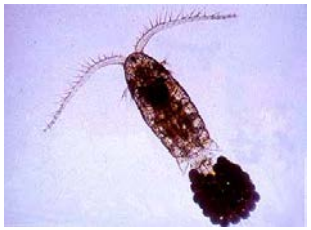
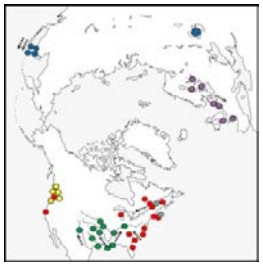
Pheno-genomics of invasive species populations: ongoing studies in *Harmonia axyridis* and *Drosophila suzukii*

Phéno-génomique des populations d'espèces envahissantes: études en cours chez *Harmonia axyridis* et *Drosophila suzukii*

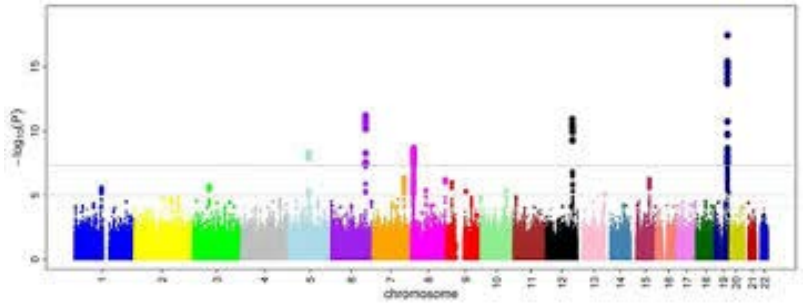
CBGP, 11/12/2018

A number of studies show substantial contemporary evolutionary changes in invasive populations

- Quantitative traits (fecundity, dispersal, adaptation to new environmental features,…)


- Genomic variation (non-random)



→ New approaches coupling phenotypic and genomics data (“pheno-genomics”)

Studies of quantitative traits in association with genomics data (pheno-genomic) at CBGP

Part I – Studies on laboratory populations (E&R): female body weight and generation time in *Harmonia axyridis* and adaptation on host-fruits in *Drosophila suzukii*

Part II – Studies on natural populations : worldwide adaptation routes in *Harmonia axyridis* and *Drosophila suzukii*

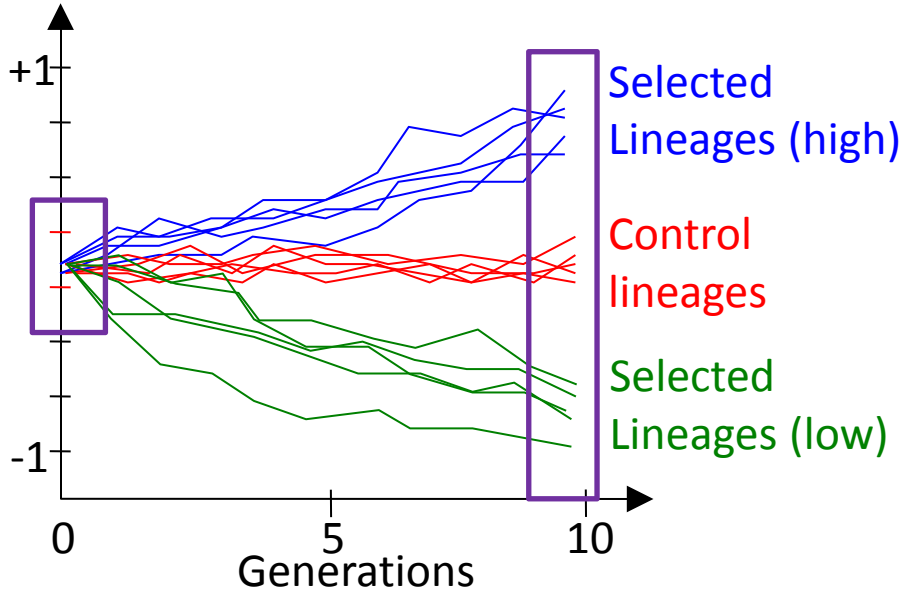
Part III - Proof of concept on a pilot trait: color polymorphism in natural (and laboratory) populations of *Harmonia axyridis*

Part I - Studies on laboratory populations (E&R): female body weight and generation time in *Harmonia axyridis* - adaptation on fruit in *Drosophila suzukii*

E&R experiments in a context of biological invasion : laboratory experiments to mimic adaptive shifts that may occur during invasion for some traits known to be associated with invasiveness

Main goals: - Measuring the phenotypic responses of both the trait under selection and other life-history traits

- Provide information on the genetic architecture of these traits (cf. coupling with NGS sequencing)

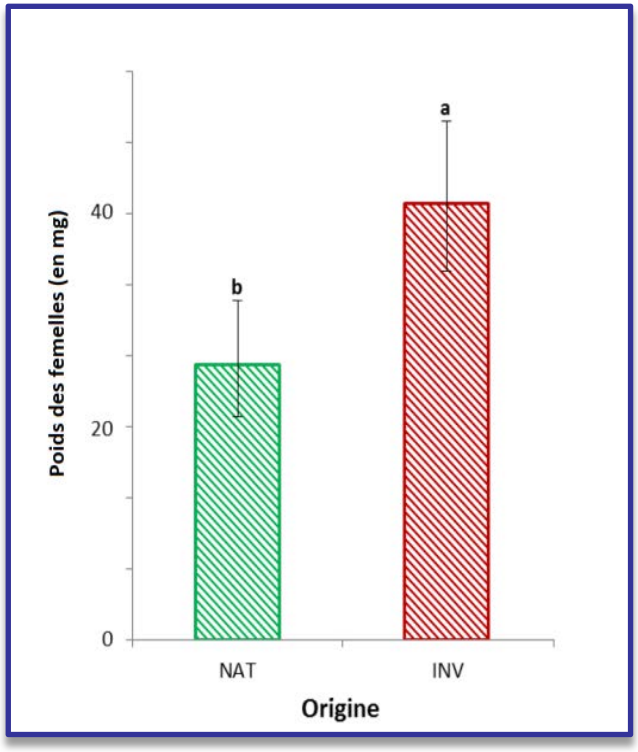


- Measure the <u>phenotypic response</u> (amplitude, speed and associated traits shifts)

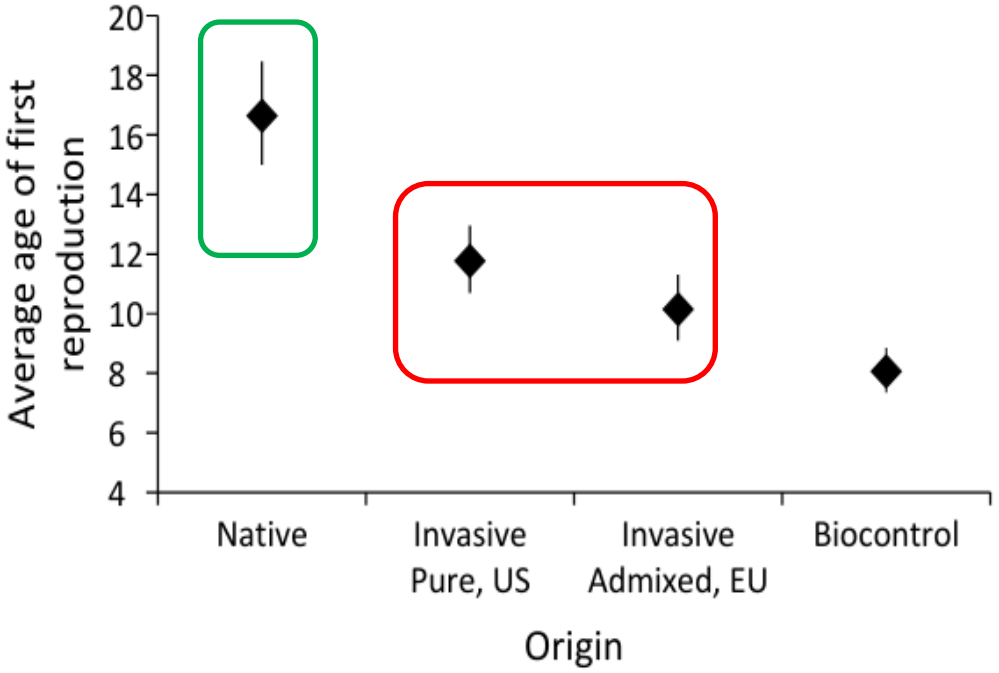
- Analyze the <u>genomic basis</u> of the phenotypic changes in evolved populations with genomic tools (NGS + BAYPASS)

Harmonia axyridis
(cf. common garden experiments)

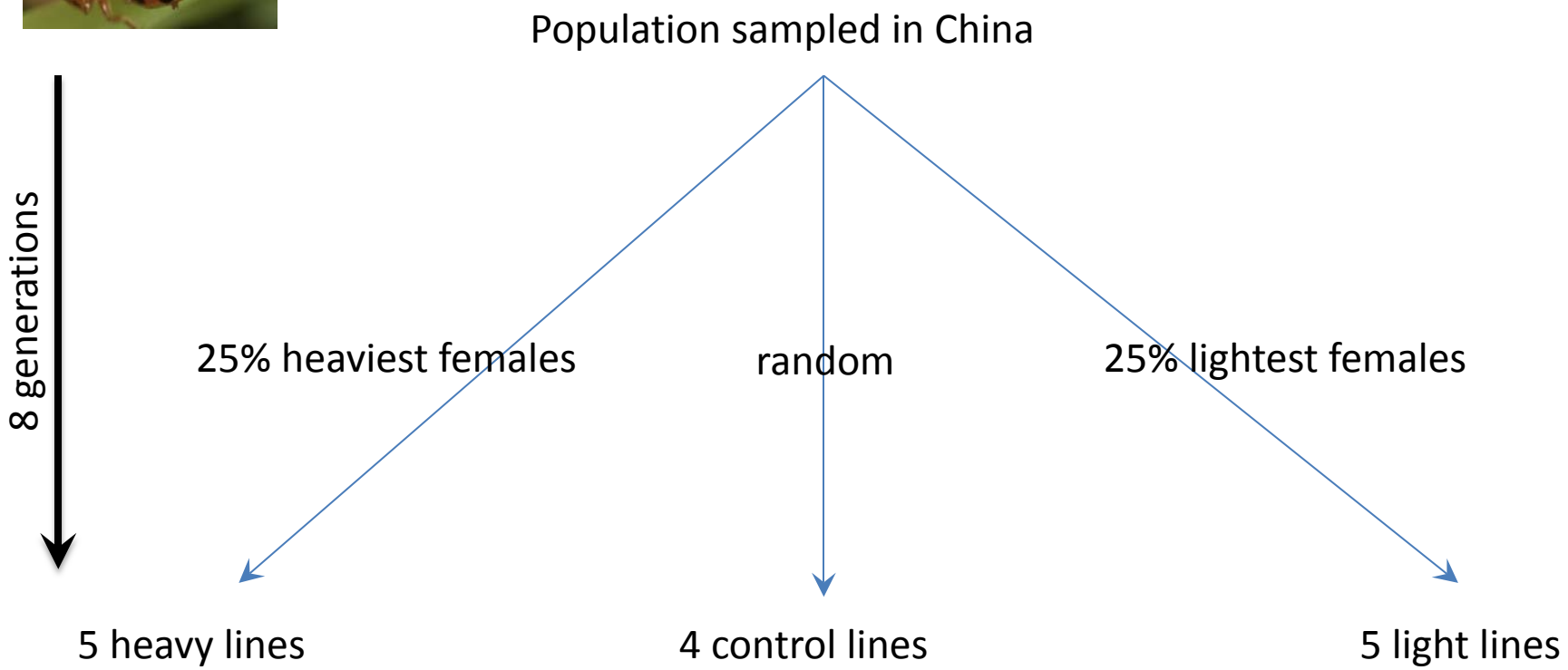
Female body weight



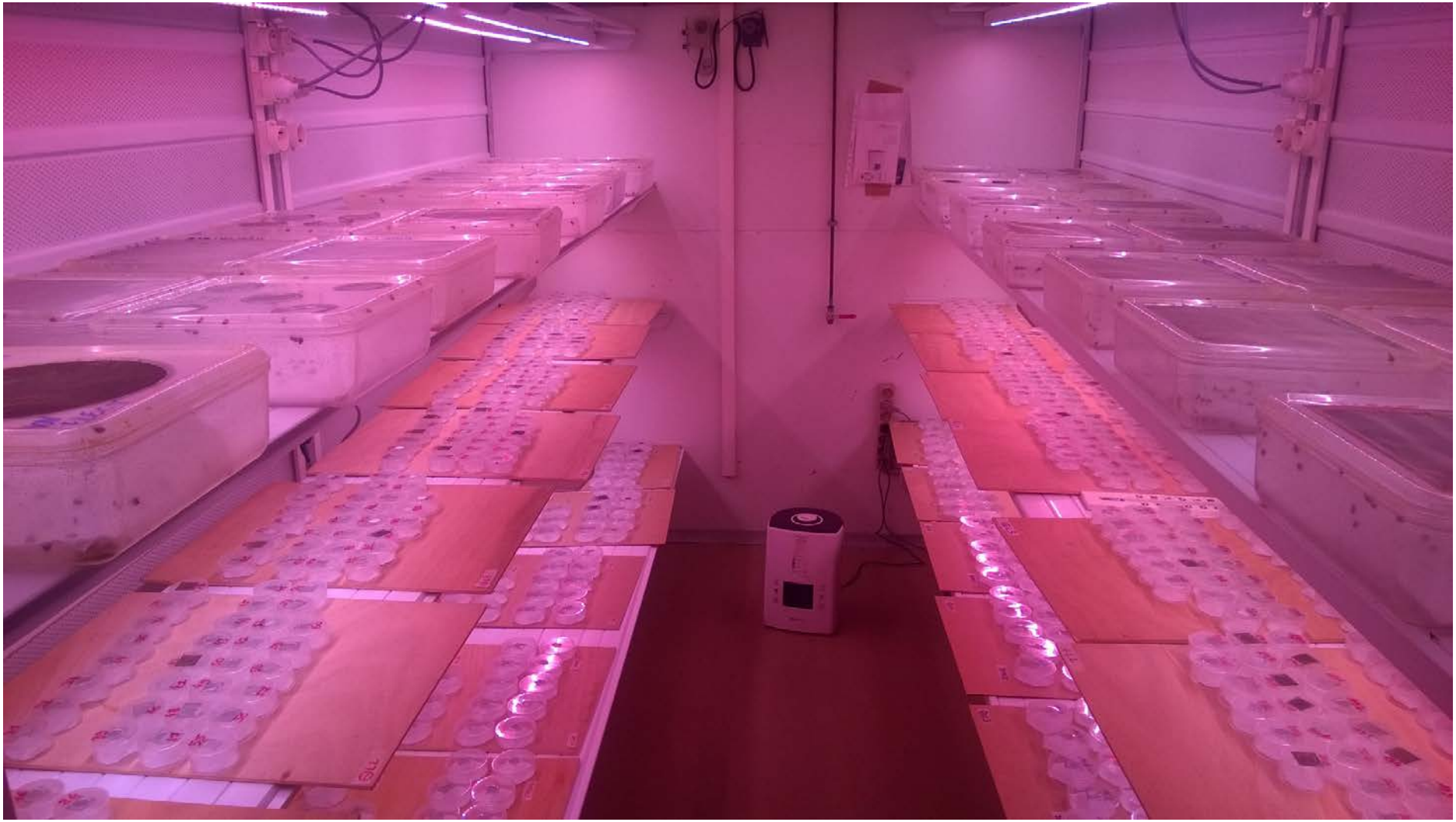
Generation time (age at first reproduction)



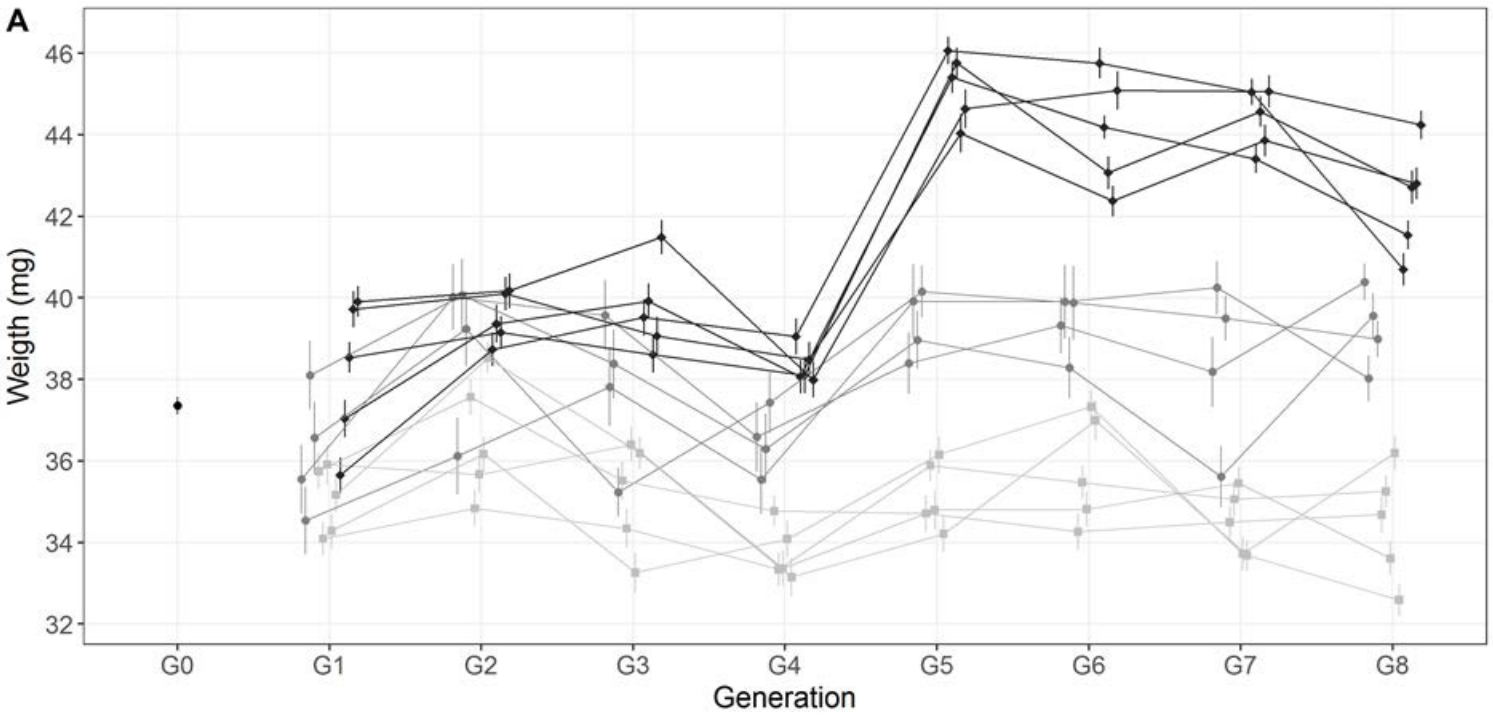
Female weight: experimental protocol – divergent selection



Nb: Each generation = 200 females and 200 males for each line (sub-pop) + Males are taken randomly



Evolution of female body mass



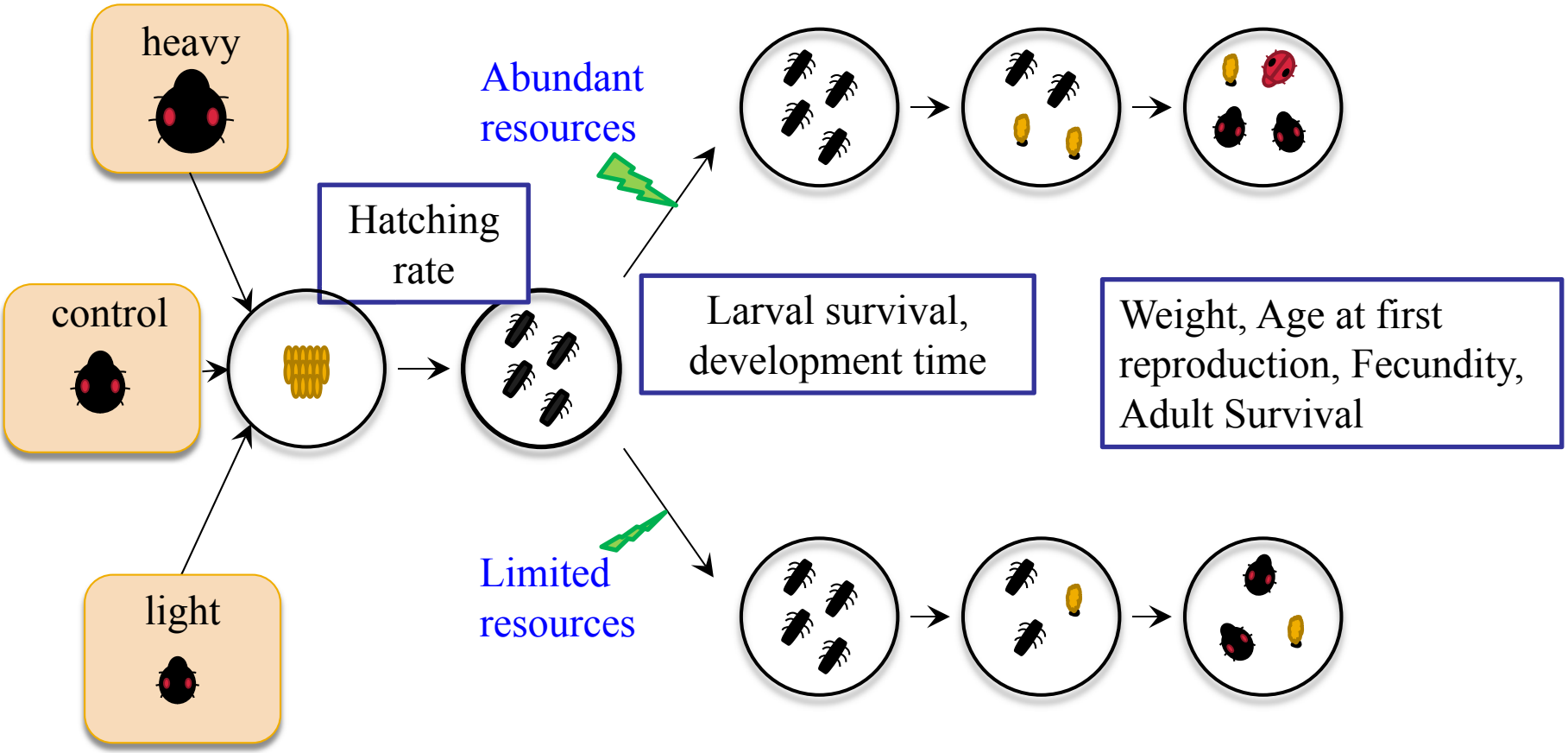
Heavy lines

Control lines

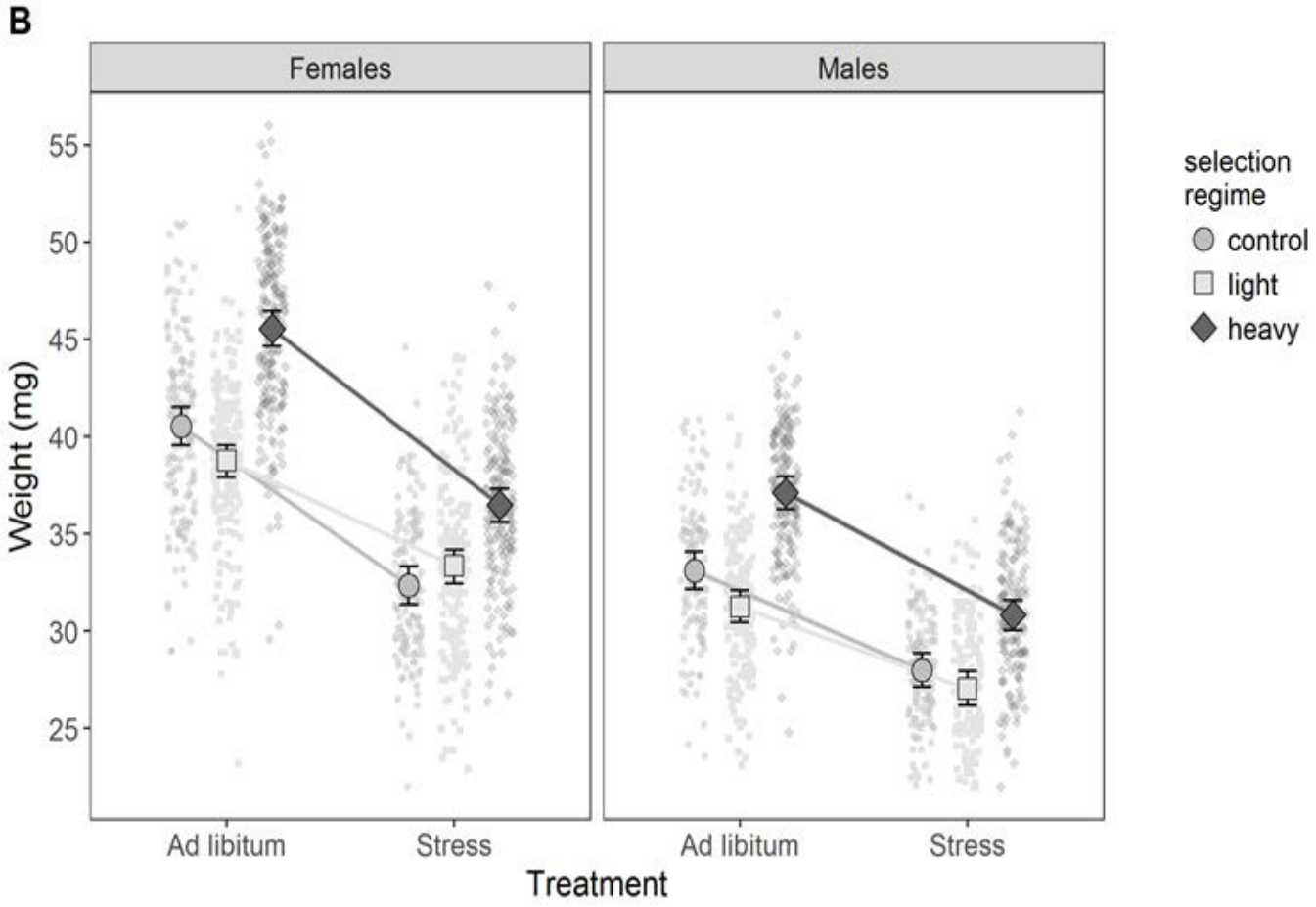
Light lines

→ *Some shifts have occurred during our experimental selection*

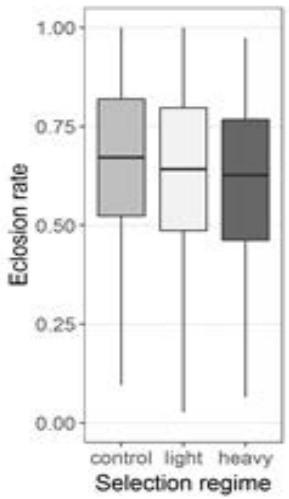
Final phenotyping: body weight and other juvenile and adult traits

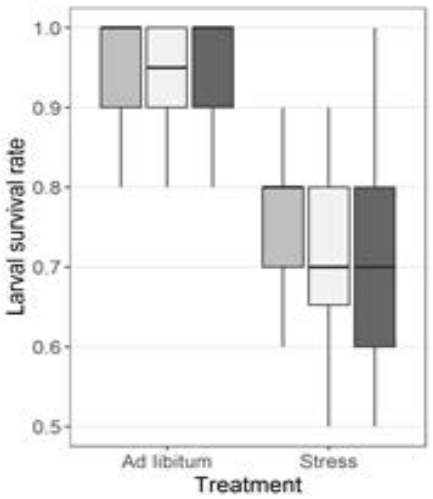


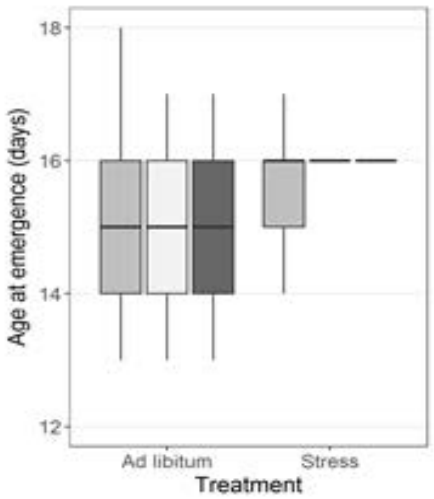
Final phenotyping: adult body weight

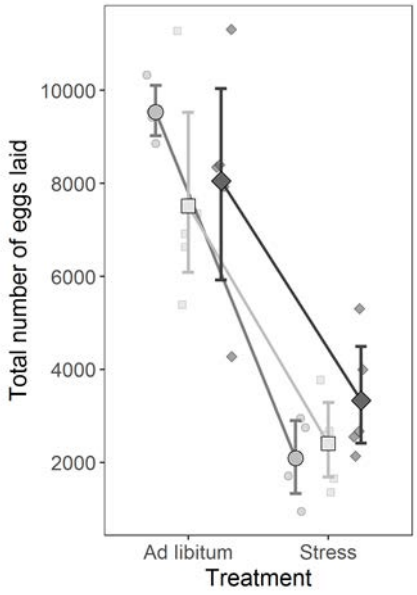


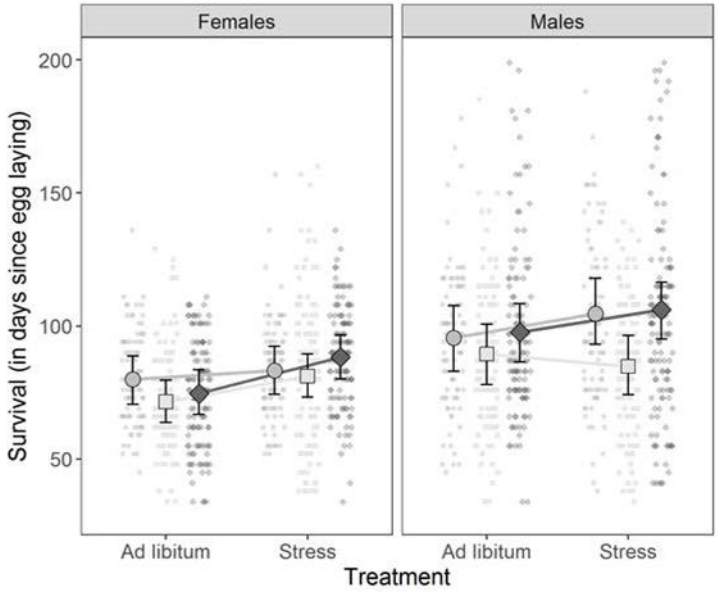
Final phenotyping: other juvenile and adult traits







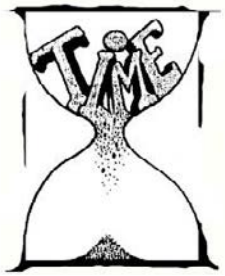




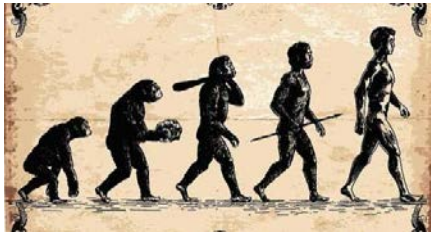
→ Limited impact on other life-history traits

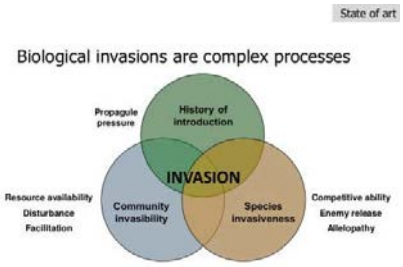
Evolution of female body mass - conclusions

- Rapid shift for the selected trait
- Limited impact on other life-history traits.
- No obvious link with the invasive syndrom displayed by invasive populations
- Potential explanations:

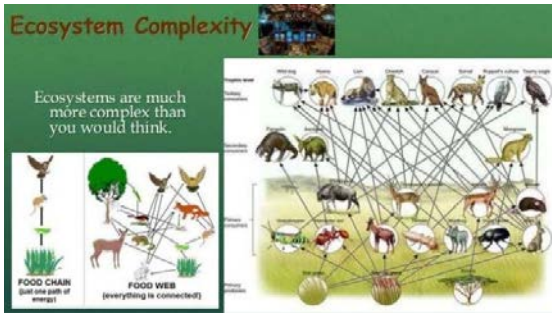


Experimental selection was too short?





Selective pressures are more complex during the course of the invasion?



Generation time: experimental protocol – directional selection

Population sampled in China

8 generations

The first twenty clutches

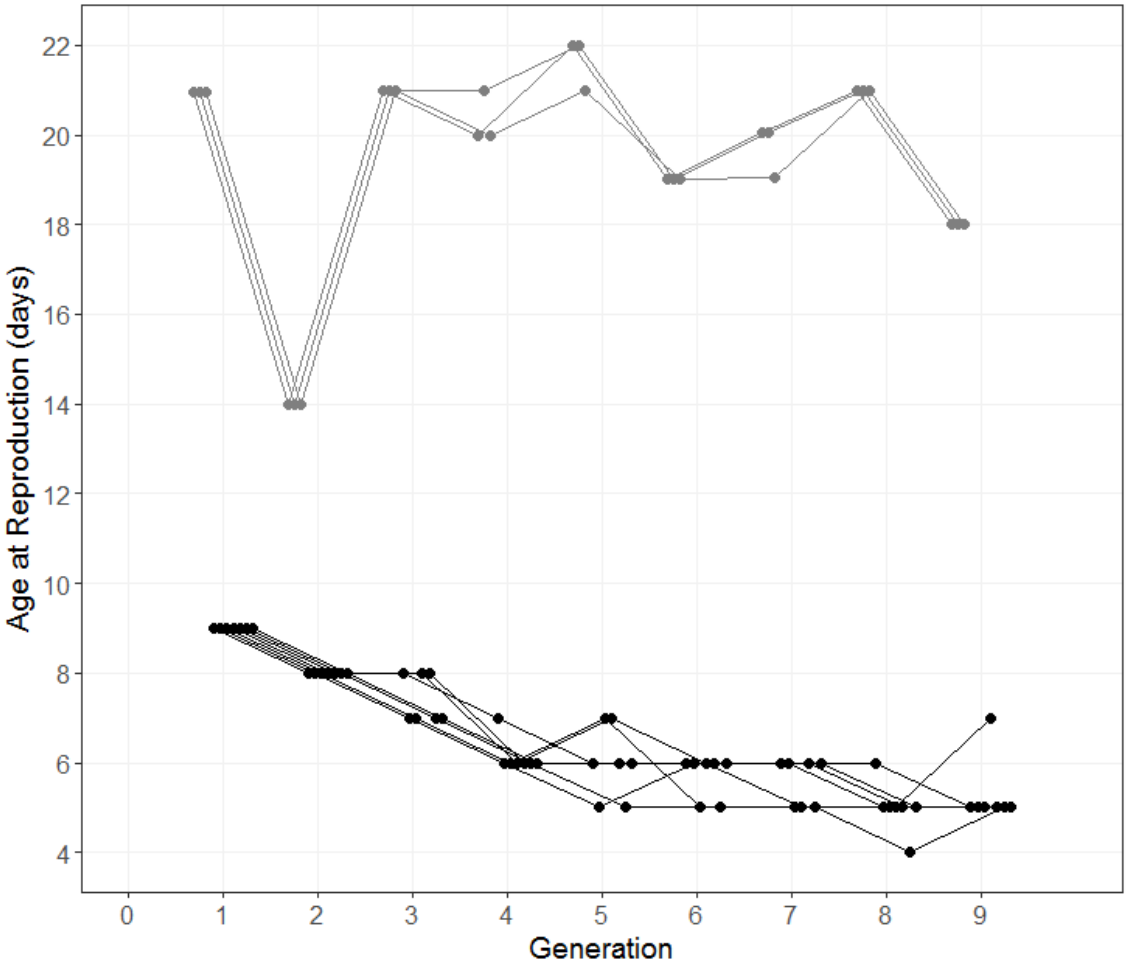
Twenty clutches after 21 days

7 short lines

3 control lines

Nb: Each generation = 100 females and 100 males for each line (sub-pop)

Evolution of age at first reproduction

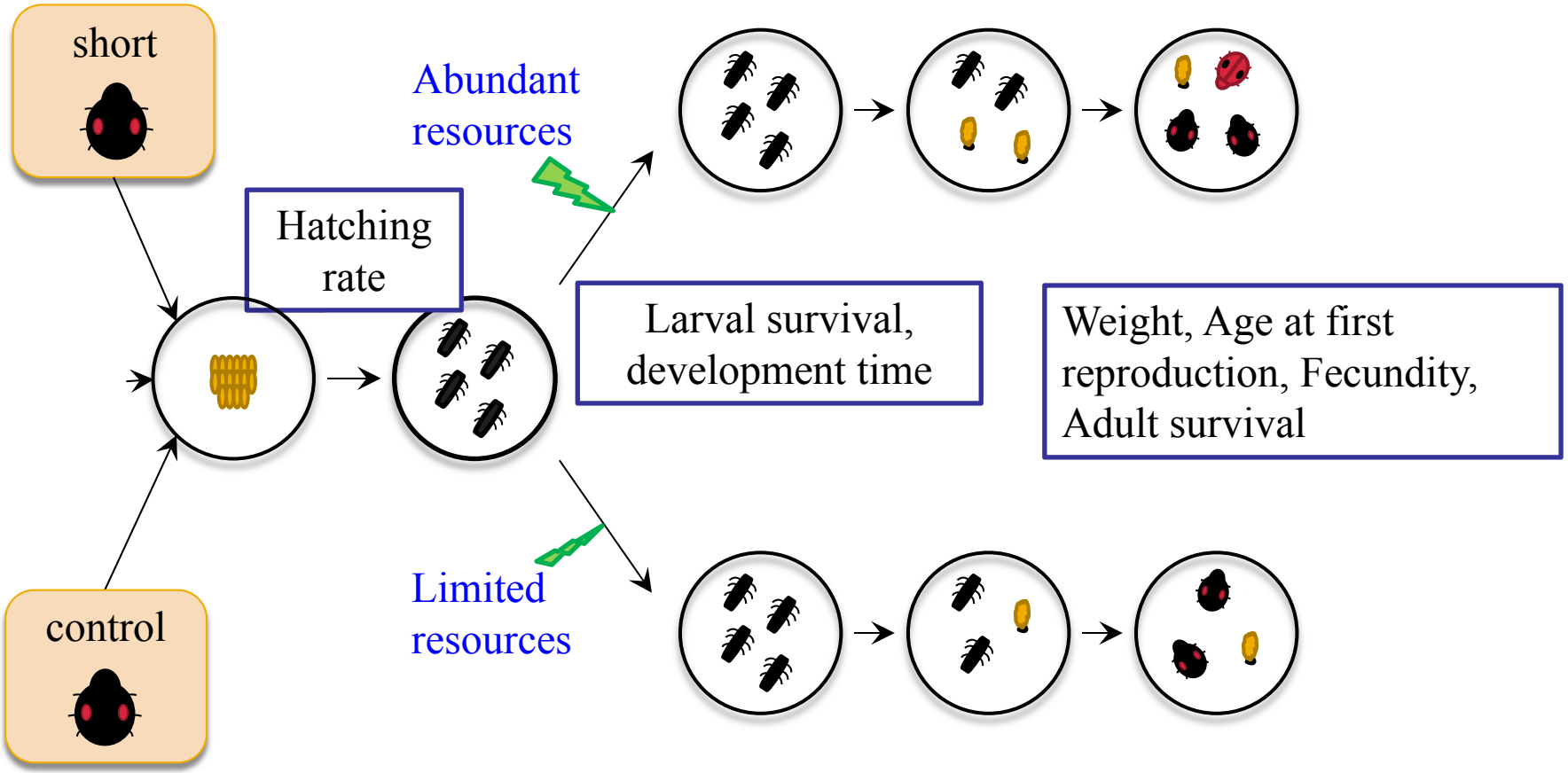


Control lines

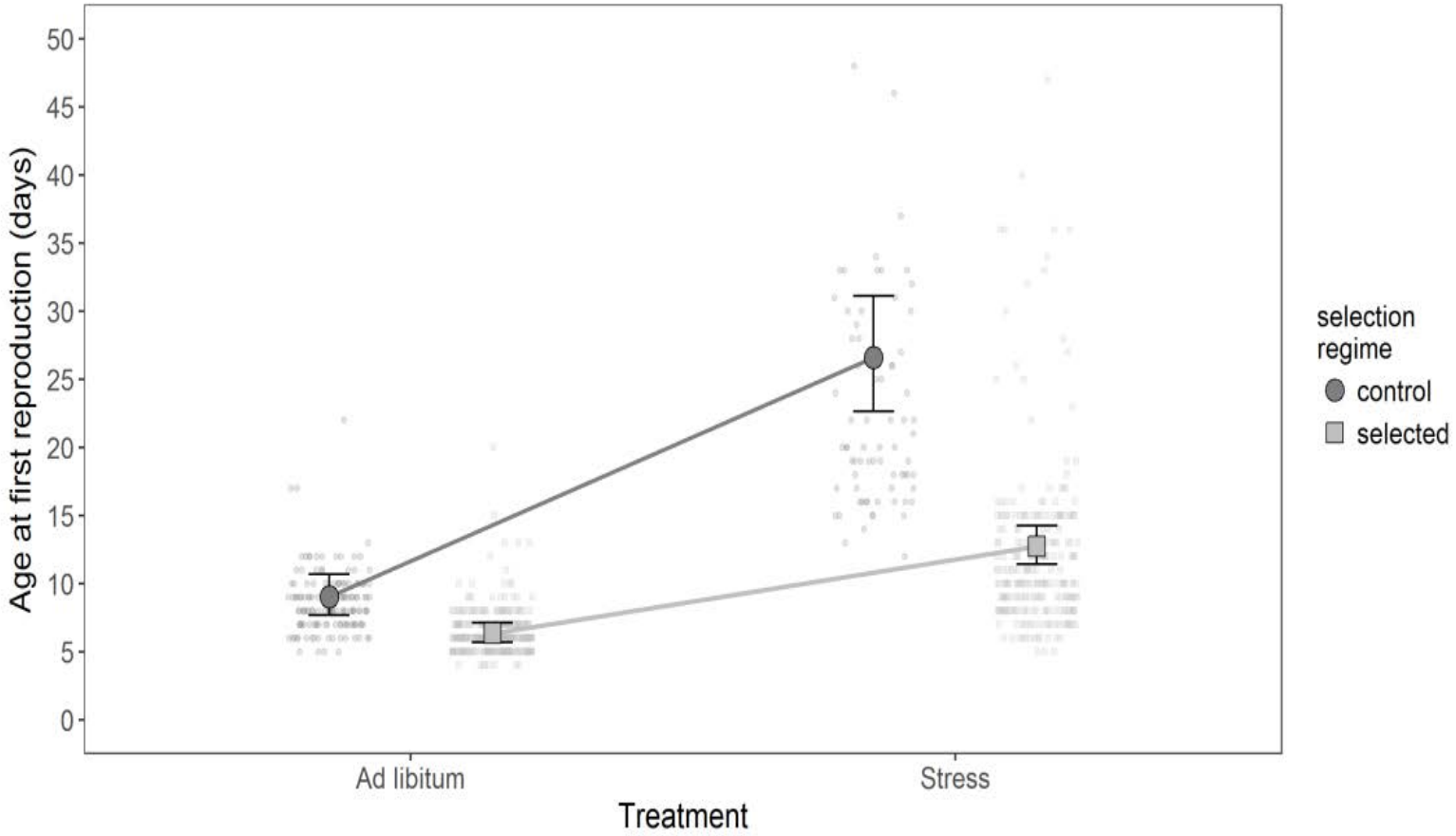
Short lines

→ *Some shifts have occurred during our experimental selection*

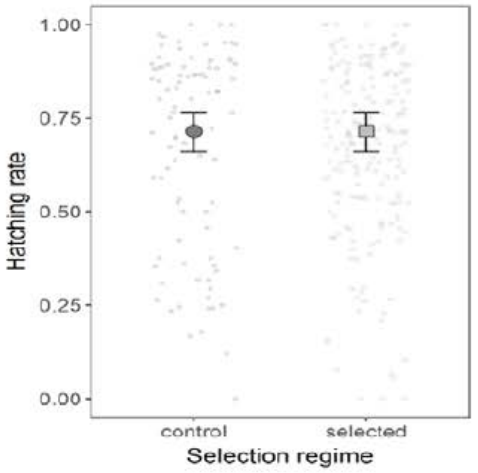
Final phenotyping: age at first reproduction and other juvenile and adult traits

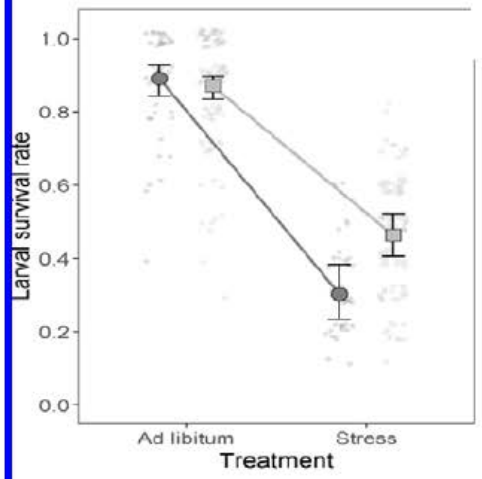


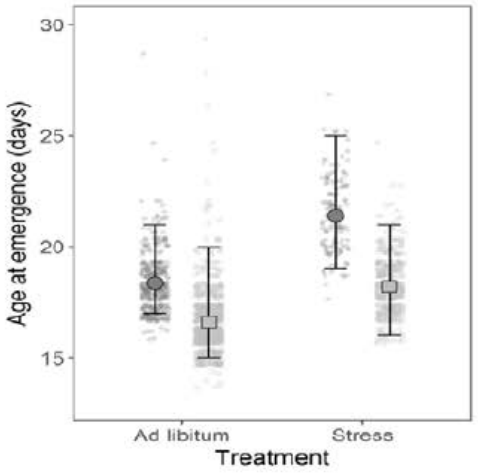
Final phenotyping: age at first reproduction

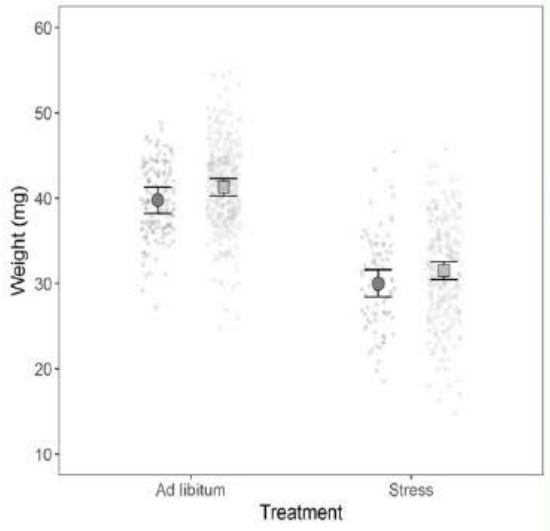


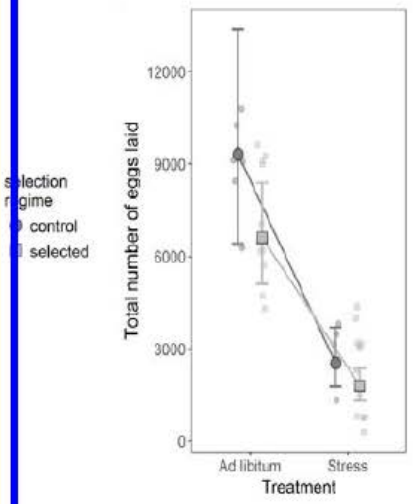
Final phenotyping of other juvenile and adult traits

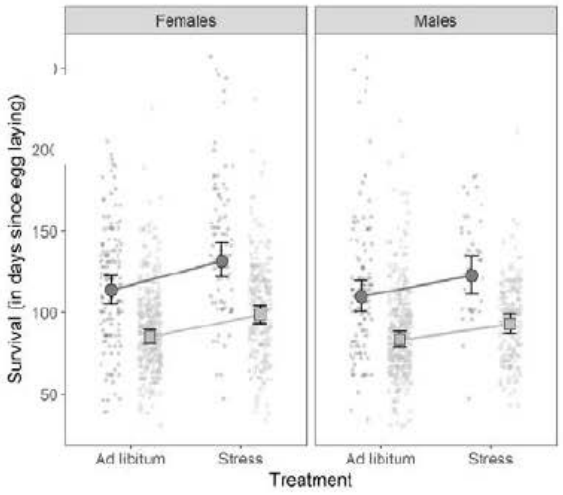












→ Substantial impact on other life-history traits

Evolution of age at first reproduction - conclusions

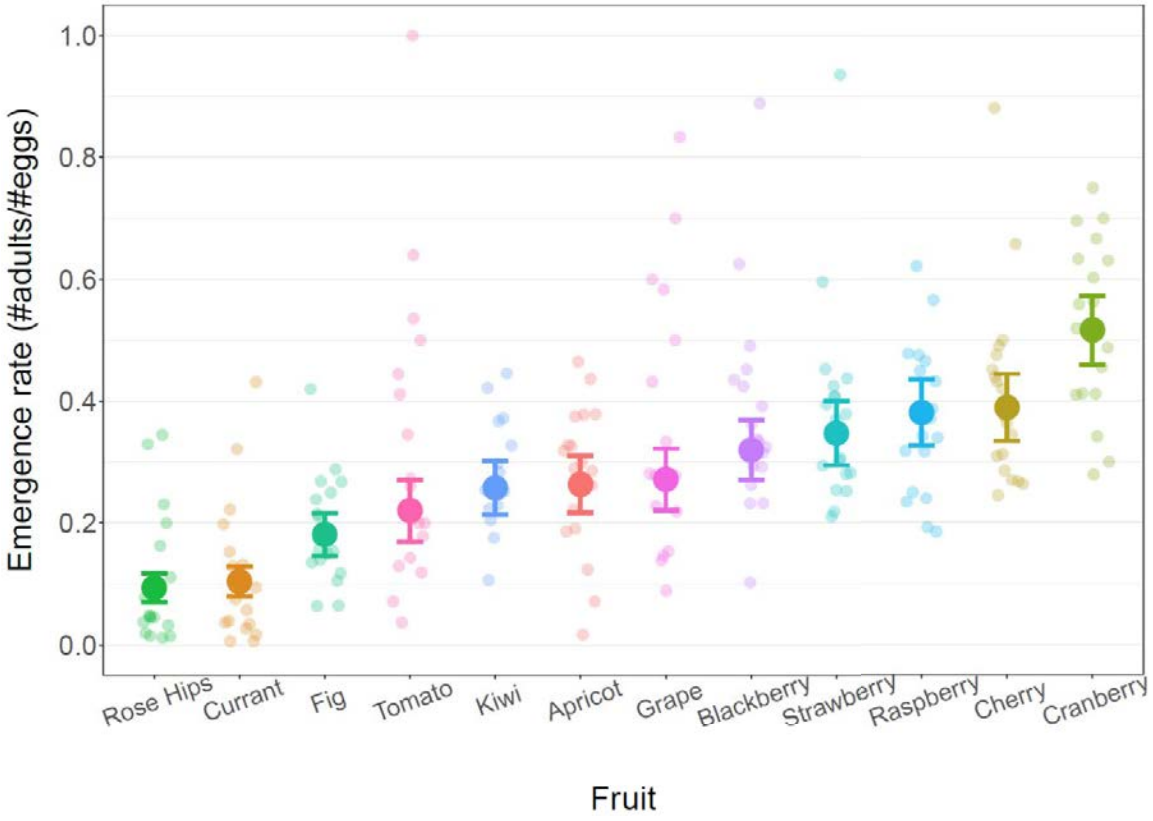
- Significant shift for the selected trait

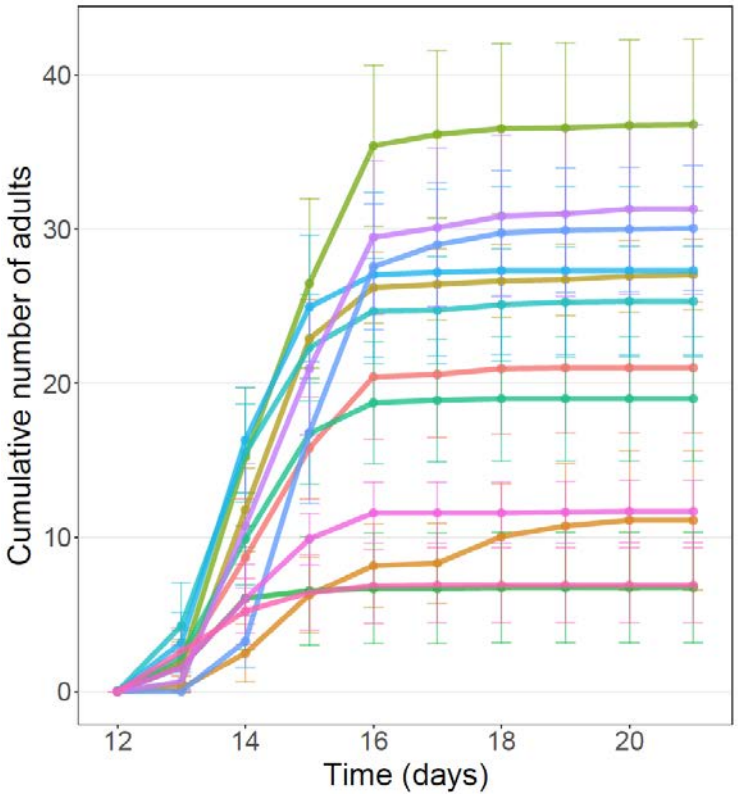
- Substantial changes on other life-history traits, especially adult traits such as fecundity and lifespan

- Life history of selected lines is (however) globally different from that of the invasive populations

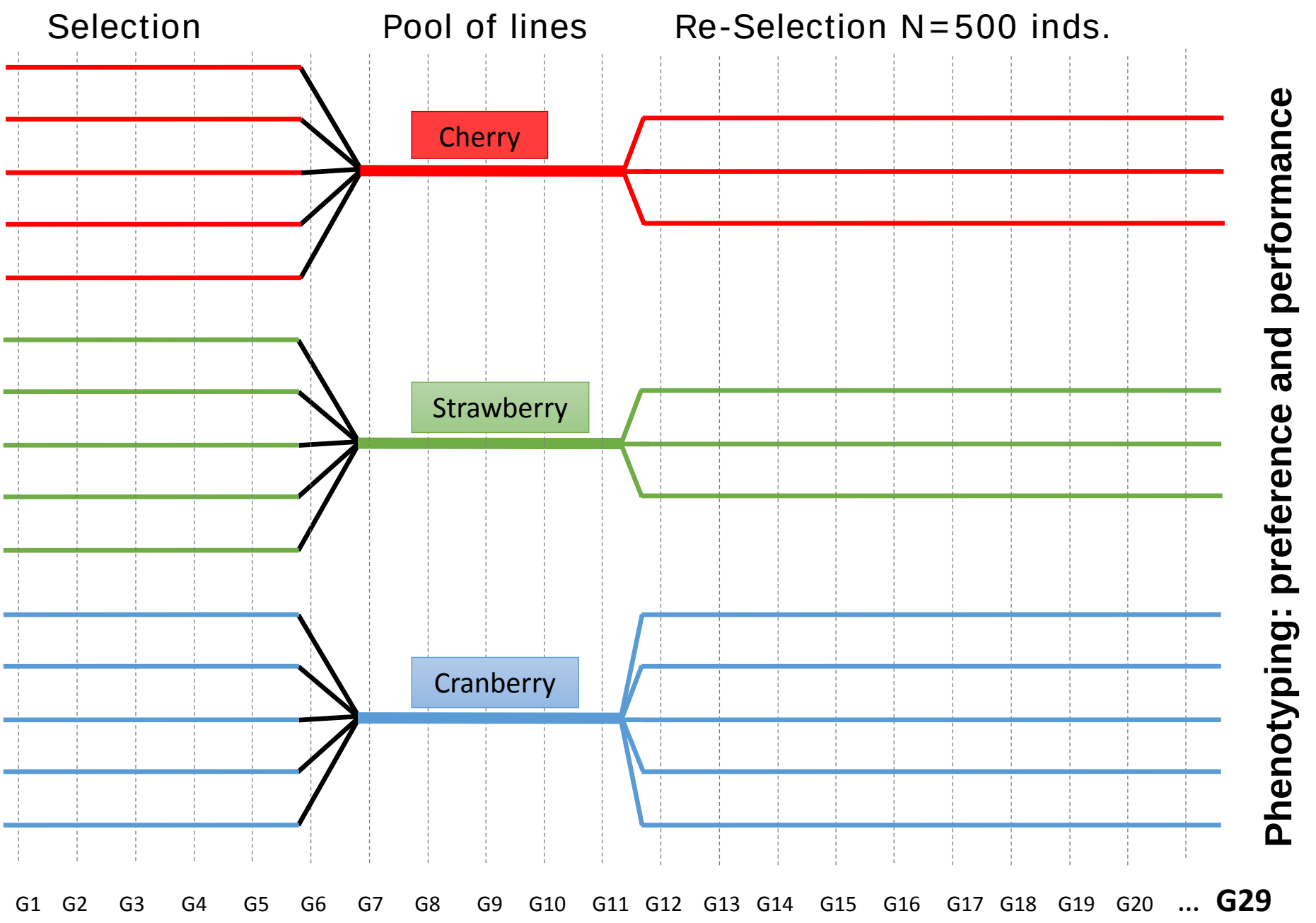
Trait	invasive pop	laboratory pop
Larval dev.	Slower	Faster
Total fecundity	higher	lower
Lifespan	longer	shorter
Weight	heavier	heavier
Early repro.	higher	higher
Age at first repro.	earlier	earlier

Drosophila suzukii:
HOST-FRUIT MATTERS !!!





Experimental selection on different media of fruit purees

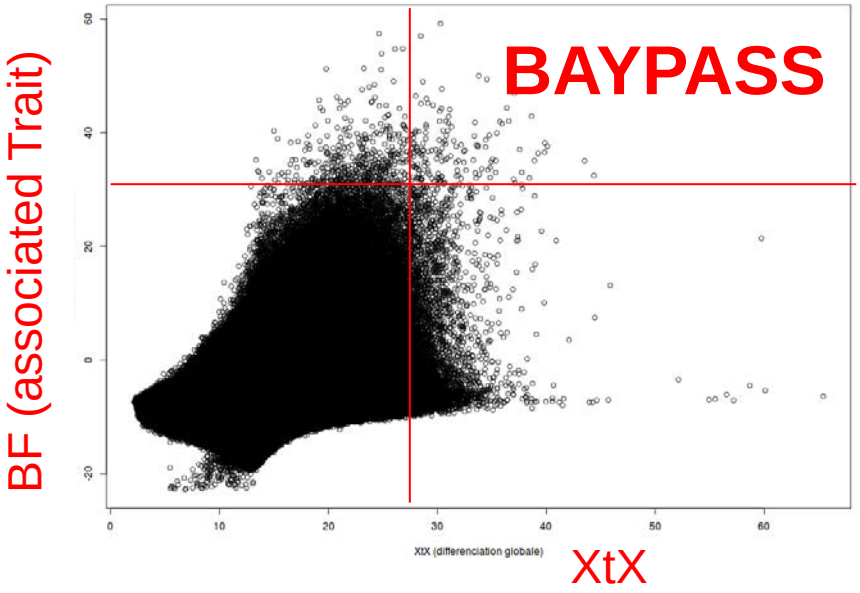

Selection
Pool of lines
Re-Selection N=500 inds.
Cherry
Strawberry
Cranberry
Phenotyping: preference and performance
G1 G2 G3 G4 G5 G6 G7 G8 G9 G10 G11 G12 G13 G14 G15 G16 G17 G18 G19 G20 ... G29

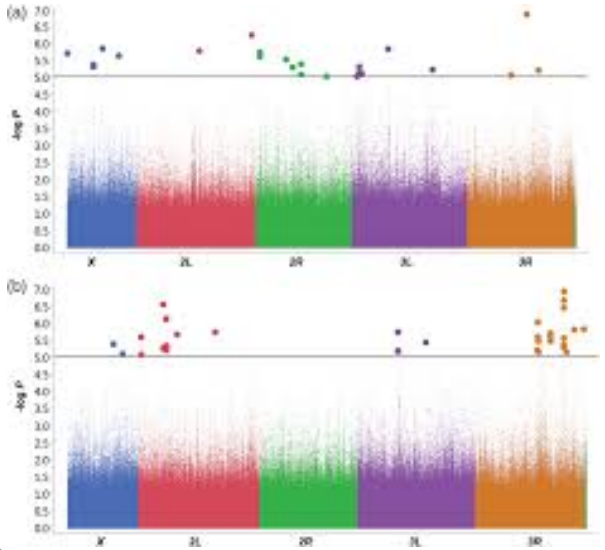
Preference

Performance

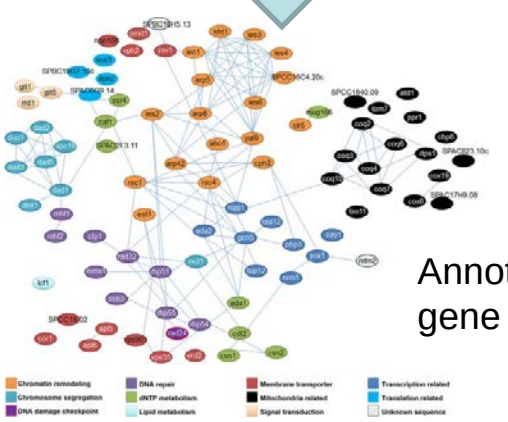

C183

WORK IN PROGRESS: NGS genotyping of selected (and control) lab populations + association study





Manhattan plot

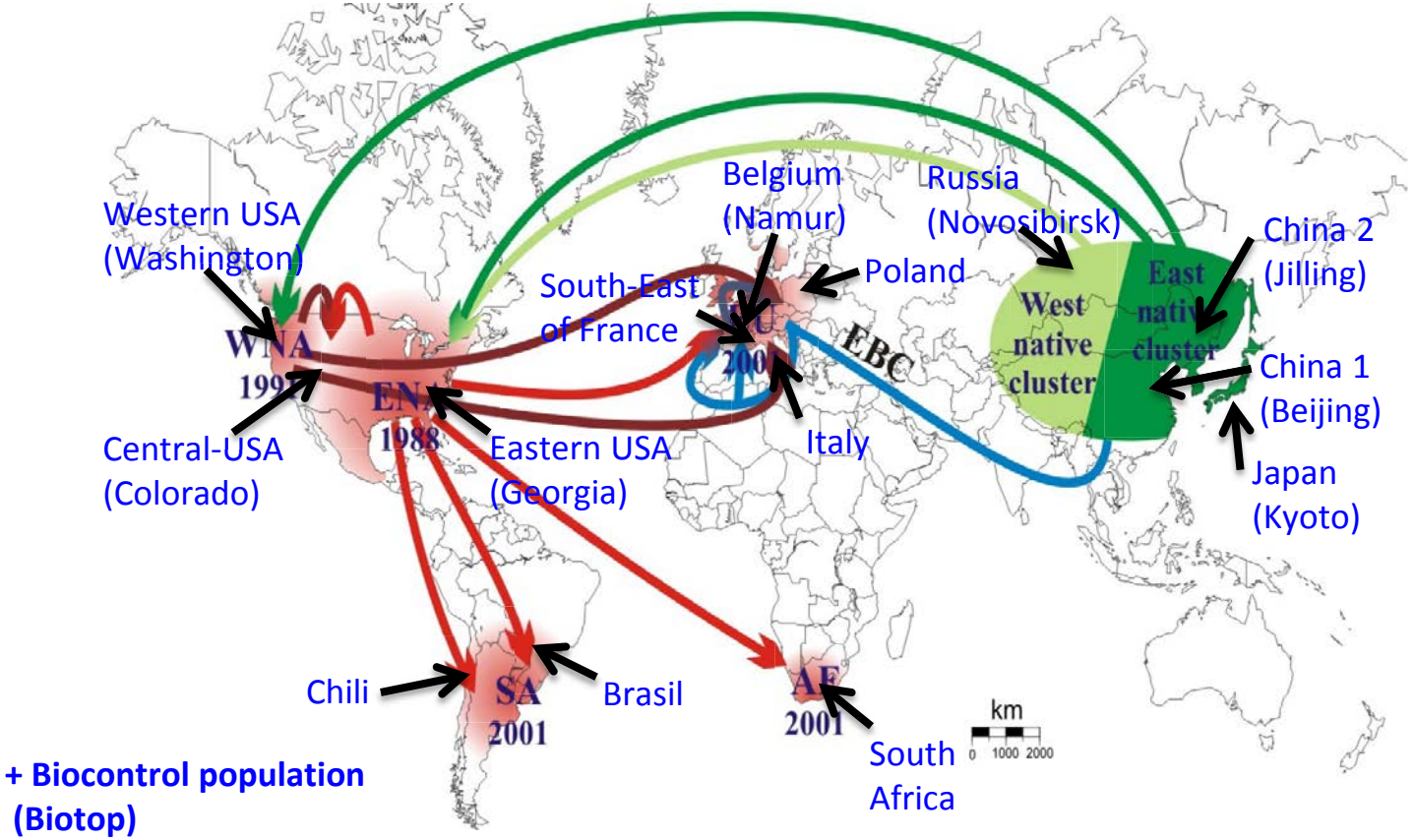


Annotated gene network

Similar genomic signal in <u>natural (native / invasive) populations</u> ?

Part II - Studies on natural populations:
worldwide adaptation routes in *Harmonia axyridis* and *Drosophila suzukii*

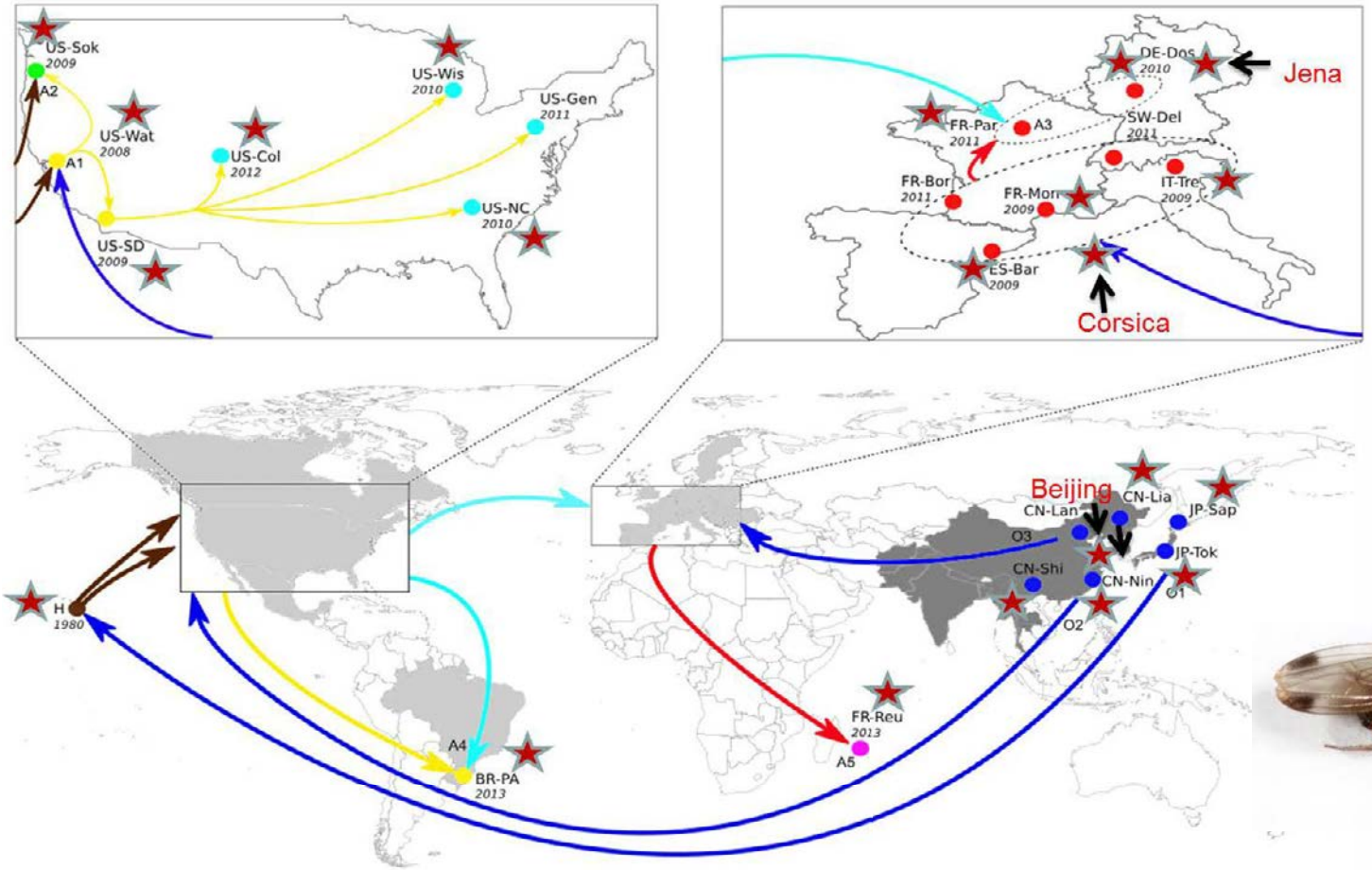
NGS: pool-seq data / 50-100 inds per pop



Quantitative co-variables

- *Native* versus *Invasive* status

- Pilot trait : discrete trait with simple genetic determinants → color polymorphism
- 19 environmental variables
- Juvenile traits: development time, larval survival, hatching rate,
  Adult traits: age of first reproduction, fecundity



Quantitative co-variables

NGS: pool-seq data / 50-100 inds per pop

- *Native* versus *Invasive* status
- 19 environmental variables
- External morphological/phenotypic traits likely associated with fitness: wing size and shape, ovipositor morphology and pigmentation

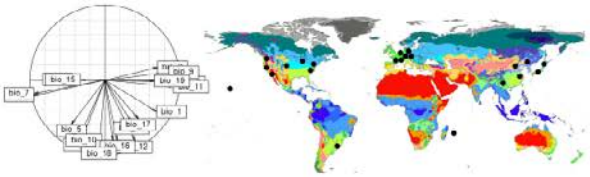
Approach

Methods

Results

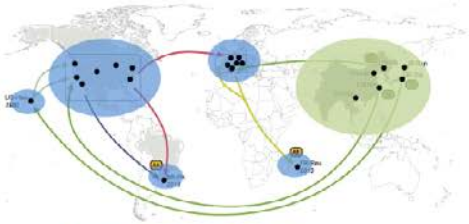
Perspectives

Genome response to climate variables



19 bioclimatic variables[1]

Genome response to invasion success



Invaded *vs* **Native** range

Whole genome pool sequencing

22 populations: 6 native and 16 invasive

60X coverage and n=50 to 100 individuals per pool

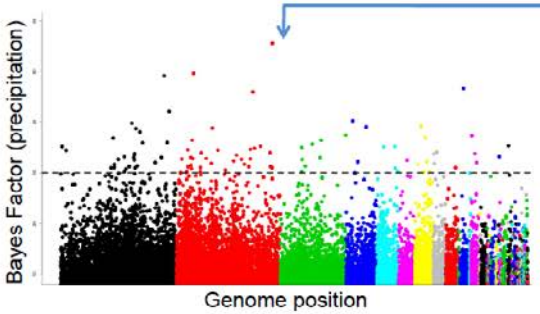
Bioinformatic treatments

Read mapping (*bwa*) on a newly developed high quality assembly[2] (total length=268 Mb, N50=2.6Mb, 546 contigs)

Variant calling (*VarScan*): ~12 x 10^6 SNP

Association with climate variables

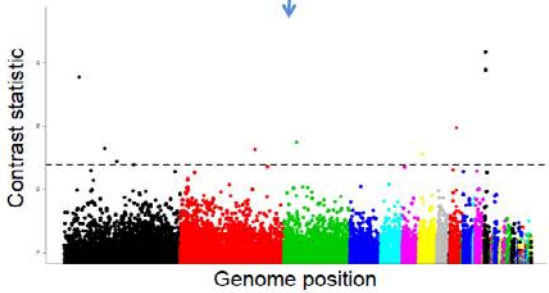
(BAYPASS software[3])



Many highly significant SNPs

Association with invasive *vs* native status

(BAYPASS software: Contrast analysis[4])



Few highly significant SNPs

Global functional analysis to identify the main physiological pathways involved in climate adaptation

↓ reverse ecology

Validation using quantitative genetics approaches

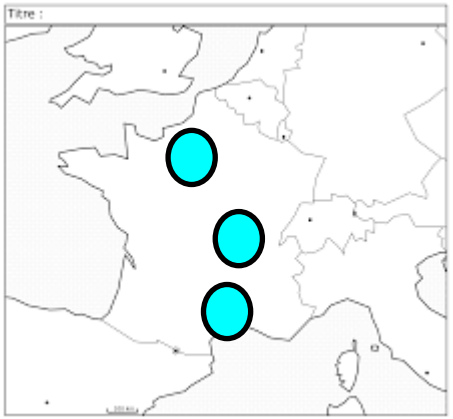
Limited number of genes with yet unknown functions

↓

Focal exploration and validation using functional genomic tools (e.g. RNAi, genome editing)

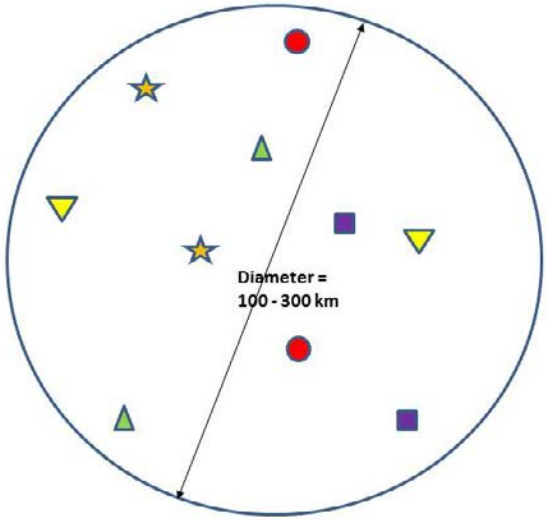
Are wild SWD populations more similar by host-fruit or by region? Can we detect genomic signals of genetic adaptation to host?

USA



France

REGION X (e.g. Wisconsin)
Not at scale!



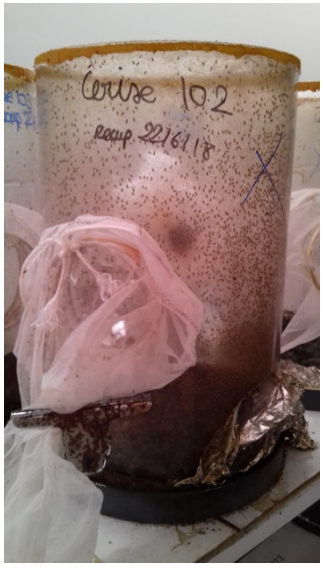
n = 50 to 150 individuals
(100 would be perfect!)

- Winter pop = *n* ind in TRAPS
- Strawberry field = *n* ind from « many fruits »
- Cherry field = *n* ind *from « many fruits »*
- Grape field = *n* ind from « many fruits »
- Blackberry (cf. wild host) = *n* ind from « many fruits »

Minimum distance between two sampled sites of the same fruit/host = ca. 50 km

NGS: pool-seq data / 50-100 inds per pop

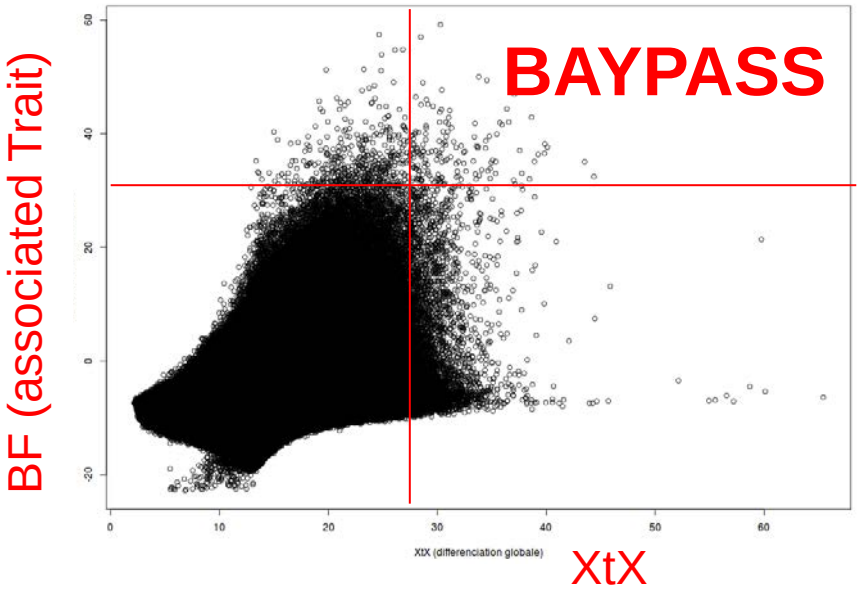
Phenotyping:
- Preference
- Performance

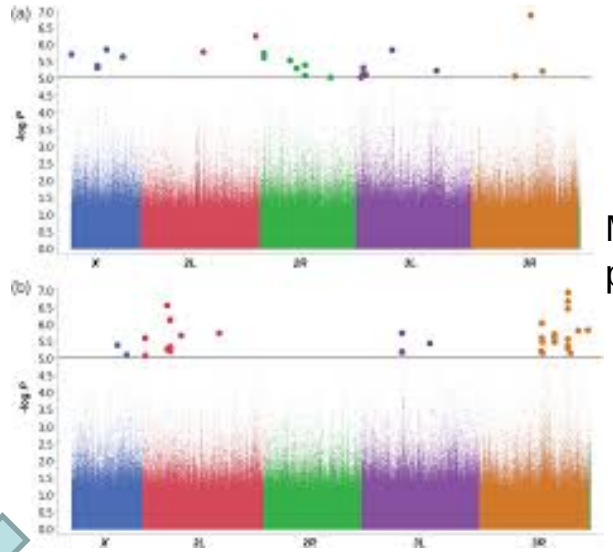


Performance

Preference

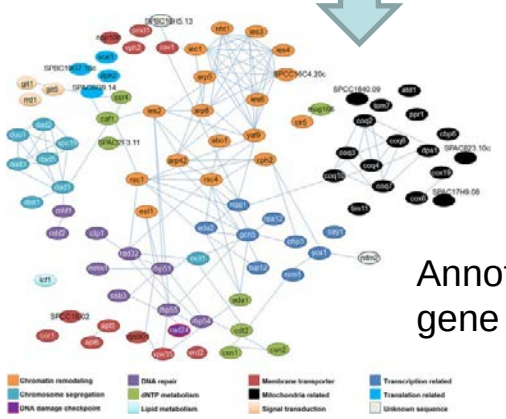
WORK IN PROGRESS: NGS genotyping of natural populations + association study





Manhattan plot

Similar genomic signal in laboratory (E&R) populations?



Annotated gene network

Common garden experiment(s) (quantitative genetics)

Candidate trait(s) ?

*« **Reverse ecology** » (Li et al. 2008)*

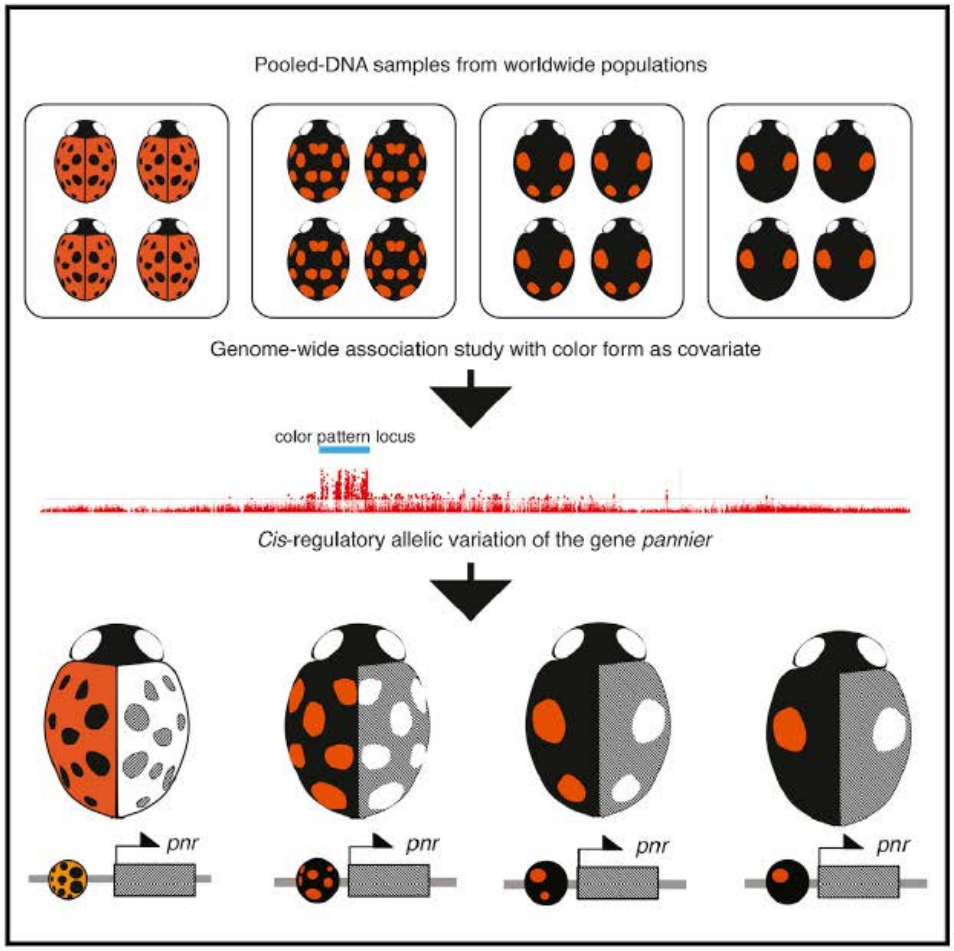
Part III - Proof of concept on a pilot trait: color polymorphism in natural (and laboratory) populations of *Harmonia axyridis*

« One gene to rule them all …»

Current Biology

The Genomic Basis of Color Pattern Polymorphism in the Harlequin Ladybird

Graphical Abstract



Authors

Mathieu Gautier, Junichi Yamaguchi, Julien Foucaud, ..., Heiko Vogel, Arnaud Estoup, Benjamin Prud'homme

Correspondence

arnaud.estoup@inra.fr (A.E.), benjamin.prudhomme@univ-amu.fr (B.P.)

In Brief

More than 200 distinct color forms have been described in natural populations of the harlequin ladybird, *Harmonia axyridis*. Gautier et al. show that this variation is controlled by the transcription factor Pannier. Pannier is necessary to produce black pigment, and its expression pattern prefigures the coloration pattern in each color form.

Highly variable (red-black) color patterns

described by the old school of taxonomic entomologists (Hemmelmann in Mader 1932)

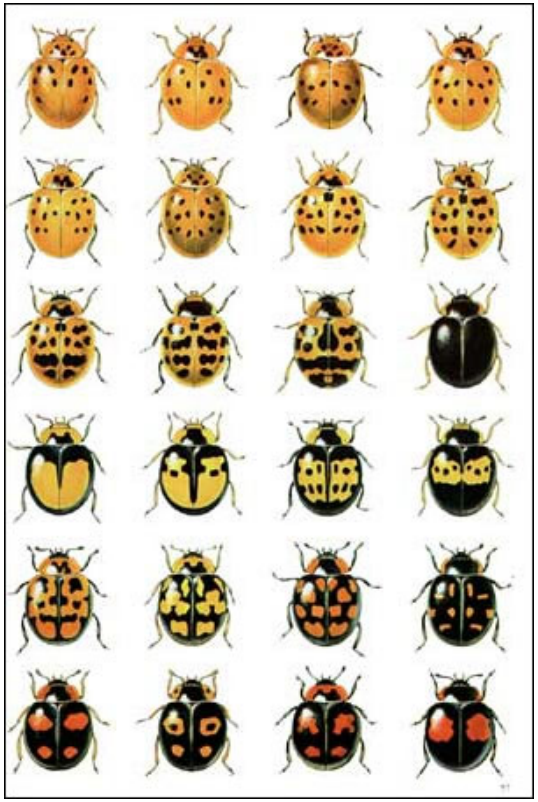
→ 200 color forms: from full red to full black

→ Several pseudo-species (genera)

…

→ Same species (geographic forms) (Dobzhansky 1924; Tan & Li 1932)

→ 15 classes of color forms



And many more …

1- Succinea (red-nSpots)





2- Conspicua (black-2Spots)



4 main color forms

3- Spectabilis (black-4Spots)



4- Axyridis (black-nSpots)



1- Succinea (S)

2- Conspicua (C)

3- Spectabilis (P)

4- Axyridis (A)

Genetic determinism: formal genetics studies

1 autosomal locus: 15 alleles = 4 frequent alleles + 11 rare alleles

→ hierarchical dominance of color patterns

C > P > A> S

Phenotypic plasticity of the S (red-nSpots) form:

T°C of larval development:

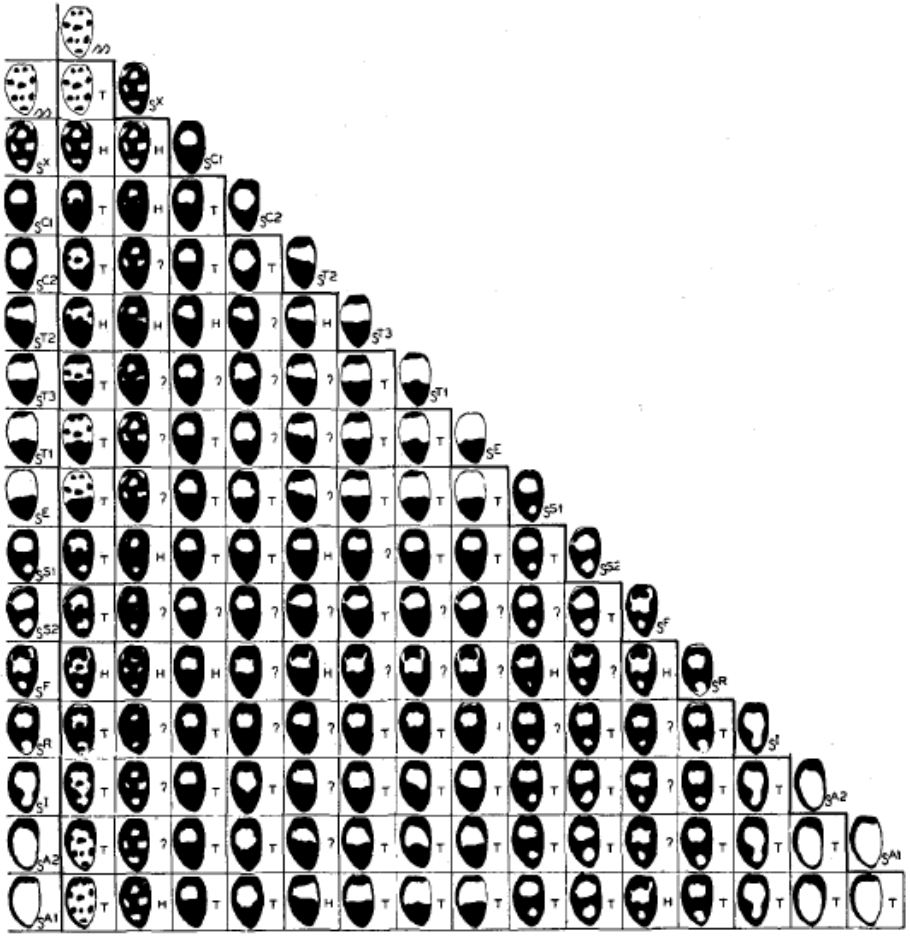
16-18°C 24°C 28°C

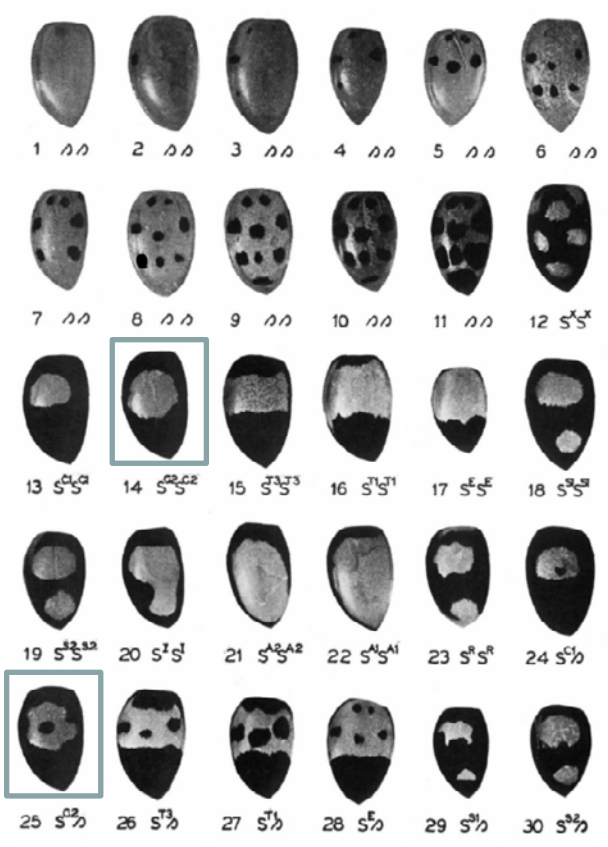
Mosaic (hierarchical) dominance of color patterns (Tan 1946 Genetics)

→15 known alleles

→ Super-imposition of « semi-transparent papers » corresponding to the phenotype of each allele

→ Obtention (or prediction) of the 105 (i.e. (15x14)/2) possible color patterns





Mosaic dominance of color patterns: observed and predicted (cf. ?) phenotypes

PLATE 1 and 2.—Figures 1–60: Pictures of left elytra of *Harmonia axyridis* showing the color patterns of various homozygotes and heterozygotes for 12 different alleles. The allelic symbols are: *s* for *succinea*, S^X for *axyridis*, S^{C1} for *conspicua*-1, S^{C2} for *conspicua*-2, S^{T2} for *transversifascia*-3, S^{T1} for *transversifascia*-1, S^E for *equicolor*, S^{S1} for *spectabilis*-1, S^{S2} for *spectabilis*-2, S^I for *intermedia*, S^{A2} for *aulica*-2, S^{A1} for *aulica*-1, and S^R for *tripunctata*.

Phenotypic (color) variation in space

→ geographical variation in the native range

FREQUENCIES OF COLOR PATTERNS (IN PER CENTS) IN *Harmonia axyridis* FROM DIFFERENT REGIONS

REGION	SUCCINEA, FRIGIDA, 19-SIGNATA	AULICA	AXYRIDIS	SPECTABILIS	CONSPICUA	UNCLASSIFIED	NUMBER EXAMINED
Altai Mountains	0.05	—	99.95	—	—	—	4,013
Yeniseisk Province	0.9	—	99.1	—	—	—	116
Irkutsk Province	15.1	—	84.9	—	—	—	73
West Transbaikalia	50.8	—	49.2	—	—	—	61
Amur Province	100.0	—	—	—	—	—	41
Khabarovsk	74.5	0.3	0.2	13.4	10.7	—	597
Vladivostok	85.6	0.8	0.8	6.0	6.8	0.1	765
Korea	81.3	—	—	6.2	12.5	—	64
Manchuria	79.7	0.5	—	11.2	8.6	—	232
North China (Peiping)	83.0	0.4	—	8.8	7.3	0.5	9,676
West China (Szechwan)	42.6	2.9	0.01	28.8	25.1	0.8	1,074
East China (Soochow)	66.6	0.6	—	16.5	16.1	0.2	6,231
Japan	27.2	—	11.0	14.3	47.4	—	154

Dobzansky 1933, 1937

Phenotypic (color) variation in time

→ seasonal/generational variation

f(M) Winter/Spring generation > f(M) Summer/Autumn generation

Winter/Spring Generation:
overwintering + reproduction April-May

Summer/Autumn generation:
summer + reproduction Sept-Oct

Field observation (counts)

Beijing (China)

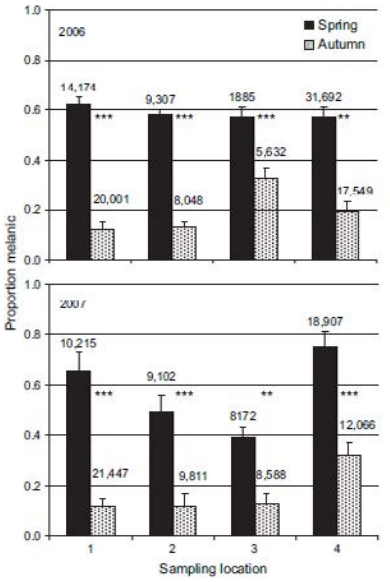


Fig. 1. Proportions of melanic (vs succinic) *Harmonia axyridis* observed at four sampling locations around Beijing, China. Each bar depicts the mean (+SEM) of five sampling dates in either spring or autumn in each of the 2 years. Numbers above bars represent total numbers of beetles observed. Asterisks indicate significant differences between absolute numbers at each site (one-way ANOVA, ***, $\alpha \leq 0.001$, **, $\alpha \leq 0.01$). 215 × 279 mm (600 × 600 DPI).

Kyoto (Japan)

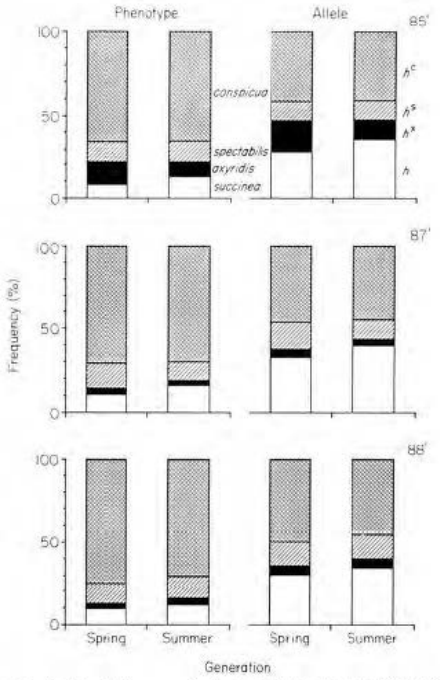
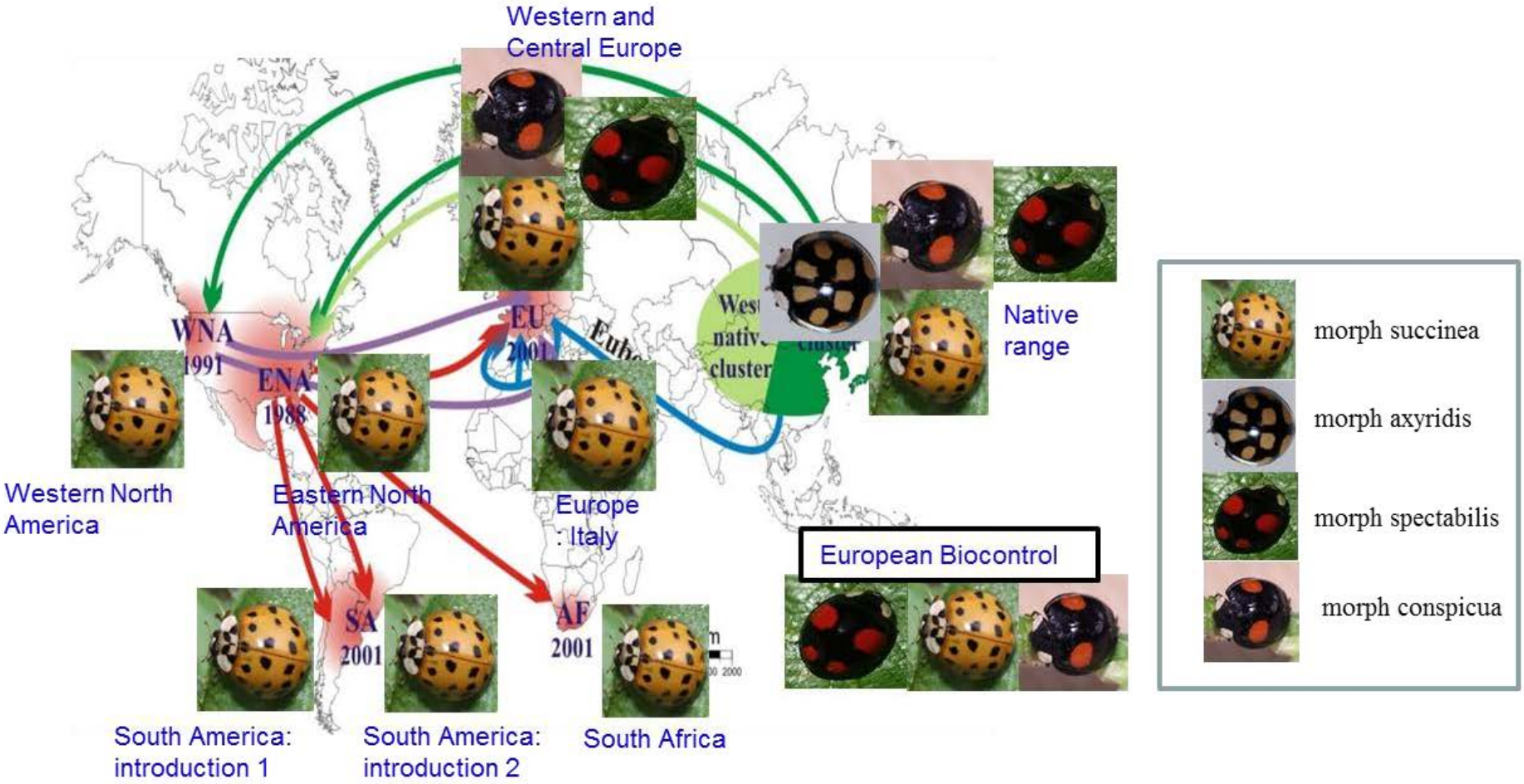


Fig. 1 Morph frequencies in phenotype and allele in spring and summer generation.

Phenotypic (color) variation and invasion history

→Historical variation = predominance of the red-nSpot form in the invasive range: random variation or selective advantage?

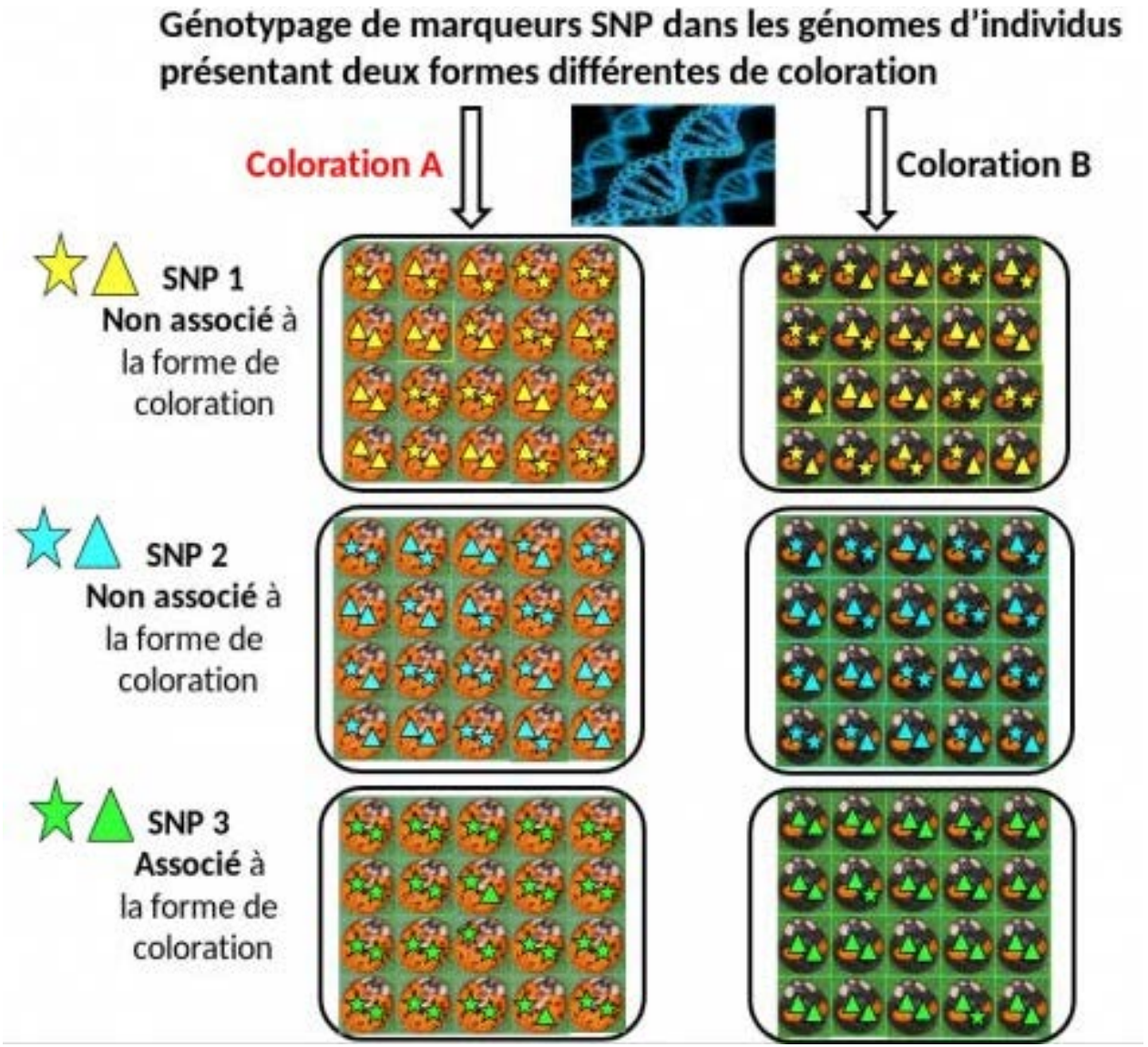


Note: Random factors (drift) / Climatic factors (thermoregulation) / Predation (red-nSpot more toxic) / Other Env. factors: eg. higher nutritional plasticity of R ind (Berkvens et al. 2007)

	Assembly *HaxR*	Assembly *HaxB4*
Data	MinION long reads (65X) Illumina PE reads (100X)	Illumina PE reads (65X) Illumina MP reads (24X)
Assembler	SMARTdenovo	ALLPATH-LG
Nb. of sequences	1,071 contigs	6,586 scaffolds
Total length (Mbp)	429	393.1
Average length (Kbp)	400.9	59.7
Max size (Kbp)	7,499	5,635
Total Ns (bp)	22	22,814,986
N50 (Kbp)	1,434	978.4
BUSCO (complete)	97.2 %	86.0 %
BUSCO (fragmented)	1.3 %	8.7 %
BUSCO (missing)	1.5 %	5.3 %

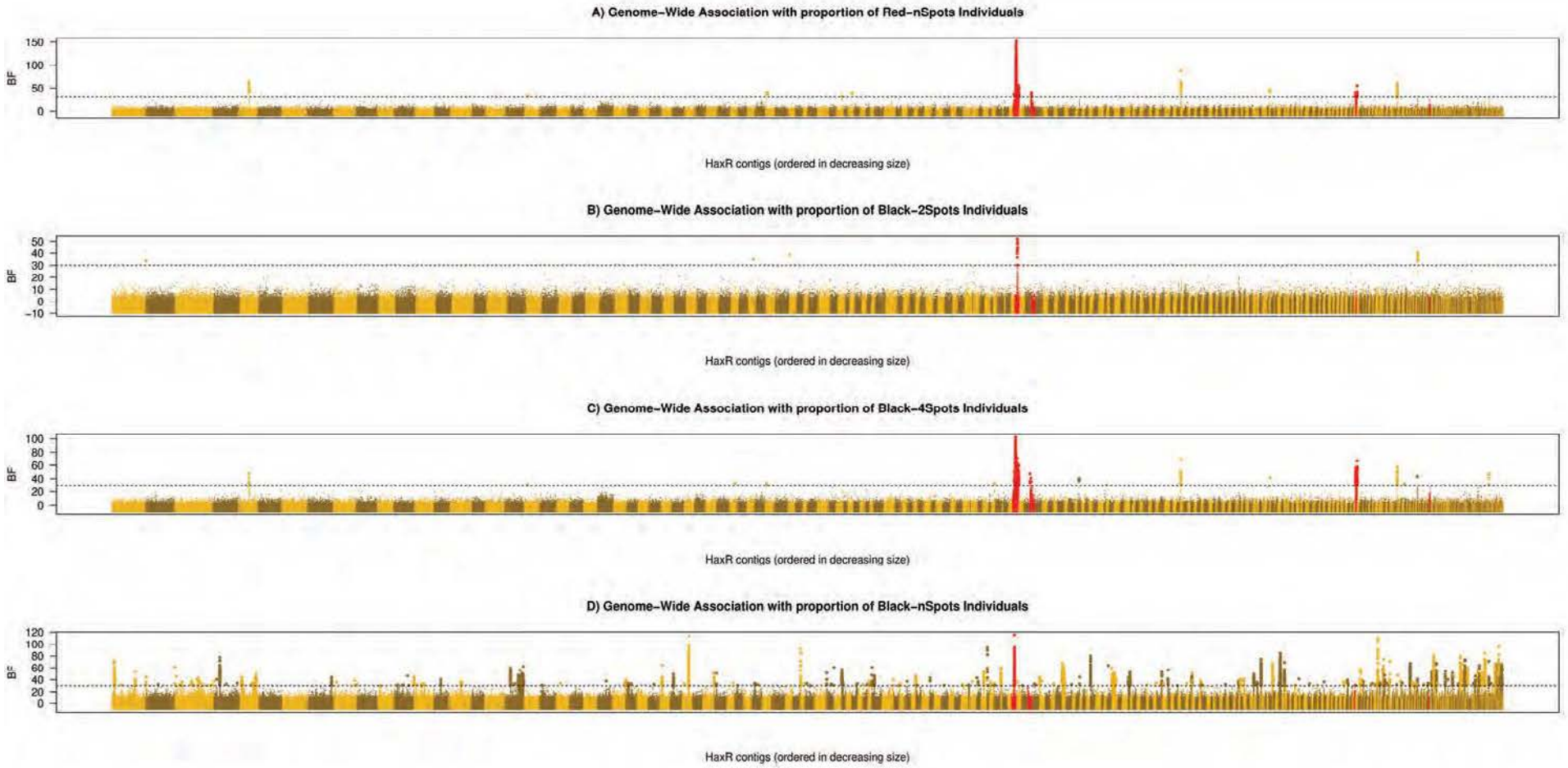
Pooled-sequencing sample code	Population sampling site		Sampling year	Colour form in the pool	No. of sequenced individuals
	Country	Region or city			
CH1-R	China	Jilin	2013	Red-nSpots only	100
CH1-B				30 Black-4Spots 28 **Black-2Spots**	58
CH2-R	China	Changchun	2015	Red-nSpots only	100
CH2-B4				Black-4Spots only	67
CH2-B2				**Black-2Spots only**	73
JP-R	Japan	Kyoto and other cities	2009	Red-nSpots only	57
JP-B4				Black-4Spots only	58
NOV-Bn	Russia	Novosibirsk	2007	Black-nSpots only	44
BRG-R	France	Bourgogne	2013	Red-nSpots only	50
BRG-B4				Black-4Spots only	50
ENA-R	USA	Georgia	2007	Red-nSpots only	45
WAS-R	USA	Washington	2007	Red-nSpots only	40
BIO-R	France	Biological control population (BIOTOP)	2012	Red-nSpots only	100
BIO-B4				Black-4Spots only	100

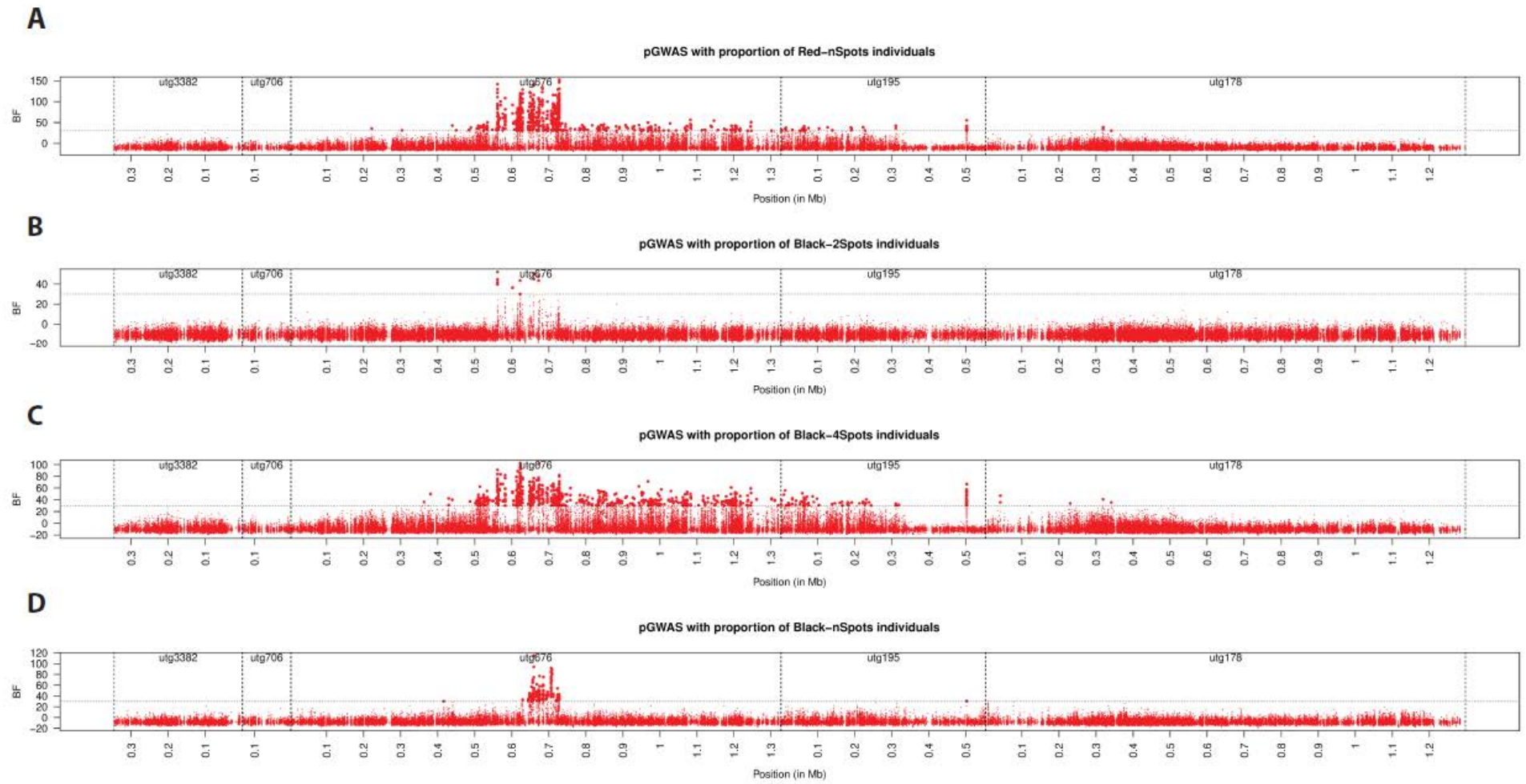
Genome-wide association study (GWAS)

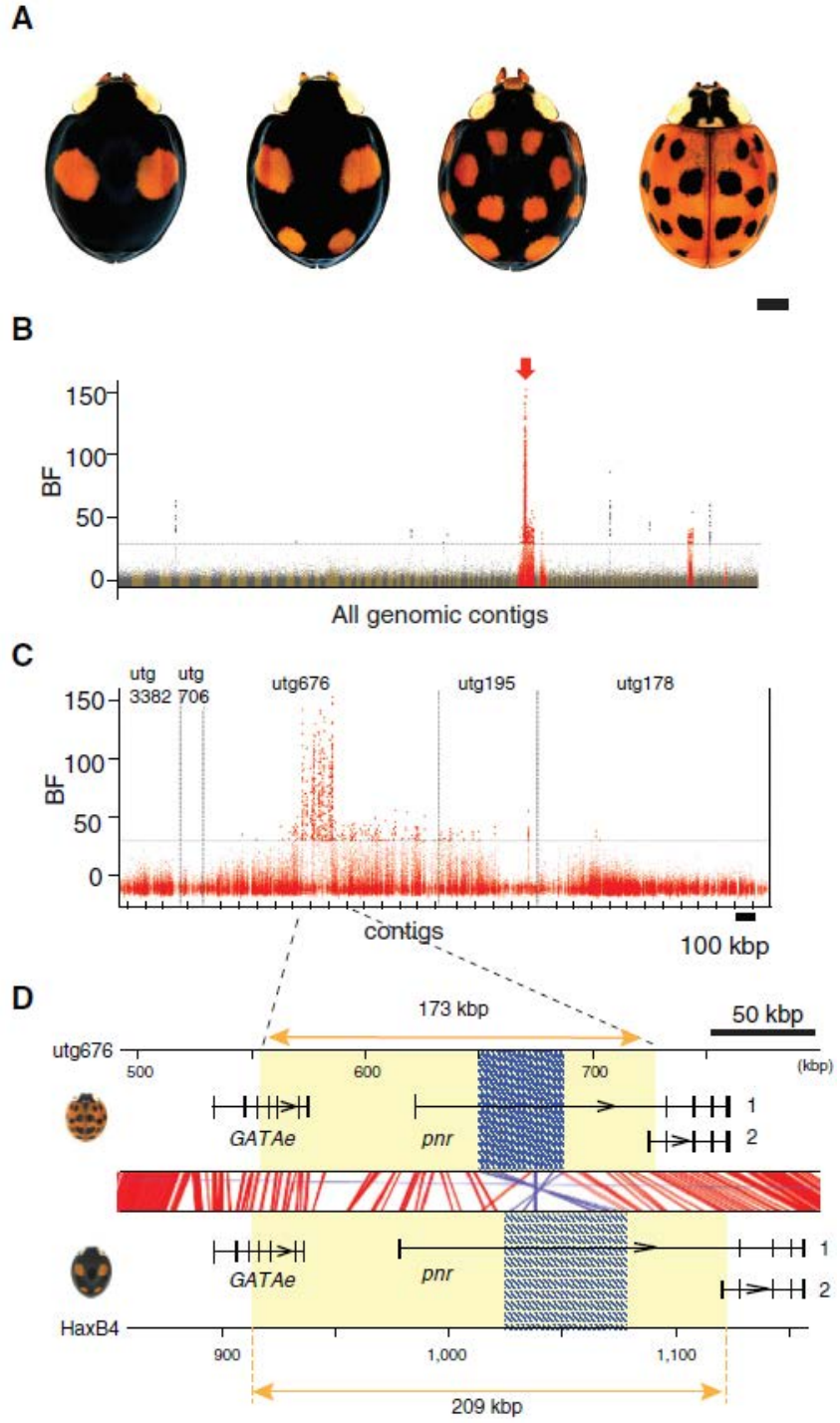


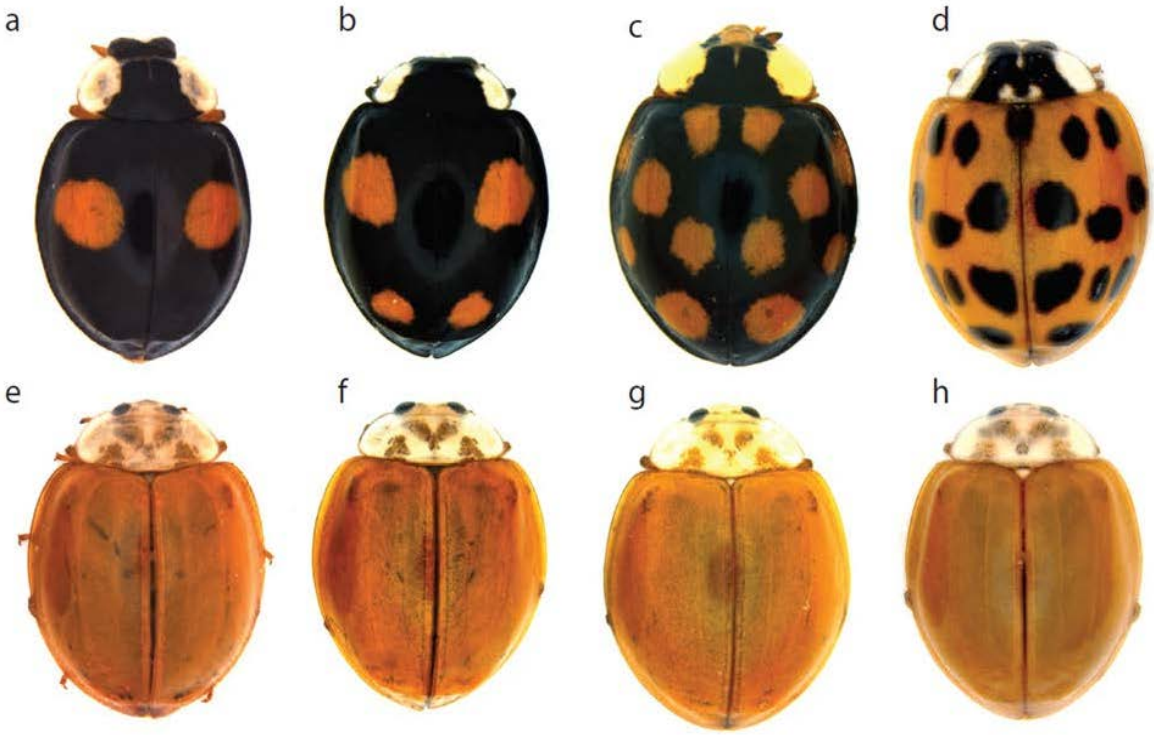
BAYPASS (Gautier et al. 2015)

→ 18,425,210 SNPs - 710 SNPs strongly associated
with the proportion of the red-nSpots form (Bayes factor > 30)

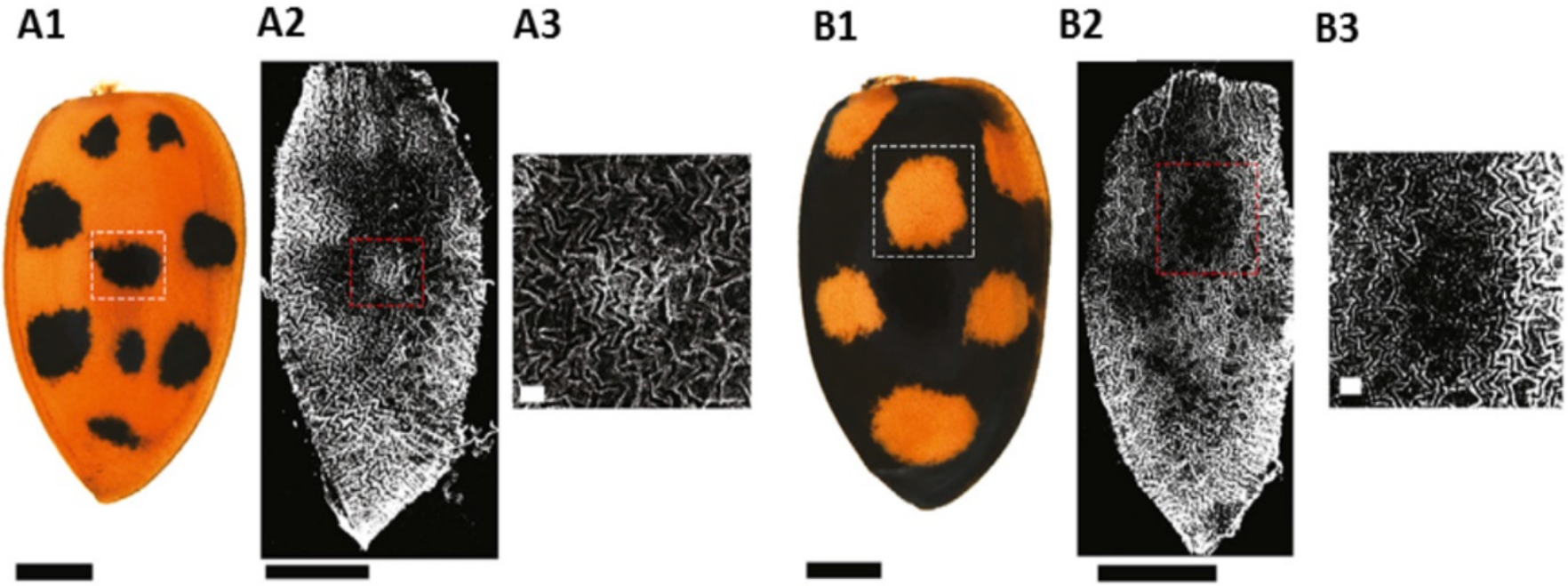



A
pGWAS with proportion of Red−nSpots individuals
utg3382
utg706
utg676
utg195
utg178
BF
Position (in Mb)
B
pGWAS with proportion of Black−2Spots individuals
C
pGWAS with proportion of Black−4Spots individuals
D
pGWAS with proportion of Black−nSpots individuals

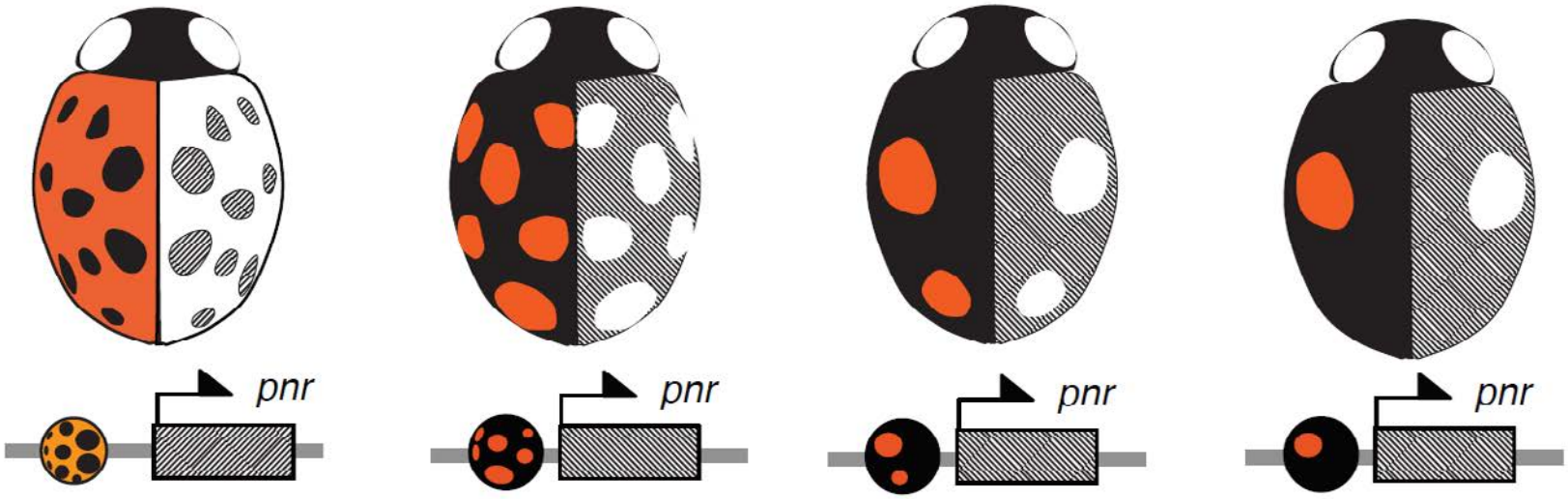

A
B
150
100
50
0
BF
All genomic contigs
C
utg 3382
utg 706
utg676
utg195
utg178
150
100
50
0
BF
contigs
100 kbp
D
173 kbp
50 kbp
utg676
500
600
700
(kbp)
GATAe
pnr
1
2
GATAe
pnr
1
2
HaxB4
900
1,000
1,100
209 kbp



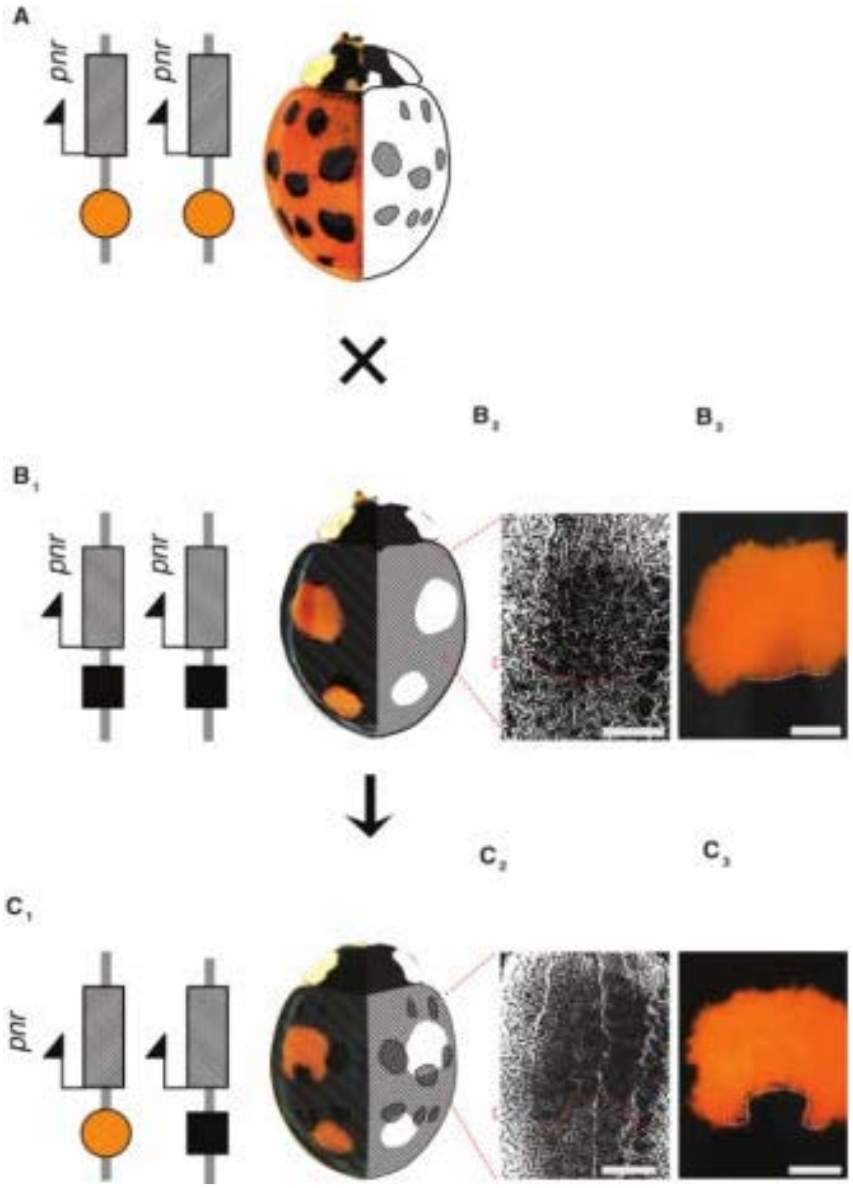
<u>Inactivation</u> of the pannier gene by RNA interference: the wild type color forms of the harlequin ladybird (from a to d) lose their black pigmentation (e-h) when the pannier gene is inactivated.



The location of the pannier protein in different cellular territories of an elytra is correlated with the spatial distribution of the black pigmentation. The gene pannier is activated (or expressed) in different cellular territories (silver areas) that prefigure the locations of black spots on the red elytra.



One (regulatory sequence) gene to rule them all: regulatory sequences (colored circles) located just upstream of the pannier coding sequence (hatched rectangle) define allelic variants

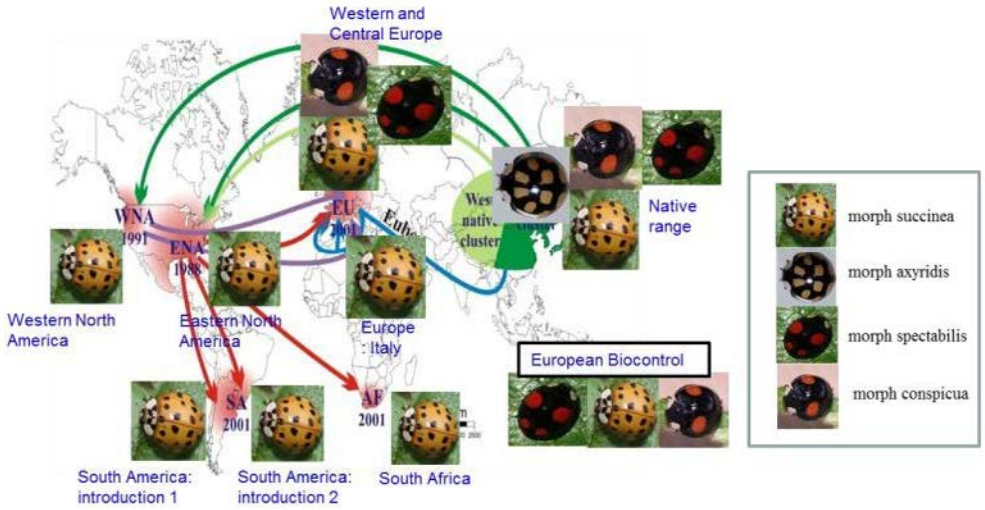


« Model » of dominance relationships among color form alleles

SOME ONGOING PERSPECTIVES

- In-depth molecular characterization of *pannier* variations, especially of its regulatory sequences (*H. axyridis* + other ladybug species)

- Test whether the quasi-fixation of a single color form (red-nSpot) in the invaded area is due to historical or demographic contingencies (simple random effects), or to a natural selection process favoring the non-melanic form during the invasion → population genomics approaches to detect selection signals on the red-nSpot haplotype(s) in invasive populations.



A very collective work!

Main collaborators: M. Gautier, A. Loiseau, B. Prud'homme, J. Yamaguchi, B. Facon, J. Foucaud, R. Hufbauer, L. Olazcuaga, B. Melbourne, H. Vogel, M Szucs, M Vahsen, C Weiss-Lehman, E. Lombaert, T. Guillemaud, V. Ravigné, R. Vitalis, N. Rode, J. Orivel, O. Rey, J. Turgeon, J-M Cornuet, C. Tayeh, G. Laugier, J. Lagnel, B. Gschloessl, C. Lee, W. Su, J. Lungren, R. Koch, I. Zakharov, P. De Clerk, L Lawson Handley,…and many other students/technicians/PhDs/scientists who helped on SEPA (+ field work)