Detecting the effects of co-adaptation in plant genomes

Françoise Budar



Institut Jean-Pierre Bourgin, INRAe Versailles

genetic interactions between nucleus and organelle genomes Cytoplasmic Male Sterilities

> Marie-Laure Martin-Magniette Institute of Plant Science, Paris-Saclay genomic Networks team MIA-Paris, Paris Statistics et genome team

> > Math For Genomics 17/06/2020



0

The plant cell metabolism is compartmentalized So is its genome



Mitochondria and chloroplasts are not functionally autonomous



The proper function of plant organelles relies on the interaction between nuclear and organelle genetic units

Mitochondrial electron transport chain

origin of subunits

- mitochondrial
- nuclear



"Co-adaptation occurs when a variation in a factor encoded by one compartment will select for a variation in a factor encoded by the other, due to physical interaction between the two factors." Rand, 2004

Cytoplasmic and nuclear genomes are expected to be co adapted

"We expect **variation in phenotypic traits** that can contribute to fitness differences between proper and impaired associations of variants." Rand 2004



Cytonuclear co adaptation : break it to see it



photo G. Pelletier

Biological questions

- Adaptive intra-specific variations in cytoplasm?
- > Which phenotypic traits?
- Impact of cytonuclear interactions?

Biological questions

- Adaptive intra-specific variations in cytoplasm?
- Which phenotypic traits?
- Impact of cytonuclear interactions?



The CYTOPHENO Project (2012-2016)

Detection of phenotypes affected by variations in organellar genomes and by the disruption of cytonuclear co adaptation



Arabidopsis cytolines: cytoplasm exchange between natural accessions



Experiment 1 Phenotyping Arabidopsis cytolines and their parents for adaptive traits in the field

questions

- 1. Do intraspecific natural variation in organelle genomes affect adaptive traits?
- 2. Are adaptive traits under the influence of cytonuclear co adaptation?

Arabidopsis adaptive traits in nature



Phenotyping of cytolines in the field

experimental design

5 Blocks



9 arrays of 66 pods/block



1 plant/pod

Randomized Complete Block Design

 $1 \operatorname{array} = 11L \times 6C$

Common garden, University of Lille (North of France)

Phenotyping of cytolines in the field

production of data



27 quantitative traits

Germination (5) Resource acquisition (3) Phenology (4) Architecture and seed dispersal (5) Fecundity (10) + survival (qualitative trait)

2745 pods sowed 2228 plants harvested

Experiment 2 Phenotyping Arabidopsis cytolines and their parents for seed physiological traits

question

Do cytonuclear interactions affect seed physiology?

E. Arc et al.



Phenotyping of seed traits : dormancy

experimental design

- 2 seed productions
- 2 plants at ≠ positions in the growth chamber for each production
- 64 genotypes (56 cytolines + 8 natural accessions)



seeds sowed in vitro and incubated at 15°C or 25°C

production of data

germination (gmax) scored after 96h, 2 technical replicates

other phenotyped seed traits



Experiment 3 Multi-omics phenotyping of Arabidopsis cytolines and their parents in two nitrogen nutrition conditions

question Does a disruption of cytonuclear coadaptation modify the molecular response of plants to a nutrition stress?

Molecular phenotyping in two nitrogen nutrition conditions



Phenotyping of seed traits : multi-omics on plants in two nitrogen nutrition conditions

experimental design

- 4 genotypes : 2 natural accessions and their reciprocal cytolines
- 2 N conditions (standard & starved)
- 6 productions (randomised designs)
 - 1 production : 4 plants/genotype/N condition



Molecular phenotyping in two nitrogen nutrition conditions



Statistical part of the project

Our motivation:

- To be involved in the project at the beginning
- A project with many interactions with biologists
- Several datasets of different nature to study the same question
- No research in statistics but opportunity to do proper statistics from A to Z

People:

- Tristan Mary-Huard
- Benjamin Vittrant (M2 internship)
- Priscilla Monfalet (M2 internship and 6 months of IE)

Our tasks:

- Design of the experiments
- Modelling of the three datasets to answer the questions
- Visualisation of the results
- Validating the biological interpretation of the results
- Writing the M&M of the papers

Phenotyping Arabidopsis cytolines and their parents

for adaptive traits in the field

questions

1. Do intraspecific natural variation in organelle genomes affect adaptive traits?

2. Are adaptive traits under the influence of cytonuclear co adaptation?

for seed physiological traits

question Do cytonuclear interactions affect seed physiology?

for molecular phenotyping in two nitrogen nutrition conditions

question Does a disruption of cytonuclear coadaptation modify the molecular response of plants to a nutrition stress?

Phenotyping Arabidopsis cytolines and their parents

for adaptive traits in the field

questions

1. Do intraspecific natural variation in organelle genomes affect adaptive traits?

2. Are adaptive traits under the influence of cytonuclear co adaptation?

for seed physiological traits

question

Do cytonuclear interactions affect seed physiology?

for molecular phenotyping in two nitrogen nutrition conditions

question Does a disruption of cytonuclear coadaptation modify the molecular response of plants to a nutrition stress?

> Several formulations of the SAME question Do the nucleus and the cytoplasm interact together ?

What is an interaction ?



24

Phenotyping of cytolines in the field

+

27 quantitative traits

Germination (5) Resource acquisition (3) Phenology (4) Architecture and seed dispersal (5) Fecundity (10)

1 qualitative trait Survival trait (1)





5 Blocks

9 arrays /block

 $1 \operatorname{array} = 11L \times 6C$

Gaussian error

 $Y_{bncij} = \mu + \alpha_b + \beta_n + \Upsilon_c + (\alpha\beta)_{bn} + (\beta\Upsilon)_{nc} + L_{i(b)} + C_{j(b)} + E_{bncij}$

Random effects within the block

$$Y_{bncij} = \mu + \alpha_b + \beta_n + \Upsilon_c + (\alpha\beta)_{bn} + (\beta\Upsilon)_{nc} + L_{i(b)} + C_{j(b)} + E_{bncij}$$

What traits are under a **genetic effect** ?

What traits are under a **cytoplasmic effect** ?

Are adaptive traits under the **influence of cytonuclear co adaptation**?

For (k,l) a couple of parents, we tested

 $[(\beta\Upsilon)_{kl} - (\beta\Upsilon)_{kk}] - [(\beta\Upsilon)_{lk} - (\beta\Upsilon)_{ll}] = 0 \quad \text{versus} \quad [(\beta\Upsilon)_{kl} - (\beta\Upsilon)_{kk}] - [(\beta\Upsilon)_{lk} - (\beta\Upsilon)_{ll}] \neq 0$

Global FDR control on all the pairs (k, l) and all the traits at 5%

Global effects of nucleus, cytoplasm, and cytonuclear interactions on phenotype

					Model t	erms					
	Blo	ck	Nucle	us	Cytop	lasm	Cytopl	asm× eus	Bloc	k × eus	Variance
Phenotypic class	F	Р	F	Р	F	Ρ	F	Ρ	F	Р	structure [†]
Germination											
Germination time	18.05	***	96.46	***	0.55	NS	2.94	***	5.36	***	hmg
Germination percentage 4 das	37.24	***	221.52	***	0.86	NS	1.72	**	4.02	***	hmg
Germination percentage 5 das	5.67	***	127.64	***	2.11	NS	2.63	***	2.25	***	htg
Germination percentage 6 das	3.04	*	125.78	***	1.3	NS	2.69	***	3.69	***	hmg
Germination percentage 13 das	15.46	***	52.55	***	4.65	***	1.7	**	2.53	***	htg
Resource acquisition											
Rosette surface area 28 das	11.45	***	259.35	***	5.29	***	2.8	***	1.82	*	hmg
Rosette perimeter 28 das	8.75	***	111.98	***	3.54	**	1.62	**	2.34	***	htg
Rosette diameter at flowering	3.57	*	33.93	***	2.59	*	1.62	**	1.83	*	hmg
Phenology											
Bolting time	1.75	NS	2648.29	***	3.1	*	2.53	***	2.8	***	htg
Flowering interval	2.38	NS	57.59	***	0.96	NS	1.55	*	2.98	***	htg
Reproductive period	5.65	**	32.1	***	1.09	NS	2.75	***	1.88	*	htg
Length of life cycle	12.7	***	273.23	***	5.21	***	5.43	***	5.02	***	htg
Architecture and seed dispersal											
Height from soil to the first fruit on the main stem [‡]	14.29	***	227.55	***	8.22	***	3.94	***	3.07	*	htg
Maximum height [*]	11.95	***	101.59	***	7.08	***	4.11	***	2.38	*	htg
Number of basal branches [‡]	0.28	NS	9.01	***	1.11	NS	0.67	NS	0.56	NS	hmg
Number of primary branches [‡]	0.78	NS	681.29	***	1.68	NS	1.87	**	1.88	NS	htg
Total number of branches [‡]	0.98	NS	491.58	***	1.71	NS	1.74	**	1.77	NS	htg
Fecundity											-
Total fruit length = proxy of total seed production [*]	4.36	**	54.04	***	0.54	NS	0.77	NS	1.04	NS	htg
Total fruit length on the main stem [‡]	0.7	NS	124.16	***	1.22	NS	0.92	NS	0.95	NS	htg
Fruit number on the main stem [‡]	1.6	NS	124.64	***	1.32	NS	0.96	NS	1.02	NS	htg
Mean fruit length on the main stem [*]	0.41	NS	210.03	***	1.28	NS	4.34	***	2.07	NS	htg
Total fruit length on primary branches [‡]	3.27	*	29.29	***	0.51	NS	0.99	NS	1.25	NS	htg
Fruit number on primary branches*	13.14	***	70.08	***	1.79	NS	1.87	**	2.42	***	htg
Mean fruit length on primary branches [*]	3.8	**	95.28	***	0.85	NS	3.76	***	1.56	NS	htg
Ratio of seeds produced on the main stem [‡]	3.84	**	31.76	***	2.33	NS	1.6	*	1.8	*	hma
Ratio of seeds produced on primary branches [‡]	3.62	*	29.91	***	2.26	NS	1.85	**	1.61	*	hma
Percentage of aborted of fruit [‡]	4.63	NS	85.51	***	1.64	NS	10.36	***	1.61	NS	htg
Survival	91.55	***	69.12	***	20.95	***	29.51	***	10.03	***	hma

Majority of adaptive traits are influenced by cytonuclear interactions

Traits	Nucleus effect	Cytoplasm effect	Cyto x nuc effect
Germination (5)	5	1	5
Resource acquisition (3)	3	3	3
Phenology (4)	4	2	4
Architecture and seed dispersal (5)	5	2	4
Fecundity (10)	10	0	6
Survival (1)	1	1	1

Sha and Ct-1 nuclear alleles have contrasted phenotype outputs in alien cytoplasmic backgrounds



Phenotyping of seed traits: dormancy Boussardon et al, 2019

Response variable: percentage of germinated seeds after 96h

- 2 seed productions
- 2 plants at ≠ positions in the growth chamber for each production
- 64 genotypes (8 nuclei and 8 cytoplasms)



seeds sowed in vitro and incubated at 15°C or 25°C

Large discrepancy observed between post-harvest times -> analysis per post-harvest time

Large number of missing data for some genotypes -> incomplete and unbalanced design

Five explicative factors (shelf, nucleus, cytoplasm, temperature) -> starting from the most complete model, nested models were sequentially fitted and the best one according to BIC was selected.

Then the contrasts relevant to the addressed question were tested p-values were adjusted using the Bonferroni procedure to control the family-wise error rate (FWER) at 5%.

Statistical analysis for a post-harvest time of 3 months

The selected model was the one with all the first and second order terms plus 4 third order interactions

- shelf x cytoplasm x nucleus (SCN)
- harvest x cytoplasm x nucleus (HCN)
- shelf x harvest x nucleus (SHN)
- harvest x temperature x nucleus (HTN)

Effect of each foreign cytoplasm (c) in each natural accession nuclear background (c', n') was tested with the null hypothesis using the contrast (C c - C c') + (CN cn' - CN c'n') $+ 1/2\Sigma t in \{1,2\} (TC tc - TC tc')$ $+ 1/2\Sigma h in \{1,2\} (HC hc - HC hc')$ $+ 1/2\Sigma h in \{1,2\} (HCN hcn' - HC hc'n')$ $+ 1/2\Sigma s in \{1,2\} (SC sc - SC sc')$ $+ 1/2\Sigma S in \{1,2\} (SCN hcn' - SC hc'n')$

Hence the cytoplasm effects in each nuclear background should be interpreted as averaged on the germination temperatures, the harvests and the shelves.

Statistical analysis for a post-harvest time of 3 months

The selected model was the one with all the first and second order terms plus 4 third order interactions

- shelf x cytoplasm x nucleus (SCN)
- harvest x cytoplasm x nucleus (HCN)
- shelf x harvest x nucleus (SHN)
- harvest x temperature x nucleus (HTN)

Effect of each foreign cytoplasm (c) in each natural accession nuclear background (c', n') was tested with the null hypothesis using the contrast (C c - C c') + (CN cn' - CN c'n') $+ 1/2\Sigma t in \{1,2\} (TC tc - TC tc')$ $+ 1/2\Sigma h in \{1,2\} (HC hc - HC hc')$ $+ 1/2\Sigma h in \{1,2\} (HCN hcn' - HC hc'n')$ $+ 1/2\Sigma s in \{1,2\} (SC sc - SC sc')$ $+ 1/2\Sigma S in \{1,2\} (SCN hcn' - SC hc'n')$

Hence the cytoplasm effects in each nuclear background should be interpreted as averaged on the germination temperatures, the harvests and the shelves.

Statistical analysis for a post-harvest time of 3 months

The selected model was the one with all the first and second order terms plus 4 third order interactions

- shelf x cytoplasm x nucleus (SCN)
- harvest x cytoplasm x nucleus (HCN)
- shelf x harvest x nucleus (SHN)
- harvest x temperature x nucleus (HTN)

An effect of the cytonuclear interacting combination was tested for all pairs of cytoplasms (c, c') and for all pairs of nuclei (n, n').

As no third order interaction involving both nucleus and cytoplasm was included in the selected model, cytonuclear interacting combinations could be tested independently from other model terms.

novel cytonuclear combinations can modify dormancy depth and release



novel cytonuclear combinations can modify germination performance



Diversité du nombre d'interactions significatives impliquant chaque fonds nucléaire et chaque fonds cytoplasmique

Chaque noyau et chaque cytoplasme est testé dans 196 combinaisons



ns

Interact° sign avec ce noyau

Interact° sign avec ce cytoplasme

new cytonuclear combination can enhance seed longevity



Molecular phenotyping in two nitrogen nutrition conditions

Chardon et al, 2020

nuclear transcriptome microarray hybridizations 26884 nuc genes

organellar transcriptomes qRT-PCR 31 mt & 80 cp genes

> proteome LC MS/MS 665 proteins

metabolome GC-TOF-MS 81 metabolites Specific pre-processing

Normalisation step to remove technical biaises

Signal analysis to answer the biological questions with an ANOVA

p-values were adjusted to control FDR at 5%

The scripts and datasets used are available as a git project https://forgemia.inra.fr/GNet/cytopheno_omics

The main task was to translate the biological questions into contrasts

Molecular phenotyping in two nitrogen nutrition conditions

Chardon et al, 2020

Each type of accumulation was modeled with a three-way ANOVA

$$Y_{ijkr} = \mu + C_i + N_j + A_k + CN_{ij} + CA_{ik} + NA_{jk} + CNA_{ijk} + E_{ijkr}$$

the nitrogen supply across both cytoplasms and both nuclei,

mean(all genotypes in nitrogen starvation) – mean(all genotypes in control nitrogen supply).

the nucleus origin across both nitrogen supplies and both cytoplasms mean(genotypes with Jea nucleus in both nitrogen conditions) – mean(genotypes with Ct-1 nucleus in both nitrogen conditions)

a cytoplasm × nucleus interaction effect across both nitrogen supplies mean(both cytolines in both nitrogen conditions) –

mean(both parental lines in both nitrogen conditions).

a cytoplasm × nuclear × nitrogen interaction effect.

[mean(both parental lines in nitrogen starvation) – mean(both parental lines in control nitrogen supply)] – [mean(both cytolines in nitrogen starvation) – mean(both cytolines in control nitrogen supply)].

 $E(Y_{entr}) = \mu + C_{c} + N_{n+} T_{t} + (C_{N})_{en} + (C_{T})_{ct} + (N_{T})_{nt} + (C_{N_{T}})_{ent}$ avec $\begin{pmatrix} C_{ct} = 0 & N_{ct} = 0 & T_{NO} = 0 & (CN)_{ct,n} = 0 & (CT)_{ct,t} = 0 \\ c, ct = 0 & (CT)_{c,NO} = 0 \end{pmatrix}$ contraintes $(NT)_{cl,t} = 0 \qquad (CNT) = 0$ Cl,n,t $C,No = 0 \qquad Cl,n,t$ C,CL,td'identifia bilité. $\begin{array}{c} c_{i}c_{t} \\ c_{i}n_{i}N_{o} = O \end{array}$ le veut dire que le modèle à

le = l'intercept = expression de Ct: Ct en condition NO C_{sea} = (Jea: Ct) No - (Ct: Ct) No = la différence d'expression entre Jea: Ct et Ct: Ct en NO NJea = (Ct: Jea) No - (Ct: Ct) No = la différence d'expression entre Ct: Jea et Ct: Ct en NO TNy = (Ct: Ct) Ny - (Ct: Ct) No = la différence d'expression entre N4 et NO pour Ct: Ct (CN) Jea, Jea (CT) Jea, Ny = la différence d'expression entre N4 et No pour Jea: Ct et la m différence (NT) Jea, Ny = la différence d'expression entre N4 et No pour Jea: Ct et la m différence (NT) Jea, Ny = la différence d'expression entre N4 et No pour Jea: Ct et la m différence (NT) Jea, Ny = la différence d'expression entre N4 et No pour Jea: Ct et la m différence (NT) Jea, Ny = la différence d'expression entre N4 et No pour Jea: Ct et la m différence (NT) Jea, Ny = la différence d'expression entre N4 et No pour Ct: Jea et la m différence (NT) Jea, Ny = la différence d'expression entre N4 et No pour Ct: Jea et la m différence (NT) Jea, Ny = la différence d'expression entre N4 et No pour Ct: Jea et la m différence (NT) Jea, Ny = la différence d'expression entre N4 et No pour Ct: Jea et la m différence (NT) Jea, Ny = la différence d'expression entre N4 et No pour Ct: Jea et la m différence (NT) Jea, Ny = la différence d'expression entre N4 et No pour Ct: Jea et la m différence d'expression entre N4 et No pour Ct: Jea et la m différence d'expression entre N4 et No pour Ct: Jea et la m différence d'expression entre N4 et No pour Ct: Jea et la m différence d'expression et Jea et la m différence pour Ct

(CNT) sea, sea, sur = [la somme des parents]-(la somme des cytolignées)] en N4- [en N6



cytolines have a sensible different molecular response to N starvation than their natural parents



cytoplasm x nucleus x nitrogen interaction (604)

nuclear DEGs

cytolines have a sensible different molecular response to N starvation than their natural parents



nuclear DEGs

general conclusions and take home messages

Interactions between nuclear and organellar genomes are worth considering

- ✓ Intraspecific variation in cytoplasm genomes is relevant for adaptive phenotypes
- \checkmark Traits impacted by cytonuclear interactions are relevant to plant breeding
- ✓ New cytonuclear genetic combination may have improved performances



- importance of open and frequent exchanges -> improvement of mutual understanding
- \succ improvement of the biologist's skills in statistical modeling \odot
- improvement of the statistician's skills in biology
- unexpected outputs from statistical analyses





acknowledgements

