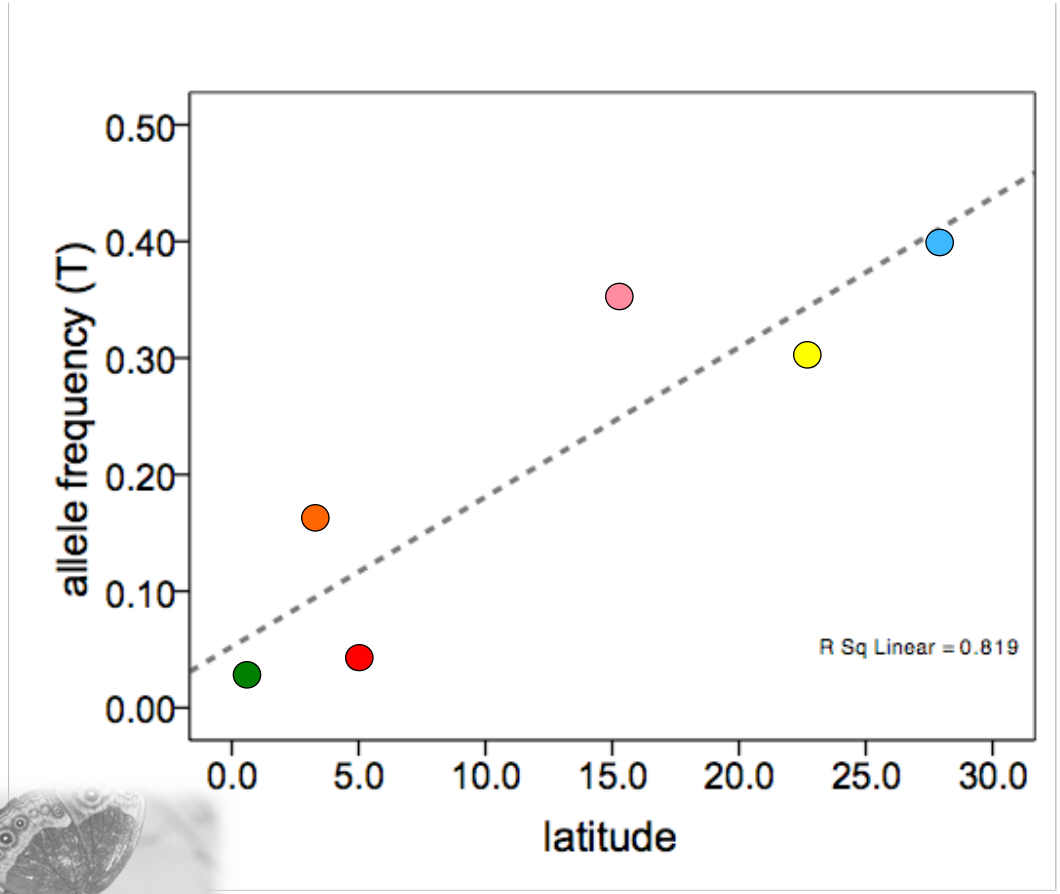
Fst outlier locus

Significant cline:

Treh

(Trehalase)





Sampled length: 931 bp

1 of 3 replacement SNPs

SNP position = 550

Fst = 0.45

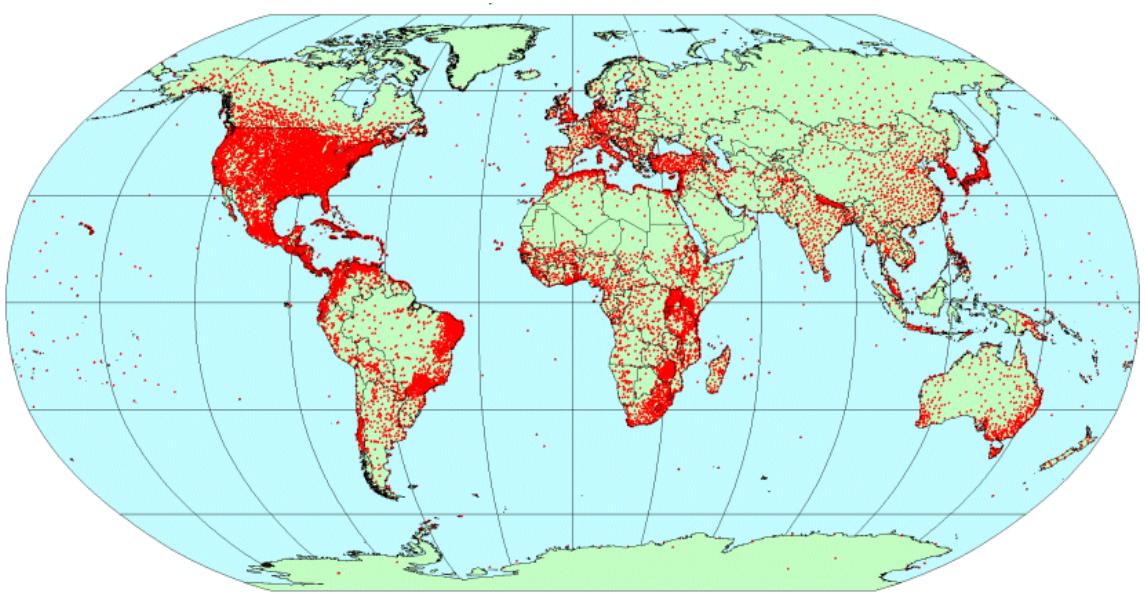
R sq = 0.905

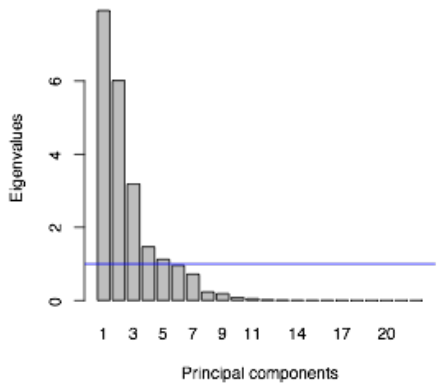
p = 0.013

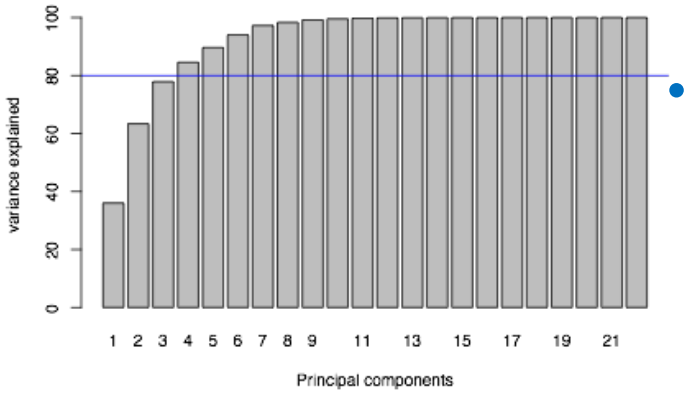
Exploiting variation; what and when?

Adaptation in European *Drosophila* populations

WorldClim Data

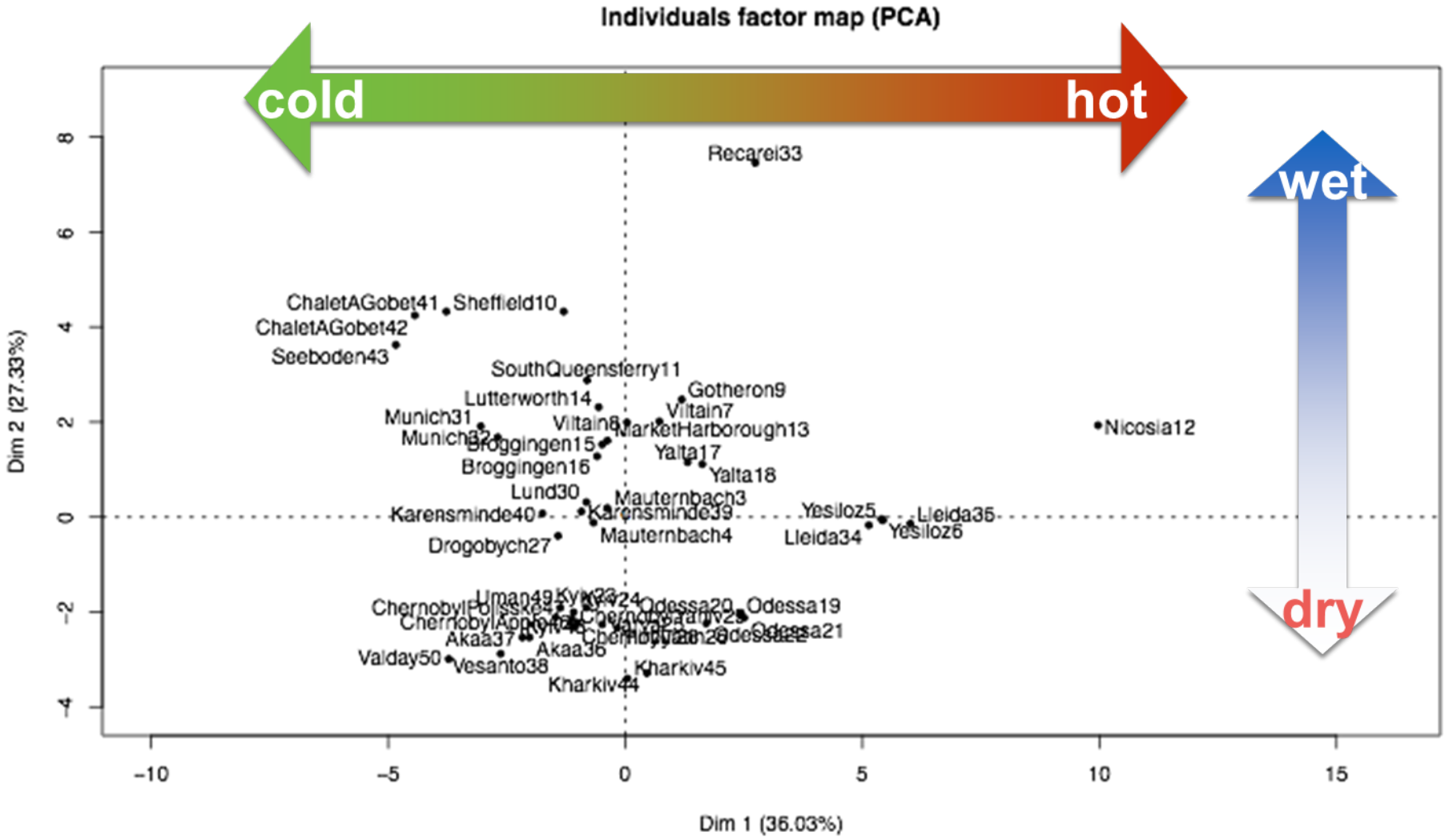






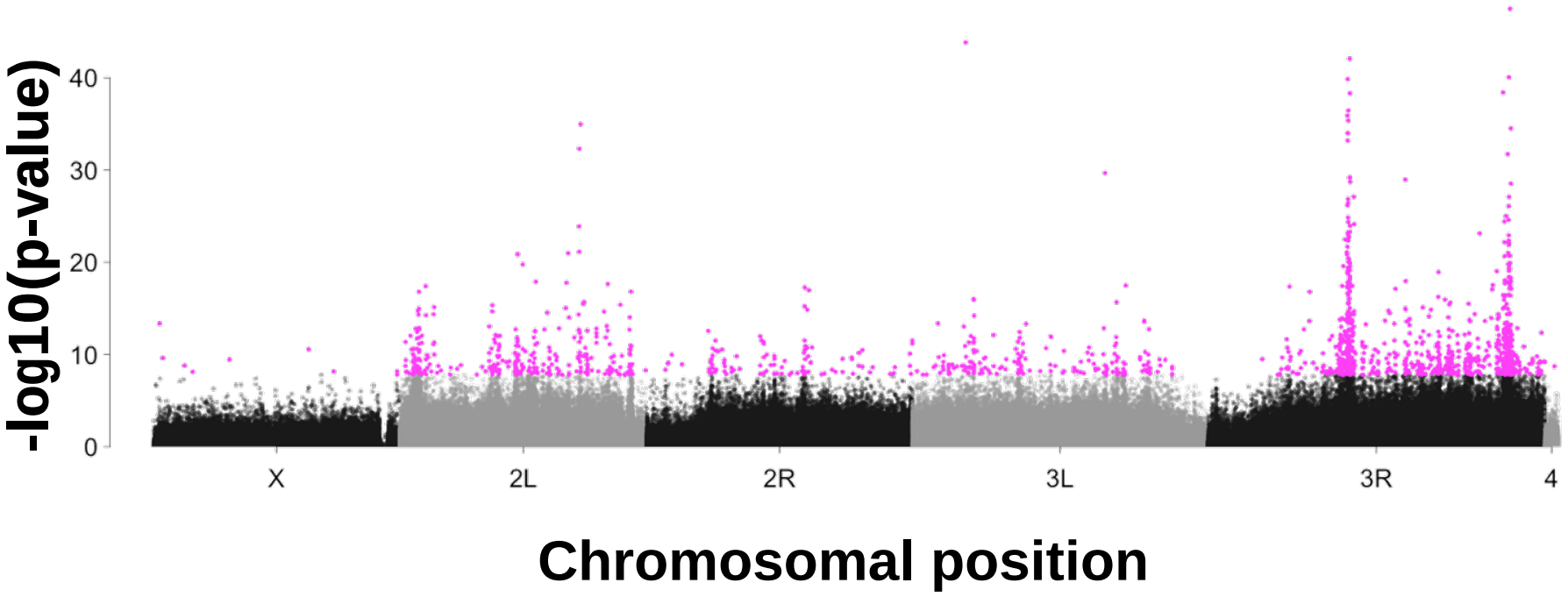
- Interpolated from data collected across 50 years (annual averages)
- 19 interpolated annual average biovariables
- 3 interpolated monthly average biovariables
- z-Transformation (mean=0; Stdev=1

Signatures of climate adaptation



Signatures of climate adaptation

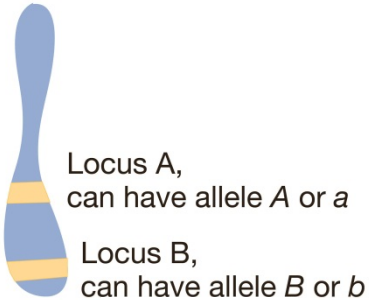
Linking SNPs to climatic variables (PC1)



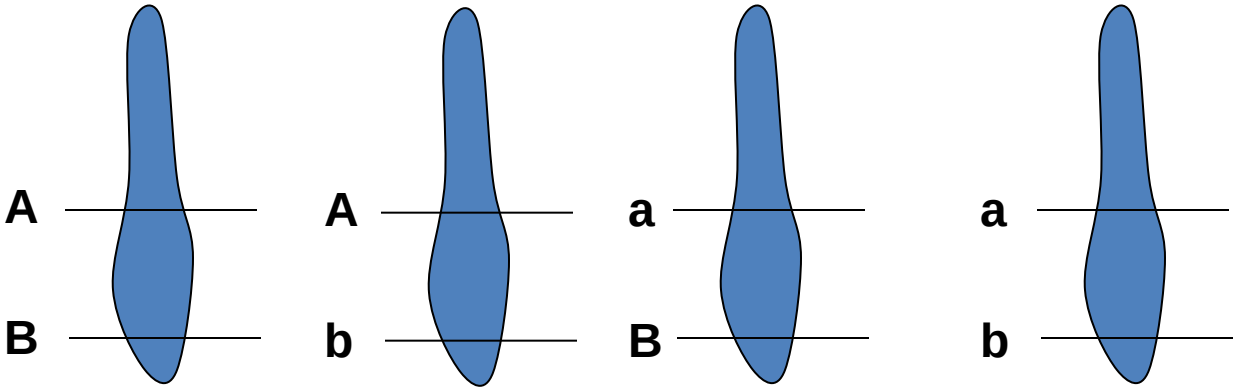
Exploiting variation; what and when?

Intermezzo – Linkage disequilibrium (LD)

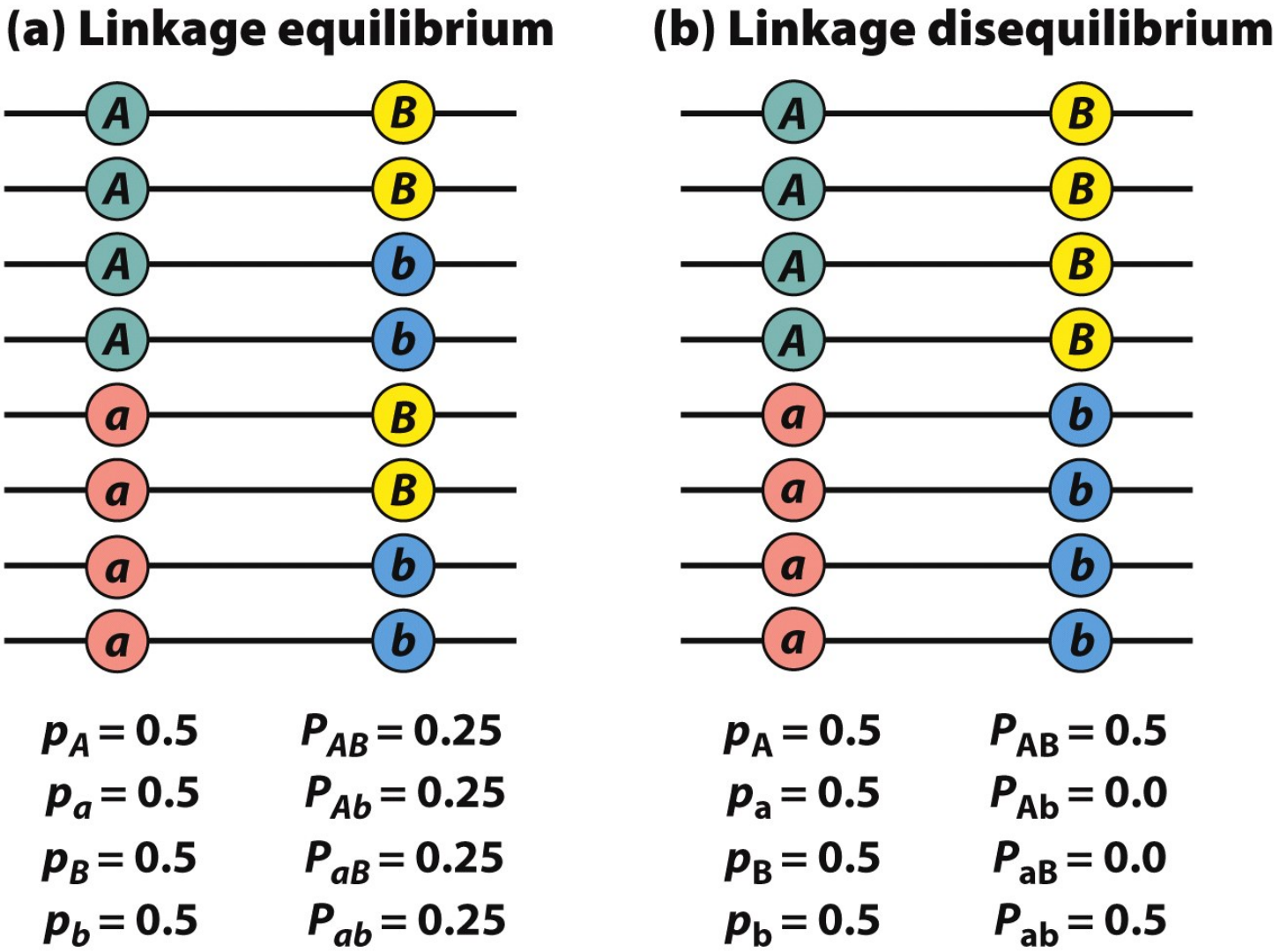
- Haploid genotype = Haplo(id)(geno)type = Haplotype
- Haplotype = the multilocus genotype of a chromosome



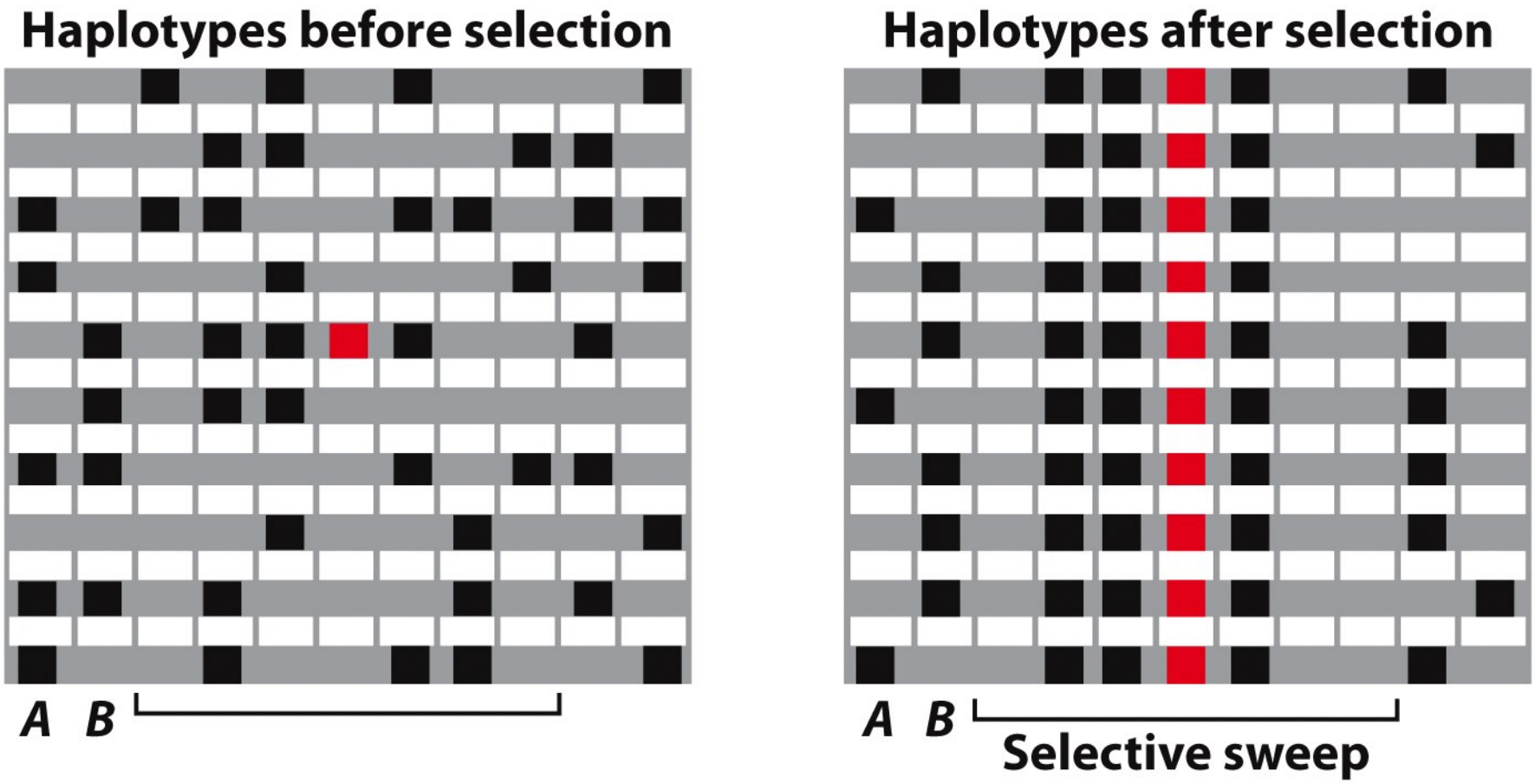
- For two loci, A and B, and 4 alleles, A, a, B, and b



Linkage disequilibrium is the nonrandom association between two loci



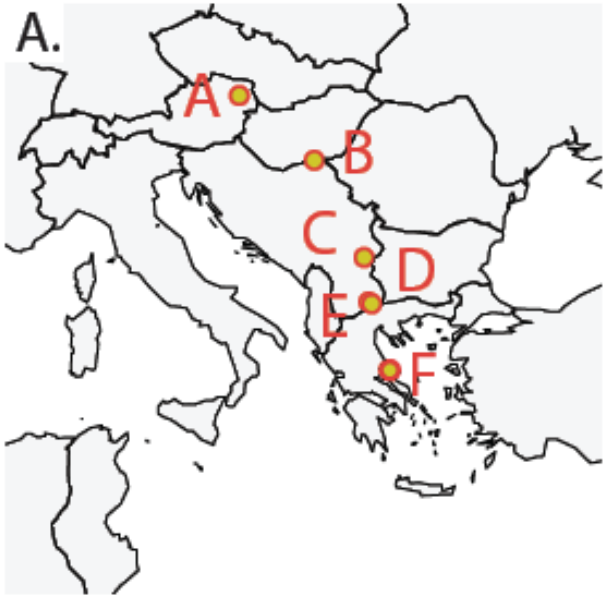
Positive selection leaves a distinct signature

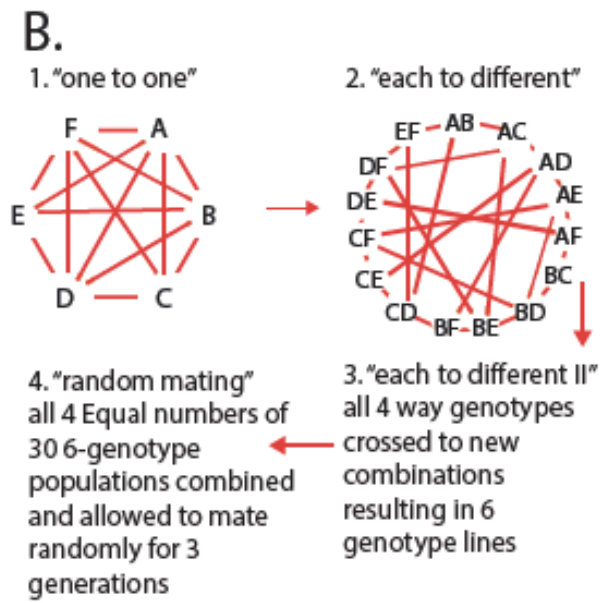


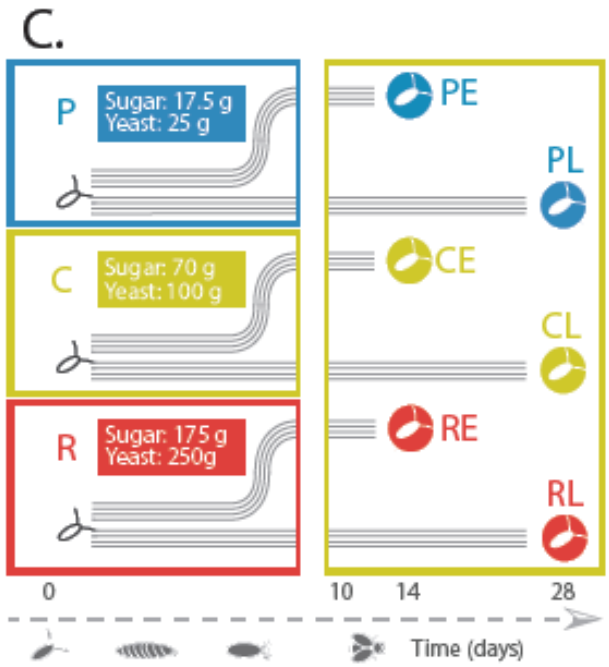
Exploiting variation; what and when?

Experimental evolution in Drosophila melanogaster

- EE well replicated (4x)
- ~170/85 generations of selection (early/late)
- Life history phenotypes measured multiple times throughout EE

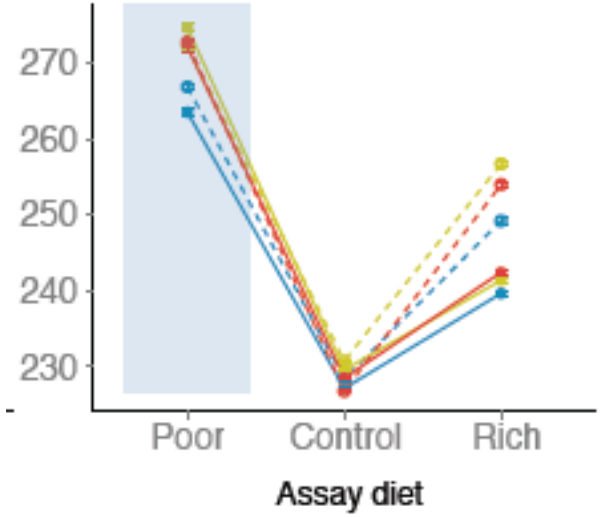




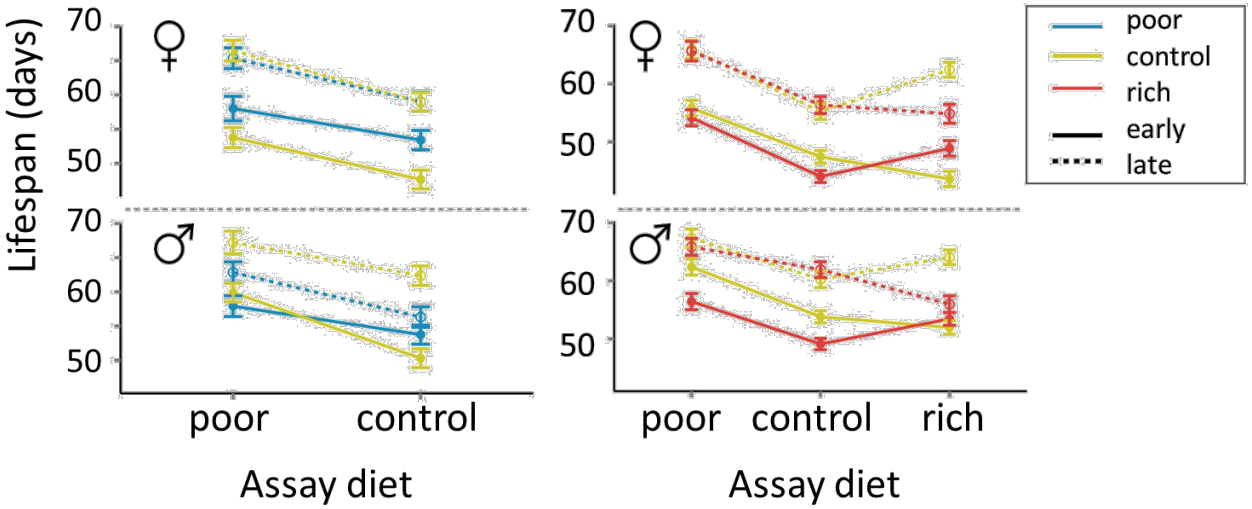


Life history phenotypes

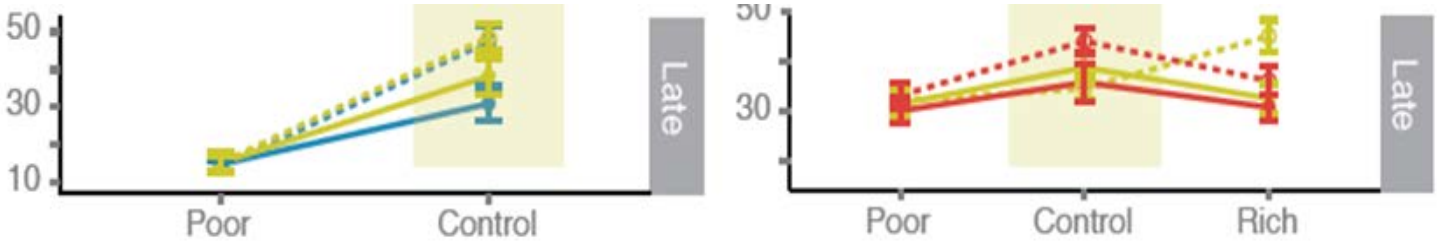
Development time



Lifespan

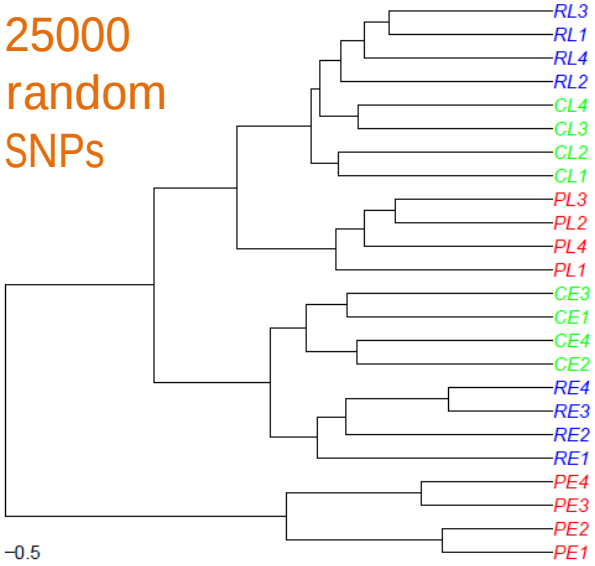


Late fecundity

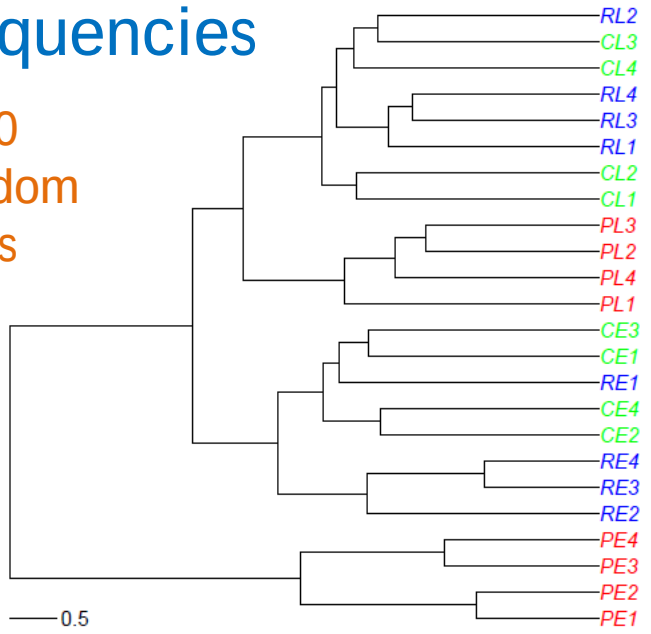


Population genetic differentiation - Euclidean distance over major allele frequencies

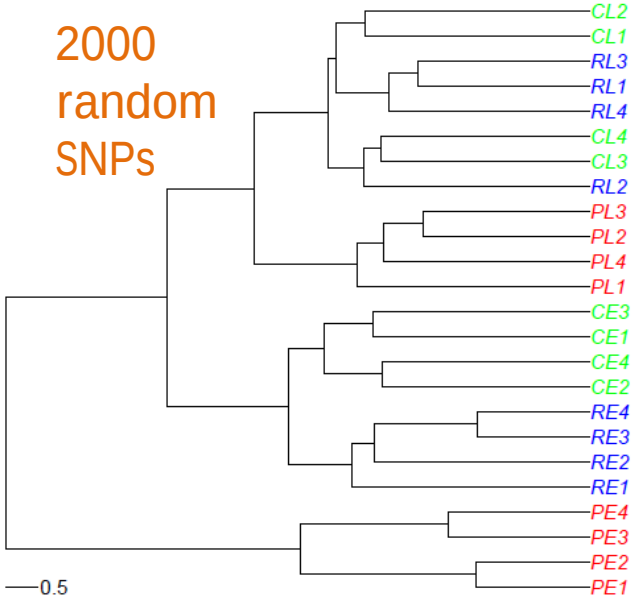
25000 random SNPs



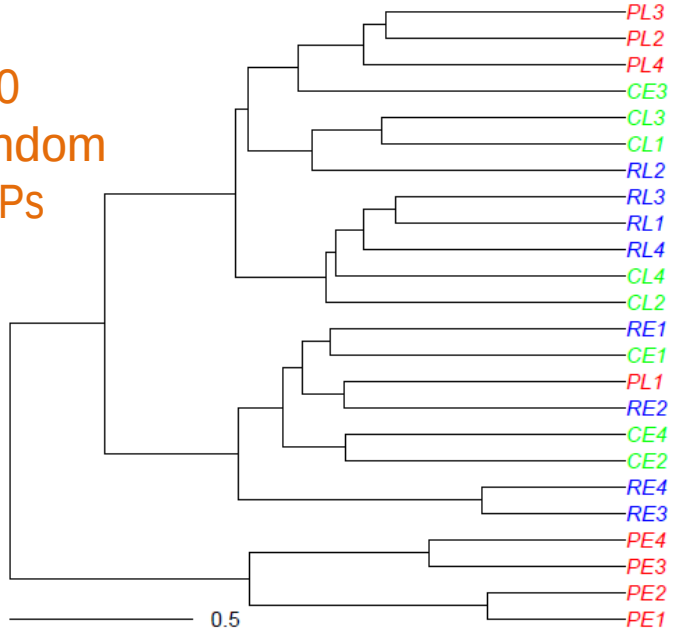
1000 random SNPs



2000 random SNPs

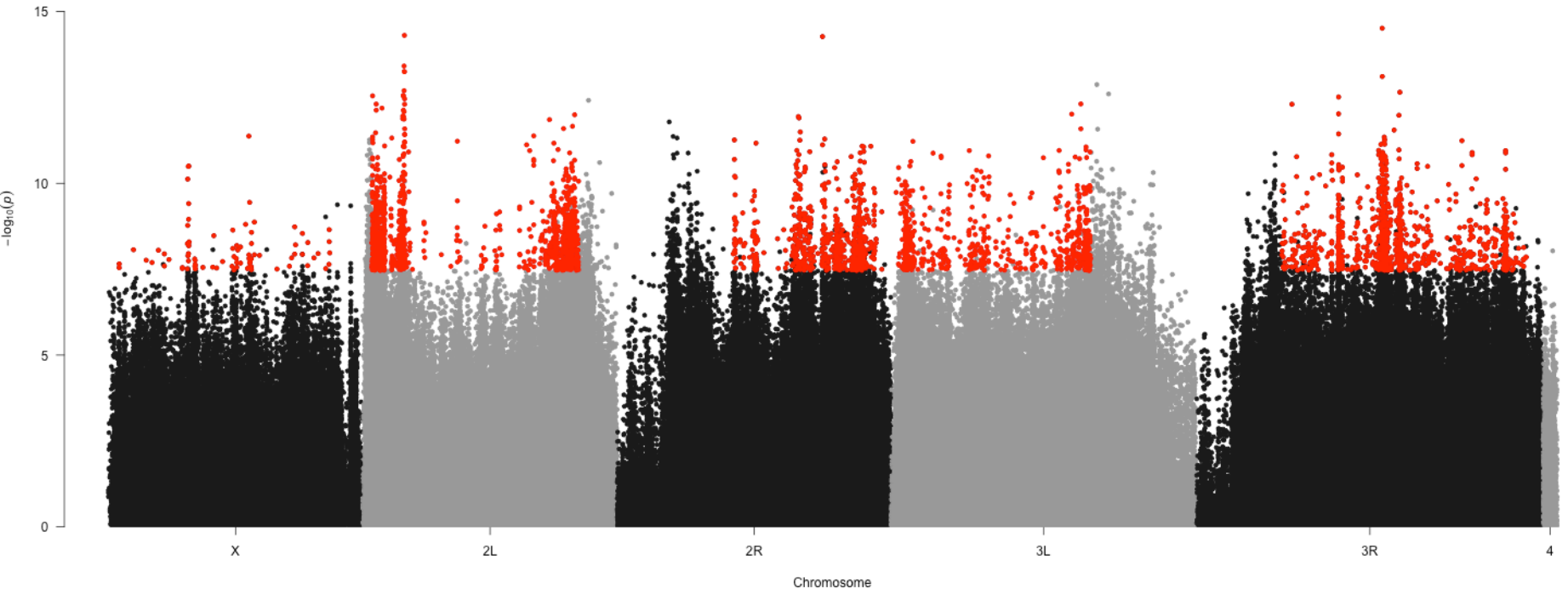


100 random SNPs



Footprints of selection – association with EE

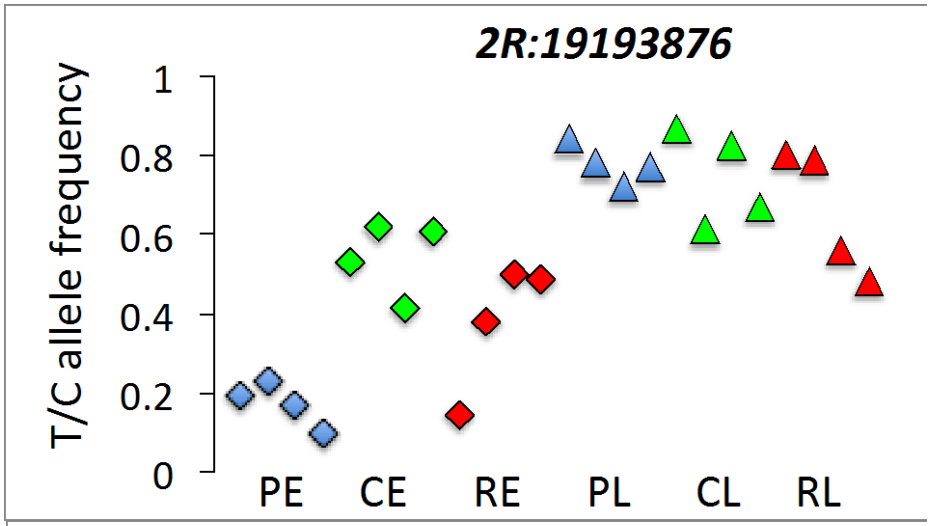
Reproductive age
3698 significant

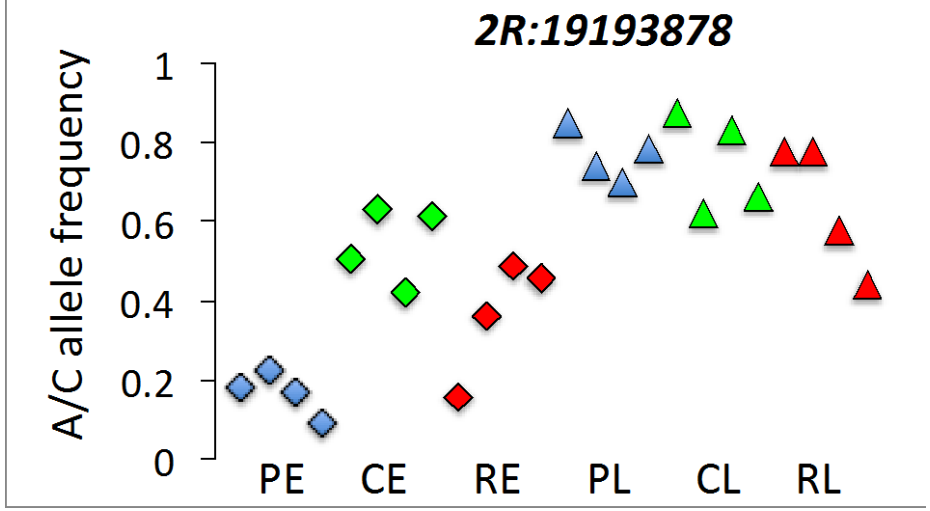


GLMM
glmer(cbind(allele,total-allele)~**diet+repr+diet:repr+(1|EEpop)**,family="binomial"

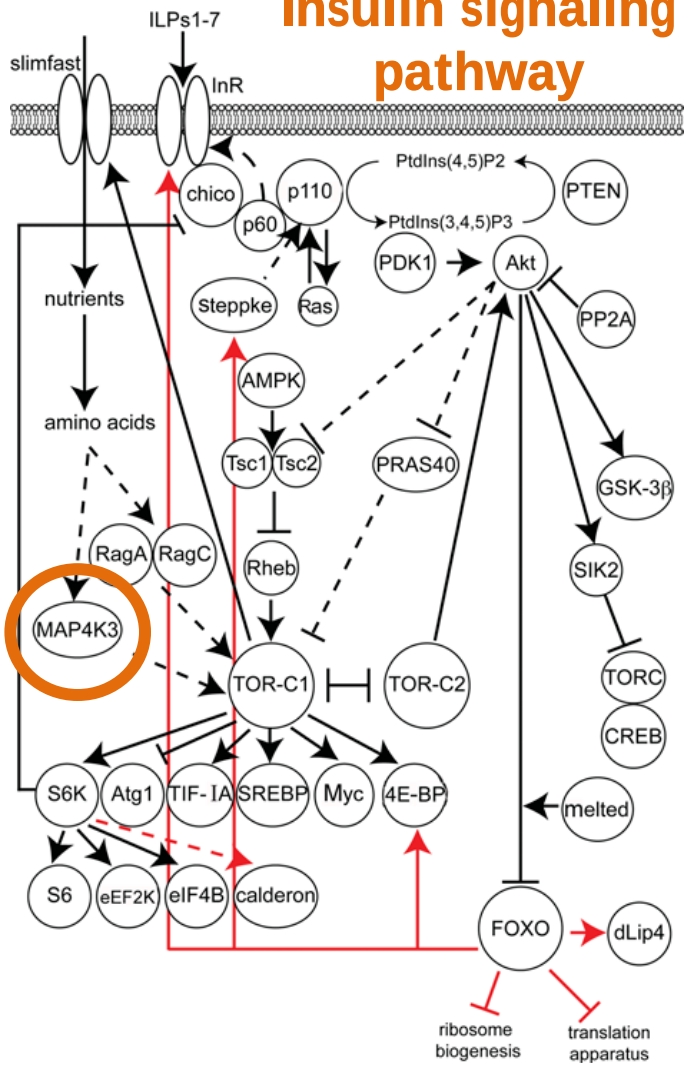
Footprints of selection – an example

SNPs in Happy Hour (*hppy*)





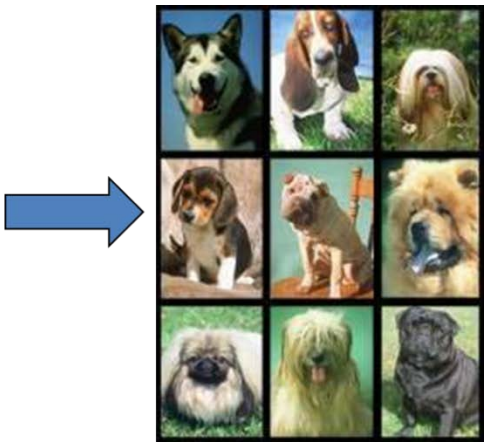
Insulin signaling pathway

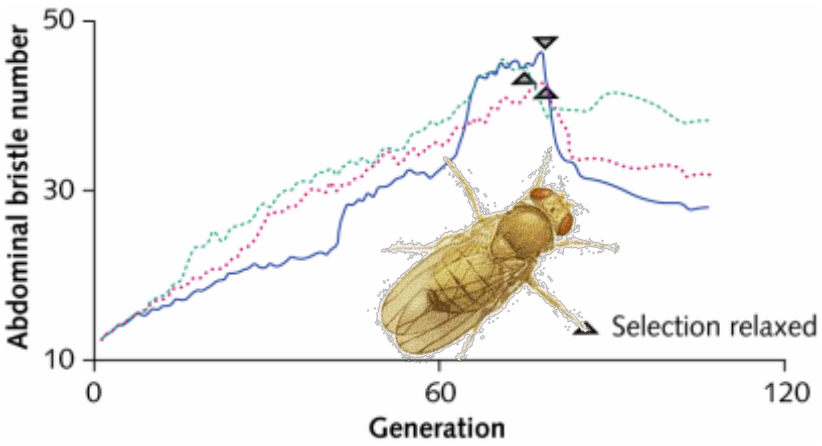


Exploiting variation; what and when?

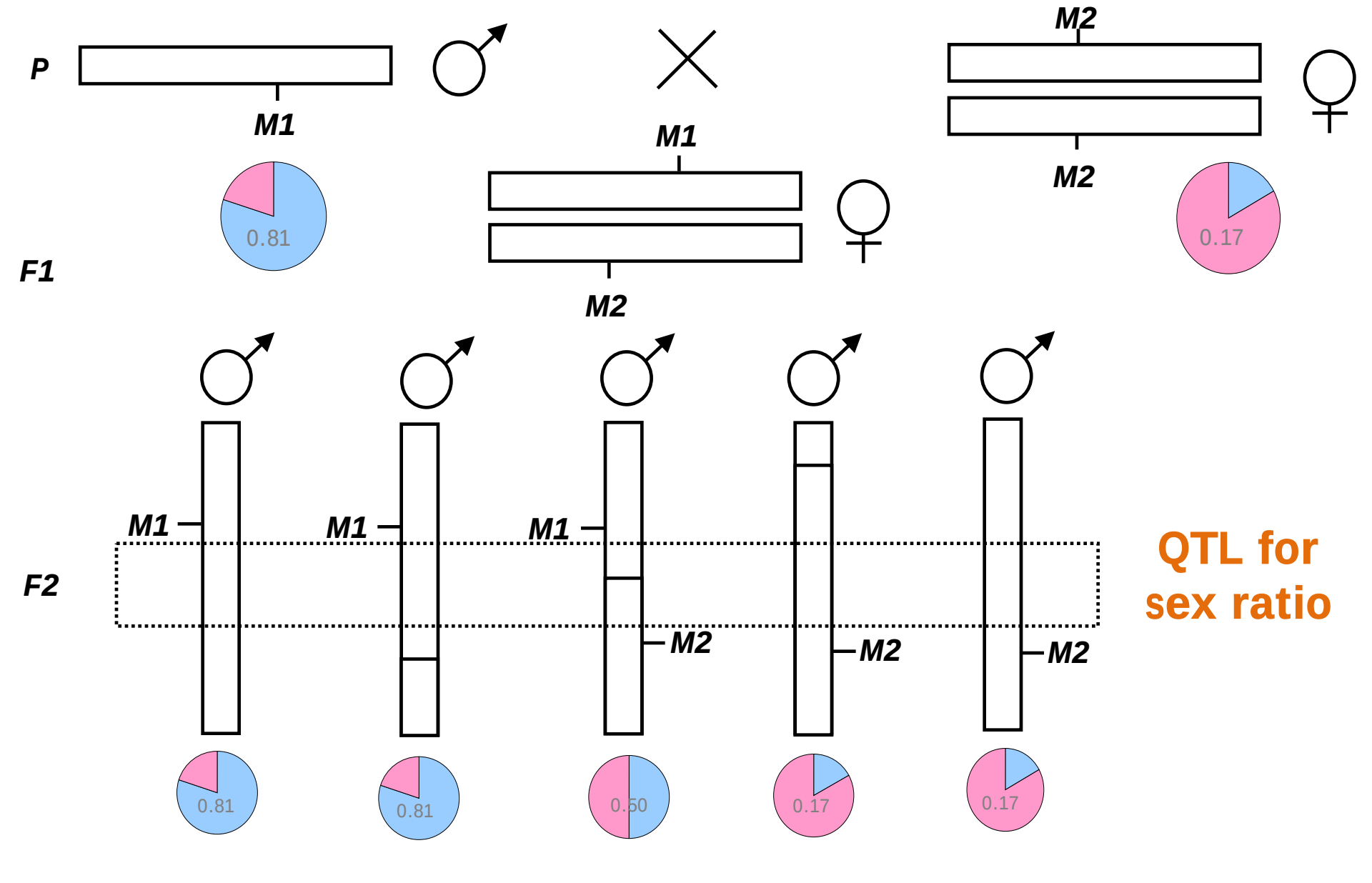
Artificial selection and QTL mapping

- Artificial selection is arguably the most straightforward to accomplish
- Response depends on heritability and selection differential (breeders equation, $R = h^2 \cdot S$)
- Proven method to go well-beyond original phenotypic variance in starting population
- Ample numbers of genetic markers facilitated QTL-mapping



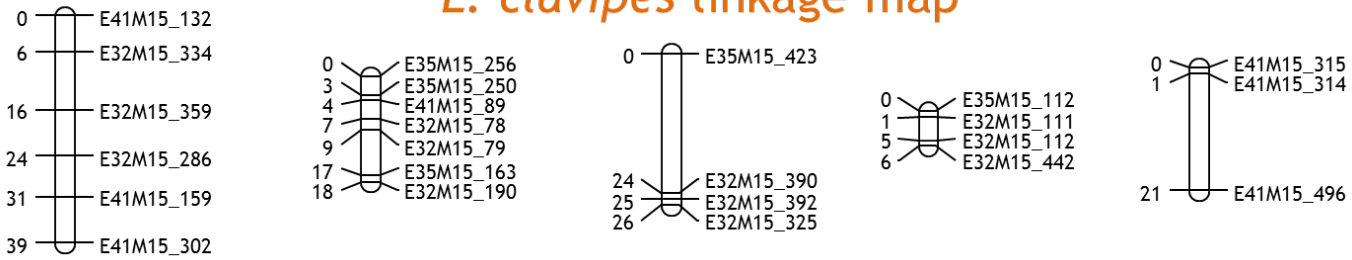


Sex ratio variation in *Leptopilina clavipes*

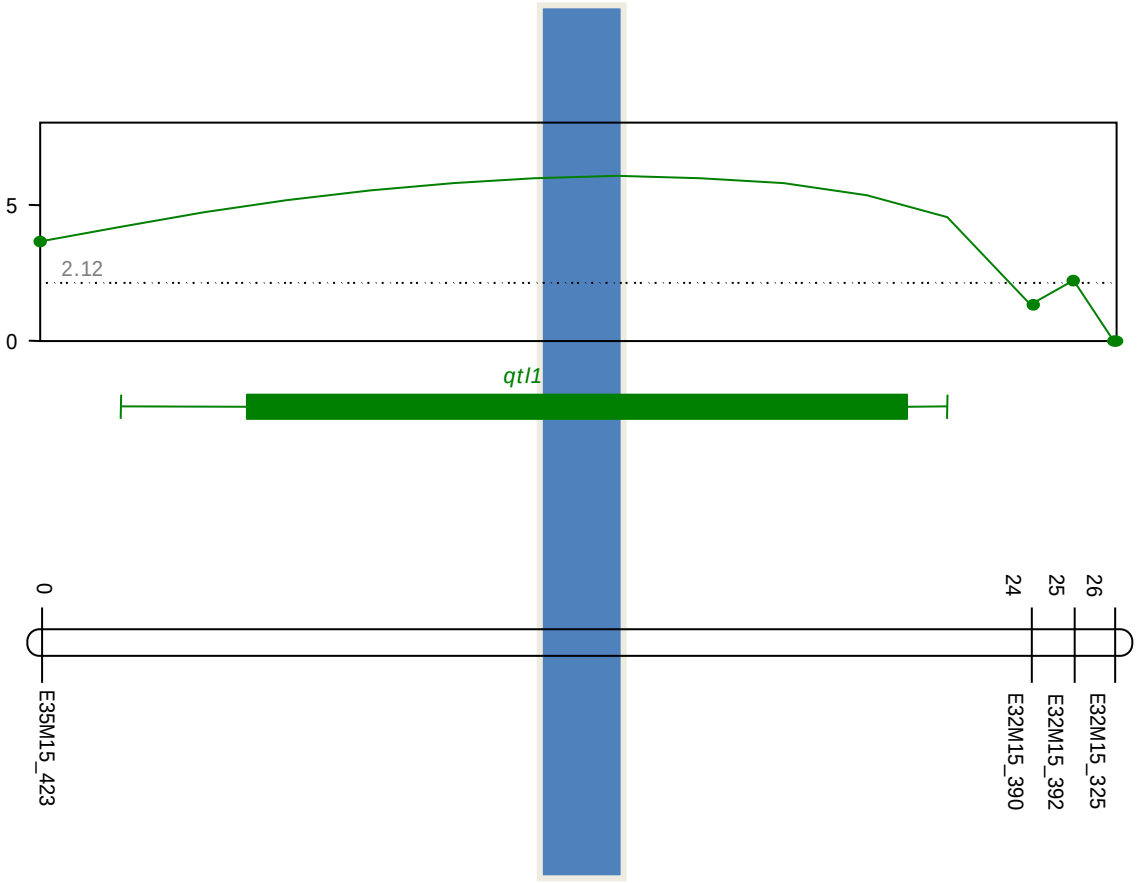


Sex ratio variation in *Leptopilina clavipes*

L. clavipes linkage map



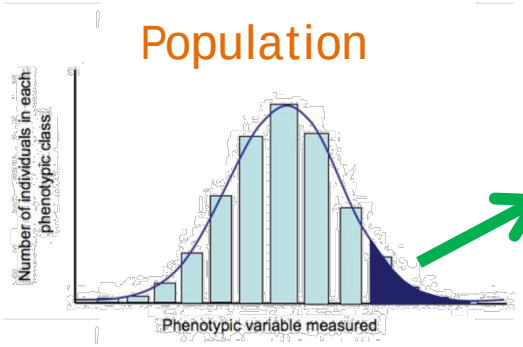
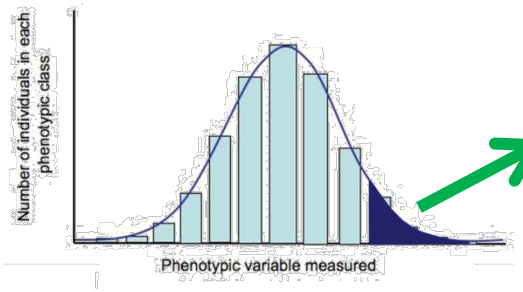
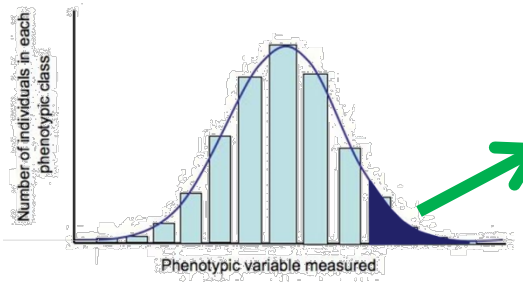
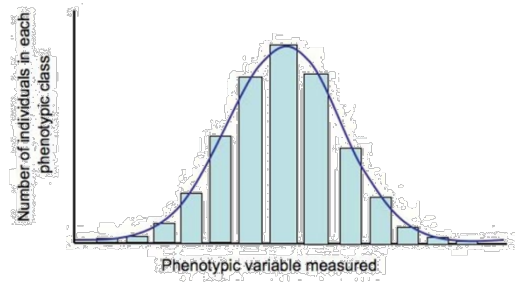
- APLP markers
- total distance: 110 cM
- 7 unlinked loci
- not saturated



Linkage group 3 (CIM mapping)

qtl1 explains 40.1% of phenotypic variance; large effect!

Outlook for genetic improvement

	Population	Information for selected individuals	Which traits/phenotypes?
Artificial selection		I. Phenotype ***(P)***	• Any
Marker Assisted Selection (MAS)		I. Phenotype ***(P)*** II. Genotype ***(G)*** for selected (QTL) markers	• Any • Traits with few QTL, e.g. resistance
Genomic selection (GS)		II. Genotype ***(G)*** for genome-wide markers	• Composite traits • Difficult to measure • Expensive to measure • …
Reference population (RS)		I. Phenotype ***(P)*** II. Genotype ***(G)*** for genome-wide markers	

Take home messages

- Genetics is central to the understanding of evolution and the generation of biodiversity
- There are ample methods to uncover genetic variation for (LH) traits relevant for biocontrol
- Genetic markers (most notably SNPs) are extremely useful for many areas in biology,
- and thus for the WPs in BINGO

Acknowledgements

People

- Jelle Zandveld, Agnieszka Doroszuk, Vicencio Oostra, Maaike de Jong, Joost van den Heuvel, Tina May, Katja Hoedjes, Thomas Flatt, Martin Kapun, Paul Brakefield, Patrícia Beldade, Bart Pannebakker, Leo Beukeboom, Linda Partridge, Bertha Koopmanschap, Frank Becker, Gabriella Bukovinszkine Kiss, Shuwen Xia, Piter Bijma, Martien Groenen, Mario Calus

Funding

NWO Nederlandse Organisatie voor Wetenschappelijk Onderzoek

Unilever

IDEAL ageing

LifeSpan development & ageing

SEVENTH FRAMEWORK PROGRAMME

MiMage

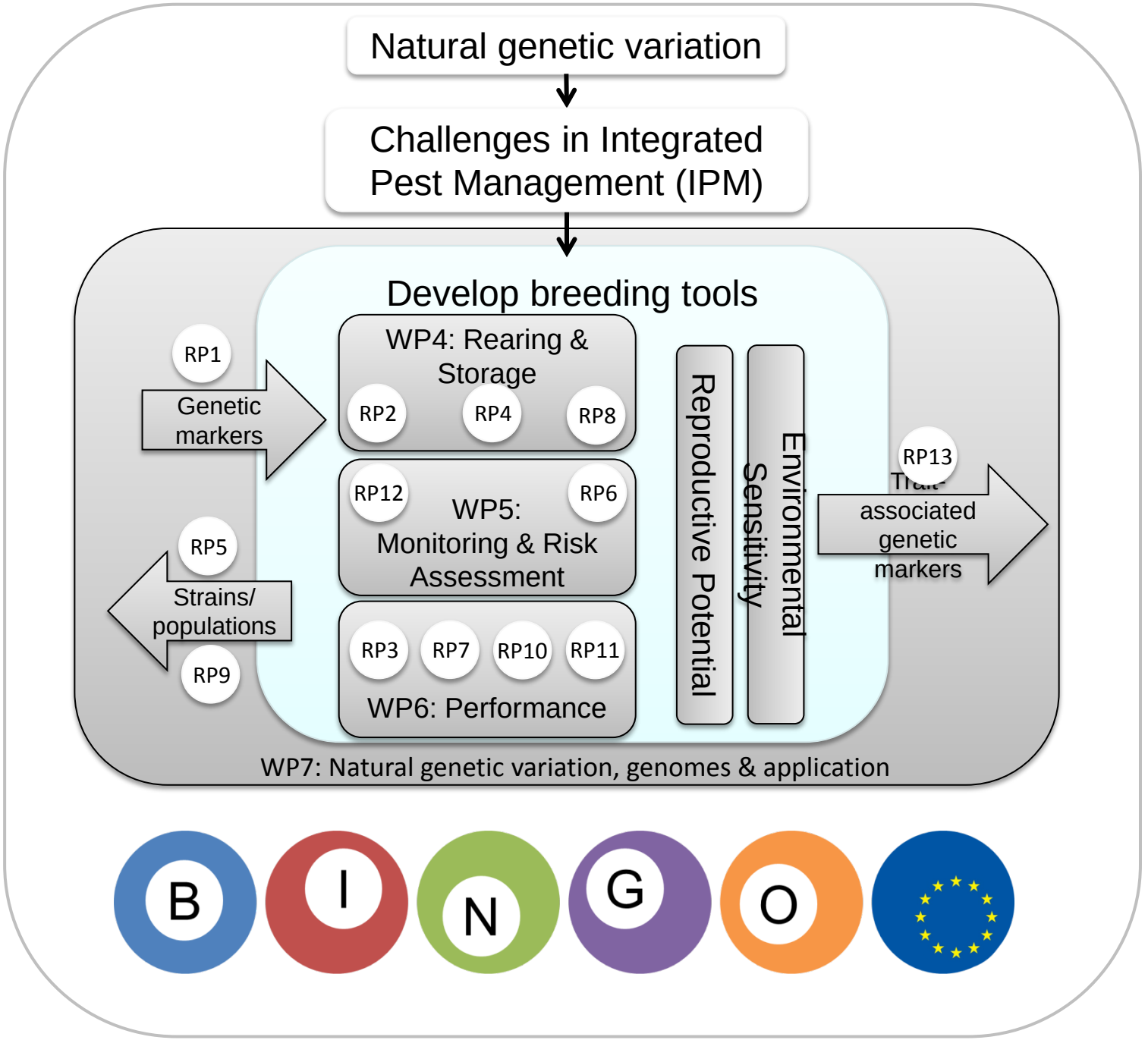
LANG LEVEN

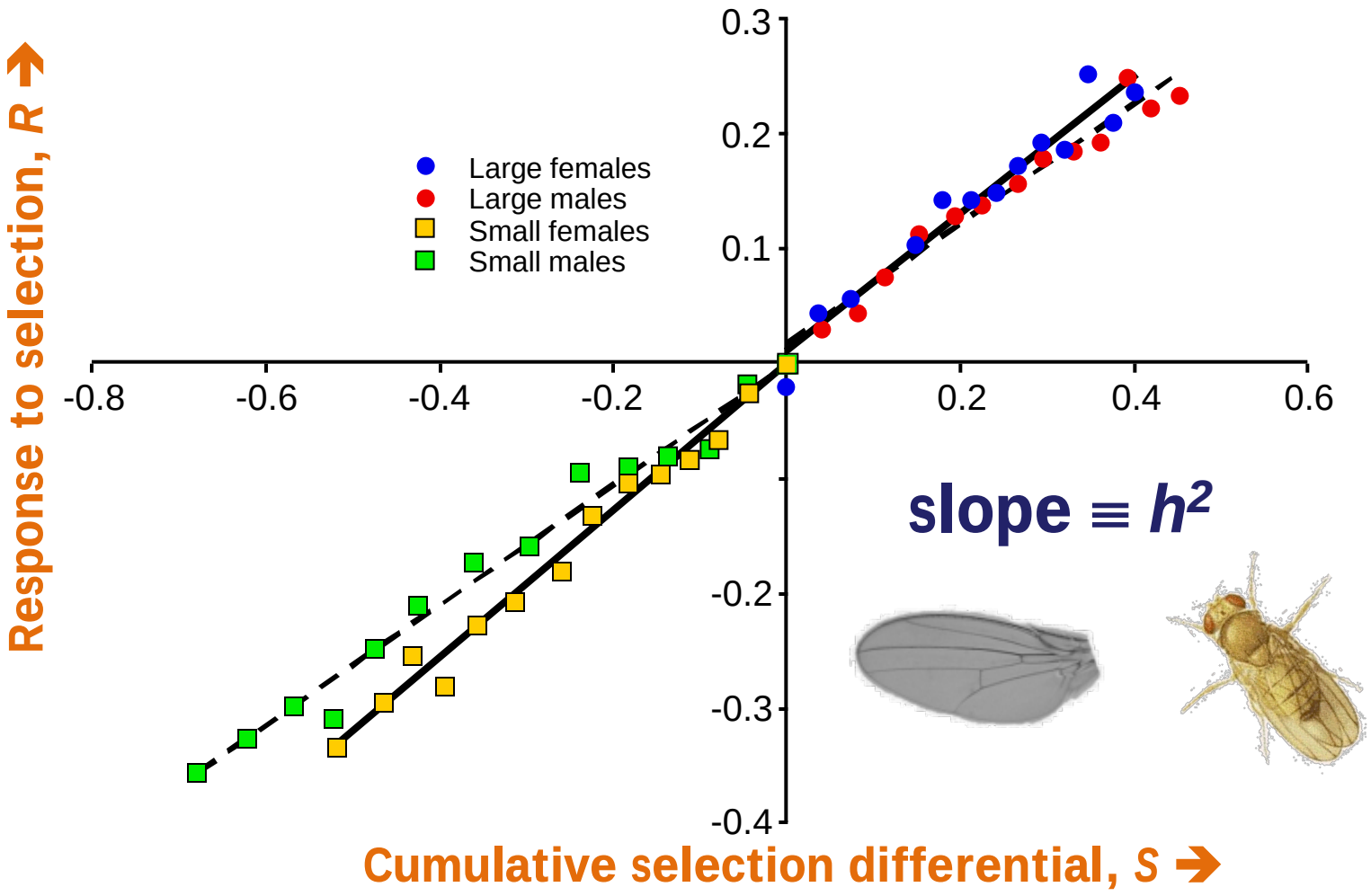
IOP Genomics

BINGO

Thank you

I am happy to take questions and comments


Natural genetic variation
Challenges in Integrated Pest Management (IPM)
Develop breeding tools
RP1
Genetic markers
WP4: Rearing & Storage
RP2
RP4
RP8
RP12
RP6
WP5: Monitoring & Risk Assessment
RP5
Strains/ populations
RP9
RP3
RP7
RP10
RP11
WP6: Performance
Reproductive Potential
Environmental Sensitivity
RP13
associated genetic markers
WP7: Natural genetic variation, genomes & application
B
I
N
G
O


Response to selection, R ➔
Large females
Large males
Small females
Small males
0.3
0.2
0.1
-0.2
-0.3
-0.4
-0.8
-0.6
-0.4
-0.2
0.2
0.4
0.6
slope ≡ h²
Cumulative selection differential, S ➔

Genetic differentiation - F_{ST}

- Unbiased F_{ST} estimation based on Weir & Cockerham (1984) *Evolution*

$$\hat{F}_{ST}^{WC} = 1 - \frac{2\frac{n_1 n_2}{n_1+n_2}\frac{1}{n_1+n_2-2}[n_1\tilde{p}_1(1-\tilde{p}_1)+n_2\tilde{p}_2(1-\tilde{p}_2)]}{\frac{n_1 n_2}{n_1+n_2}(\tilde{p}_1-\tilde{p}_2)^2+\left(2\frac{n_1 n_2}{n_1+n_2}-1\right)\frac{1}{n_1+n_2-2}[n_1\tilde{p}_1(1-\tilde{p}_1)+n_2\tilde{p}_2(1-\tilde{p}_2)]}, \quad (6)$$

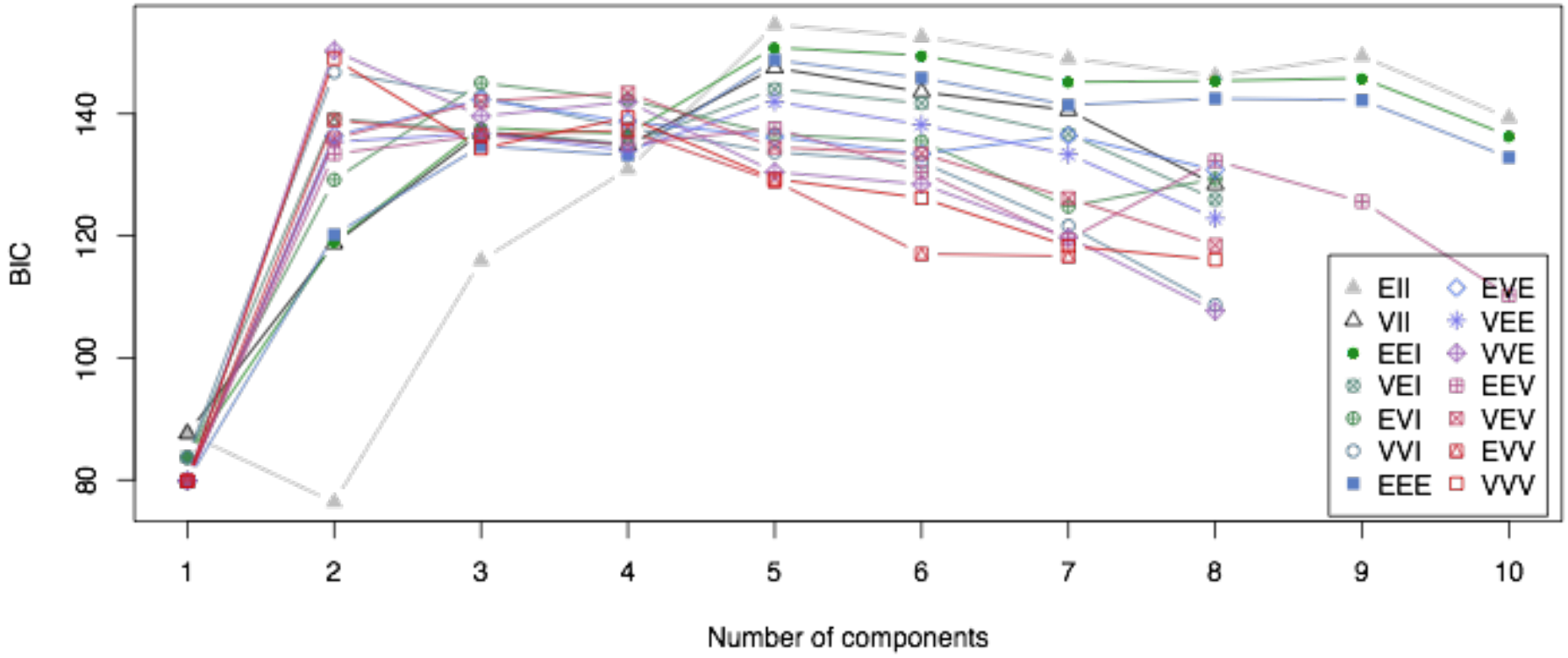
n1 = sample size pop 1
n2 = sample size pop 2
p1 = estimated allele frequency of allele A in pop 1
p2 = estimated allele frequency of allele A in pop 2

Genetic differentiation - PCA

Principal components analysis (PCA)

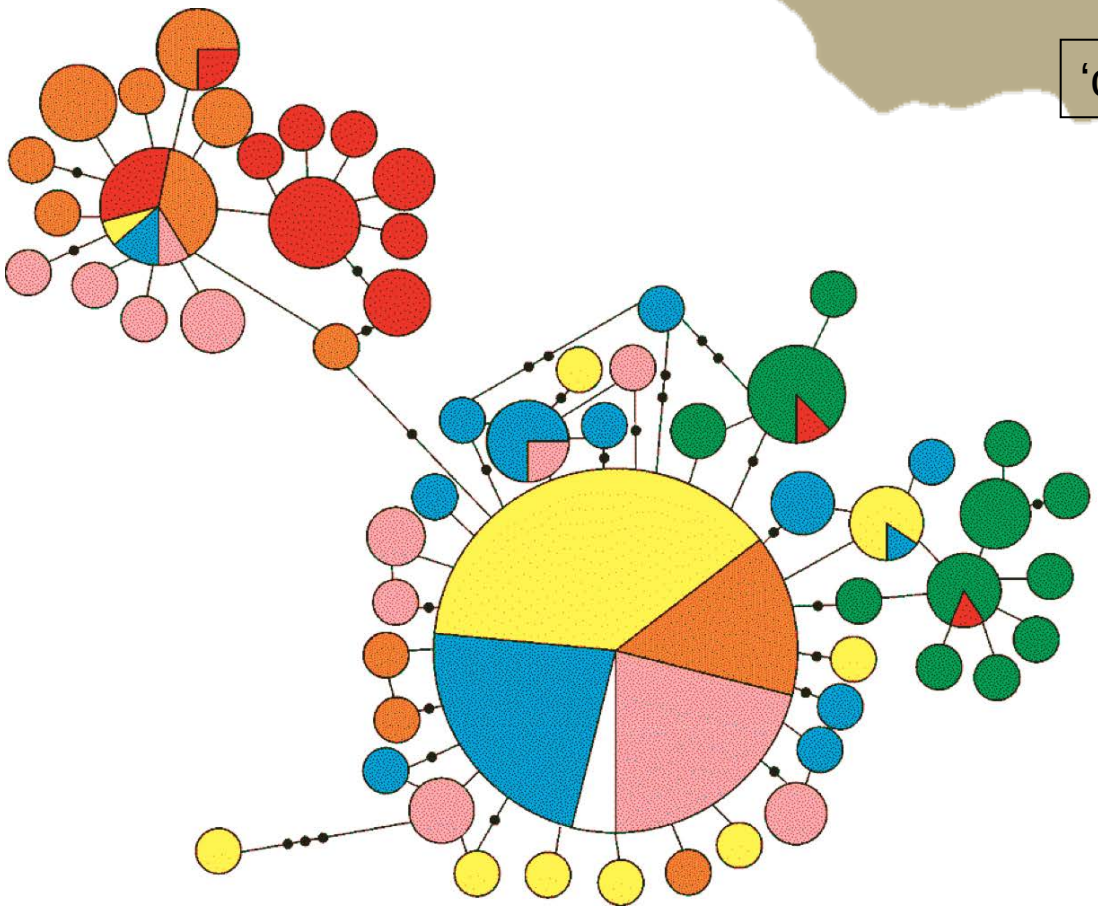
- Based on allele frequencies of intronic SNPs
- using the *R* package *LEA*
- Cluster analysis to identify the optimal number of sub-clusters
  - Based on the best fit (using BIC) of multiple hierarchical models (*R* package: *mclust*)
  - K-means method to identify optimal clustering for *k* clusters (*R* package *kmeans*)

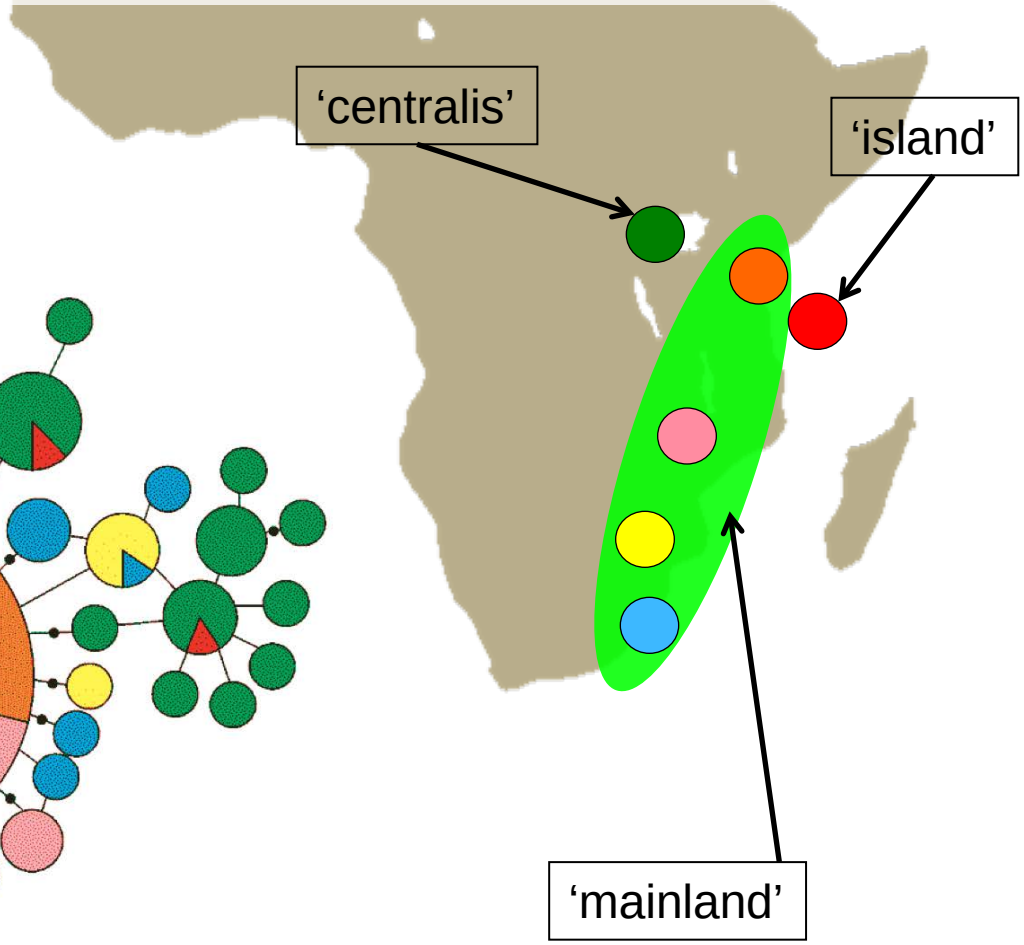
Optimal number of clusters



MtDNA COI

-little geographic structure among ‘mainland’ populations
-more differentiation for ‘centralis’ and ‘island’





Candidate genes for **thermal** adaptation:

***Metabolic** genes*:
Life history and other fitness traits closely linked with metabolic processes

***Pigmentation** genes*:
Wing/body melanization involved in UV protection and regulation of body temperature

***Heat shock protein** genes*:
Cell regulation & stress responses, including thermal stress

Study clinal variation in candidate genes

19 genes putatively involved in **thermal** adaptation:

metabolic genes:

Glycolytic pathway:
Gapdh2
GlyP
Tpi
Treh
UGPase

Lipid pathway:
Desat1
TAG Lipase
LpR
ApolPre
Lipase like
Vitellogenin

Other:
Gdh
Cat

pigmentation genes:

black
Catsup
Ddc
light
yellow

heat shock protein genes:

Hsp23
Hsp60
Hsp68
Hsp83
Hsc70-3
Hsc70-4

In addition, inclusion of genes for which we do not expect clinal variation:

Developmental genes:

Apc, Distalless, Engrailed, Ovo, Wingless

- Important in embryogenesis and throughout development
- No association with thermal adaptation

Methods

Candidate genes
sequence lengths: 500-1500bp
PCRs: pools per 3 individuals
Illumina solexa 75bp sequencing
positions on sequence and allele frequencies for each SNP

analysis:
reconstruction of genotype files (no linkage, assuming HWE)
Clinal variation: regression allele frequencies with latitude
fdist method (Beaumont & Nichols), 4 mainland populations
R-package DetSel (Vitalis *et al*.), pair-wise, parameters sensitivity

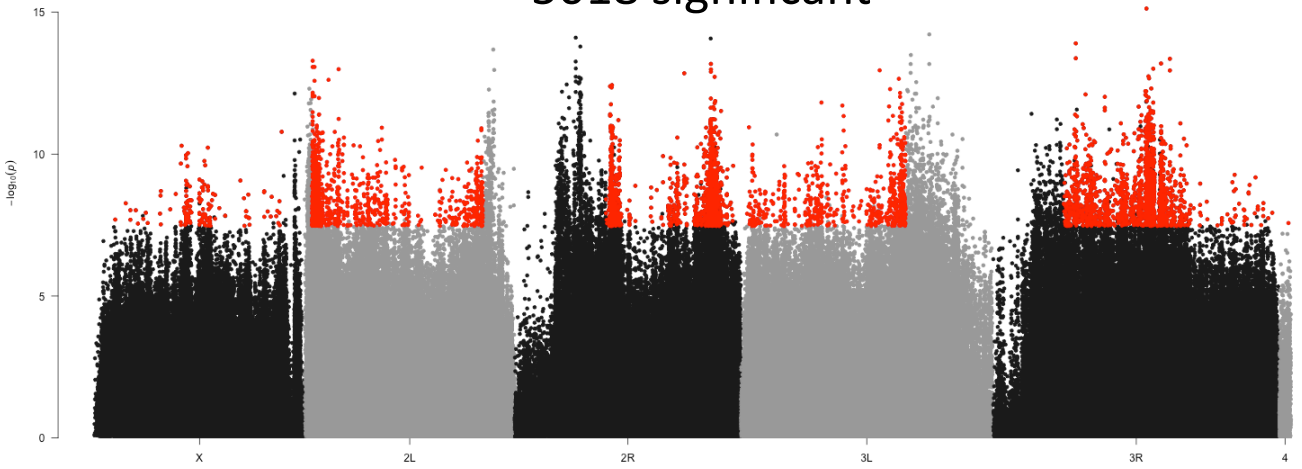
General genetic distance

- Poor lines are very different from the other lines, both for short- and long-lived
- Genetic divergence larger in short-lived comparison of food types (for instance CE vs RE) compared to long-lived
- Effect of lifespan a bit larger than food for C and R lines, but else effects of food and lifespan quite similar

GLMM

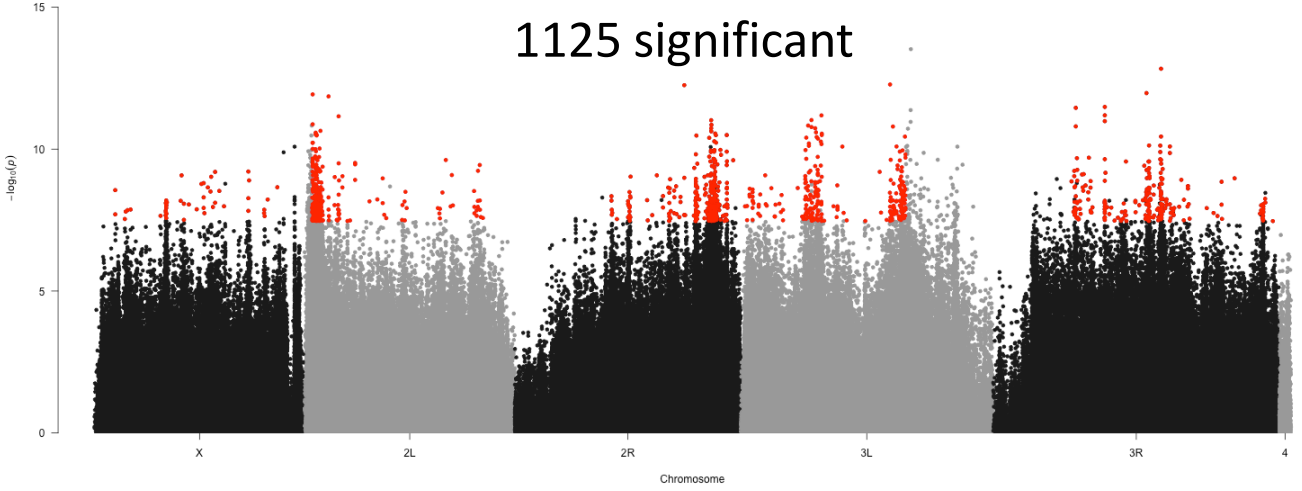
Developmental diet

5018 significant

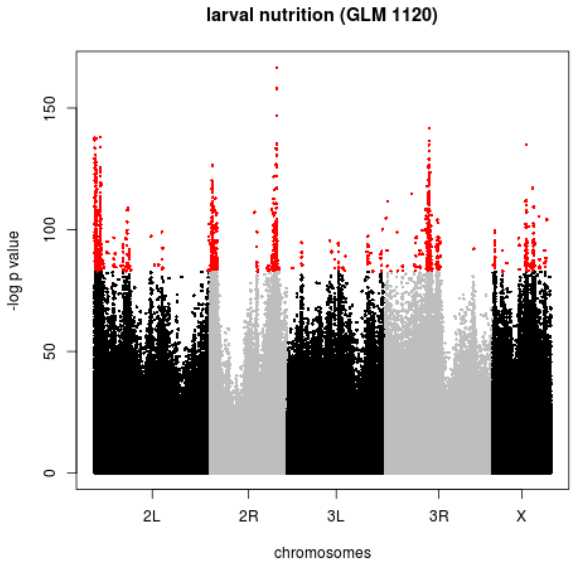


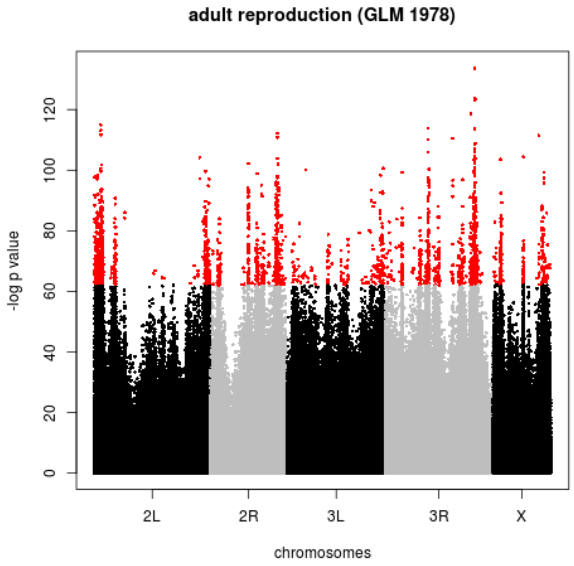
Interaction

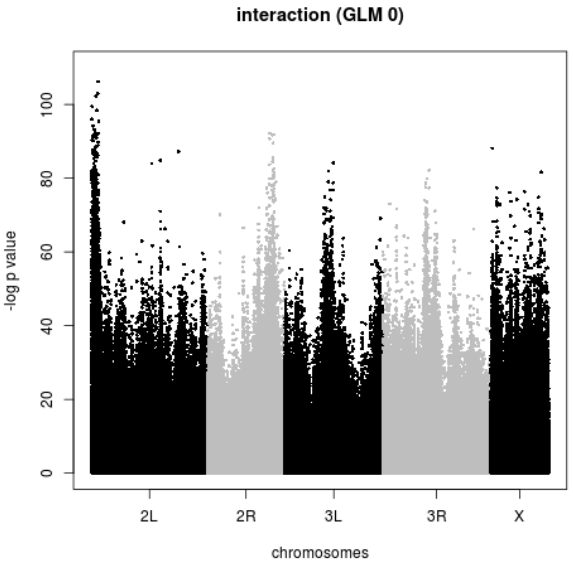
1125 significant

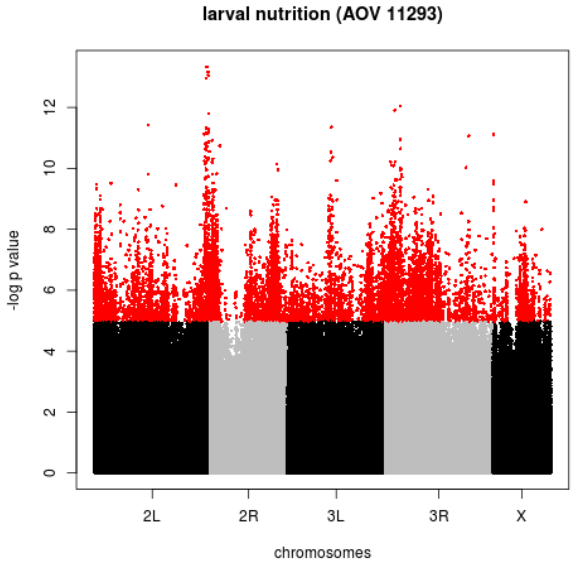


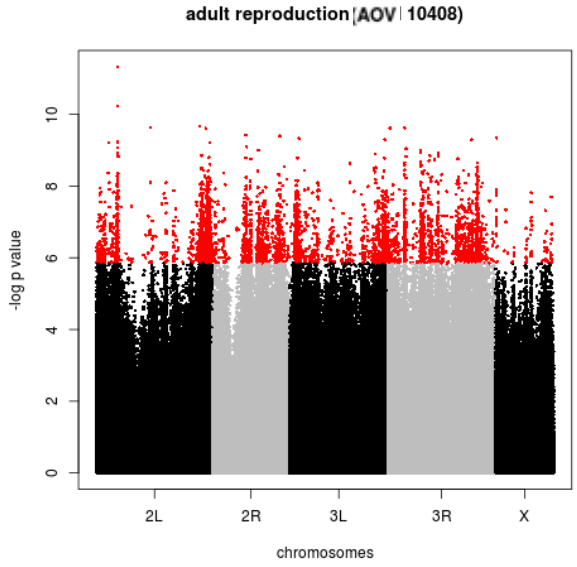
FDR = 0

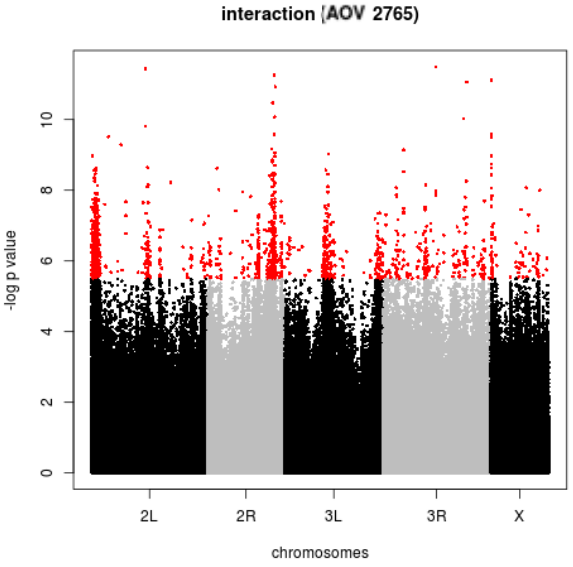

larval nutrition (GLM 1120)
-log p value
0
50
100
150
2L
2R
3L
3R
X
chromosomes

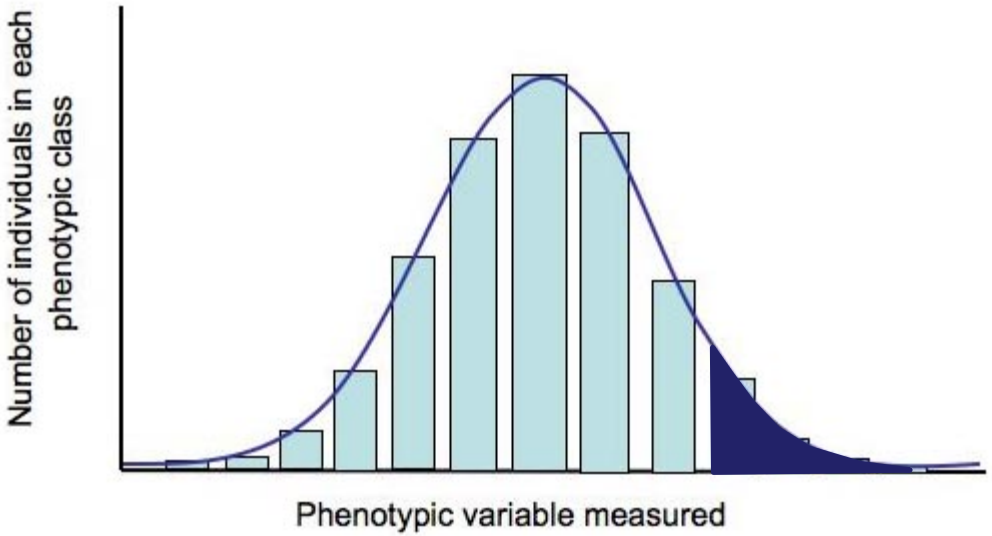

adult reproduction (GLM 1978)
-log p value
0
20
40
60
80
100
120
2L
2R
3L
3R
X
chromosomes

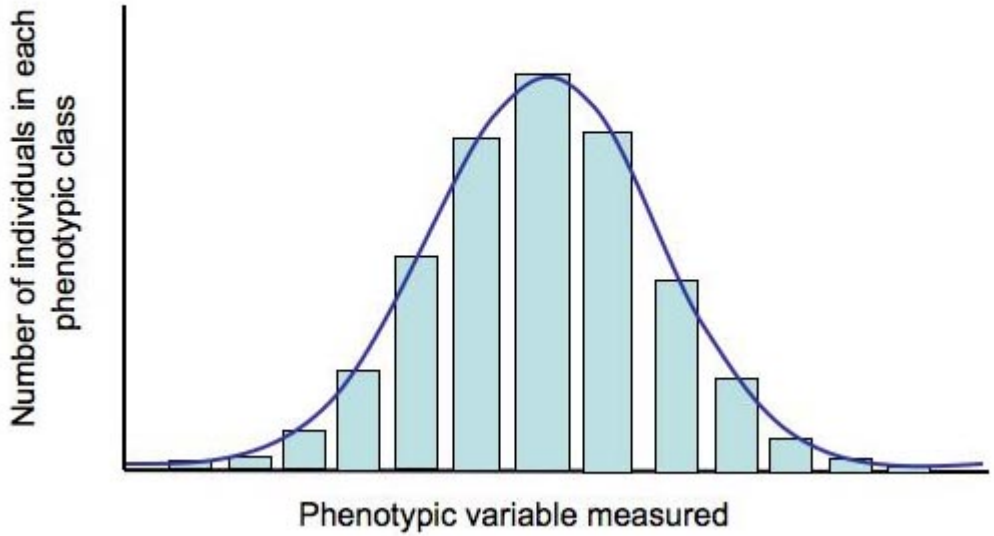

interaction (GLM 0)
-log p value
0
20
40
60
80
100
2L
2R
3L
3R
X
chromosomes


larval nutrition (AOV 11293)
-log p value
0
2
4
6
8
10
12
2L
2R
3L
3R
X
chromosomes


adult reproduction (AOV 10408)
-log p value
0
2
4
6
8
10
2L
2R
3L
3R
X
chromosomes


interaction (AOV 2765)
-log p value
0
2
4
6
8
10
2L
2R
3L
3R
X
chromosomes


Number of individuals in each phenotypic class
Phenotypic variable measured


Number of individuals in each phenotypic class
Phenotypic variable measured