



**HAL**  
open science

# Régulations chromatinienne et transcriptionnelles impliquées dans le cycle de vie du puceron du pois

Gautier Richard

► **To cite this version:**

Gautier Richard. Régulations chromatinienne et transcriptionnelles impliquées dans le cycle de vie du puceron du pois. Génétique des plantes. Agrocampus Ouest, 2017. Français. ⟨NNT : 2017NSARB130⟩. ⟨tel-01937224⟩

**HAL Id: tel-01937224**

**<https://theses.hal.science/tel-01937224v1>**

Submitted on 28 Nov 2018

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

# AGRO CAMPUS

---

## OUEST

**Gautier RICHARD** • 20 octobre 2017

Thèse AGROCAMPUS OUEST  
sous le label de l'Université Bretagne Loire  
pour obtenir le grade de  
**DOCTEUR D'AGROCAMPUS OUEST**  
Spécialité Génétique, génomique et bio-informatique

ÉCOLE DOCTORALE • Écologie, Géosciences, Agronomie  
et Alimentation (EGAAL)  
LABORATOIRE D'ACCUEIL • INRA - Institut de génétique,  
environnement et protection des plantes (UMR IGEPP)

## Régulations chromatinienne et transcriptionnelles impliquées dans le cycle de vie du puceron du pois

**Maria MANZANARES-DAULEUX**

Professeure, AGROCAMPUS OUEST, UMR INRA - AO - UR1 IGEPP /  
*présidente*

**Sébastien BLOYER**

Professeur, Université Paris Sud / *rapporteur*

**Nadia PONTS**

Chargée de recherche, INRA Bordeaux Aquitaine / *rapporteur*

**Emmanuelle JACQUIN-JOLY**

Directrice de recherche, INRA Versailles Grignon / *examinatrice*

**Stéphanie JAUBERT-POSSAMAI**

Chargée de recherche, INRA PACA / *examinatrice*

**Denis TAGU**

Directeur de recherche, UMR INRA - AO - UR1 IGEPP / *directeur de  
thèse*

**Gaëli LE TRIONNAIRE**

Chargée de recherche, UMR INRA - AO - UR1 IGEPP / *co-encadrant*

**Thèse de Doctorat** AGROCAMPUS OUEST  
pour obtenir le titre de DOCTEUR  
d'AGROCAMPUS OUEST Spécialité Biologie  
et Agronomie

École Doctorale : Vie-Agro-Santé  
Laboratoire d'accueil : UMR Institut de  
Génétique, Environnement et Protection des  
Plantes (IGEPP, INRA, Agrocampus Ouest,  
Université Rennes 1), Ecologie et Génétique  
des Insectes (EGI)

# Régulations chromatinienne et transcriptionnelles impliquées dans le cycle de vie du puceron du pois

Gautier RICHARD • 20 Octobre 2017

Présentée devant la commission d'examen composée de :

Présidente : Maria Manzanares

Rapporteuse : Nadia Ponts

Rapporteur : Sébastien Bloyer

Examinatrice : Emmanuelle Jacquin-Joly

Examinatrice : Stéphanie Jaubert-Possamai

Directeur de thèse : Denis Tagu

Maître de thèse : Gaël Le Trionnaire

Professeure, Agrocampus Ouest, UMR IGEPP, Rennes

Chargée de Recherche, INRA, UR MycSA, Bordeaux

Professeur, Université Paris-Sud, UMR I2BC CNRS, Paris

Directrice de recherche, INRA, UMR iEES, Versailles

Chargée de recherche, INRA, UMR ISA, Sophia Antipolis

Directeur de Recherche, INRA, UMR IGEPP, Rennes

Chargé de Recherche, INRA, UMR IGEPP, Rennes

# Remerciements

Je tiens tout d'abord à remercier l'ensemble des membres du jury et mes deux rapporteurs, Nadia Ponts et Sébastien Boyer, ainsi que l'ensemble des membres de mon comité de thèse pour le soutien qu'ils m'ont apporté durant mes 3 années de thèse.

Bien évidemment, j'adresse un immense merci à mes encadrants, Denis Tagu et surtout Gaël Le Trionnaire qui a su me supporter, me motiver dans les coups durs, et me transmettre son énergie, sa passion de la science et bien plus encore durant ces 3 ans et demi (il ne faut pas oublier le stage de M2 !).

Merci à mes trois techniciennes préférées, Sylvie T., Nathalie et Sylvie H., et merci à mes deux bioinformaticiens préférés, Fabrice et Anthony, sans qui cette thèse n'aurait pas pu être possible.

Un immense merci à mes parents, Corinne et Olivier, et à mon frère Thibaud, sans eux et l'amour qu'ils m'ont apporté, je ne serais pas là où j'en suis actuellement. Merci à mes beaux-parents Daniel et Dominique et bien évidemment à ma chérie d'amour, Laure, pour tout le soutien et l'amour qu'ils ont su m'apporter durant ma thèse et l'ensemble de mes études.

Un immense merci à tous les collègues et amis de l'INRA : Benoît (tu me manqueras mon pote, mais on est jamais très loin t'inquiète ! Toujours chaud pour une partie de babyfoot !), Lydie, Yoann A., Yoann B., Julien, JB, Camille, David, Marion, Paola, Morgane H., Thibaut et Thibault (les 3 orthographes dans la même page !).

Un grand merci à tous mes potes Rennais pour tous ces moments de rigolade et d'amitié : David, Morgane R., Florent, Étienne, FX, Geoffroy, Thomas, Heathcliff, Axel et Ludo.

Un bisou sur la truffe et une caresse aux différents animaux de compagnie de la famille : Tessa, Moumousse (vous nous manquerez à tous), Lilou, et plus récemment Odin (pour les titounets), Choupi et Pompon (pour les chats !).

Merci à mes sponsors : Spotify, Discord, Steam, Bonzini et René Pierre.


Merci à Risitas et son inspiration quotidienne, et la chancla.

# Table des matières

"	INTRODUCTION GENERALE.....	4
'''	SYNTHESES BIBLIOGRAPHIQUES.....	8
!	<i>Les pucerons : des ravageurs majeurs des cultures agronomiques.....</i>	8
	<i>Cycle de vie et alternance du mode de reproduction.....</i>	10
	<i>Traits d'histoire de vie et plasticité phénotypique.....</i>	13
	<i>Le chromosome X chez les pucerons : cytologie et compensation.....</i>	20
	<i>Bases moléculaires du polyphénisme de reproduction.....</i>	21
.	<i>Unités de transcriptions.....</i>	25
	<i>Chromatine et régions régulatrices longue distance.....</i>	27
	<i>Accessibilité de la chromatine.....</i>	31
	<i>Modifications post-traductionnelles des histones.....</i>	35
	<i>Structure tridimensionnelle de la chromatine : définition et méthodes d'études.....</i>	40
	<i>Méthylation de l'ADN.....</i>	42
	<i>ARN non-codants.....</i>	45
	<i>Épigénétique chez les pucerons.....</i>	49
/	<i>Définition de la compensation de dose.....</i>	50
	<i>Les mécanismes épigénétiques participant à la compensation de dose chez les organismes modèles.....</i>	52
	<i>Étude de la compensation de dose chez les insectes non-drosophilides.....</i>	56
5	<i>Le polyphénisme de castes chez l'abeille.....</i>	57
	<i>Le polyphénisme de castes chez les fourmis.....</i>	59
	<i>Le polyphénisme de phase chez le criquet.....</i>	60
'''	OBJECTIFS ET STRATEGIES.....	62


V. MATERIEL ET METHODES.....	64
! /v - lq lhb 5- a ! m9w9[ . lh[hDLv - 9	73
1. Compensation de dose.....	<i>Error! Bookmark not defined.</i>
2. Polyphūnisme de reproduction.....	65
. Chwa ! [59I B59! {{lq m95 lq h[! mhb hCVW9D- [! mhW9[9a 9b m{ {- l' L5 b {9v - 9b4! D9I! - m [59. lq C lW9[9v、	75
1. Protocole utilisū.....	68
2. Mūthodes d'analyse bioinformatique de donnūes FA'RE-seq.....	70
VI. CHAPITRE " : COMPENSATION DE DOSE ET ACCESSIBILITE SEXE-SPECIFIQUE DU CHROMOSOME X CHEZ ACYRTHOSIPHON P'SUM.....	74
! lq m95- / mhb	76
1. Contexte.....	74
2. Rappel des objectifs et stratūgie.....	75
. t w9a l9w! w m l [9	78
/ 5 lq / - {{ lhb	80
1. Absence de donnūes de FA'RE-seq chez les femelles sexuūes.....	108
2. Challenge bioinformatique : des donnūes de qualitū moyenne.....	110
3. Identification de longs ARN non-codants mōes-biaisū.....	112
VII. CHAPITRE " : REGULATIONS CHROMATIENNES IMPLIQUEES DANS LE POLYPHENISME DE REPRODUCTON DU PUCERON DU PO'S.....	115
! lq m95- / mhb	117
1. Contexte.....	115
2. Rappel des objectifs et stratūgie.....	115
. {9/ hb5 ! w m l [9	118
/ W9[- [m m{ / ha t [9a 9b m l W9{	120
1. RNA-seq : un potentiel manque de profondeur de sūquenūage par rapport aux donnūes FA'RE-seq.....	149
2. Identification de sites de fixation de facteurs de transcription.....	149

V". DISCUSSION GENERALE..... 154

!  [! a 9m h59 C lW9v d! b D9{hb/hb1 9b9b{ 9m [la Lp9{ ..... 154


1. Choix de la méthode FA"RE-seq vis-à-vis du DNase-seq et de l'ATAC-seq..... 154

2. L'analyse de données FA"RE-seq..... 157

.  [9 D9bha 95 d a b9! bbh{ hb 9m - b! {{9a . [! D9lb/ha t[9m{ ..... 162

1. Le génome d *A. pisum* : état de l'art..... 162

2. L'incidence des limites du génome d *A. pisum* sur l'analyse de données de régulation de la transcription..... 163

/  / D[959' l959{ t- /9whb{ d h[ d I 9b l a 99m /ha t9b{! hb 595h{9 ..... 170

1. Un mécanisme de compensation de dose permettrait d'équilibrer la transcription du chromosome X vis-à-vis des autosomes chez les pucerons mâles ..... 170

2. Le polyphénisme de reproduction des pucerons est potentiellement associé à différentes voies de signalisation neuro-endocrine ..... 172

V"". PERSPECTIVES ET CONCLUSION..... 178

"X. REFERENCES ..... 179

## 1. Introduction g n rale

Les organismes vivants doivent constamment et rapidement faire face aux fluctuations constantes de leur environnement biotique et abiotique pour assurer leur survie. Si le patrimoine g n tique dont ils h ritent est fixe, il leur est n anmoins possible de moduler leur ph notype pour mieux s'adapter aux contraintes environnementales. Cette modulation du ph notype peut notamment se faire par l'expression diff rentielle rapide et coordonn e de diff rents sets de g nes,   savoir des programmes g n tiques alternatifs. Il leur est ainsi possible de produire   partir d'un g notype unique une continuit  de ph notypes sp cifiquement adapt s   un ensemble de conditions environnementales diff rentes et fluctuantes. Un tel m canisme adaptatif est commun ment appel  plasticit  ph notypique et g n re des normes de r action, d finies comme la production par un g notype donn  d'une gradation continue de ph notypes en r ponse   la modulation d'un facteur environnemental [Griffiths *et al.*, 2000].

Chez l'homme la plasticit  ph notypique peut par exemple permettre d'adapter la musculature d'un individu selon le r gime d'exercices auquel il est confront  en g n rant une continuit  de ph notypes musculaires adapt s   ces diff rents efforts, et ce   partir d'un g notype fixe [Hoppeler et Fl ck, 2002]. Ainsi, en r ponse   des exercices d'endurance et de faible intensit , un individu pr sentera des changements qualitatifs de fibre musculaire caract ris s par une augmentation du volume des structures transportant et consommant l'oxyg ne, tandis qu'un individu r alisant des exercices intenses de force brute comportera une musculature caract ris e par une quantit  plus importante de prot ines contractiles [Hoppeler *et al.*, 2011] comme l'actine et la myosine [Barden *et al.*, 1988]. Ces ph notypes musculaires diff rents et adapt s au type d'effort sont ainsi permis par la mise en place de programmes g n tiques.

La plasticit  ph notypique est  galement importante chez les v g taux qui sont des organismes immobiles pour qui une adaptation aux conditions environnementales est donc particuli rement cruciale [Gratani, 2014]. Chez la plante mod le *Arabidopsis thaliana*, en cas de d veloppement en milieu salin, un g notype donn  produira une continuit  de ph notypes suivant la concentration en sel du milieu. Ces ph notypes correspondent notamment   une r duction globale de la taille de la plante, une floraison tardive, une diminution du nombre de ramifications, ainsi que des sections entre n uds et des inflorescences plus courtes [Kooke *et al.*, 2015], l'ensemble de ces ph notypes adaptatifs permettant ainsi   la plante de r sister   des concentrations salines  lev es.

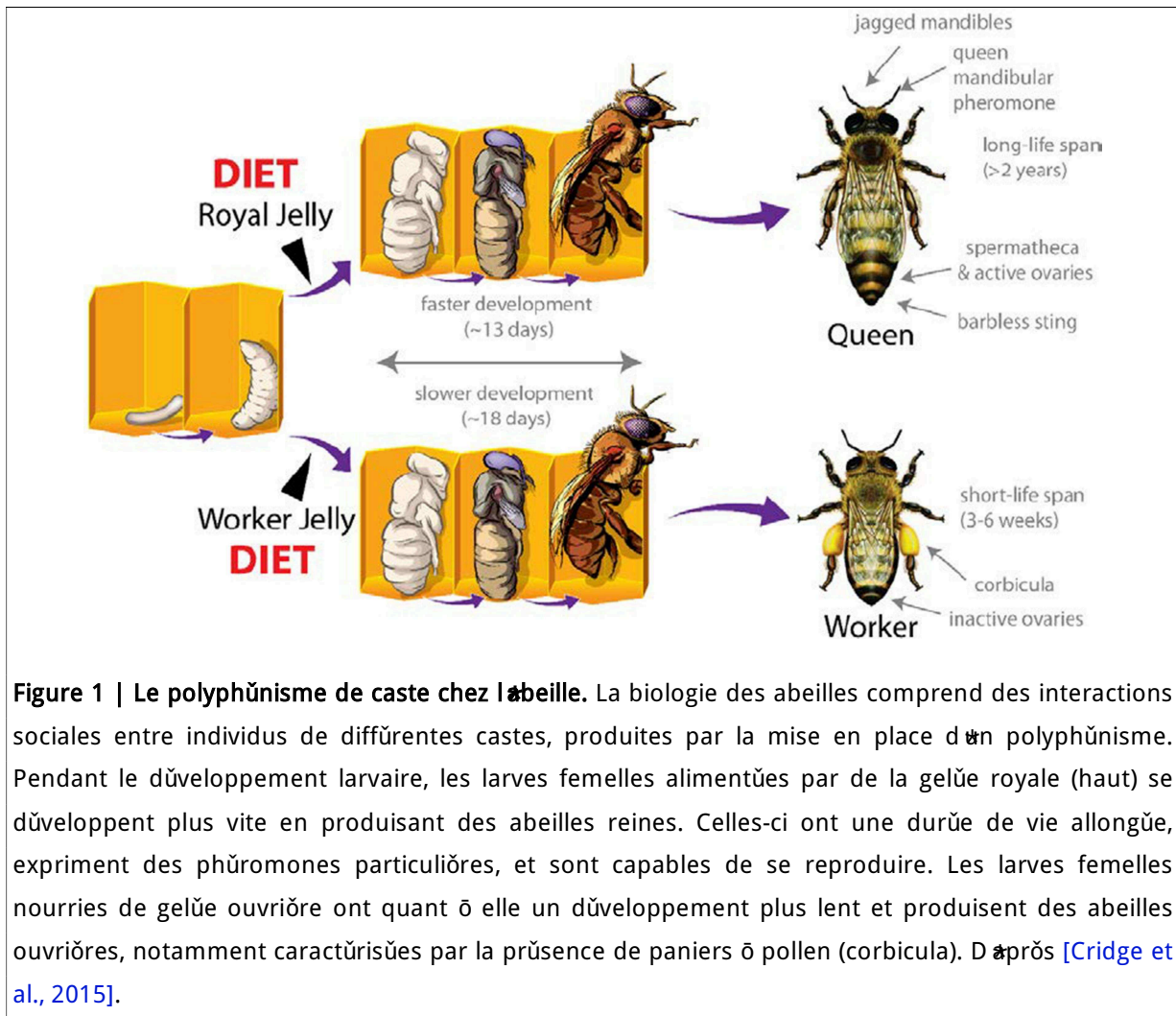
Les organismes ectothermes, comme par exemple les insectes, adaptent la taille de leur corps   la temp rature de leur environnement [Stevenson, 1985]. Cette plasticit  ph notypique se traduit par une augmentation de la taille du corps d'un individu adulte lorsque celui-ci est expos    de basses temp ratures au cours de son d veloppement, et r sulte d'une augmentation de la taille des cellules plus que d'une augmentation de leur nombre. Ce m canisme permet de limiter les pertes de chaleur dues   de faibles temp ratures en r duisant proportionnellement la surface de contact du corps des

ectothermes avec l'air froid environnant [French *et al.*, 1998]. Chez la Drosophile, la taille du corps et la taille des ailes doivent rester proportionnelles afin de maintenir une stabilitŃ lors du vol. Ainsi, les Drosophiles ŃlevŃes Ń de faibles tempŃratures prŃsentent des ailes plus grandes que chez des Drosophiles ŃlevŃes dans des tempŃratures mŃdianes [Stanley, 1935], et cette plasticitŃ phŃnotypique est associŃe Ń une augmentation de la taille des cellules constituant les ailes [De Moed *et al.*, 1997].

Concernant ces exemples, et dans la majoritŃ des cas observŃs dans la nature, la plasticitŃ phŃnotypique est continue. Dans d'autres cas plus rares et spectaculaires appelŃs polyphŃnismes [Mayr, 1963], la plasticitŃ phŃnotypique est discrŃte, avec la production d'un nombre dŃfini de phŃnotypes par un seul et mŃme gŃnotype en rŃponse Ń un facteur environnemental donnŃ. Les exemples les mieux caractŃrisŃs chez les insectes correspondent au polyphŃnisme de caste chez les abeilles (Figure 1), comme par exemple *Apis mellifera* [Evans et Wheeler, 1999], ou chez les fourmis telles que *Camponotus floridanus* [Bonasio, 2014]. Ainsi, les colonies sont composŃes d'individus ayant diffŃrents phŃnotypes discrets associŃs Ń des fonctions spŃcifiques au sein de la ruche ou de la fourmiliŃre. Ces castes peuvent par exemple correspondre Ń des ouvriŃres, des soldats, des exploratrices ainsi qu'Ń des reines, lesquelles sont les seules capables de se reproduire au sein de la ruche ou de la fourmiliŃre, et donc d'assurer la survie de la colonie. Chez ces deux organismes, le rŃgime nutritif auquel sont soumises les larves est le facteur environnemental majeur permettant d'orienter leur devenir et faŃ onner leur phŃnotype Ń l'Ńge adulte. Ainsi chez les abeilles, selon que les larves soient nourries avec de la gelŃe de type royale ou commune, qui diffŃrent notamment par leur composition en Acide (E)-10-hydroxy-2-dŃcŃnoŃque (10HDA) [Viuda-Martos *et al.*, 2008], des programmes gŃnŃtiques alternatifs vont Ńtre mis en place et aboutir Ń l'expression des phŃnotypes reine ou ouvriŃre, respectivement. Selon deux conditions nutritives distinctes, deux phŃnotypes prŃsentant des caractŃristiques morphologiques et fonctionnelles trŃs diffŃrentes sont donc produits Ń partir d'un mŃme gŃnotype (Figure 1).

De faŃ on plus gŃnŃrale, les polyphŃnismes chez ces insectes font d'abord intervenir la perception d'un signal environnemental par l'organisme suivie de sa transduction jusqu'Ń des tissus cibles, et aboutissant Ń une reprogrammation de l'expression des gŃnes au niveau de ces tissus. Chez les fourmis et les abeilles, il a ŃtŃ montrŃ que la mise en place de ces reprogrammations transcriptionnelles rapides impliquait des rŃgulations ŃpigŃnŃtiques. Ces derniŃres sont dŃfinies comme une modification de l'expression hŃrŃditaire des gŃnes sans modification de la sŃquence d'ADN, gŃnŃralement en rŃponse Ń l'environnement [Waddington, 1942, 2012]. D'aprŃs l'exemple prŃcŃdemment dŃcrit, les mŃcanismes ŃpigŃnŃtiques peuvent par exemple participer au changement du phŃnotype d'un individu, en rŃponse Ń des stimuli environnementaux et sans modification de son gŃnotype. Elle implique diffŃrents mŃcanismes tels que la mŃthylation de l'ADN (cytosines notamment), les modifications post-traductionnelles des histones, ou encore l'expression de certains ARN non-codants. Chez *A. mellifera*, un niveau de mŃthylation diffŃrentiel a par exemple ŃtŃ identifiŃ entre des larves nourries avec de la gelŃe royale et d'autres nourries avec de la gelŃe commune [Foret *et al.*, 2012]. Chez *C. floridanus*, un niveau diffŃrentiel global d'acŃtylation de la lysine 27 des histones H3 (H3K27ac) a par exemple ŃtŃ identifiŃ entre les castes ouvriŃres majeure et mineure [Simola *et al.*, 2013b], qui se diffŃrencient par leur comportement et les tŃches qu'elles accomplissent au sein de la

fourmili re. En revanche, la contribution de m canismes  pig n tiques dans la mise en place du polyph nisme de reproduction chez les pucerons n'a jusqu'alors pas  t   tudi e. On peut ainsi  mettre l'hypoth se que la signalisation neuroendocrine d'une diminution de la photop riode au sein d'individus asexu s peut  tre   l'origine d'une modification rapide de l'organisation de la chromatine des tissus cibles  ici, les lign es germinales embryonnaires   l'aide de r gulations  pig n tiques permettant ainsi la mise en place des ph notypes sexu s.



**Figure 1 | Le polyph nisme de caste chez l'abeille.** La biologie des abeilles comprend des interactions sociales entre individus de diff rentes castes, produites par la mise en place d'un polyph nisme. Pendant le d veloppement larvaire, les larves femelles aliment es par de la gel e royale (haut) se d veloppent plus vite en produisant des abeilles reines. Celles-ci ont une dur e de vie allong e, expriment des ph romones particuli res, et sont capables de se reproduire. Les larves femelles nourries de gel e ouvri re ont quant   elle un d veloppement plus lent et produisent des abeilles ouvri res, notamment caract ris es par la pr sence de paniers   pollen (corbicula). D'apr s [Cridge et al., 2015].

De fa on concomitante, le polyph nisme de reproduction des pucerons a pour cons quence la production de trois morphes sexuels au cours du cycle de vie annuel des pucerons : des femelles asexu es vivipares, des femelles sexu es ovipares et des m les. Ces trois morphes sont g n tiquement identiques   l'exception de leurs chromosomes sexuels, puisque les deux types de femelles comportent deux chromosomes X, tandis que les m les n'en comportent qu'un [Blackman, 1987]. Les pucerons correspondent ainsi   une esp ce h t rogam tique m le, comportant un syst me chromosomique du type XX / XO. Un syst me d s quilibr  comme celui-ci pose la question

inhŕrente ō toute espŕce hŕtŕrogamŕtique : qu'en est-il de la rŕgulation du niveau global de l'expression des gŕnes portŕs par le(s) chromosome(s) X vis-ŕ-vis des gŕnes portŕs par les autosomes chez les mŕles et les femelles, puisque ces derniers n'ont pas le mŕme nombre de allŕles pour chaque gŕne sur le chromosome X ? Chez plusieurs organismes hŕtŕrogamŕtiques, des mŕcanismes dits de compensation de dose ont ŕtŕ identifiŕs, impliquant gŕnŕralement des modifications ŕpigŕnŕtiques permettant d'ŕquilibrer la transcription des gŕnes portŕs par le chromosome X entre mŕles et femelles. On sait que chez le puceron du pois, les gŕnes du seul chromosome X des mŕles sont en moyenne plus exprimŕs que ceux portŕs par les deux X des femelles pour des gŕnes autosomaux globalement exprimŕs de maniŕre ŕquivalente [Jaquiŕry *et al.*, 2013]. On peut donc ici ŕmettre l'hypothŕse qu'il existerait un mŕcanisme global de rŕgulation de l'expression des gŕnes portŕs par le chromosome X, putativement de compensation de dose, impliquant des rŕgulations ŕpigŕnŕtiques chez les pucerons.

Les travaux rŕalisŕs au cours de ma thŕse visent donc ō ŕtudier la rŕsultante des mŕcanismes ŕpigŕnŕtiques sur l'accessibilitŕ de la chromatine et la transcription des gŕnes au niveau du gŕnome de *Acyrtosiphon pisum* (puceron du pois) - espŕce aphidienne modŕle utilisŕe pour cette ŕtude - dans le cadre : i) de la potentielle compensation de dose pouvant avoir lieu au niveau des gŕnes portŕs par le chromosome X des mŕles et des femelles, et ii) de la modification des programmes gŕnŕtiques ō l'origine de la trajectoire de dŕveloppement sexuŕ ou asexuŕ des embryons dans le cadre du polyphŕnisme de reproduction.

Nous avons dans un premier temps ŕtudiŕ les profils d'expression du chromosome X entre mŕles et femelles (par RNA-seq) puis le niveau global d'accessibilitŕ de la chromatine de ce chromosome entre mŕles et femelles, via notamment l'utilisation de la technique du FA<sup>3</sup>RE-seq (Formaldehyde Assisted Resolution of Regulatory Elements suivi d'un sŕquenŕage haut-dŕbit) afin de tester l'hypothŕse de l'existence d'un mŕcanisme global de compensation de dose. Dans un second temps, nous avons comparŕ par RNA-seq les transcriptomes d'embryons ō devenir sexuŕ ou asexuŕ puis ŕtudiŕ le niveau global d'ouverture de leur chromatine par FA<sup>3</sup>RE-seq afin d'identifier et caractŕriser des rŕgions du gŕnome diffŕrentiellement accessibles et potentiellement impliquŕes dans la mise en place des phŕnotypes alternatifs.

Afin de rŕpondre ō ces questions nous introduirons tout d'abord les ŕlŕments de bibliographie en relation avec ces diffŕrentes questions, puis nous prŕsenterons une vision synthŕtique des techniques molŕculaires et bioinformatiques communes ō l'ensemble de ces questions. Par la suite, nous prŕsenterons et discuterons les principaux mŕthodes et rŕsultats dŕveloppŕs durant cette thŕse pour conclure sur leurs implications dans le domaine de l'ŕtude des mŕcanismes ŕpigŕnŕtiques chez les pucerons dans le cadre de son cycle de vie et de son mode reproduction.

## ". Synthèses bibliographiques

### A. Biologie des pucerons

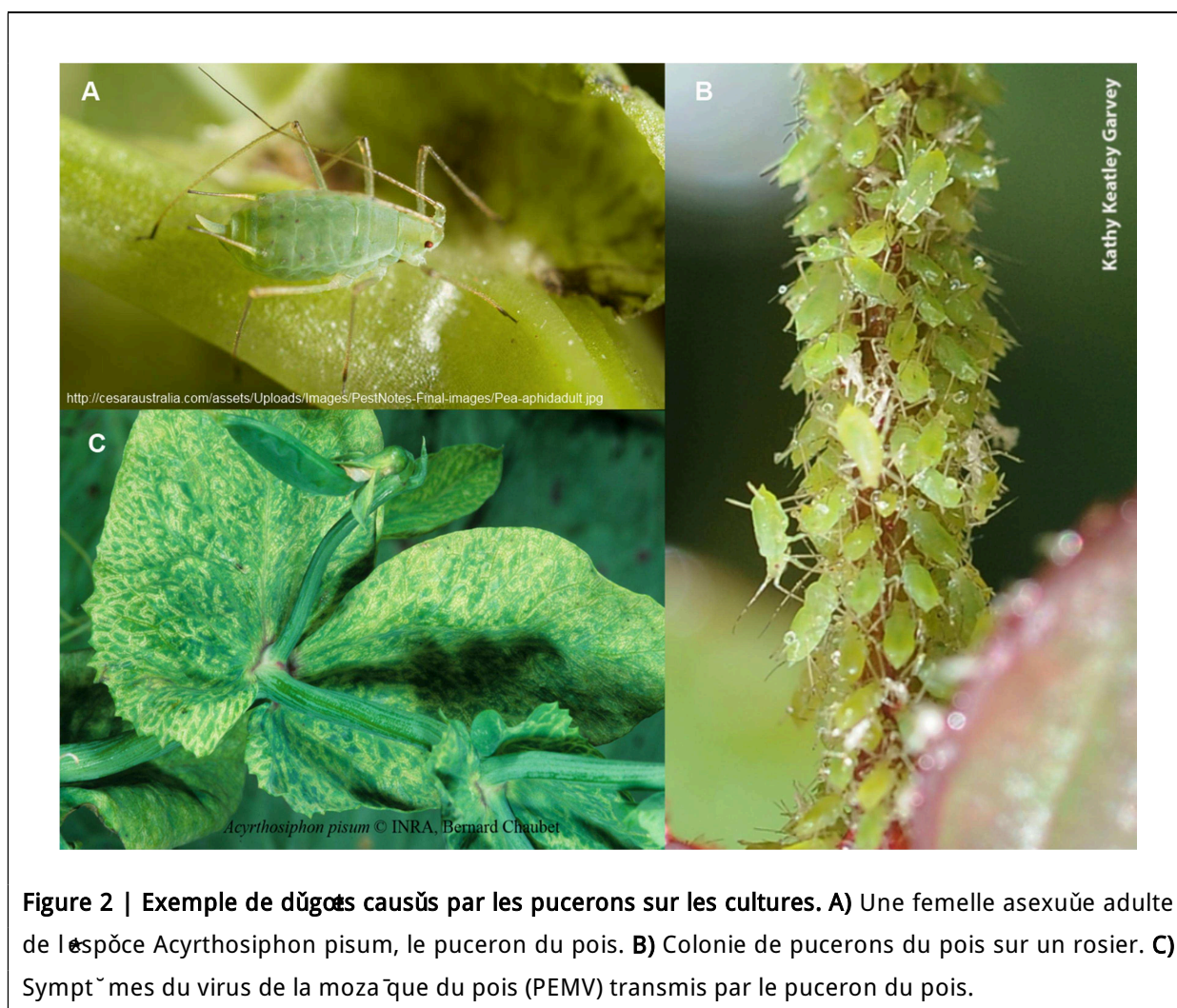
#### 1. Les pucerons : des ravageurs majeurs des cultures agronomiques

Les pucerons sont des insectes hımıptöres de la famille des *Aphididae* comptant plus de 4700 espèces identifiées ö ce jour. Ce sont des insectes piqueurs-suceurs de plante se nourrissant exclusivement de la sève ũlaborũe issue du phloöme. Ces insectes sont ö l'origine de dũgöts importants sur des plantes de grandes cultures (cũrũales, fabacũes v), sur certaines plantes horticoles et sur des arbres forestiers ou fruitiers. Ces insectes provoquent deux grands types de dũgöts sur les plantes au dũpend desquelles ils se nourrissent (Figure 2).

##### a. Dũgöts directs

Le stylet des pucerons est l'organe qui permet leur nutrition : cette structure pũnötre les tissus vũgũtaux jusqu'au phloöme et permet de prũlever la sève ũlaborũe, provoquant ainsi des dommages directs sur la plante (blessure, dũtournement de ressources nutritives). Avant de passer en phase de nutrition ö proprement parler pouvant durer de quelques heures ö plusieurs jours, le puceron effectue de nombreuses piqũres d'essai afin de valuer l'acceptabilitũ de l'hôte vũgũtal. Si l'hôte convient, le stylet du puceron va atteindre le phloöme en s'infiltant entre la paroi des cellules vũgũtales de l'epiderme, afin d'initier la phase de nutrition permettant une absorption de nombreux photosynthũtats. Ces derniers ũtant essentiels au dũveloppement de la plante, les pucerons peuvent ainsi entraıner une croissance ralentie, une mauvaise fructification, ou la production de grains ou de fruits de taille plus faible et aux valeurs nutritionnelles et commerciales amoindries. En outre, au cours de la piqũre, les pucerons sũcrötent une salive liquide et une salive visqueuse, accompagnũes de composũs chimiques et de peptides agissant sur la plante en perturbant ses rũactions de dũfense, empũchant notamment la rũaction normale des tubes criblũs du phloöme en cas de blessure, ö savoir une occlusion dũpendante du calcium [Tjallingii et Esch, 1993 ; Miles, 1999 ; Will *et al.*, 2007].

Enfin, le mode de nutrition des pucerons entraıne l'excrũtion d'un miellat composũ principalement de fructose et glucose extraits du phloöme et non assimilũs par le puceron [Lamb, 1959]. Ce miellat peut d'une part entraıner le dessũchement d'organes qui en sont recouverts, et il peut d'autre part promouvoir le dũveloppement de champignons causant des fumagines, rũduisant ainsi l'activitũ photosynthũtique des feuilles atteintes [Dedryver, 2007].



### b. Dommages indirects

Les pucerons provoquent également des dommages indirects à leurs plantes hôtes par la transmission de phytovirus lors de leur phase de nutrition, en se dupliquant et se nourrissant d'une plante à l'autre. Les symptômes causés par ces infections sont dépendants du virus transmis et sont très divers, pouvant aller d'une simple décoloration des feuilles (chlorose) à la mort de la plante. Par exemple, le virus de la jaunisse nanisante de l'orge (BYDV) est transmis par plus de 25 espèces de pucerons [Du *et al.*, 2007] et est responsable de pertes de récoltes allant de 10 à 85% par parcelle [Brault *et al.*, 2007]. Les phytovirus circulant au sein du phloème de la plante peuvent être transmis par les pucerons de manière circulante ou non-circulante.

Les virus non-circulant s'attachent au niveau d'une structure protéique particulière du stylet du puceron : l'acrostyle [Uzest *et al.*, 2010]. Au cours des piqûres d'essai ou de nutrition, les virus non-circulant s'y attachent dès la pénétration du stylet dans les tissus de la plante hôte, pour être ensuite relâchés lors des piqûres d'épreuve suivantes, et potentiellement sur une nouvelle plante. Il existe ainsi deux stratégies d'accrochage des virus sur le stylet des pucerons [Brault *et al.*, 2010]. Dans le

cadre de la stratégie dite <sup>2</sup> capsid P, notamment utilisée par les Cucumovirus, l'accrochage du virus au stylet se fait par l'interaction directe entre les protéines constituant la capsid du virus et des récepteurs aphidiens situés sur la cuticule du stylet. Dans le cadre de la stratégie dite <sup>2</sup> helper P, l'accrochage du virus se fait par l'intermédiaire d'un facteur viral assistant l'interaction entre la capsid et un récepteur situé sur le stylet du puceron. Cette seconde stratégie est par exemple utilisée par les Caulimovirus et les Potyvirus [Ng et Falk, 2006].

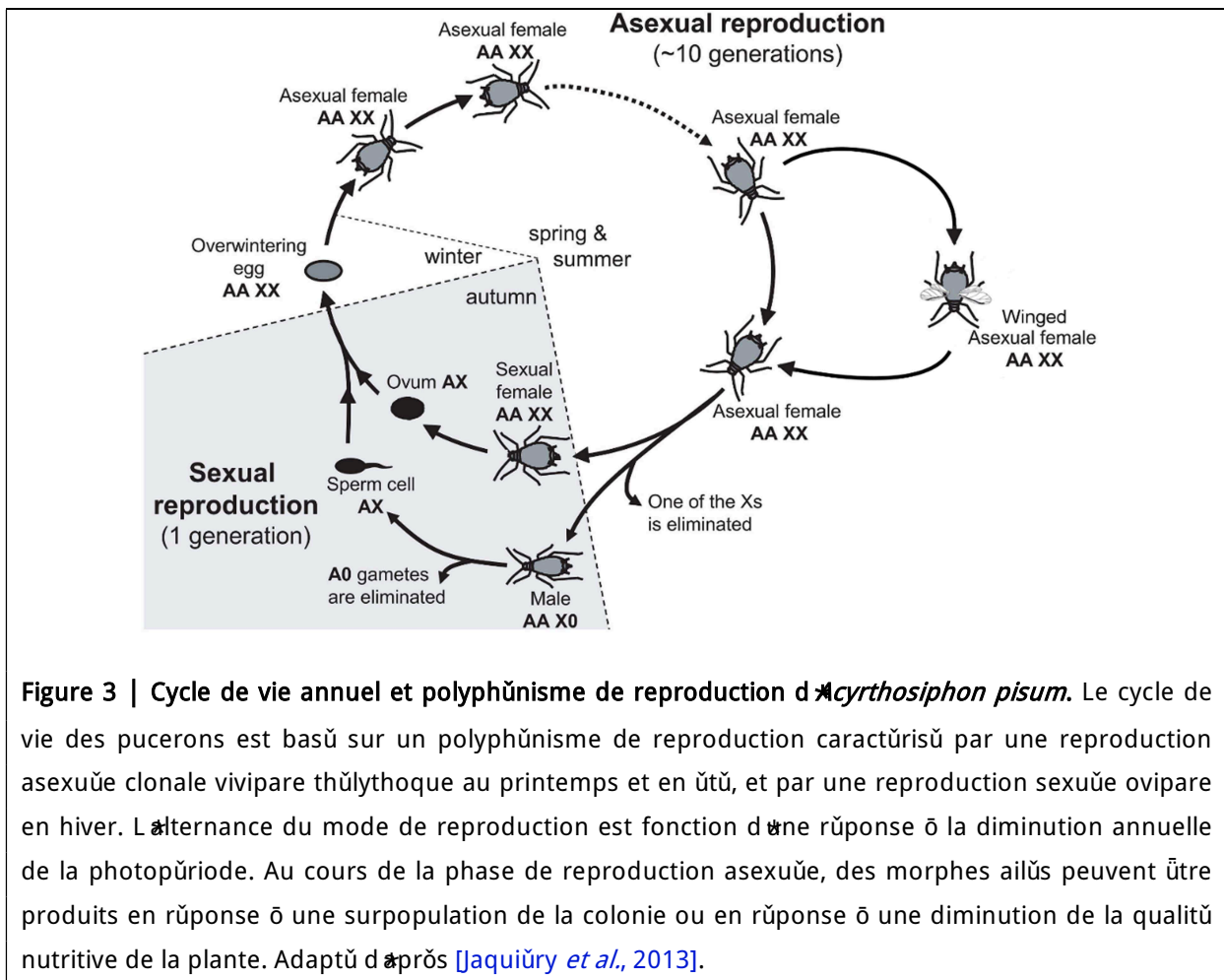
Certains virus sont transmis selon un mode circulant [Ng et Perry, 2004], auquel cas ils transitent à l'intérieur du puceron (intestin, hémolymphe) avant d'atteindre les glandes salivaires où les virions s'accumulent au niveau des espaces intercellulaires connectés à la lumière des conduits salivaires afin d'infecter une nouvelle plante via la salive émise par le puceron lors de piqûres de nutrition ou de essai [Ammar et Nault, 1985 ; Brault *et al.*, 2010]. Les virus transmis selon ce mode circulant peuvent être propagatifs ou non-propagatifs selon qu'ils se multiplient ou non à l'intérieur du puceron [Brault *et al.*, 2010]. Les Rabdovirus sont des virus circulant propagatifs qui libèrent leur génome viral au niveau du cytoplasme des cellules de l'intestin pour se répliquer de manière lente, sans causer de dommages au puceron. Après leur multiplication au niveau des cellules de l'intestin, les particules virales s'échappent en traversant le plasmalemme basal. Une fois libérés dans l'hémolymphe, les virions ont accès à différents organes, dont notamment les organes reproducteurs qui leur permettent une transmission verticale depuis la femelle infectée vers ses descendants [Sylvester, 1969 ; Nault, 1997]. Ce processus se déroule lentement compte tenu du nombre de barrières physiologiques à traverser (gaine ovarienne et épithélium folliculaire). Parmi les autres organes accessibles aux virions depuis l'hémolymphe, on compte le cerveau, les ganglions nerveux, l'épiderme, les tissus conducteurs et de réserve, la rutine, les muscles ainsi que la trachée [Ammar et Nault, 1985]. Les virus non-propagatifs adoptent le même trajet de circulation au sein des pucerons, sans toutefois se multiplier [Brault *et al.*, 2010].

## 2. Cycle de vie et alternance du mode de reproduction

Il existe principalement deux types de cycles de vie chez les pucerons : certains comportent une parthénogénèse cyclique (CP pour <sup>2</sup> Cyclic Parthenogenesis) tandis que d'autres comportent une parthénogénèse obligatoire (OP pour <sup>2</sup> Obligatory Parthenogenesis).

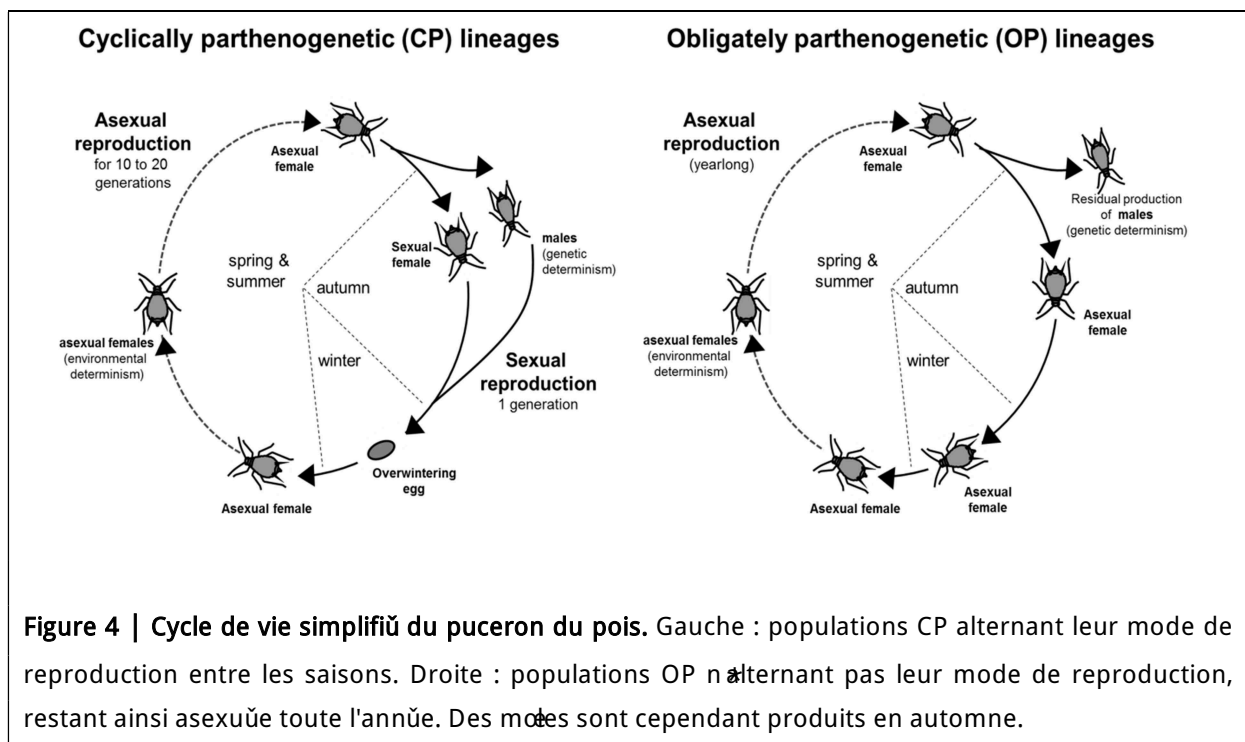
La plupart des espèces et populations de pucerons sont CP, et alternent au cours de l'année un mode de reproduction parthénogénétique clonal avec une reproduction sexuée, et sont ainsi dites holocycliques (Figure 3). Au printemps et en été, les populations sont constituées de femelles à reproduction parthénogénétique apomictique théllytoque vivipare. Ce mode de reproduction correspond ainsi à une reproduction asexuée, sans fécondation et sans méiose *sensu stricto* (apomixie), ne produisant que des femelles (théllytoque) se développant dans l'abdomen de leur mère sous forme d'embryons (viviparité). La rapidité de production des descendants (jusqu'à 100 individus produits par adulte) confère aux pucerons une forte capacité de colonisation de leur environnement. A la fin de l'été ces femelles vont percevoir la diminution de la photopériode - ainsi qu'une diminution

de la température - en mesurant la durée de la nuit (scotophase) et en percevant une accumulation d'un nombre minimum de scotophases longues consécutives. En réponse à ces stimuli, les femelles parthénogénétiques vont produire dans leur descendance des femelles sexuées ovipares et des mâles. Après accouplement, les femelles sexuées vont pondre des œufs fertilisés correspondant à une forme d'adaptation aux hivers rigoureux, puisque ces œufs sont capables de résister à de faibles températures, normalement létales pour les adultes et les larves. Un œuf ainsi pondu va rester en diapause pendant environ 90 jours durant lesquels l'embryon contenu dans cet œuf se développe de manière ralentie [Shingleton *et al.*, 2003]. A la fin de cette période de diapause, c'est-à-dire à l'arrivée du printemps, les œufs vont éclore en donnant naissance à des femelles dites <sup>2</sup> fondatrices <sub>P</sub> qui sont des femelles parthénogénétiques vivipares produisant uniquement des femelles asexuées. En conditions contrôlées, ces <sup>2</sup> fondatrices <sub>P</sub> sont incapables de répondre à une photopériode courte et ainsi de produire des descendants sexués pendant plusieurs générations. C'est <sup>2</sup> l'effet fondatrice <sub>P</sub> [Lees, 1960], qui limite probablement la production de morphes sexués au début du printemps lorsque la photopériode est encore courte et potentiellement inductrice.



**Figure 3 | Cycle de vie annuel et polyphénisme de reproduction de *Myrthosiphon pisum*.** Le cycle de vie des pucerons est basé sur un polyphénisme de reproduction caractérisé par une reproduction asexuée clonale vivipare thélytoque au printemps et en été, et par une reproduction sexuée ovipare en hiver. L'alternance du mode de reproduction est fonction d'une réponse à la diminution annuelle de la photopériode. Au cours de la phase de reproduction asexuée, des morphes ailés peuvent être produits en réponse à une surpopulation de la colonie ou en réponse à une diminution de la qualité nutritive de la plante. Adapté d'après [Jaquiury *et al.*, 2013].

De nombreuses espèces de pucerons comportent des populations (ou génotypes) présentant un cycle parthénogénétique obligatoire (OP, Figure 4). Ces clones sont caractérisés par une réponse altérée à la diminution de la photopériode, car ne produisant dans leur descendance que des femelles asexuées vivipares, voire quelques mâles [Simon *et al.*, 2002, 2010]. De par l'absence de production d'ufs, ces lignées OP sont sensibles au froid : elles sont donc contre-sélectionnées dans les climats présentant des hivers rigoureux. C'est pourquoi l'on retrouve principalement ces lignées dans des régions géographiques présentant des hivers doux ou une absence d'hiver, où les populations OP présentent un avantage démographique certain vis-à-vis des lignées CP [Rispe et Pierre, 1998 ; Dedryver *et al.*, 2001]. Réciproquement, les lignées CP sont majoritaires au niveau des régions géographiques froides, tandis que les deux types de lignées coexistent dans des régions tempérées présentant des hivers aux températures fluctuantes d'une année sur l'autre [Dedryver *et al.*, 2001 ; Simon *et al.*, 2002]. Alors que le passage de la reproduction clonale à la reproduction sexuée chez les lignées CP est déclenché par des changements de la photopériode, la perte de reproduction sexuée est supposée génétique chez les lignées OP, puisque les changements de conditions environnementales n'ont aucun effet sur leur phénotype reproducteur [Simon *et al.*, 2010 ; Dedryver *et al.*, 2013 ; Jaquiéry *et al.*, 2014].



**Figure 4 | Cycle de vie simplifié du puceron du pois.** Gauche : populations CP alternant leur mode de reproduction entre les saisons. Droite : populations OP n'alternant pas leur mode de reproduction, restant ainsi asexuée toute l'année. Des mâles sont cependant produits en automne.

### 3. Traits d'histoire de vie et plasticité phénotypique

#### a. Bactéries endosymbiotiques

Les pucerons comportent de nombreux traits d'histoire de vie qui expliquent leurs remarquables capacités d'adaptation aux contraintes de leur environnement proche. Ils comportent ainsi des bactéries endosymbiotiques de différentes espèces et genres, leur permettant par exemple une meilleure adaptation aux plantes hôtes au dépend desquelles ils se nourrissent. *Buchnera aphidicola* est la seule bactérie endosymbiotique obligatoire chez de nombreuses espèces de pucerons. Cette symbiose obligatoire est caractérisée par la localisation intracellulaire de *Buchnera aphidicola* au niveau du cytoplasme des bactériocytes (adipocytes spécialisés contenant des bactéries endosymbiotiques) et par la transmission verticale de cette bactérie de la mère à ses descendants [Douglas, 1998]. Cette symbiose permet à la fois aux pucerons et aux bactéries de se nourrir efficacement. La sève élaborée dont les pucerons se nourrissent est dépourvue de certains acides aminés essentiels, comme le tryptophane ; ces acides aminés sont synthétisés par *Buchnera* [Lai et al., 1994] qui assure également le recyclage de l'azote, et qui en échange reçoit des composés carbonés et azotés de la part du métabolisme du puceron hôte. Les pucerons comportent également un certain nombre de bactéries endosymbiotiques secondaires, leur conférant une grande variété de traits phénotypiques. Actuellement, 8 espèces de bactéries secondaires ont été identifiées chez le puceron du pois *Acyrtosiphon pisum* : cinq Gammaproteobactéries qui sont *Hamiltonella defensa*, *Regiella insecticola*, *Serratia symbiotica* [Moran et al., 2005], *Rickettsiella viridis* [Tsuchida et al., 2010, 2014] et PAXS (Pea Aphid X-type Symbiont) [Guay et al., 2009] ; deux Alphaproteobactéries du genre *Rickettsia* [Chen et al., 1996] et *Wolbachia* [Russell et al., 2013] ; et *Spiroplasma* [Fukatsu et al., 2001] appartenant aux Mollicutes. Ces bactéries sont localisées dans différentes parties du puceron du pois telles que l'hémolymphe ou les bactériocytes [Moran et al., 2005]. Ces bactéries symbiotiques facultatives sont impliquées dans divers processus biologiques tels qu'une valeur adaptative (*fitness*) accrue sur des plantes hôtes spécifiques [Tsuchida et al., 2004], des modifications de la couleur du corps [Tsuchida et al., 2010], une tolérance accrue à la chaleur [Montllor et al., 2002], ou encore une protection contre des ennemis naturels [Oliver et al., 2003 ; Scarborough et al., 2005]. Bien que transmises de manière verticale comme *Buchnera aphidicola*, ces bactéries symbiotiques facultatives peuvent aussi être transmises horizontalement entre pucerons de la même espèce ou de espèces différentes [Henry et al., 2013 ; Gauthier et al., 2015]. Notons que seul un nombre réduit d'espèces de bactéries endosymbiotiques cohabitent au sein d'un même puceron, et que différentes associations de ces bactéries avec *Buchnera aphidicola* peuvent se trouver au sein d'une même colonie.

### b. Plasticité phénotypique et polyphénisme chez les pucerons

D'une manière générale, un organisme ayant un génotype donné est capable de produire un continuum de phénotypes en réponse à une condition environnementale particulière (plasticité phénotypique continue). Mais chez certains organismes, un même génotype est capable de produire des phénotypes discrets extrêmement différents en réponse aux fluctuations de facteurs environnementaux. Ce phénomène correspond au polyphénisme (détaillé en E. de cette synthèse bibliographique) présent chez de nombreux insectes dont les pucerons. Chez ces derniers, il existe trois types de polyphénisme : le polyphénisme de dispersion, de caste et de reproduction. Cette plasticité phénotypique discrète explique en grande partie la persistance de ces ravageurs au sein des agroécosystèmes ainsi que l'importance des dégâts causés sur les cultures.

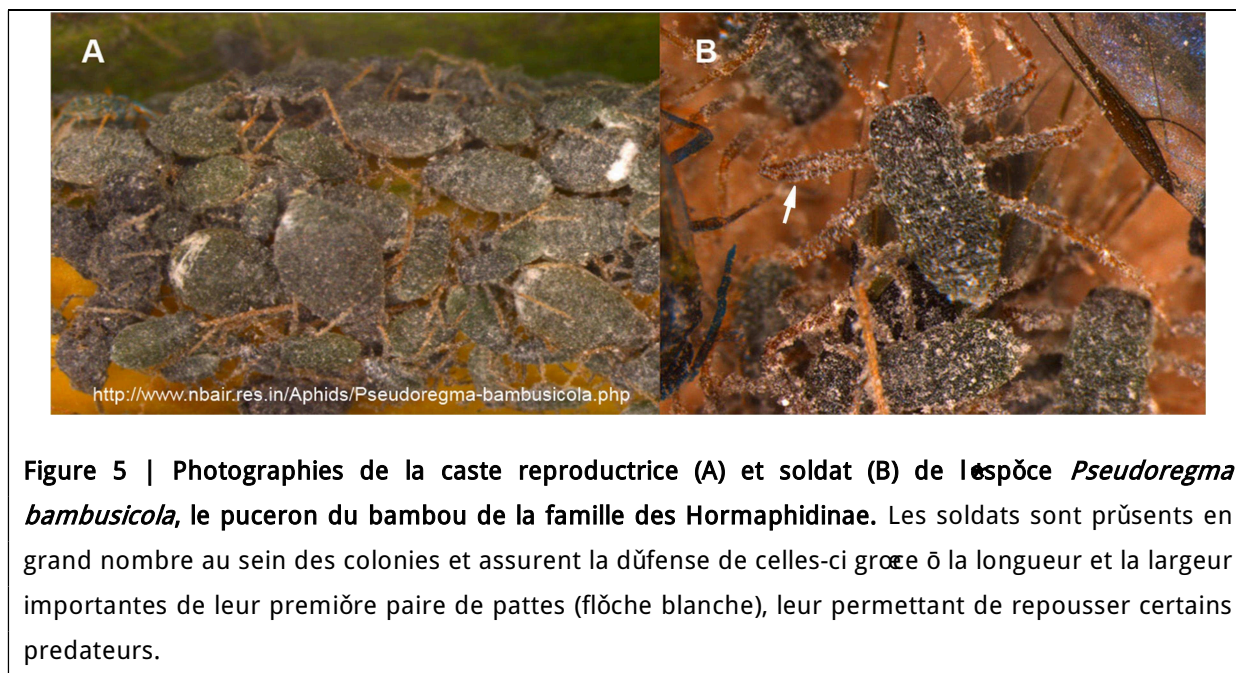
#### **Polyphénisme de dispersion**

Les pucerons sont capables de répondre à une diminution de la qualité nutritive de la plante et/ou à une surpopulation causée par leurs congénères sur une même plante. Les femelles perçoivent ces signaux et y répondent en produisant dans leur descendance des femelles ailées, capables de coloniser de nouvelles plantes en assurant ainsi le maintien de la population sur une autre plante présentant une meilleure qualité nutritive [Dixon et Howard, 1986 ; Zera et Denno, 1997]. Ce type de polyphénisme a lieu au cours de la phase de reproduction asexuée des pucerons et leur confère une importante capacité de colonisation. Une récente étude a mis en évidence que des gènes impliqués dans la voie hormonale de l'ecdysone sont sous-exprimés chez les femelles productrices d'individus ailés [Vellichirammal *et al.*, 2017]. Ces gènes présentent des homologues chez la Drosophile dont *ultraspiracle* et *ecdysone receptor* (EcR), deux gènes codant pour des protéines impliquées dans la formation hétérodimérique du récepteur à l'ecdysone, ainsi que des gènes dont l'expression est induite par l'ecdysone, dont *broad* et *ftz-fl*. De plus, des individus ayant subi des injections d'ecdysone produisent moins d'individus ailés, tandis que les individus traités avec un antagoniste d'un récepteur à l'ecdysone ou traités par RNAi ciblant le gène codant pour ce récepteur à l'ecdysone produisent davantage d'individus ailés dans leur descendance vis-à-vis des individus contrôle. Ceci démontre un lien de causalité entre le polyphénisme de dispersion des pucerons et la voie hormonale de l'ecdysone.

#### **Polyphénisme de caste**

Chez les insectes sociaux, certains individus sont spécialisés dans la reproduction, tandis que d'autres ne produisent que peu de - voire aucun - descendants. Ces morphes peu fertiles, voire infertiles, sont caractérisés par un comportement et des traits physiologiques distincts des individus reproducteurs, en étant généralement spécialisés dans des fonctions altruistes. Ces phénotypes altruistes infertiles sont produits en réponse à des facteurs environnementaux précis et correspondent à des polyphénismes de castes, notamment présents chez certains pucerons [Aoki, 1977]. Certaines espèces de pucerons produisent en effet des individus sacrifiant leur capacité de reproduction afin d'améliorer la survie de leurs congénères. Ces individus sont considérés comme des <sup>2</sup> soldats <sup>3</sup> puisque leur rôle social primaire correspond à la défense de la colonie (Figure 5), même si certains de ces individus peuvent en plus comporter des rôles domestiques [Aoki, 1980 ; Benton et Foster, 1992]. Actuellement,

60 espèces de pucerons productrices de soldats ont été identifiées parmi deux sous-familles de pucerons (Pemphiginae et Hormaphidinae) [Aoki, 1987 ; "t", 1989 ; Stern et Foster, 1996 ; Fukatsu *et al.*, 2005 ; Pike et Foster, 2008]. Chez *Tuberaphis styraci*, la production de soldats dans la descendance des femelles est corrélée positivement avec la densité en individus et la taille de la colonie. De plus, la perception de ces signaux environnementaux semble être à la fois pré et post-natale, indiquant donc une forme de transmission de ces signaux de la mère à sa descendance [Shibao *et al.*, 2010].



**Figure 5 | Photographies de la caste reproductrice (A) et soldat (B) de l'espèce *Pseudoregma bambusicola*, le puceron du bambou de la famille des Hormaphidinae. Les soldats sont présents en grand nombre au sein des colonies et assurent la défense de celles-ci grâce à la longueur et la largeur importantes de leur première paire de pattes (flèche blanche), leur permettant de repousser certains prédateurs.**

### Polyphénisme de reproduction

Comme décrit plus en amont, les pucerons comportent un polyphénisme de reproduction au cours de leur cycle de vie caractérisé par un basculement d'une reproduction asexuée vers une reproduction sexuée en réponse à la diminution de la photopériode à la fin de l'été. En réponse à ce stimulus, les femelles asexuées parthénogénétiques vivipares perçoivent la diminution de la photopériode, intègrent ce signal, puis le transduisent via le système neuroendocrine aux organes cibles, les embryons. Ces derniers y répondent par une modification de leur trajectoire développementale, et notamment un changement du devenir de leur lignée germinale, passant d'ovocytes diploïdes qui se développent en embryons dans le cadre de la reproduction asexuée à une lignée germinale haploïde (gamètes) produisant soit des ovocytes pour les futures femelles ovipares, soit des spermatozoïdes pour les mâles dans le cadre de la reproduction sexuée. Le génome des embryons est donc plastique, car capable de mettre en place différents programmes génétiques en réponse au signal neuroendocrine afin de produire ces trois morphes discrets.

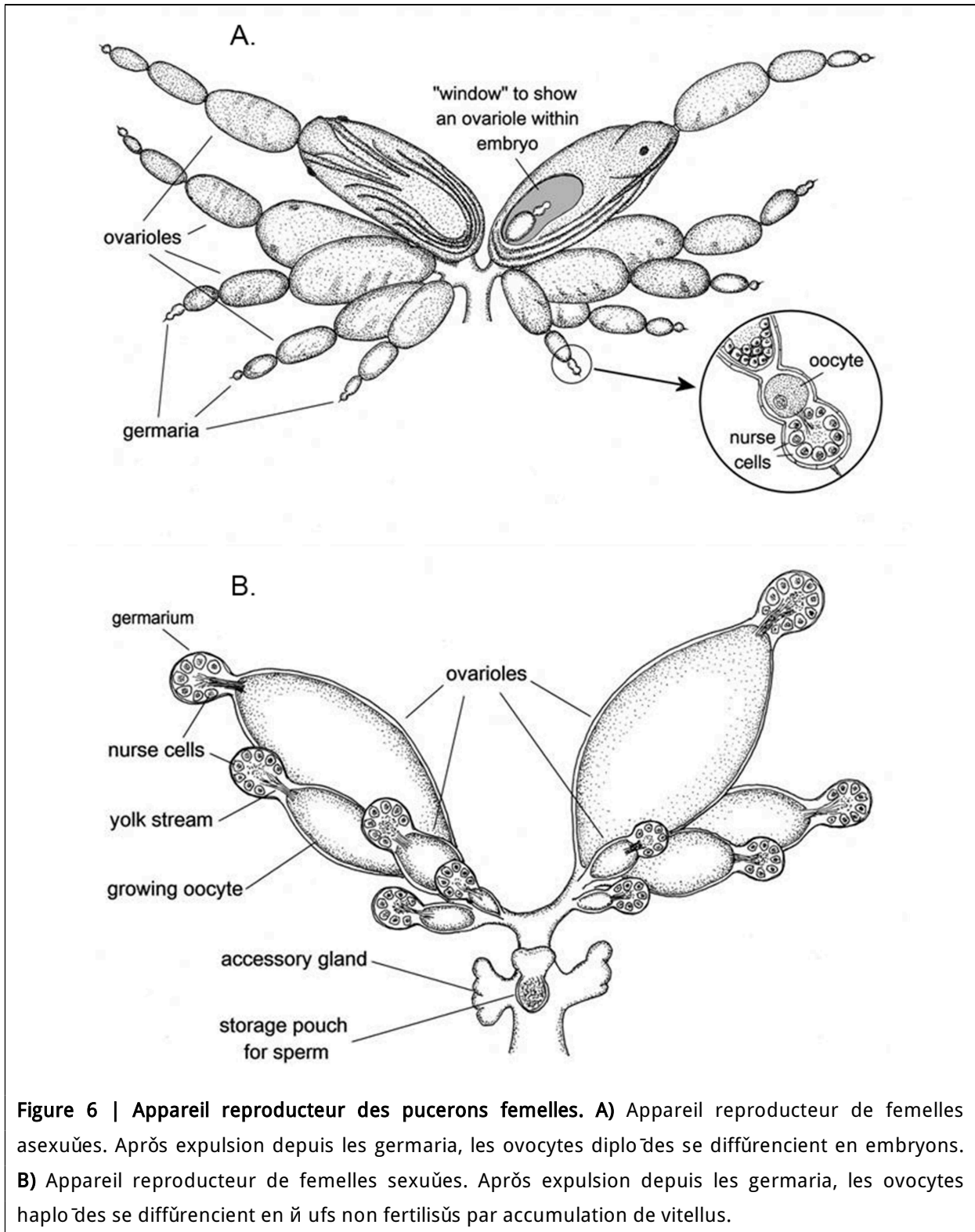
Chacun de ces morphes est adapté à l'environnement au sein duquel ils sont produits, avec des femelles parthénogénétiques adaptées à une reproduction et une colonisation rapide des agroécosystèmes dans le cadre de conditions favorables (printemps et été), et des mâles et femelles sexués adaptés à un mode de survie de la colonie dans le cadre de conditions hivernales défavorables. Une résultante de ce polyphénisme de reproduction est que ces trois morphes présentent des appareils reproducteurs distincts.

### c. L'appareil reproducteur des femelles asexués et sexués

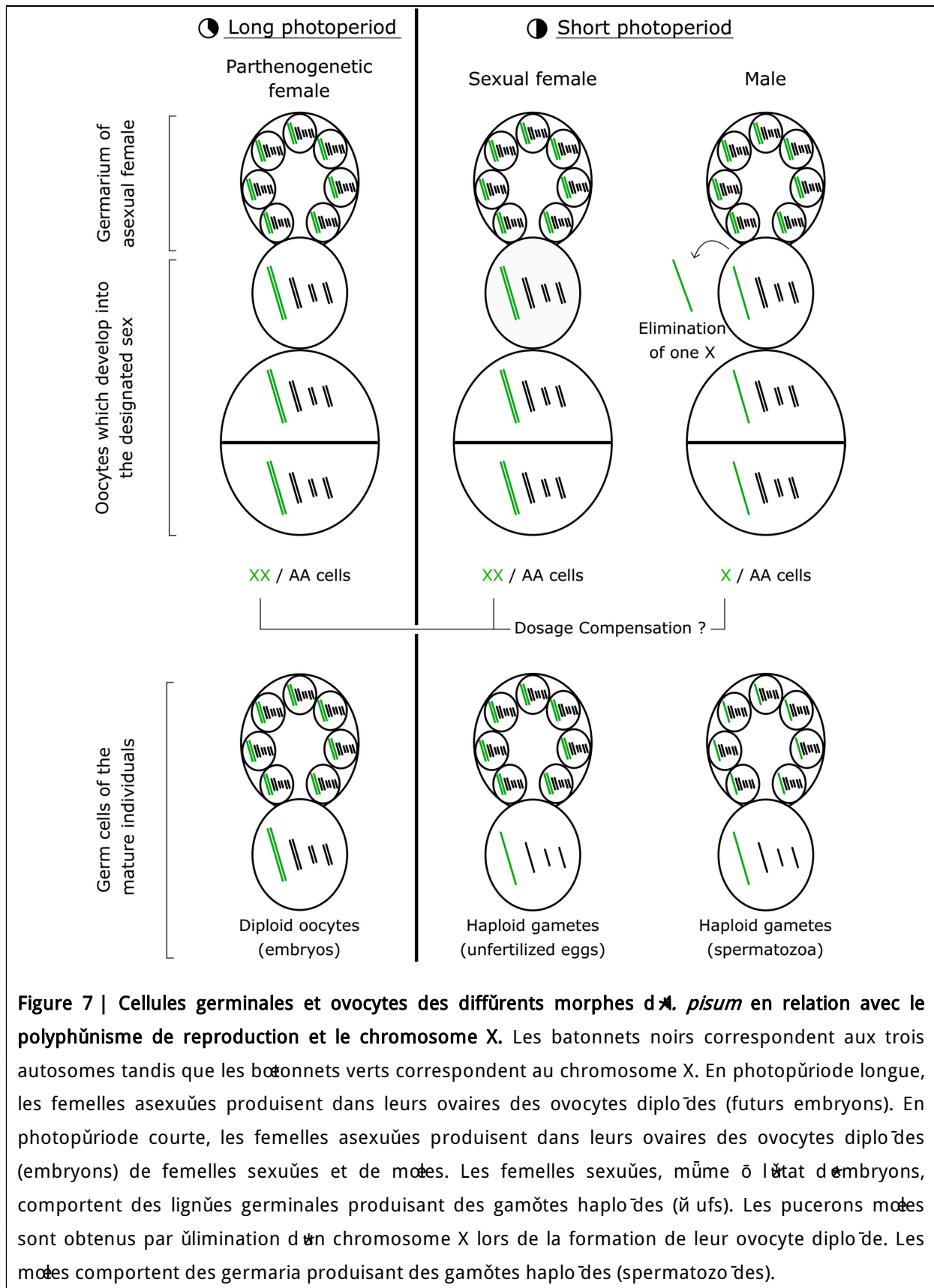
Les pucerons femelles présentent deux ovaires constitués de plusieurs ovarioles, chacun étant constitué d'un germarium et d'ovocytes haploïdes se différenciant en œufs non-fertilisés chez les femelles ovipares ou d'ovocytes diploïdes se différenciant en embryons chez les femelles parthénogénétiques (Figure 6). Les ovarioles des pucerons sont polytrophiques, ainsi les ovocytes sont accompagnés de cellules nourricières. Chez ces deux morphes, les cellules germinales à la base de chaque ovariole forment un kyste sphérique, appelé germarium, constitué de 32 cellules ovogoniales interconnectées. Ces 32 cellules ovogoniales se différencient soit en ovocyte, soit en cellule nourricière, avec un rapport de un pour un [Blackman, 1978].

En plus du type d'ovocytes produits par ces deux morphes, une autre différence se situe au niveau des cellules nourricières, puisque chez les femelles sexués, le noyau de ces cellules augmente considérablement de taille par endomitose permettant ensuite la production d'œufs non fertilisés, ce qui n'est pas le cas chez les femelles asexués. En plus de cette différence majeure, l'appareil reproducteur des femelles sexués comporte une spermathèque permettant de stocker la semence des mâles pour la fécondation des ovocytes haploïdes, et ainsi de former des œufs fécondés prêts à l'oviposition.

Dans le cas de l'ovogénèse des femelles sexués, les futurs ovocytes entrent en méiose " au sein du germarium, et sont donc haploïdes en fin de méiose. Les ovocytes ainsi formés restent bloqués en métaphase " avant d'entrer en phase de croissance, permettant l'accumulation de vitellus dans leur cytoplasme grâce aux cellules nourricières. Les chromosomes restent condensés jusqu'à la fécondation des œufs par les spermatozoïdes localisés dans la spermathèque lors du passage de l'ovocyte haploïde dans l'oviducte. Ces œufs fécondés sont ensuite déposés sur la plante hôte (oviposition). Dans les jours suivant la ponte, l'œuf se mûrit par la formation d'une cuticule sursable sombre. La bande germinale de l'embryon s'invagine dans le vitellus deux à trois jours après la ponte, puis l'inversion des axes ainsi que la segmentation complète de l'embryon sont terminés 15 jours après l'oviposition [Miura *et al.*, 2003]. L'œuf rentre ensuite en diapause pendant 90 jours durant lesquels l'embryon se développe de manière ralentie [Shingleton *et al.*, 2003].



Concernant l'embryogénèse chez les femelles asexuées, les futurs ovocytes situés dans le germarium sont bloqués en prophase et sont ensuite expulsés du germarium dans une chambre folliculaire. Ils y subissent une méiose "modifiée" ne présentant pas de phase de réduction chromosomique caractéristique de la méiose", permettant la formation d'ovocytes diploïdes. Les chromosomes de ces ovocytes sont ensuite rapidement condensés. Une seule division de maturation a alors lieu, ayant pour effet la libération d'un globule polaire. L'ovocyte diploïde subit ensuite plusieurs divisions mitotiques initiant le développement embryonnaire. Au sein de l'ovariole, dès qu'un ovocyte est expulsé en dehors du germarium, un autre ovocyte prend sa place, formant ainsi une chaîne d'embryons en cours de développement dont les embryons les plus développés progressent au fur et à mesure vers l'orifice génital situé au niveau de la partie postérieure du puceron. Une fois un embryon parvenu au dernier stade embryonnaire, il est expulsé sur la plante hôte, et est alors considéré comme une larve de premier stade (L1). Environ six embryons sont produits par chacun des 16 ovarioles, permettant ainsi la production d'environ 100 embryons par une femelle asexuée adulte. Le développement des embryons a été finement caractérisé en 20 stades embryonnaires [Miura *et al.*, 2003]. Dès les premiers stades embryonnaires, des cellules germinales se développent au sein des embryons. Ainsi, au sein d'une mère de génération 1, se trouve des embryons de génération 2, et eux-mêmes comportent des cellules germinales ainsi que des ovocytes en cours de différenciation, constituant la génération 3 et montrant bien le phénomène de télescopage des générations caractéristique chez les pucerons parthénogénétiques [Miura *et al.*, 2003]. Ceci a été démontré par l'utilisation d'anticorps ciblant deux protéines codées par des gènes marqueurs des cellules germinales de la *Drosophile* et conservés chez le puceron, à savoir *vasa* et *nanos*. Ces anticorps marquent en effet des cellules germinales primordiales se trouvant au sein des embryons, et ce à partir du 5<sup>ème</sup> stade de développement [Chang *et al.*, 2004]. Durant les stades tardifs de développement, à partir du stade 16, les diverses parties du corps comme les muscles, les yeux et les appendices sont déjà observables. En fonction de la photopériode, les ovocytes formés au sein des embryons peuvent soit correspondre à de futures femelles parthénogénétiques, à de futures femelles sexuées ou à de futurs mâles (Figure 7). Les ovocytes diploïdes à l'origine d'individus mâles sont obtenus par élimination d'un chromosome X via un mécanisme encore mal caractérisé. Les cellules somatiques des mâles comportent ainsi trois paires d'autosomes et un seul chromosome X, tandis que les femelles (sexuées ou asexuées) comportent trois paires d'autosomes et deux chromosomes X par cellule somatique. Il existe donc un déséquilibre du nombre de allèles associés au chromosome X entre mâles et femelles, vis-à-vis des autosomes. Des dysfonctionnements cellulaires peuvent survenir suite à un tel déséquilibre qui doit convenir de contrebalancer *via* des mécanismes dits de compensation de dose. Avant de rentrer plus dans le détail de ces mécanismes, nous présenterons les spécificités du chromosome X des pucerons.



**Figure 7 | Cellules germinales et ovocytes des différents morphes de *P. sativum* en relation avec le polyphénisme de reproduction et le chromosome X.** Les batonnets noirs correspondent aux trois autosomes tandis que les batonnets verts correspondent au chromosome X. En photopériode longue, les femelles asexuées produisent dans leurs ovaires des ovocytes diploïdes (futurs embryons). En photopériode courte, les femelles asexuées produisent dans leurs ovaires des ovocytes diploïdes (embryons) de femelles sexuées et de mâles. Les femelles sexuées, même à l'état d'embryons, comportent des lignes germinales produisant des gamètes haploïdes (œufs). Les pucerons mâles sont obtenus par élimination d'un chromosome X lors de la formation de leur ovocyte diploïde. Les mâles comportent des germaria produisant des gamètes haploïdes (spermatozoïdes).

#### 4. Le chromosome X chez les pucerons : cytologie et compensation

Au cours du siècle dernier, de nombreuses espèces de pucerons ont été karyotypées, montrant une grande variabilité du nombre et de la morphologie de leurs chromosomes de manière intra et interspécifique [Kuznetsova et Shaposhnikov, 1973 ; Blackman, 1980]. Les pucerons possèdent des chromosomes holocentriques qui ne présentent donc pas de véritables centromères, comportant ainsi une activité transcriptomique le long de l'ensemble de leurs chromosomes [Blackman, 1987]. Alors que chez les organismes à chromosomes monocentriques la position du centromère est une caractéristique d'identification importante, l'absence de centromère des chromosomes holocentriques chez les pucerons les rendent difficilement identifiables s'ils présentent des tailles similaires. Par conséquent, la cytogénétique chez les pucerons est longtemps limitée au comptage et au tri des chromosomes selon leur taille, à l'identification des chromosomes X grâce à la différence du nombre de ce chromosome entre mâles et femelles, et à la localisation de la région organisatrice du nucléole (NOR) par marquage à l'argent. Le NOR correspond à un cluster de gènes codant pour les ARN ribosomiques 18S, 5,8S et 28S, et ces régions ont été étudiées chez les pucerons en tant que marqueurs chromosomiques [Novotný et al., 2011]. Les pucerons sont en effet des espèces hétérogamétiques mâles de type XX/X0, comme par exemple le nématode *Caenorhabditis elegans*. Les pucerons mâles ne possèdent donc qu'un unique chromosome X et sont produits par des femelles parthénogénétiques XX (Figure 7). Au cours de ce processus de production d'individus mâles, l'un des chromosomes X est éliminé afin de générer des ovocytes diploïdes comportant deux sets d'autosomes et un seul chromosome X (Figure 7).

Alors que les autosomes se séparent de manière indépendante, les deux chromosomes X sont ici reliés entre eux par une extrémité, contrairement à une mitose ordinaire. Un des deux chromosomes X se sépare de son complément et sa chromatide s'effondre dans les cellules filles [Orlando, 1974 ; Blackman et Hales, 1986]. Il a été démontré que les chromosomes X de prométaphase sont reliés par un étroit pont ADN-protéine qui est fortement marqué au nitrate d'argent et qui correspondrait au NOR [Mandrioli et al., 1999]. Par conséquent, le NOR est supposé jouer un rôle dans la ségrégation des chromosomes X dans les ovocytes destinés à donner des individus mâles. Il est également important de noter que chez les pucerons, le chromosome X tend à représenter une proportion importante du génome comme c'est le cas chez *Diuraphis noxia* (35%) [Novotný et al., 2011] et *Acyrtosiphon pisum* (31%) [Richard et al., 2017 ; Jaquiere et al., 2017].

Tout comme chez de nombreux organismes hétérogamétiques mâles, il existe un déséquilibre dans le nombre d'allèles associés au chromosome X entre mâles et femelles, induisant de ce fait un différentiel du niveau d'expression des gènes liés au X. Au cours de l'évolution, certaines espèces hétérogamétiques ont acquis des mécanismes de compensation de dose (détaillés en ".C.) permettant d'assurer un niveau égal de transcription des chromosomes X de chaque sexe comparativement aux autosomes. Ils impliquent généralement des mécanismes épigénétiques, augmentant ou diminuant l'accessibilité de la chromatine ainsi que l'expression des gènes du chromosome X chez les mâles ou les femelles. Ces mécanismes ont notamment été étudiés chez certains organismes modèles, dont la *Drosophila* chez qui un complexe protéique (le DCC, ou complexe de compensation de dose)

orchestre la compensation de dose chez les mâles [Conrad et Akhtar, 2012]. Afin d'étudier la présence potentielle de tels mécanismes chez les pucerons, des études visant à caractériser les profils d'expression du chromosome X entre mâles et femelles (sexuées et asexuées) ont été effectuées [Jaquiury *et al.*, 2013 ; Pal et Vicoso, 2015 ; Richard *et al.*, 2017] (voir V.). Il a été identifié que le seul chromosome X des mâles est en moyenne plus exprimé que les deux chromosomes X des femelles sexuées et asexuées, témoignant d'un potentiel mécanisme de compensation de dose sous-jacent, ou tout du moins d'un mécanisme global de régulation de la transcription du X chez un des deux sexes [Jaquiury *et al.*, 2013].

Ces problèmes de déséquilibre du nombre d'allèles associés au chromosome X entre mâles et femelles découlent du mécanisme permettant la production des mâles par élimination d'un des deux chromosomes X. Ceci se fait au cours du basculement du devenir des embryons et de leur lignée germinale dans le cadre du polyphénisme de reproduction (Figure 7). Les mécanismes moléculaires découlant de la perception de la diminution de la photopériode et permettant le basculement du devenir des embryons ont été étudiés chez les pucerons, et principalement chez le puceron du pois *Acyrtosiphon pisum*. Nous allons à présent décrire ces mécanismes moléculaires.

## 5. Bases moléculaires du polyphénisme de reproduction

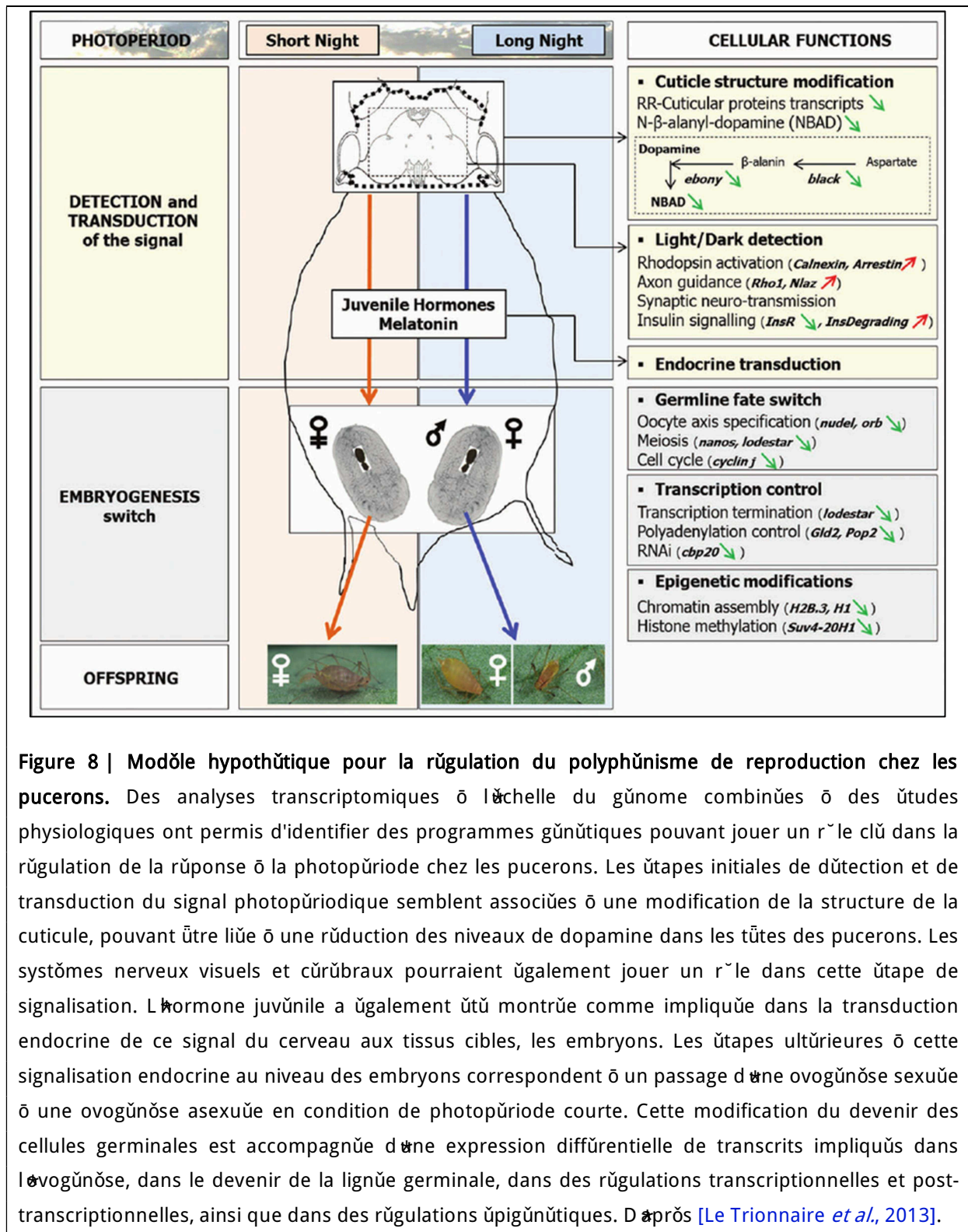
Le polyphénisme de reproduction chez le puceron peut se décomposer en plusieurs phases. La première phase consiste en une perception du signal photopériodique et plus précisément de la longueur de la nuit (scotophase) [Lees, 1973]. Le mécanisme précis de l'horloge circadienne permettant potentiellement une mesure de la photopériode et de la scotophase est encore inconnu chez les pucerons, mais il semble que les gènes *timeless (tim)* et *period (per)* y soient impliqués chez *A. pisum* puisque leur expression est corrélée positivement et oscille de manière robuste avec la luminosité sur une période de 24 heures [Barberó *et al.*, 2017]. Ces deux gènes sont des acteurs majeurs de l'horloge circadienne chez *D. melanogaster* en formant un hétérodimère TIM/PER dont la concentration est corrélée avec un rythme circadien de 24 heures [Gekakis *et al.*, 1995 ; Zeng *et al.*, 1996].

La température influe également sur le basculement du mode de reproduction en accentuant le polyphénisme de reproduction lorsque les températures sont moyennes (15-20, C) comparativement aux faibles températures (11, C) ou aux températures trop élevées (25, C), ces dernières supprimant complètement le basculement du mode de reproduction [Lees, 1959 ; Chapman, 1998]. Cependant le signal photopériodique en conditions contrôlées est le seul paramètre nécessaire et suffisant à l'initiation du polyphénisme de reproduction. Le déclenchement de ce processus se fait selon deux critères : i) une longueur de nuit critique (CNL), consistant en une durée minimum d'obscurité ininterrompue, et ii) un nombre de jours minimum consécutifs (<sup>2</sup> required day number  $P_r$  ou RDN) comportant cette CNL. Par exemple, chez *Megoura viciae*, la CNL est de 7 heures et la RDN est de 5 jours à 15, C [Vaz Nunes M et Hardie, 2000]. Les durées respectives de ces deux facteurs varient d'une espèce à voire d'un clone à puceron à l'autre. De plus, ces paramètres dépendent de la température, avec par exemple chez *Megoura viciae* des CNL de 6h, 7h ou 8h chez des femelles

asexués placés respectivement à 12.5, C, 15, C ou 17.5, C, et une RDN de 5 jours ou 4.5 jours à 15, C et 17.5, C, respectivement [Vaz Nunes M et Hardie, 2000].

Plusieurs études ont permis de mieux comprendre les mécanismes cellulaires et moléculaires sous-jacents au polyphénisme de reproduction, en commençant par s'intéresser à la première phase du polyphénisme de reproduction : la perception et la transduction du signal photopériodique aux tissus cibles (Figure 8). Le signal photopériodique est tout d'abord perçu par le cerveau du puceron (les photorécepteurs photopériodiques ne sont pas connus) puis il est transmis aux tissus cibles, les ovarioles de la femelle asexuée. Le groupe de cellules neuro-sécrétrices " (NSC Group ") semble être impliqué dans la perception du signal photopériodique ou dans la transmission de ce signal vers les tissus cibles, car une fois le NSC Group " micro-cautérisé, il s'est avéré que les pucerons étaient incapables de répondre aux changements photopériodiques [Steel et Lees, 1977]. La nature de la molécule permettant la transduction du signal photopériodique est encore inconnue, mais il est fait l'hypothèse qu'elle pourrait circuler par les projections axonales des cellules du NSC Group " jusqu'aux ovarioles où des récepteurs pourraient être localisés. Une étude comparant les transcriptomes de têtes de *A. pisum* placés en conditions de jours courts ou jours longs a permis d'identifier environ 600 transcrits différentiellement exprimés [Le Trionnaire et al., 2007]. Parmi ces transcrits se trouvent des gènes codant des protéines impliquées dans la modification de la structure de la cuticule, dans la signalisation nerveuse et dans la signalisation neuro-endocrine, ainsi que des enzymes impliquées dans les voies de signalisation de la dopamine et de l'insuline [Le Trionnaire et al., 2009]. En outre, il a été montré que le génome du puceron contient 10 gènes codant l'insuline, annotés "nsuline Related Protein ("RP) 1 à 10, ces derniers étant particulièrement exprimés au niveau de la tête de *M. persicae* et *A. pisum* [Huybrechts et al., 2010].

L'hormone juvénile est également impliquée dans la signalisation de la photopériode. L'application ectopique de Kinoprène, un analogue de l'hormone juvénile, sur la cuticule de pucerons placés en jours courts produisant donc normalement des individus sexués, reverse le devenir des embryons, ces derniers se développant alors en individus asexués [Corbitt et Hardie, 1985 ; Hardie, 1987]. Ceci suggère une implication de l'hormone juvénile dans la transduction du signal photopériodique. La mélatonine serait aussi impliquée de manière potentielle dans la transduction du signal photopériodique, puisque les gènes arylalkylamine N-acetyltransferase (AANAT) 1 et 3 ont été identifiés comme surexprimés au niveau de têtes de pucerons placés en jours courts par rapport à des têtes de pucerons placés en jours longs chez *A. pisum* [Barberó et al., 2013].



La deuxième phase de la réponse photopériodique correspond à un changement du devenir des embryons en réponse à la transduction du signal précédemment décrit, permettant la mise en place des programmes génétiques alternatifs associés à la production d'embryons asexués ou sexués (Figure 8). Ces programmes génétiques ont été identifiés par comparaison des transcriptomes d'embryons sexués et asexués à l'aide de puces à ADNc [Gallot *et al.*, 2012]. La comparaison des transcriptomes des deux types d'embryons a révélé que 33 transcrits sont différenciellement exprimés. Parmi ces transcrits, sept sont potentiellement impliqués dans l'épigénose dont *orb* et *nanos*, deux gènes homologues de la Drosophile impliqués dans le développement des cellules germinales, cinq dans la régulation post-traductionnelle et trois dans la régulation du cycle cellulaire dont *cyclin J*, un homologue notamment impliqué dans la mitose au niveau des embryons chez la Drosophile. Enfin, il a été identifié quatre transcrits différenciellement exprimés codant des protéines impliquées dans des régulations épigénétiques : i) Suppressor Of Variegation 4-20 Homolog 1 (*suv(4)-20H1*), une histone méthyltransférase responsable du dépôt de la marque H4K20me3 chez les mammifères [Neguembor *et al.*, 2013]. ii) Ubiquitin Like With PHD And Ring Finger Domains 1 (*uhrf1*), une protéine impliquée dans le maintien de la méthylation de l'ADN au cours de sa réplication, et dans la reconnaissance de la marque histone H3K9me3 ainsi que du domaine H3R2 pour le recrutement d'enzymes régulatrices de la structure de la chromatine [Bostick *et al.*, 2007 ; Karagianni *et al.*, 2008 ; Rajakumara *et al.*, 2011 ; Felle *et al.*, 2011 ; Guan *et al.*, 2013]. iii) Les transcrits codant les histones linker H1 ainsi que iv) les transcrits codant pour des histones variants H2B.3 ont aussi été identifiés comme différenciellement exprimés. Ceci constitue une première piste indiquant que la mise en place des programmes génétiques permettant le basculement du mode de reproduction chez les pucerons impliquerait des modifications épigénétiques au niveau de la chromatine des embryons.

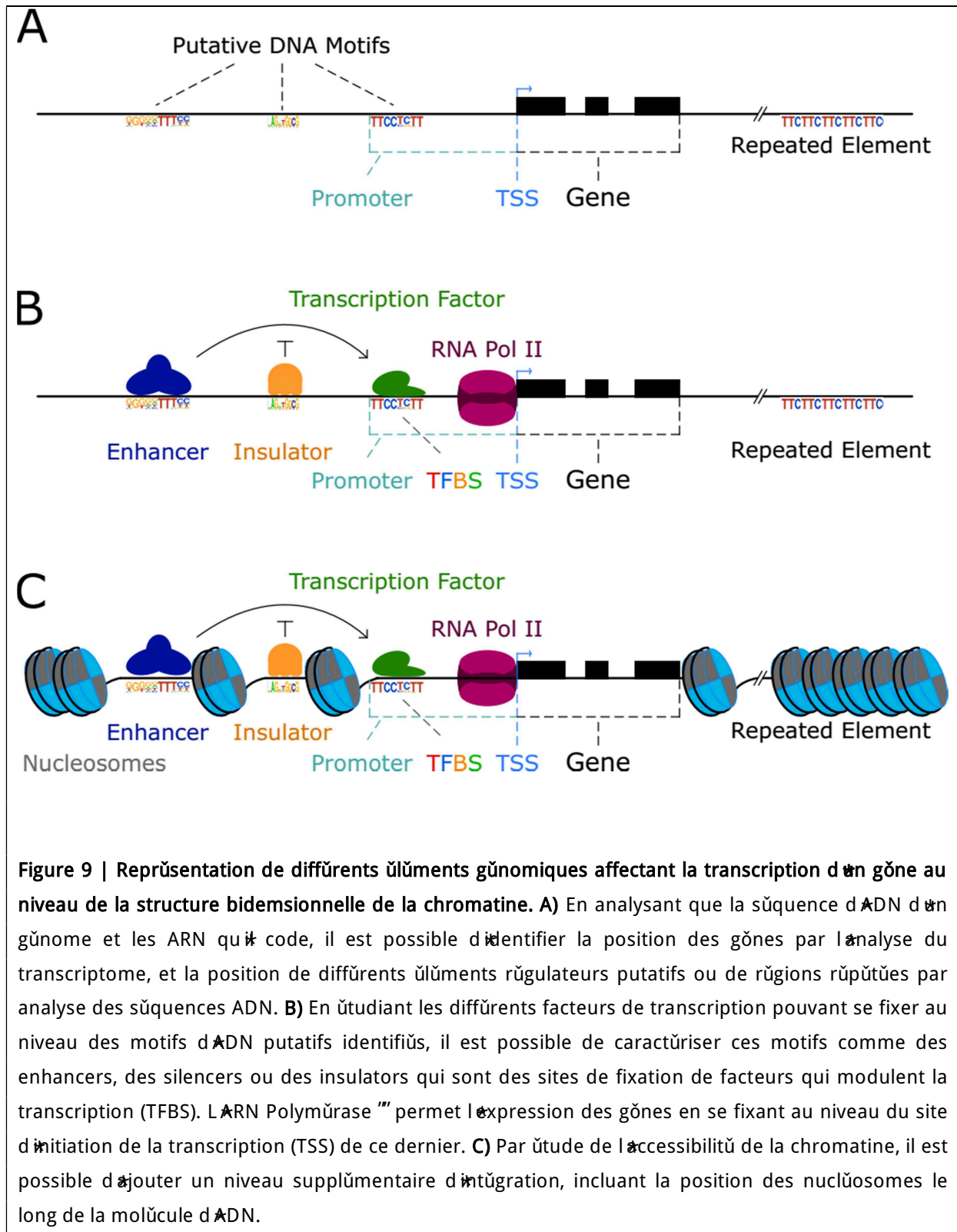
En résumé, les pucerons présentent un polyphénisme de reproduction en réponse à une diminution de la photopériode, qui est à l'origine de la production d'individus mâles ne présentant qu'un seul chromosome X comparé aux femelles qui en comportent deux, nécessitant une potentielle compensation de dose. Le polyphénisme de reproduction est basé sur une modification de la trajectoire développementale des embryons et notamment du devenir des cellules germinales. Il a été montré que certains gènes impliqués dans des régulations épigénétiques sont différenciellement exprimés entre embryons à devenir sexué ou asexué. En outre, la compensation de dose implique également des régulations épigénétiques, comme cela a été montré chez la Drosophile ou l'Homme [Conrad et Akhtar, 2012 ; Brockdorff et Turner, 2015] (détailé en ").C). Des mécanismes épigénétiques seraient donc impliqués au niveau des deux éléments du cycle de reproduction des pucerons auxquels nous nous intéressons dans cette thèse : le polyphénisme de reproduction, et l'éventuelle compensation de dose de l'expression du chromosome X entre mâles et femelles. Nous allons donc maintenant définir et décrire l'épigénétique et ses mécanismes associés.

## B. Mécanismes épigénétiques et structure de la chromatine

### 1. Unités de transcriptions

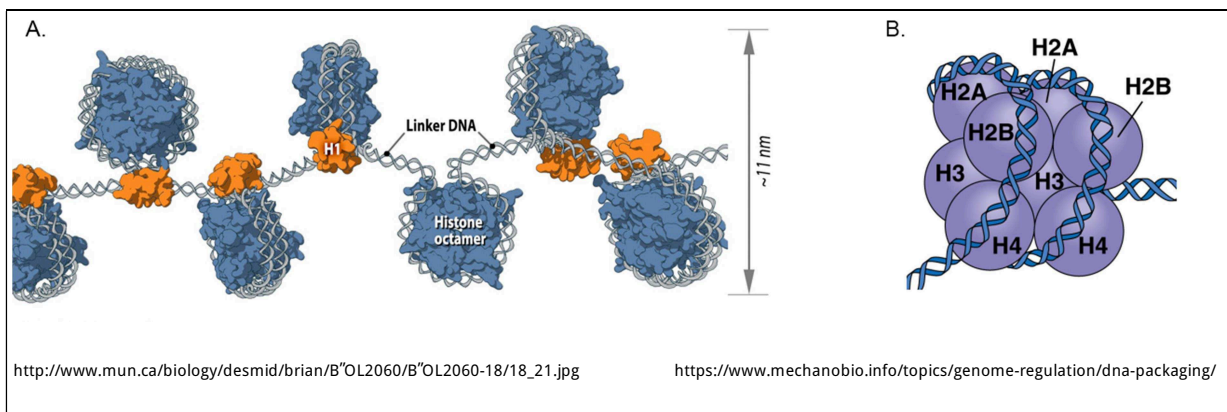
Avant d'introduire le concept d'épigénétique, il est nécessaire de bien définir les éléments fonctionnels d'un génome aussi bien au niveau des unités de transcription que dans les régions intergéniques.

Les génomes sont composés de différents éléments fonctionnels. En utilisant des techniques de séquençage ADN haut-débit et à l'aide d'outils bioinformatiques, il est possible d'assembler (même partiellement) la séquence d'un génome en chromosomes de manière linéaire au niveau bidimensionnel. L'analyse de données transcriptionnelle (RNA-seq) permet ensuite d'identifier la position des gènes, et ainsi définir la position de leurs exons, introns, et des régions transcrites non traduites en 5' et 3' (5' UTR et 3' UTR) (Figure 9). Grâce à l'identification de ces éléments du génome, il est ensuite possible de déterminer la position putative des sites d'initiation de la transcription (TSS). La position des TSS peut être définie avec précision à l'aide de techniques telles que le Global Run-On sequencing (GRO-seq) [Gardini, 2017], basées sur le séquençage et le mapping de transcrits naissants en cours de synthèse par l'ARN Polymérase " (RNAP "). De cette manière, il est possible d'identifier les core promoteurs, ces derniers étant définis comme une région de taille variable, en amont du TSS (selon l'orientation en 5' ou 3' de ce dernier). Les régions promotrices, ainsi que des régions en aval du TSS, contiennent de nombreux éléments nécessaires à la transcription du gène, comme la boîte TATA, une région riche en AT comportant généralement le motif TATAWAAR, permettant la fixation de la TATA Binding Protein (TBP). Cette dernière, accompagnée d'autres protéines, participe à l'assemblage du complexe de pré-initiation de la transcription (P<sup>IC</sup>) [Struhl *et al.*, 1998]. Bien que extrêmement connue, l'initiation de la transcription d'un gène à l'aide de la boîte TATA correspond plus à une exception qu'à la règle, du moins chez les mammifères, où seulement 10% à 20% des promoteurs en contiennent une [Gershenson *et al.*, 2004 ; Cooper *et al.*, 2006]. D'autres éléments permettent l'initiation de la transcription comme l'élément initiateur ("nr), contenu dans 40% à 60 % des promoteurs chez les mammifères [Lagrange *et al.*, 1998 ; Yang *et al.*, 2007 ; Filipowicz *et al.*, 2008], correspondant aux motifs YYANWYY chez l'Homme et TCAKTY chez la Drosophile, avec A correspondant au TSS [Smale et Baltimore, 1989].



## 2. Chromatine et régions régulatrices longue distance

Au sein du noyau interphasique, de nombreuses protéines s'associent à l'ADN afin d'initier la transcription des gènes en ARNm, éléments nécessaires à leur traduction en protéine. D'autres protéines permettent quant à elles une structuration et une compaction des plusieurs mètres d'ADN composant le génome d'un organisme, en permettant à l'ensemble de cette information génétique d'être contenue dans un noyau de quelques micromètres sous forme de chromatine. Le terme chromatine vient du latin *khroma* (couleur), tout d'abord défini par Walther Flemming en 1882 comme une structure ayant la propriété de se colorer en absorbant des colorants basophiles [Flemming, 1882 ; Paweletz, 2001]. La chromatine est actuellement définie comme l'enroulement de deux boucles de 73 paires de base (pb) d'ADN (soit 146 pb) autour d'un octamère de protéines histones, aussi appelées core histones [Luger *et al.*, 1997], chacune présente en deux copies. Un nucléosome est ainsi composé de deux histones H2A, deux histones H2B, deux histones H3 et deux histones H4 [Luger *et al.*, 1997] ainsi que de 146 nucléotides (Figure 10). Chez la plupart des espèces, une famille supplémentaire de protéine histone est présente dans la chromatine avec une stœchiométrie de 1 pour 1 vis-à-vis des nucléosomes : les Histones Linker (ou H1). Ces protéines ne partagent pas d'homologie de structure avec les core histones [Allan *et al.*, 1980 ; Syed *et al.*, 2010] et se lient à la partie extérieure du nucléosome, permettant de stabiliser l'ancrage de l'ADN autour du nucléosome.



**Figure 10 | Le nucléosome, structure de base de la chromatine. A)** La chromatine est constituée d'ADN double brin s'enroulant autour des nucléosomes en étant maintenus par des histones H1. Entre chaque nucléosome se trouve une portion d'ADN libre (Linker DNA). **B)** Chaque nucléosome consiste en 146 nucléotides ainsi que d'un octamère d'histones constitué d'une paire de chaque type de protéine histone : H2A, H2B, H3 et H4.

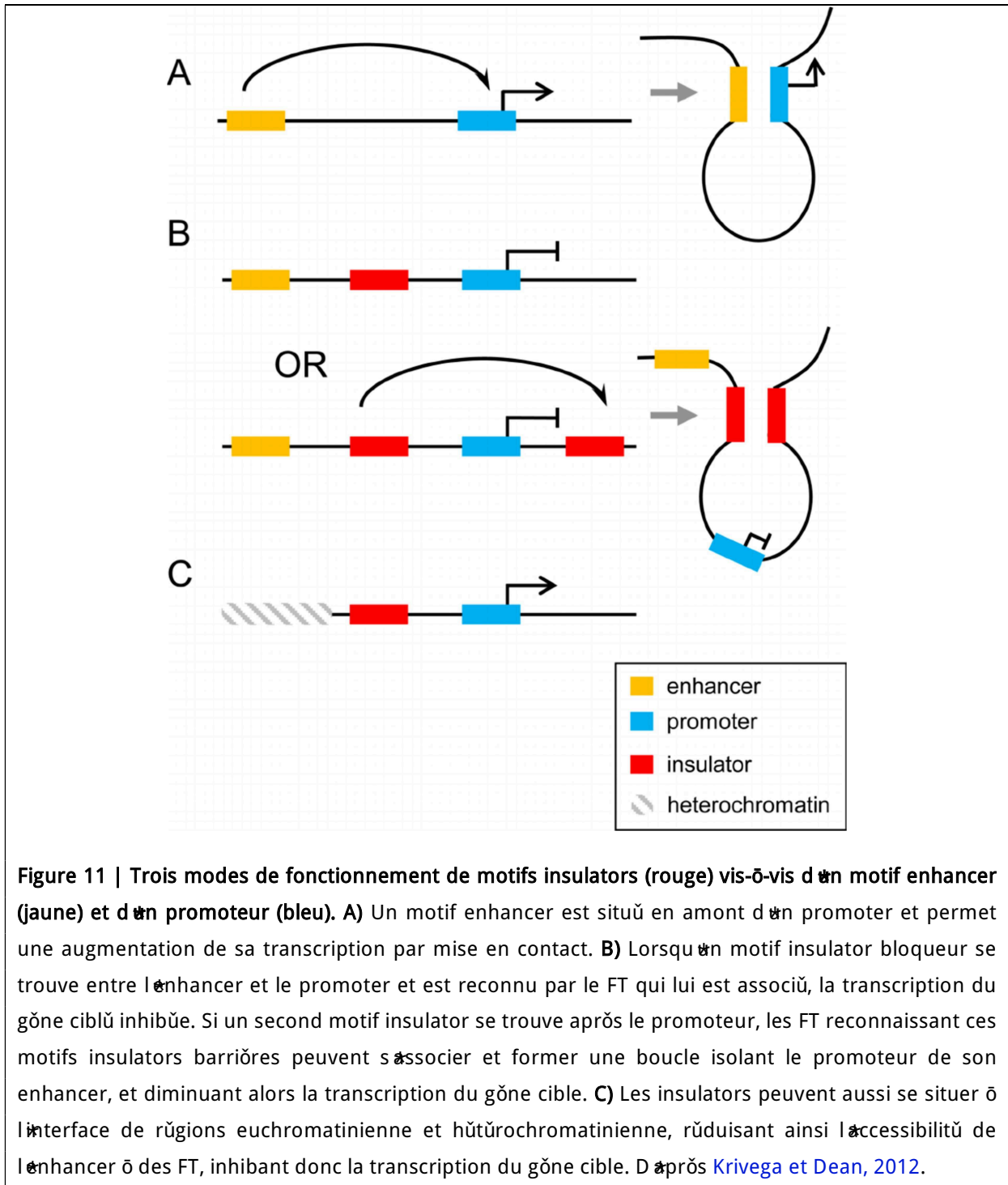
Outre les protéines histones permettant de structurer la chromatine ou la RNAP III permettant d'initier la transcription, il existe d'autres protéines qui permettent de réguler la transcription des gènes. Ces protéines appelées facteurs de transcription (FT) se fixent au niveau de différents motifs d'ADN spécifiques, correspondant à des sites de fixation de facteurs de transcription (TFBS). Certains de ces motifs sont des enhanceurs et sont capables, par recrutement d'un FT, d'augmenter le taux de formation et de fixation des complexes de pré-initiation au niveau des core promoteurs. Ces motifs enhanceur peuvent se retrouver de manière proximale en amont ou en aval des TSS au sein des promoteurs, des exons, des introns, ou encore des régions 5' et 3' UTR. Il existe également des enhanceurs distaux, se trouvant jusqu'à 10 kbp ou 100kbp du TSS qui les régulent, respectivement chez la Drosophile et les mammifères [Levine et Tjian, 2003 ; Birney *et al.*, 2007]. Le mécanisme précis par lequel ces éléments régulent l'activité transcriptionnelle fait encore débat, mais il est certain que le déclenchement de l'activité d'un enhanceur nécessite souvent la fixation de plusieurs facteurs de transcription au niveau de motifs *cis*-régulateurs de l'enhancer [Riethoven, 2010]. La première séquence ADN enhanceur a été identifiée dans le génome du virus simien 40 (SV40). Il a été montré que lorsque cette séquence est associée à des promoteurs, elle stimule la transcription du gène associé indépendamment de son orientation, et même depuis des positions situées en aval du gène [Banerji *et al.*, 1981].

En plus des motifs enhanceurs, il existe des motifs silencers (Figure 9) impliqués dans l'inhibition de la transcription des gènes. Les silencers sont également des séquences d'ADN spécifiques reconnues par des protéines, mais sont cependant moins bien décrits que les enhanceurs, notamment car ils ont été découverts de manière plus tardive. On distingue cependant deux classes de silencers : i) les *silencer elements* correspondant à des motifs courts et position-indépendants (qui n'ont pas besoin d'être situés de manière particulière vis-à-vis d'autres motifs) retrouvés en amont des TSS. Ils sont reconnus par des facteurs de transcription répresseurs capables d'interférer avec l'assemblage du complexe de pré-initiation de la transcription, permettant ainsi d'inhiber la transcription du gène cible ; ii) Les éléments régulateurs négatifs (NRE) correspondent quant à eux à des silencers position-dépendants qui empêchent passivement la fixation de facteurs de transcription au niveau de motifs régulateurs en *cis* [Riethoven, 2010]. Les NRE sont retrouvés à la fois en amont et en aval des TSS, et parfois même au niveau des introns et des exons [Ogbourne et Antalis, 1998]. Le Repressor Element-1 Silencing Transcription Factor/Neuron-Restrictive Silencer Factor (REST/NRSF) est un facteur de transcription repressif jouant un rôle clé dans la différenciation des neurones chez les mammifères [Chong *et al.*, 1995 ; Schoenherr et Anderson, 1995]. La réduction de l'expression de REST/NRSF dans les cellules souches a pour effet d'augmenter l'expression de nombreux gènes neurones-spécifiques portant la séquence silencer RE-1 dans leur promoteur, reconnue par REST/NSRF. Par ce processus, REST/NSRF est supposé orchestrer divers processus critiques pour la neurogenèse, la synaptogenèse ou encore la transmission synaptique [Ballas et Mandel, 2005 ; Ooi et Wood, 2007 ; D'Alessandro *et al.*, 2009].

Enhancers et silencers peuvent agir sur de multiples gènes en même temps, mais certaines de ces interactions peuvent être gênantes dans le cadre du maintien d'un programme génétique donné. Ainsi, une troisième classe de séquences d'ADN régulatrices existe afin de prévenir ces interactions :

les insulators (Figure 9). Deux types d'insulators ont été identifiés à ce jour : les motifs insulators bloquant l'activité des motifs enhancers, et les motifs insulators barrière [Gaszner et Felsenfeld, 2006]. La première catégorie protège la transcription des gènes des enhancers en interférant avec l'interaction enhancer-promoteur en étant situé entre ces deux éléments. Par exemple, le CCCTC-binding factor (CTCF) est une protéine doigt de zinc conservée chez l'ensemble des vertébrés et exprimée de manière ubiquiste [Filippova *et al.*, 1996]. CTCF se fixe sur le motif insulateur CCCTC et prévient les interactions entre promoteurs et enhancers par blocage de l'activité enhancer, par exemple au niveau de l'extrémité 5' du locus de la globine chez le poulet [Bell *et al.*, 1999]. Chez l'Homme et la souris, CTCF permet l'isolement du gène codant la globine vis-à-vis de motifs enhancers voisins, suggérant également un rôle d'insulator barrière pour le motif insulator CCCTC [Ristimäki *et al.*, 1991 ; Bell *et al.*, 1999]. Ce type de motif insulator permet également de protéger le génome de la diffusion de l'hétérochromatine constitutive en se situant au niveau des interfaces euchromatine  $\Rightarrow$  hétérochromatine [Riethoven, 2010] (Figure 11).

La structure de la chromatine est en effet dynamique et peut être plus ou moins compactée sous forme d'euchromatine ou d'hétérochromatine selon son niveau d'accessibilité, définie et modélisée par un ensemble de mécanismes épigénétiques. C'est en 1942 que Waddington a pour la première fois employé le terme <sup>2</sup> épigénétique, qui a alors défini comme un changement du phénotype d'un individu n'étant pas caractérisé par une modification du génotype [Waddington, 1942, 2012]. Aujourd'hui, nous savons que les mécanismes épigénétiques permettent la transmission héréditaire de programmes d'expression de gènes en modifiant l'architecture de la chromatine, et notamment son accessibilité, sans altérer la séquence d'ADN sous-jacente [Allis et Jenuwein, 2016].



### 3. Accessibilité de la chromatine

#### a. Accessibilité de la chromatine à l'échelle bidimensionnelle

Des expériences dans les années 1970 chez le canard et le poulet ont montré, respectivement par hybridation ARN-ADN ou par digestion à la nucléase, des degrés de compaction différents de l'ADN eucaryote lors de comparaison de régions génomiques activement transcrites et de régions transcriptionnellement silencieuses [Axel *et al.*, 1973 ; Weintraub et Groudine, 1976]. La chromatine est ainsi une structure permettant non seulement de condenser l'ensemble d'un génome eucaryote au sein du noyau, mais aussi de moduler localement l'accessibilité de l'ADN pour sa transcription, sa recombinaison, sa réparation et sa réplication [Venkatesh et Workman, 2015]. Bien que la structure des nucléosomes puisse paraître rigide d'un point de vue cytologique, la position de ces unités constituantes de la chromatine est dynamique et permet de moduler la compaction de la chromatine [Radman-Livaja et Rando, 2010]. On peut distinguer l'euchromatine, correspondant à des régions accessibles et transcrites, et l'hétérochromatine, dense et transcriptionnellement silencieuse [Luger *et al.*, 2012 ; Milon *et al.*, 2014]. Il existe bien évidemment des états chromatiniens intermédiaires et flexibles pour lesquels l'accessibilité de la chromatine et la transcription sont régulées par différents mécanismes. Un des mécanismes permettant cette fluidité de la chromatine est l'échange de histones, impliquant la suppression ou le remplacement par le jeu de différents types de variants de histones. Ce mécanisme d'échange, aussi connu sous le nom de histone turnover, a diverses implications dans la composition, la structuration ou la fonction de différentes régions génomiques. Par exemple, une augmentation du taux d'échange des histones peut augmenter l'accessibilité d'une région génomique à certains facteurs nucléaires comme la RNAP III, permettant de fait la transcription de cette région. En revanche, si ces histones sont remplacées par des variants pouvant bloquer ce mécanisme d'échange, alors l'accessibilité et la transcription de cette région diminuera [Talbert et Henikoff, 2010 ; Venkatesh et Workman, 2015], comme c'est par exemple le cas de l'histone macroH2A activement déposée sur le chromosome X inactif et condensé des mammifères en remplacement des histones H2A [Costanzi et Pehrson, 1998] (détaillé en 3.1.C). Ces changements d'accessibilité peuvent également être régulés par des protéines régulatrices considérées comme des facteurs pionniers à l'origine de cascades de régulation comme Forkhead box protein A1 (FOXA1, aussi connu sous le nom HNF3X) [Zaret et Carroll, 2011]. Ces protéines sont supposées se lier à l'ADN associé aux nucléosomes et recruter des facteurs nucléaires additionnels afin de modifier la compaction de la chromatine. De manière additionnelle, ces facteurs pionniers peuvent rentrer en compétition avec les nucléosomes pour la liaison à l'ADN et ainsi augmenter l'accessibilité de la chromatine [Miller et Widom, 2003 ; Mirny, 2010].

Des changements dans la structure de la chromatine ont été montrés comme impliqués dans de nombreux processus biologiques, y compris dans certains dysfonctionnements cellulaires. Chez l'Homme par exemple, des mutations au niveau de gènes codant pour des facteurs nucléaires modulateurs de la chromatine, et affectant le positionnement des nucléosomes, se sont révélées impliquées dans la formation de cancers [Gaspar-Maia *et al.*, 2009 : 1 ; Hargreaves et Crabtree, 2011 ; Schwartzenuber *et al.*, 2012]. Il peut donc être intéressant de comparer des profils d'accessibilité de

la chromatine à l'échelle de génomes entiers de cellules placées dans différentes conditions afin d'identifier les régions différentiellement ouvertes accompagnant le développement de maladies, la différenciation de cellules ou encore l'intégration de signaux environnementaux [Tsompana et Buck, 2014]. De grands projets collaboratifs comme ENCODE [Consortium, 2012] ont ainsi largement participé au développement de ces profils chromatiniens et des méthodes permettant de les étudier.

#### b. Méthodes d'études

L'analyse de la structure et de la dynamique de la chromatine est possible grâce aux avancées en génomique et à l'accès à des génomes séquencés et assemblés. Plusieurs techniques permettent une analyse bidimensionnelle de la structure de la chromatine, en décrivant les régions ouvertes (ADN accessible) ou fermées (ADN non-accessible). Concernant l'analyse de la chromatine ouverte, d'abord en nucléosomes, il existe le DNase-seq [Weintraub et Groudine, 1976 ; Song et Crawford, 2010 ; John *et al.*, 2013], le FAIRE-seq [Nagy *et al.*, 2003 ; Giresi *et al.*, 2007 ; Giresi et Lieb, 2009 ; Simon *et al.*, 2012, 2013] et ATAC-seq [Buenrostro *et al.*, 2013b, 2015], tandis que le MNase-seq permet l'étude la position des nucléosomes, et donc des régions fermées de la chromatine [Noll, 1974 ; Cui et Zhao, 2012 ; Rizzo et Sinha, 2014 ; Tsompana et Buck, 2014]. Bien que ces techniques reposent sur des principes de biologie moléculaire différents, elles permettent d'identifier des régions en communs en plus de régions spécifiques à chaque méthode. La description rapide ainsi que les avantages et inconvénients de ces méthodes sont présentés dans le Tableau ". La méthode FAIRE ayant été largement utilisée au cours de ma thèse, une description plus poussée de cette méthode est proposée en Matériel et Méthodes ("".)

#### c. Identification d'éléments régulateurs par l'étude de l'accessibilité de la chromatine

L'utilisation de ces méthodes permet d'une part d'étudier les restructurations de la chromatine à l'échelle du génome, et d'autre part l'identification d'éléments régulateurs putatifs tels que les enhanceurs, ou d'une manière plus générale les sites putatifs de fixation de facteurs de transcription (TFBS). Ceci est rendu possible du fait de la corrélation positive entre l'accessibilité de la chromatine et la présence de TFBS, l'ADN devant être accessible pour permettre la reconnaissance et la fixation du FT sur le motif d'ADN qui lui est associé.

Tableau "

**Outils actuels d'étude de l'accessibilité de la chromatine à l'échelle du génome (adapté de Tompa et Duij, 2014)**

	Nombre de cellules	Type de séquençage	Approche utilisée	Cible génomique	Considérations expérimentales	Références
<b>MNase-seq</b>	1 à 10 million de cellules	Paired-end ou Single-end	La MNase digère l'ADN au niveau de sites situés de part et d'autre des nucléosomes, permettant d'extraire l'ADN lié à ces complexes protéiques	Cartographie l'ensemble des nucléosome du génome de manière qualitative et quantitative	<ol style="list-style-type: none"> <li>1. De nombreuses cellules sont nécessaires</li> <li>2. Titrations enzymatiques laborieuses</li> <li>3. Dégère les régions régulatrices actives, rendant leur identification possible seulement de manière indirecte</li> <li>4. Demande 150 à 200 millions de lectures pour une étude standard du génome humain</li> </ol>	Noll, 1974 ; Rizzo et Sinha, 2014 ; Cui et Zhao, 2012
<b>DNase-seq</b>	1 à 10 million de cellules	Paired-end ou Single-end	La DNase " coupe l'ADN au niveau de régions accessibles où l'ADN n'est pas protégé et permet l'extraction de chromatine ouverte	Cartographie de la chromatine ouverte	<ol style="list-style-type: none"> <li>1. De nombreuses cellules sont nécessaires</li> <li>2. Préparation longue et complexe des échantillons</li> <li>3. Titrations enzymatiques laborieuses</li> <li>4. Demande 20 à 50 millions de lectures pour une étude standard du génome humain</li> </ol>	Weintraub et Groudine, 1976 ; Song et al., 2010 ; John et al., 2013
<b>FAIRE-seq</b>	100 000 à 10 millions de cellules	Paired-end ou Single-end	Le FAIRE est basé sur la création de liaisons covalentes entre ADN et nucléosomes grâce au formaldéhyde, puis sur une séparation de l'ADN ouvert soniqué de l'hétérochromatine par extraction au phénol-chloroforme	Cartographie de la chromatine ouverte	<ol style="list-style-type: none"> <li>1. Faible ratio signal/bruit, rendant difficile l'interprétation des données bioinformatiques</li> <li>2. Les résultats dépendent grandement de l'efficacité de fixation au formaldéhyde</li> <li>3. Demande 20 à 50 millions de lectures pour une étude standard du génome humain</li> </ol>	Giresi et al., 2007 ; Giresi et Lieb, 2009 ; Simon et Lieb, 2009 ; Simon et al., 2013 ; Simon et al., 2013 ; Nagy et al., 2003
<b>ATAC-seq</b>	500 à 500 000 cellules fraîchement isolées	Paired-end	Des noyaux non fixés sont taggés in vitro avec des adaptateurs pour NGS au niveau de régions accessibles par un transposase Tn5 purifiée	Cartographie de la chromatine ouverte, des facteurs de transcription et des nucléosomes	<ol style="list-style-type: none"> <li>1. Contamination des données par de l'ADN mitochondrial</li> <li>2. Outils d'analyse bioinformatique encore immatures</li> <li>3. Demande 60 à 100 millions de lectures pour une étude standard du génome humain</li> </ol>	Buenrostro et al., 2013

ATAC: assay for transposase-accessible chromatin; DNase " : deoxyribonuclease " ; FAIRE: formaldehyde-assisted isolation of regulatory elements; MNase: micrococcal nuclease.

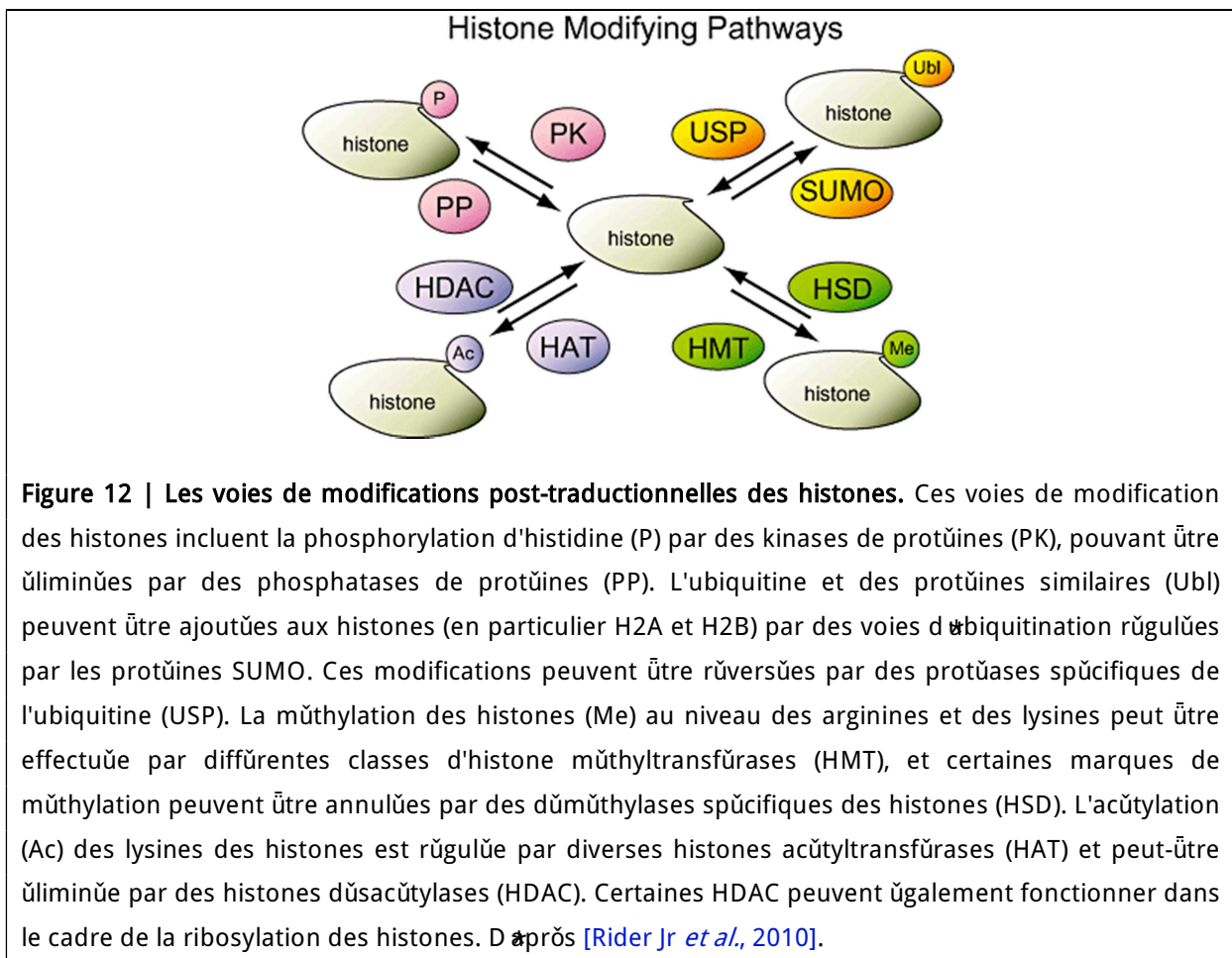
Le FA<sup>RE</sup>-seq et le MA<sup>NE</sup>-seq ont par exemple été utilisés chez le parasite responsable du paludisme chez l'Homme, *Plasmodium falciparum*, un parasite protozoaire au génome riche en AT [Ponte *et al.*, 2010]. Chez cet organisme, et comparativement aux eucaryotes non-protzoaires, les facteurs de transcription sont sous-représentés compte tenu de la taille du génome, et les mécanismes sous-jacents associés à la régulation de la transcription sont controversés. Les méthodes de FA<sup>RE</sup>-seq, permettant d'extraire les régions libres de nucléosomes, et de MA<sup>NE</sup>-seq, permettant l'extraction de l'ADN associé aux nucléosomes, ont donc été mises en œuvre afin d'étudier l'implication des modulations de l'accessibilité de la chromatine en relation avec la régulation de l'expression des gènes au cours du cycle d'infection de ce parasite. Il a été montré que l'accessibilité globale de la chromatine subit des changements drastiques au cours du cycle de vie du parasite, contrastant avec les modifications plus ciblées de l'accessibilité de la chromatine chez les eucaryotes non-protzoaires. Par exemple, l'accessibilité globale de la chromatine augmente au cours du cycle de vie de *P. falciparum* après leur invasion des globules rouges, puis se compacte à nouveau avant le début du prochain cycle. L'occupation des promoteurs par les nucléosomes au niveau local suit ces modifications globales d'accessibilité de la chromatine, avec cependant des exceptions au niveau des gènes *var* impliqués dans la virulence du parasite. Ces résultats montrent que les processus dirigeant l'expression des gènes chez *Plasmodium falciparum* sont différents de ceux décrits chez les eucaryotes non-protzoaires, impliquant ici des restructurations globales de la chromatine au cours du cycle de vie de ce protiste [Ponte *et al.*, 2010].

Le FA<sup>RE</sup>-seq et l'ATAC-seq ont également été utilisés dans le cadre de l'identification de TFBS et d'enhancers lors du développement de tumeurs *in vivo* chez l'Homme [Davie *et al.*, 2015]. 3778 régions présentant une chromatine hautement accessible au cours du développement d'une tumeur ont ainsi été identifiées comme suractivées. L'application de méthodes bioinformatiques de recherche de motifs a ensuite permis l'identification d'un enrichissement de la fixation putative de deux facteurs de transcription au niveau de ces régions : Stat92E et AP-1. L'utilisation d'un mutant Stat92E a ensuite résulté en un sauvetage du phénotype tumoral. Cette étude démontre que l'utilisation de méthodes de séquençage des régions accessibles de la chromatine, couplées à des méthodes bioinformatiques de recherche et d'enrichissement de motifs constituent un moyen efficace pour identifier les éléments régulateurs du génome tels que des enhancers ou des silenciers [Davie *et al.*, 2015].

## 4. Modifications post-traductionnelles des histones

### a. Définition des marques histones

D'une manière générale, les nucléosomes empêchent la transcription de l'ADN en réduisant son accessibilité aux facteurs nucléaires notamment par obstruction physique, ou par flexion de l'ADN [Luger *et al.*, 1997] lorsqu'ils sont en grande densité. Les histones composant les nucléosomes peuvent comporter de nombreuses modifications post-traductionnelles (PTMs) pouvant influencer la compaction ou l'accessibilité de la chromatine par différents mécanismes [Lawrence *et al.*, 2016]. Classiquement, ces PTMs comprennent l'acétylation, la méthylation, la phosphorylation, l'ubiquitinylation, la sumoylation, l'ADP ribosylation et la désamination [Kouzarides, 2007]. Le dépôt de l'ensemble de ces modifications des histones étant orchestré par un ensemble de familles d'enzymes (Figure 12). Plus récemment, d'autres modifications, telles que la propionylation et la butyrylation, ont également été décrites [Kebede *et al.*, 2015]. Les modifications post-traductionnelles des histones actuellement bien connues concernent principalement leur région N-terminale, correspondant à la queue des histones, c'est-à-dire les acides aminés accessibles à la surface du nucléosome, comme par exemple les Lysines 4, 9, 14, 23, 27, 36 de l'histone H3, ou encore les Lysines 5, 8, 12, 16, 20 de l'histone H4 [Luger *et al.*, 1997].



En plus des queues des histones, les domaines globulaires des histones peuvent aussi recevoir un certain nombre de modifications post-traductionnelles [Cosgrove *et al.*, 2004 ; Tropberger et Schneider, 2013]. Ces domaines globulaires forment le noyau du nucléosome, et leur partie externe en contact direct avec l'ADN présente un intérêt particulier vis-à-vis de l'étude des marques pouvant modifier ces régions du nucléosome. La spectrométrie de masse a permis d'identifier de nouvelles PTMs touchant le domaine globulaire du nucléosome [Cerbo *et al.*, 2014].

#### b. Modes d'action des marques histones

Certaines modifications apportées aux acides aminés composant les queues des histones affectent directement les interactions entre les nucléosomes. C'est par exemple le cas de H4K16ac qui réduit la compaction de la chromatine [Shogren-Knaak *et al.*, 2006], et par conséquent augmente la transcription des gènes à la fois *in vitro* et *in vivo*. Le dépôt de la marque H4K16ac est catalysé par la protéine MOF, également nommée KAT8, codée par le gène *males absent on the first (mof)*, tout d'abord identifiée dans le génome de *D. melanogaster*, puis retrouvée par homologie de séquence au sein du génome de nombreux mammifères. MOF est une histone acetyl transférase (HAT) [Hilfiker *et al.*, 1997] comportant un domaine de fixation sur la chromatine (Chromatin Binding Domain). Elle est impliquée dans la compensation de dose chez *D. melanogaster* en s'intégrant au complexe protéique Male Sex-Lethal (MSL). Chez les mammifères, cette protéine est une sous-unité du complexe Non-Specific-Lethal (NSL), contrôlant l'initiation de la transcription de nombreux gènes domestiques et mitochondriaux [Mendjan *et al.*, 2006 ; Cai *et al.*, 2010 ; Prestel *et al.*, 2010 ; Raja *et al.*, 2010 ; Lam *et al.*, 2012 ; Feller *et al.*, 2012].

Concernant les modifications post-traductionnelles touchant la partie globulaire des histones, certaines peuvent avoir des effets directs sur la transcription comme H3K122ac. Les Lysines 115 et 122 des histones H3 ont toutes deux été identifiées comme de nouveaux sites d'acétylation au niveau de la surface latérale du nucléosome entrant en contact avec l'ADN, correspondant plus particulièrement à l'endroit où la force de liaison entre ADN et histone est la plus forte [Cocklin et Wang, 2003 ; Hall *et al.*, 2009 ; Manohar *et al.*, 2009]. Le dépôt de cette marque est catalysé par la Bromodomain protein 4 (BRD4) par un mécanisme inconnu. Cette HAT est par exemple impliquée dans le développement de cancers et dans certaines maladies auto-immunes chez l'homme [Devaiah *et al.*, 2016]. H3K122ac co-localise avec des marques caractéristiques des gènes actifs (H3K4me3) et des enhanceurs (H3K4me1 et H3K27ac) [Tropberger *et al.*, 2013]. H3K122ac est ainsi retrouvée enrichie au niveau de régions pauvres en nucléosomes et proches des TSS où les nucléosomes sont absents afin d'initier la transcription. Ces résultats sont plus que simplement corrélatifs, puisque il a été montré que H3K122ac promeut l'éviction des nucléosomes et augmente la transcription de manière directe lors d'expérimentations *in vitro* [Tropberger *et al.*, 2013 ; Tropberger et Schneider, 2013]. Dans le cadre de l'induction de l'expression de gènes par l'hypoxogène, la marque H3K122ac est déposée en moyenne 10 minutes après ajout d'hypoxogène comparé à la marque H3K4me3, (pourtant considérée comme une marque histone inhérente à l'activation de la transcription des gènes) déposée au mieux 30 à 40 minutes après ajout d'hypoxogène [Tropberger *et al.*, 2013]. La marque H3K122ac touchant la partie

globulaire du nucléosome pourrait ainsi constituer à l'avenir un nouveau contrôle de l'activité des gènes, pouvant être plus précis d'un point de vue temporel que les marques histones actuellement utilisées dans ce domaine.

D'autres marques peuvent quant à elles agir de manière indirecte sur la structure de la chromatine par le recrutement de protéines régulatrices. Ces dernières peuvent par exemple bloquer l'accès aux complexes de remodelage [Margueron *et al.*, 2005], ou influencer le recrutement de modificateurs de la chromatine ou des facteurs de transcription [Vettese-Dadey *et al.*, 1996 ; Clements *et al.*, 2003]. C'est notamment le cas de la reconnaissance de H3K9me3 par Heterochromatin Protein 1 (HP1) dans le cadre de l'inhibition de la transcription d'éléments répétés [Lehnertz *et al.*, 2003].

### c. Méthodes d'études

Les marques histones sont principalement étudiées grâce à l'immunoprécipitation de la chromatine (ChIP) [Turner, 2001], aujourd'hui généralement couplée à du séquençage haut-débit (ChIP-seq) ou de la PCR quantitative. La méthode ChIP permet, à l'aide d'anticorps dirigés contre des modifications post-traductionnelles d'histone, d'immunoprécipiter l'ADN associé à ces histones modifiées, ou encore d'étudier les sites de fixations de protéines se liant à l'ADN (comme les facteurs de transcription). Il existe plusieurs types de protocoles permettant de conduire une ChIP, principalement la XChIP et la Native ChIP (NChIP) [Turner, 2001]. Ces deux techniques diffèrent principalement dans la préparation de la chromatine avant immunoprécipitation et présentent chacune des avantages et des inconvénients.

La XChIP est basée tout d'abord sur une lyse cellulaire puis sur une fixation de l'ADN aux protéines du nucléosome par du formaldéhyde, c'est-à-dire la formation de liaisons covalentes entre ces deux éléments à l'aide de formaldéhyde. Cette étape est suivie d'une fragmentation de l'ADN par sonication. Cette méthode présente les avantages de pouvoir étudier à la fois des protéines histones et non-histones, requiert une plus faible quantité de cellules que le NChIP, et permet de minimiser les risques de réarrangements de la chromatine au cours de la préparation de la chromatine et de sa précipitation, l'ADN et les nucléosomes étant fixés par des liaisons covalentes créées par le formaldéhyde. En revanche, la fixation au formaldéhyde peut entraîner l'enrichissement d'événements de fixation transitoires, impliquant donc la détection de faux positifs lors des analyses en aval des données issues de XChIP.

Dans le cas du NChIP, l'étape de crosslinking est omise ce qui permet la préparation de la chromatine native pour une fragmentation de l'ADN par digestion enzymatique après une lyse cellulaire. Le NChIP présente l'avantage d'une spécificité sans doute accrue des protéines reconnues par les anticorps, la plupart d'entre eux étant dirigés contre des protéines non-fixées au formaldéhyde. En outre l'immunoprécipitation donne généralement des rendements bien supérieurs au XChIP, évitant de ce fait le recours à des étapes d'amplification. En revanche, le NChIP n'est applicable qu'aux protéines histones, présente des possibilités de réarrangement de la chromatine au cours de l'expérimentation et donc des résultats potentiellement moins répétables. Enfin la fragmentation de l'ADN par digestion

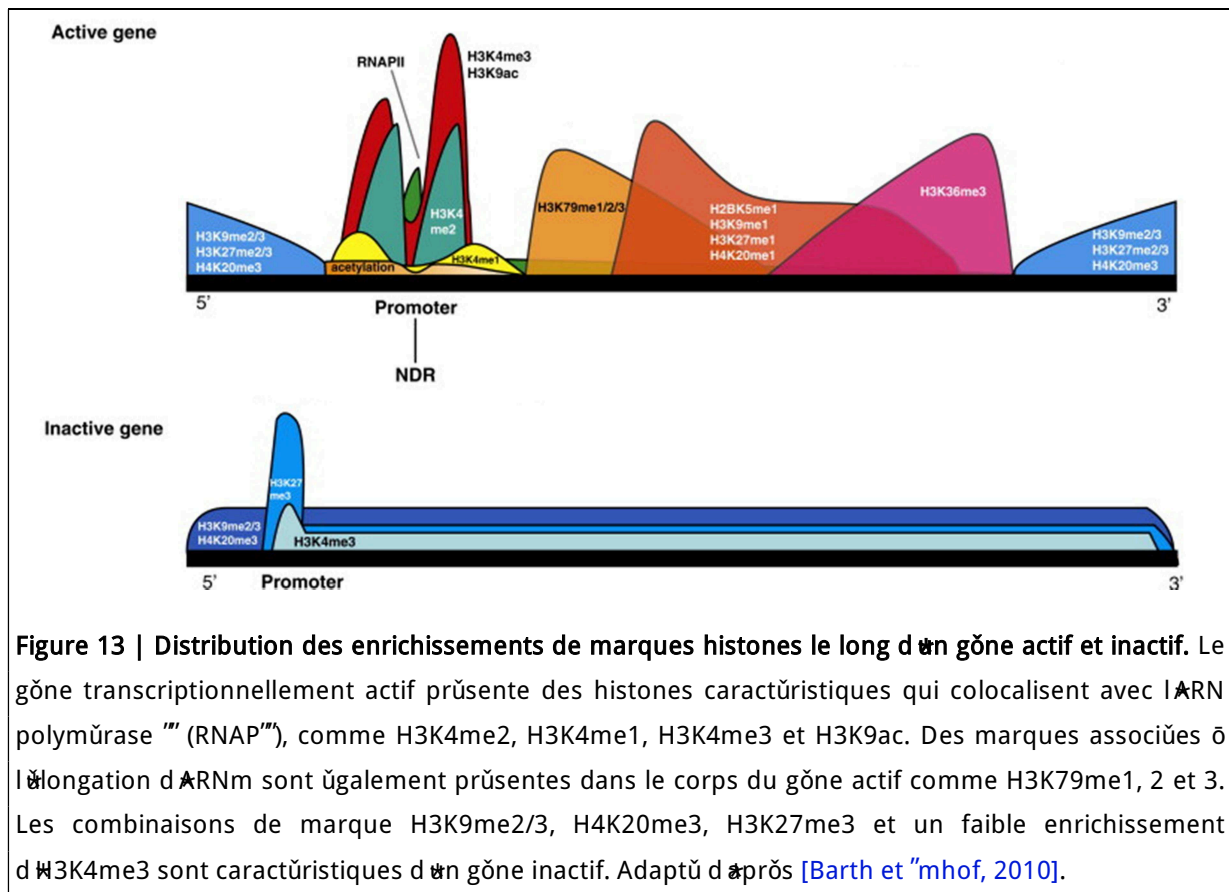
enzymatique peut être sélective envers certaines séquences de la chromatine, qui peuvent alors être surreprésentées [Turner, 2001].

#### d. Combinaison de marques et états chromatinien

Au sein des génomes eucaryotes, différents états chromatiniens coexistent et sont définis notamment par des combinaisons de marques histones. Par exemple chez *D. melanogaster*, 9 états chromatiniens ont été définis à l'aide de l'étude de 18 marques histones. Ces dernières ont également été corrélées à d'autres données génomiques, comme la position de autres protéines non-histones comme les FT, les régions accessibles de la chromatine par DNase-seq, la cartographie des transcrits naissants par GRO-seq, ou encore les longs et petits ARN non-codants [Kharchenko *et al.*, 2011]. L'ensemble de ces informations a permis de dresser une carte génomique et épigénétique complète du génome de la Drosophile, permettant de distinguer les caractéristiques des différents éléments du génome (chromosomes, gènes, éléments régulateurs et autres domaines fonctionnels). Le premier état chromatinien le plus distinct est associé aux régions transcriptionnellement actives et proches des TSS et des promoteurs, enrichi en marques histones actives telles que H3K4me2/3 et H3K9ac. L'activité d'élongation des ARN est quant à elle associée à l'enrichissement de la marque H3K36me3 au sein d'un 2<sup>nd</sup> état chromatinien retrouvé préférentiellement au niveau des régions exoniques des gènes transcrits. Le 3<sup>ème</sup> état chromatinien est retrouvé particulièrement au niveau des régions introniques et se distingue par l'enrichissement des marques H3K27ac, H3K4me1 et H3K18ac. Le 4<sup>ème</sup> état chromatinien est apparenté au 3<sup>ème</sup> avec un enrichissement de H3K36me1 et une déplétion de H3K27ac au niveau des régions introniques. Le 5<sup>ème</sup> état chromatinien est retrouvé spécifiquement sur le chromosome X et se distingue par l'enrichissement élevé de la marque H4K16ac, accompagné d'un enrichissement de la marque H3K36me3, ainsi que d'autres marques appartenant au 2<sup>ème</sup> état chromatinien d'élongation. Ce 5<sup>ème</sup> état chromatinien est notamment retrouvé au niveau du seul chromosome X des cellules mâles et est associé au mécanisme de compensation de dose retrouvé chez ces derniers [Larschan *et al.*, 2007]. Le 6<sup>ème</sup> état chromatinien est caractérisé par des répressions de la transcription par des complexes Polycomb et l'enrichissement de la marque H3K27me3. Les domaines hétérochromatiniens péri-centromériques et le chromosome 4 sont quant à eux caractérisés par un enrichissement des marques répressives H3K9me2 et H3K9me3, constituant le 7<sup>ème</sup> état chromatinien du génome de la Drosophile [Eissenberg et Reuter, 2009]. Le 8<sup>ème</sup> état chromatinien est également associé aux marques H3K9me2/me3, mais est retrouvé particulièrement au niveau du chromosome X [Eissenberg et Reuter, 2009]. Enfin, le 9<sup>ème</sup> état chromatinien est caractérisé par des états silencieux de la chromatine présentant l'enrichissement des marques histone H3K27me3 et H3K23ac, séparant divers clusters de chromatine ouverte.

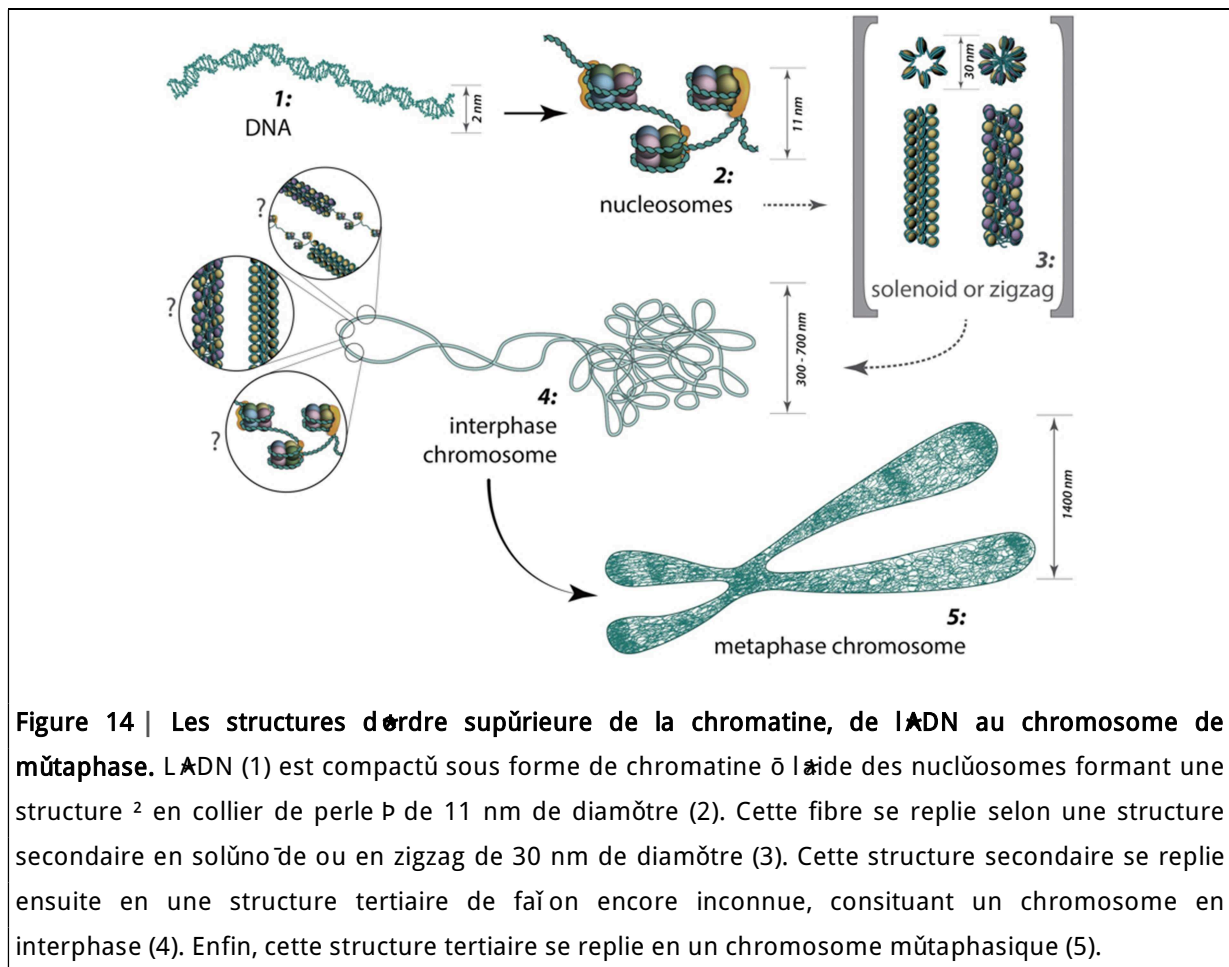
Les marques histones ainsi que leurs fonctions ayant globalement été conservées au cours de l'évolution, certains de ces états chromatiniens sont retrouvés chez d'autres organismes comme l'Homme, où une étude plus précise a permis l'identification de 51 états chromatiniens au travers du génome humain [Ernst et Kellis, 2010]. Par exemple 11 états chromatiniens de l'Homme partagent de fortes similitudes avec l'état chromatinien 1 du génome de *D. melanogaster* concernant les régions

présentant une activité transcriptionnelle importante accompagnée d'un enrichissement de H3K4me3. Chez la plante modèle *A. thaliana*, 42 états chromatiniens ont d'abord été identifiés [Luo *et al.*, 2013], puis réduits au nombre de neuf [Sequeira-Mendes *et al.*, 2014], comme chez *D. melanogaster*. Concernant la marque H3K4me3 par exemple, parmi les neuf états chromatiniens chez *A. thaliana*, deux présentent un enrichissement de H3K4me3 au niveau des sites d'initiation de la transcription [Sequeira-Mendes *et al.*, 2014]. En comparant ces états chromatiniens entre espèces, il est possible de définir avec précision la conservation des caractéristiques de différentes modifications post-traductionnelles des histones, notamment en terme de positionnement dans le génome vis-à-vis des unités transcriptionnelles et régulatrices (Figure 13).



## 5. Structure tridimensionnelle de la chromatine : définition et méthodes d'études

Au-delà de sa structure bidimensionnelle, la chromatine possède des structures d'ordre supérieur (Figure 14) comme une organisation spatiale tridimensionnelle au sein du noyau. *In vitro*, la structure primaire de la chromatine correspond à une organisation en <sup>2</sup> collier de perles par où se enchainent ADN libre et nucléosomes, observable au microscope électronique ou à force atomique dans des solutions possédant de faibles forces ioniques [Olins et Olins, 1974 ; Daban, 2011]. Dans des solutions contenant des cations divalents comme Mg<sup>2+</sup>, cette structure primaire s'arrange de manière spontanée en une structure secondaire plus dense, correspondant à la fibre de chromatine de 30nm de diamètre [Thoma *et al.*, 1979 ; Widom, 1986]. D'autres structures secondaires peuvent être formées à l'aide de protéines <sup>2</sup> architectes à non-histones comme les protéines nucléaires dites High Mobility Group (HMGN) ou les protéines du groupe Polycomb (PcG) [Luger et Hansen, 2005]. Sous certaines conditions physiologiques, ces structures secondaires peuvent se replier en une structure tertiaire formant des fibres de 100 à 130nm de diamètre [Belmont et Bruce, 1994 ; Hansen, 2002]. Ces structures peuvent enfin se compacter sous forme de chromosomes métaphasiques.



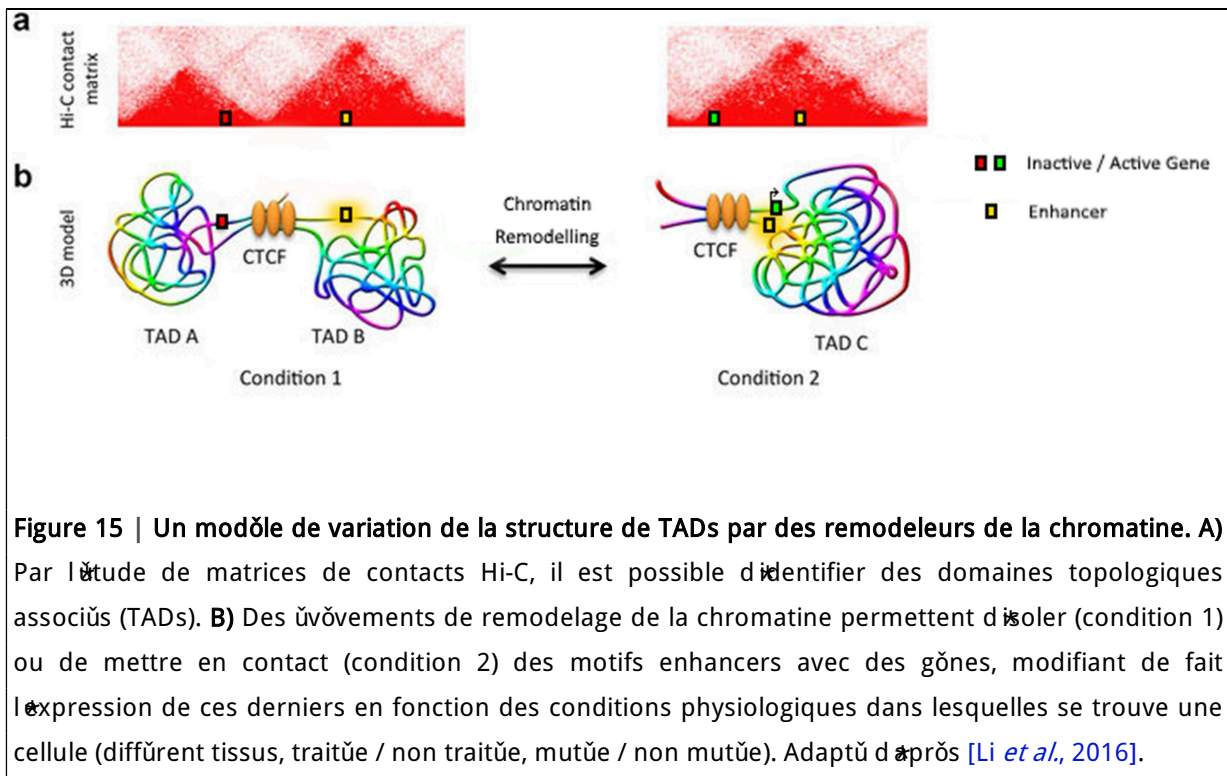
**Figure 14 | Les structures d'ordre supérieure de la chromatine, de l'ADN au chromosome de métaphase.** L'ADN (1) est compacté sous forme de chromatine à l'aide des nucléosomes formant une structure <sup>2</sup> en collier de perles à de 11 nm de diamètre (2). Cette fibre se replie selon une structure secondaire en solénoïde ou en zigzag de 30 nm de diamètre (3). Cette structure secondaire se replie ensuite en une structure tertiaire de façon encore inconnue, constituant un chromosome en interphase (4). Enfin, cette structure tertiaire se replie en un chromosome métaphasique (5).

Au sein du noyau, la structure tridimensionnelle de la chromatine permet de mettre en contact des régions pourtant très éloignées d'un même chromosome, voire des régions de différents chromosomes, en formant des domaines topologiquement associés (TADs), permettant une mise en place efficace de la régulation transcriptionnelle de nombreuses régions parfois distantes de plusieurs millions de nucléotides [Nora *et al.*, 2012]. Un TAD est défini comme une région montrant des niveaux élevés de contacts entre les séquences d'ADN au sein de cette région, et de faibles contacts, à longue ou courte distance, avec les régions voisines [Pombo et Dillon, 2015]. Les TADs prendraient la forme de boucles stables, définies par les protéines CTCF et cohésines [Stevens *et al.*, 2017].

Les TADs ont été étudiés à l'aide de méthodes de capture de la conformation des chromosomes comme le 3C [Dekker *et al.*, 2002] dont de nombreuses variantes existent, comme par exemple le plus récent Hi-C [Lieberman-Aiden *et al.*, 2009]. Le principe général du 3C est basé sur une fixation au formaldéhyde (cross-linking) de cellules suivie d'un isolement et d'une purification de la chromatine, suivie enfin de sa digestion par une enzyme de restriction. Pour le 3C, les enzymes utilisées sont choisies afin de cibler des contacts ADN-ADN entre deux régions d'intérêt et de libérer les fragments [Dekker, 2006]. Plusieurs paramètres sont à considérer afin de choisir la bonne enzyme de restriction, notamment la résolution à laquelle on veut cartographier les contacts et la distribution homogène des sites de coupure de l'ADN par l'enzyme de restriction au sein de la région étudiée [Naumova *et al.*, 2012]. Les fragments ainsi libérés sont ensuite ligués en anneaux, puis la fixation au formaldéhyde est annulée (reverse cross-linking). Cette étape de ligation est suivie d'une PCR quantitative afin de déterminer l'abondance des produits de ligation après purification de l'ADN. L'abondance ainsi mesurée correspond directement à la fréquence d'interaction des deux régions liguées. Ce principe peut être couplé à d'autres méthodes comme du séquençage afin d'en augmenter le débit ou l'échelle. Avec la méthode Hi-C [Lieberman-Aiden *et al.*, 2009], l'étape de ligation implique l'ajout d'un nucléotide marqué à la vitamine B (biotinylé) à la jonction de la ligation, permettant ainsi une purification sélective des produits de ligation, suivie d'un séquençage haut-débit [Belton *et al.*, 2012]. Ce couplage aux technologies NGS permet au Hi-C d'explorer sans *a priori* et à l'échelle du génome entier les interactions entre l'ensemble des régions du génome à l'aide de matrices de contact (Figure 15). Cette exhaustivité de l'Hi-C vis-à-vis des autres méthodes de capture de la conformation des chromosomes, telles que le 4C (Circularized Chromosome Conformation Capture) ou le 5C (Carbon Copy Chromosome Conformation Capture), permet d'explorer à la fois les propriétés biophysiques de la chromatine et les implications biologiques des structures tridimensionnelles de la chromatine au sein du noyau [Belton *et al.*, 2012]. En outre, les fragments d'ADN issus du Hi-C ont été montrés comme particulièrement adaptés à l'assemblage de génomes *de novo* [Bickhart *et al.*, 2017 ; Dudchenko *et al.*, 2017], ou à l'amélioration ou la correction de l'assemblage de génomes préexistants. Une étude Hi-C à l'échelle du génome de la souris a ainsi identifié 2200 TADs occupant 91% du génome, avec une taille moyenne de 880kb [Dixon *et al.*, 2012]. Cette étude a aussi montré que les frontières des TADs sont riches en gènes domestiques. En outre, la structure et l'organisation des TADs sont conservés entre différents types de cellules, et sont partiellement conservés entre les génomes de la souris et de l'homme [Dixon *et al.*, 2012]. La structure des TADs a également été étudiée chez *D. melanogaster*, et

il apparaît que ces derniers partagent beaucoup de caractéristiques avec les TADs des vertébrés [Sexton *et al.*, 2012 ; Hou *et al.*, 2012 ; Van Bortle *et al.*, 2014]. Les TADs de la Drosophile sont cependant plus petits, reflétant éventuellement le génome plus compact de cet insecte. Aucun TADs n'a pour le moment été observé chez *Saccharomyces cerevisiae* [Pombo et Dillon, 2015].

Notons que la modification de la structure des TADs est à l'origine de la mise en place de programmes génétiques spécifiques, par exemple entre cellules de tissus distincts, en permettant la mise en contact de motifs enhancer ou silencer voire insulator avec des promoteurs de gènes dont l'expression en est alors modifiée (Figure 15). Les TADs constituent ainsi, par rapport aux marques histones et à l'accessibilité de la chromatine, un niveau supplémentaire de régulation de la transcription des gènes. D'autres mécanismes épigénétiques permettent également de modifier l'expression des gènes en impactant la structure de la chromatine, comme la méthylation de l'ADN et les ARN non-codants.



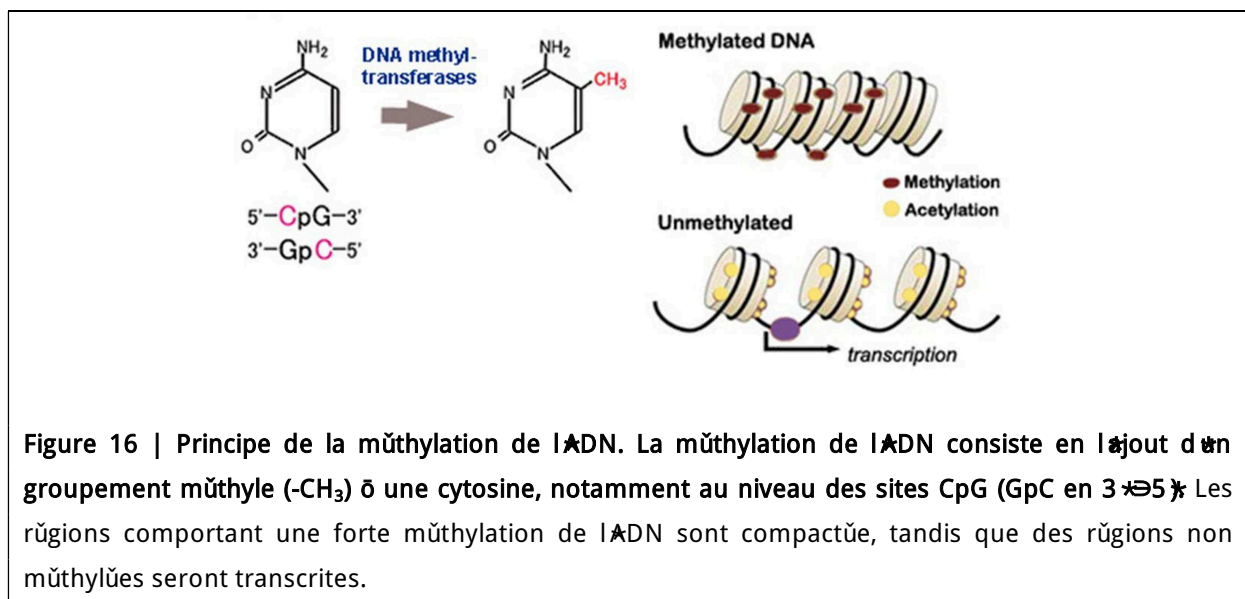
**Figure 15 | Un modèle de variation de la structure de TADs par des remodeleurs de la chromatine. A)** Par l'étude de matrices de contacts Hi-C, il est possible d'identifier des domaines topologiques associés (TADs). **B)** Des évènements de remodelage de la chromatine permettent d'isoler (condition 1) ou de mettre en contact (condition 2) des motifs enhancers avec des gènes, modifiant de fait l'expression de ces derniers en fonction des conditions physiologiques dans lesquelles se trouve une cellule (différents tissus, traitée / non traitée, mutée / non mutée). Adapté d'après [Li *et al.*, 2016].

## 6. Méthylation de l'ADN

### a. Définition et mécanismes

La méthylation de l'ADN est un mécanisme épigénétique correspondant à un processus biologique selon lequel un groupement méthyle est ajouté de façon covalente à une cytosine, donnant ainsi une 5-méthylcytosine (5mC) [Bestor, 1990 ; Bird et Wolffe, 1999 ; Jaenisch et Bird, 2003] (Figure 16). Ce processus de méthylation de l'ADN est assuré par un ensemble d'enzymes appelées ADN-

méthyltransférases (DNMT) [Goll et Bestor, 2005], utilisant du S-adenosyl méthionine comme origine de la méthylation. On retrouve principalement deux types de DNMT au travers de l'ensemble du règne vivant animal, végétal, bactérien et fongique notamment [Ehrlich *et al.*, 1985 ; Magill et Magill, 1989 ; Bestor, 2000 ; Law et Jacobsen, 2010]. Les DNMT1 sont considérées comme des méthyltransférases de maintenance ayant pour rôle de créer des groupements méthyle au niveau de brins d'ADN nouvellement synthétisés lors de la réplication de l'ADN, afin de conserver les profils de méthylation lors des divisions cellulaires [Goll et Bestor, 2005]. Les DNMT3 ont quant à elles un rôle de méthylation *de novo* de l'ADN [Goll et Bestor, 2005]. La DNMT2 n'est aujourd'hui plus considérée comme une véritable ADN méthyltransférase, car elle est impliquée dans la méthylation de la cytosine 38 des ARN de transfert au cours de la traduction des ARNm en protéines [Wang *et al.*, 2006]. La méthylation de l'ADN a pour effet de modifier les propriétés physiques de l'ADN en promouvant la formation de l'hétérochromatine par repositionnement des nucléosomes [Pennings *et al.*, 2005], empêchant la liaison des facteurs nucléaires à l'ADN. La méthylation de l'ADN peut également être reconnue par des protéines spécifiques telles que les methyl-CpG binding proteins (MBPs), capables de recruter des complexes protéiques additionnels dans le but de réguler l'activité transcriptionnelle [Fournier *et al.*, 2012]. On suppose également que la méthylation de l'ADN peut altérer l'affinité de l'ADN pour les nucléosomes [Collings *et al.*, 2013]. En outre, la méthylation de l'ADN est un processus biologique réversible. En effet, les enzymes de la famille des Ten Eleven Translocation (TET) oxydent les 5-méthylcytosine (5mC), générant ainsi des 5-hydroxyméthyle (5hmC), 5-formyl (5fC) et 5-carboxylcytosine (5caC) [Guo *et al.*, 2011 ; He *et al.*, 2011 ; Shen *et al.*, 2013]. La fonction biologique de ces cytosines modifiées reste incertaine, mais il apparaît qu'elles permettent la restauration de cytosines non modifiées [Spruijt *et al.*, 2013 ; Wu et Zhang, 2014].



### b. Méthodes d'étude

La méthylation de l'ADN peut être étudiée actuellement grâce aux techniques de Methylated DNA Immunoprecipitation suivie d'un séquençage haut-débit de l'ADN (MeDIP-seq) [Taiwo *et al.*, 2012] ou de Bisulfite Sequencing (BS-seq) [Li et Tollefsbol, 2011 ; Takayama *et al.*, 2014]. Le MeDIP-seq est basé sur l'utilisation d'anticorps ciblant les 5mc et sur le séquençage de cet ADN méthylé immunoprécipité. Le BS-seq est quant à lui basé sur l'utilisation du bisulfite ayant pour effet de convertir les 5mc en uracile. La comparaison entre les séquences de l'ADN traité au bisulfite et de l'ADN contrôle non traité permet alors de déterminer quelles cytosines sont méthylées à l'échelle du génome entier. Le développement d'une approche biochimique plus sensible et résolutive a permis de montrer que certains génomes précédemment considérés comme n'ayant pas de méthylation de l'ADN en comportent en réalité mais de manière très limitée, comme cela a été le cas pour *D. melanogaster* [Lyko *et al.*, 2000 ; Krauss et Reuter, 2011 ; Takayama *et al.*, 2014 : 2]. Cette nouvelle méthode (MeDIP-BS-seq) combine les deux protocoles précités ; elle permet de séquencer de l'ADN traité au bisulfite et préalablement enrichi en 5mc grâce à une étape d'immunoprécipitation utilisant des anticorps anti-5mc. L'ensemble de ces méthodes permet des études comparatives du méthylome entre différentes espèces, différents tissus ou mutants d'une même espèce, en étudiant le niveau global de méthylation, en analysant la distribution de cette méthylation entre régions géniques et intergéniques, ou encore en distinguant des régions différenciellement méthylées (DMRs).

### c. Méthylation et transcription de l'ADN

Ces méthodes ont par exemple permis de démontrer que les sites préférentiellement méthylés ne sont pas identiques entre les différents ordres du vivant, présentant chacun des contextes de méthylation différents. Chez la plupart des vertébrés, l'ADN est principalement méthylé au sein des dinucléotides CpG (ou CG), avec par exemple une méthylation de 70% à 80% des sites CG chez les mammifères. Quelques régions appelées îlots CpG (aussi retrouvés chez les plantes) et se trouvant au niveau du site TSS de gènes sont constitutivement non-méthylés, permettant la transcription de ces gènes. Chez les plantes et les champignons, en plus de la méthylation des régions CpG, une grande partie de la méthylation de l'ADN se produit également au niveau des nucléotides CHG ou CHH (où H est un nucléotide différent de G) [Selker, 1997 ; Lister *et al.*, 2008 ; Feng *et al.*, 2010 ; Jeon *et al.*, 2015].

Ces méthodes ont également montré que le pourcentage de cytosines méthylées varie de manière importante d'une espèce à l'autre. Par exemple, les vertébrés ont souvent un pourcentage élevé de méthylation des dinucléotides CpG (70% dans le génome Humain), tandis que certains invertébrés comme les diptères (*D. melanogaster*, *Anopheles gambiae*), certains hyménoptères (*Aphidius ervi*) et coléoptères (*Tribolium castaneum*) présentent une absence globale ou de faibles niveaux de méthylation des sites CpG de l'ordre du pourcent [Suzuki et Bird, 2008 ; Law et Jacobsen, 2010 ; Bewick *et al.*, 2017]. Ces variations peuvent également se retrouver au sein d'un même ordre. Par exemple au sein des Blattodea, *Reticulitermes flavipes* présente un niveau de méthylation du génome d'environ 5% tandis qu'une espèce voisine, *Reticulitermes virginicus* présente une absence globale de méthylation. Ce différentiel du niveau de méthylation peut s'expliquer par une présence / absence des

DNMT, ainsi que par un différentiel du nombre de copies de ces protéines au sein du génome [Bewick *et al.*, 2017]. Par exemple, l'ensemble des diptères présente une absence de protéines de la famille des DNMT1 et 3, corrélée avec une absence ou une très faible méthylation globale du génome, tandis que d'autres espèces comme le lépidoptère *Bombyx mori* présente un niveau de méthylation du génome de l'ordre de 1% ainsi que la présence de DNMT1 dans son génome [Bewick *et al.*, 2017]. Ceci témoigne d'une grande diversité de la présence, de l'absence ou du nombre de copies des DNMT et donc des proportions de 5mC au sein des génomes au travers de l'arbre du vivant. Le cas des pucerons sera traité plus spécifiquement dans la suite du manuscrit.

La méthylation de l'ADN joue un rôle important dans la régulation de l'expression des gènes et est généralement associée à une inhibition de l'expression des gènes, par exemple chez les mammifères dans l'inactivation du chromosome X au cours de la méiose [Bird, 1992 ; Jones et Takai, 2001]. En outre, la méthylation anormale de l'ADN est une caractéristique de nombreuses maladies comme certains cancers [Robertson, 2005], la schizophrénie [Gavin et Sharma, 2010] et les troubles du spectre autistique [Jiang *et al.*, 2004 ; Nagarajan *et al.*, 2006]. La méthylation de l'ADN s'avère également impliquée dans des formes de polyphénismes, notamment chez l'abeille, les fourmis et les criquets (détaillé en ").D).

## 7. ARN non-codants

### a. Petits ARN non-codants

#### **Définition**

Les petits ARN non codants (sncARN) sont définis chez les eucaryotes comme des ARN d'environ 20 à 30 nucléotides. Il existe plusieurs types de sncARN dont la définition repose sur de nombreux éléments tels que leur taille, leur biogénèse, leur mode d'action ou le type de protéine Argonaute (Ago) avec lesquelles ils interagissent. Le premier sncARN, lin-4, a été identifié par criblage génétique chez *C. elegans* et joue un rôle dans la régulation négative du niveau de traduction de la protéine LIN-14 entre le dernier stade embryonnaire et le premier stade larvaire [Lee *et al.*, 1993 ; Wightman *et al.*, 1993]. Depuis plusieurs années, le nombre de petits ARN caractérisés ne cesse d'augmenter grâce au séquençage haut-débit d'ARN triés par taille (smallRNA-seq) [Lagos-Quintana *et al.*, 2001 ; Lau *et al.*, 2001 ; Lee et Ambros, 2001]. Les avancées récentes dans le domaine du séquençage et de la prédiction bioinformatique [Lai *et al.*, 2003 ; Nam *et al.*, 2005 ; Li *et al.*, 2006 ; Huang *et al.*, 2007] des sncARN a permis la découverte de nouveaux sncARN, même faiblement représentés. Les fonctions des sncARN sont très diverses, allant de la formation de hétérochromatine à la déstabilisation et à la régulation traductionnelle des ARN messagers [Chu et Rana, 2007 ; Filipowicz *et al.*, 2008]. Ils participent également à des processus biologiques très divers dont notamment le développement, la différenciation des cellules, leur prolifération et leur mort, le silencing des transposons ainsi qu'au système immunitaire anti-viral. Au moins trois classes de sncARN sont actuellement caractérisées chez

les animaux : les micro-ARN (miARNs), les petits ARN interférant endogènes (endo-siARNs ou esiARNs ou siARN), et les ARN interagissant avec les protéines Piwi (piARN). Les miARN étant des éléments régulateurs post-transcriptionnels, seuls les piARNs et les siARNs sont considérés comme des régulateurs épigénétiques, et c'est donc de ces derniers que nous allons maintenant nous intéresser.

### Les petits ARN interférant

Les siARNs sont des sncARN de 21 à 24 nt produits à partir de longs ARN double brins (dsARN), considérés comme des précurseurs et issus de transposons, de transgènes, d'éléments répétés (siARN endogènes) ou de virus (siARN exogènes). Les dsARN précurseurs des siARN produits par ces éléments endogènes ou exogènes sont coupés par la Ribonucléase RNase "Dicer en duplex de siARNs de 22 à 23 nt de long comportant deux nucléotides cohésifs en 3' [Zamore *et al.*, 2000 ; Parrish *et al.*, 2000 ; Elbashir *et al.*, 2001 ; Bernstein *et al.*, 2001 ; Song *et al.*, 2003]. Les duplex de siARN sont pris en charge par des protéines de la famille AGO se associant à la partie 3' des siARN à l'aide du domaine M<sup>1</sup>D, et à la partie 5' grâce à un domaine P<sup>1</sup>W-AGO-ZW<sup>1</sup>LLE (PAZ) [Ma *et al.*, 2004, 2005 ; Wang *et al.*, 2009]. Les siARNs ainsi chargés vont guider la protéine AGO ainsi que des complexes protéiques associés (Dicer, R2D2) au niveau d'ARN complémentaires de la séquence du siARN qui sont ainsi marqués pour du silencing transcriptionnel, ou une répression traductionnelle [Hammond *et al.*, 2001 ; Hutvagner et Zamore, 2002 ; Meister *et al.*, 2004]. Chez les plantes, *Schizosaccharomyces pombe*, *Tetrahymena thermophila* et *Caenorhabditis elegans*, les protéines AGO guidées par des siARN recrutent une ARN polymérase ARN dépendante (RdRP), utilisant l'ARN cible comme base pour synthétiser un dsRNA converti par la protéine Dicer en de nouveaux siARNs, amplifiant ainsi l'effet RNAi de ces derniers [Dalmay *et al.*, 2000 ; Sijen *et al.*, 2001 ; Baulcombe, 2004 ; Motamedi *et al.*, 2004].

En terme de fonctions biologiques en lien avec des régulations épigénétiques, chez la plante *Arabidopsis thaliana*, des siARNs de 24 nucléotides régulent la méthylation de l'ADN *de novo* ainsi que sa maintenance au niveau des sites CHH, en étant dépendant de deux RNA polymérases " (RNA Pol " ) spécifiques des plantes : RNA Pol "V et RNA Pol V [Ream *et al.*, 2009 ; Matzke et Mosher, 2014]. Chez *S. pombe*, une délétion d'un des gènes ago1+, dcr1+ ou rdp1+ encodant respectivement pour des protéines AGO, Dicer et RdRP entraîne une diminution de la formation d'hétérochromatine au niveau des régions péri-centromériques, accompagnée d'une réduction globale de la méthylation de la Lysine 9 des histones H3 (H3K9me1/2/3). Certains organismes comme les mammifères ou *D. melanogaster* ne semblent pas présenter de RdRP. Cependant ces organismes présentent une autre classe de petits ARN non-codants, les P<sup>1</sup>W<sup>1</sup>-interactions RNAs (piARNs) [Holoch et Moazed, 2015].

### Les piwi interacting RNA

Les éléments génétiques mobiles tels que les transposons sont une menace constante pour l'intégrité du génome. Les piARN protègent notamment les cellules germinales contre les transposons au cours de la formation de gamètes et lors de la fécondation afin de limiter la dissémination de ces éléments dans le génome d'organismes animaux très divers comme les mouches, les poissons ou les

mammifères. Les piARNs sont des sncARN de 25 à 33 nucléotides de long, dépendant de la classe de protéine P<sup>W</sup> à laquelle ils s'associent [Meister, 2013]. Les piARNs dérivent de différents transposons dans le génome, correspondant alors à des clusters de piARNs. Ces sncARN sont caractérisés par une grande diversité de séquences couvrant de larges portions de ces différents transposons [Siomi *et al.*, 2011]. Les clusters de piARN sont transcrits de manière sens ou anti-sens et les longs ARN simple brins produits à partir de ces clusters servent de base de production pour les piARNs. Deux étapes principales sont impliquées dans la production des piARNs. La voie primaire de synthèse est chargée de produire les piARNs primaires dont la production est ensuite doublée par un cycle d'amplification appelé la boucle ping-pong [Siomi *et al.*, 2011]. Au cours de la voie de synthèse primaire de biogenèse des piARNs, les longs ARN issus des transposons sont d'abord clivés par la nucléase Zucchini (Zuc) [Psaro *et al.*, 2012 ; Nishimasu *et al.*, 2012 ; Voigt *et al.*, 2012], générant la terminaison 5' des piARNs primaires. Les étapes de maturation de ces piARNs primaires ne sont pas encore bien connues, bien que différents acteurs aient été identifiés [Siomi *et al.*, 2011 ; Czech et Hannon, 2016]. Au même titre que les siARNs, les piARNs contribuent à des régulations épigénétiques, notamment en contribuant à l'hérédité épigénétique. Cette idée a été mise pour la première fois dans une étude démontrant la nécessité des piARNs au cours de la reproduction, en assurant le silencing des transposons et la fertilité de la descendance [Brennecke *et al.*, 2008]. Chez la Drosophile au niveau de l'hétérochromatine, P<sup>W</sup> peut être dirigée contre une région d'ADN génomique ciblée par le piARN auquel elle est associée. Lors de la reconnaissance de cette région, P<sup>W</sup> recrute le remodelleur de la chromatine HP1 X qui recrute quant à lui l'histone méthyltransférase Su(var)3-9 qui va alors déposer un groupe méthyle sur la Lysine 9 de l'histone H3 proche de cette région [Brower-Toland *et al.*, 2007 ; Huang *et al.*, 2013]. Comme expliqué précédemment, cette marque va inhiber la fixation de l'ARN Pol II, inhibant de fait la transcription de la région complémentaire du piARN. Au niveau de l'euchromatine, P<sup>W</sup> est capable de cibler des transcrits naissant par complémentarité avec la séquence du piARN et de promouvoir le dépôt de H3K9me1, soit par recrutement de HP1 X puis Su(var)3-9, soit par recrutement direct de cette dernière ; ceci permet une fois à nouveau d'inhiber la transcription de la région ciblée, et cela démontre le lien existant entre piARNs et mécanismes épigénétiques [Brower-Toland *et al.*, 2007 ; Huang *et al.*, 2013 ; Ross *et al.*, 2014].

#### b. Longs ARN non-codants

Certains ncARN font plus de 200 nucléotides et ne codent pas pour des protéines et sont ainsi considérés comme de longs ARN non-codants (lncARN). Ces derniers ont été découverts grâce aux avancées dans le domaine du séquençage haut-débit et de l'analyse des données de RNA-seq [Birney *et al.*, 2007 ; Ulveling *et al.*, 2011 ; Rinn et Chang, 2012 ; Guttman et Rinn, 2012]. Ainsi, des transcrits ayant des propriétés similaires aux ARN messagers [Guttman *et al.*, 2009 ; Derrien *et al.*, 2012] ont été identifiés, sans toutefois avoir la capacité d'être traduits en protéines [Mercer *et al.*, 2008 ; Clark *et al.*, 2012 ; Yoon *et al.*, 2015]. De par leur définition large sans lien entre leur structure et leur fonction, les lncARNs sont hétérogènes en terme de biogenèse [Zhang *et al.*, 2014], de stabilité [Kaushik *et al.*, 2013], d'abondance [Guttman *et al.*, 2009 ; Derrien *et al.*, 2012] et d'évolution [Hiriart et Verdel, 2013 ; Diederichs, 2014 ; Saayman *et al.*, 2014 ; Tsai *et al.*, 2015], et sont ainsi supposés hétérogènes dans

leur mode de fonctionnement [Khandelwal *et al.*, 2015]. Bien qu'il soit difficile de déterminer la proportion des lncARNs qui est fonctionnelle, des dizaines de lncARN (parmi les milliers qui sont annotés dans les génomes) sont aujourd'hui caractérisés comme jouant un rôle dans la régulation de programmes géniques [Ma *et al.*, 2013 ; Fatica et Bozzoni, 2014].

Concernant leur biogénèse, les lncARN partagent de nombreuses caractéristiques avec celle des ARN messagers, comme leur synthèse par la RNAP II. Certains lncARNs ou classes de lncARN sont en revanche régulés différemment des ARNm tant au niveau de leur biogénèse, que de leur maturation ou de leur dégradation. Certains lncARN peuvent être synthétisés de manière sens ou anti-sens par la RNAP II au niveau de régions intergéniques, exoniques, au niveau de gènes, de promoteurs, de enhancers. Au niveau de régions introniques, ils sont synthétisés par la RNAP polynucléaire V au polypeptide unique (spRNAP V) [Khandelwal *et al.*, 2015]. Concernant la localisation des lncARN dans la cellule, certains d'entre eux peuvent se retrouver au niveau de la chromatine, de domaines subnucléaires, du nucléoplasme ou du cytoplasme, et comportent donc des localisations bien plus variées que les ARNm [Quinn et Chang, 2016]. L'un des exemples parmi les mieux caractérisés de l'implication de lncARN dans des mécanismes épigénétiques correspond au rôle du Xist, un transcript spécifique (Xist), orchestrant l'inactivation d'un des deux chromosomes X (XC) des femelles chez les mammifères [Partridge *et al.*, 2002 ; Verdell *et al.*, 2004]. Le Xist fut tout d'abord identifié car il est transcrit au niveau du X inactif, mais pas au niveau du X actif [Till *et al.*, 2007 ; El-Shami *et al.*, 2007]. Puis il a été démontré que sa séquence ne possédait pas de cadre ouvert de lecture [Motamedi *et al.*, 2004 ; El-Shami *et al.*, 2007], qu'il était localisé dans le noyau [Bähler *et al.*, 2006] en formant un unique compartiment nucléaire recouvrant le X inactif [Bähler, 2009]. La délétion du Xist a ensuite permis de montrer un échec de l'initiation du XC [Sugiyama *et al.*, 2007 ; Motamedi *et al.*, 2008]. La modification de la séquence du Xist par délétion entraîne une suppression du rôle de silencing de la transcription [Moazed, 2011] tandis que la modification de la localisation du Xist au niveau nucléaire empêche également son silencing. En outre, son expression ectopique sur d'autres chromosomes est suffisante à entraîner un silencing de l'expression des gènes de ces chromosomes [Horn *et al.*, 2005 ; Hong *et al.*, 2005]. Ces études ont donc montré que le lncARN Xist est une molécule d'ARN fonctionnelle requise pour le silencing de la transcription des gènes associés au chromosome X chez les mammifères, en étant associé à des enrichissements de marques histones comme H3K27me3, et donc à des mécanismes épigénétiques. Tout comme le Xist, d'autres lncARN ont de larges effets sur l'expression des gènes, suggéré par diverses études de perte de fonction [Zhang *et al.*, 2008 ; Gerace *et al.*, 2010 : 4 ; Li *et al.*, 2011 ; Zaratiegui *et al.*, 2011]. C'est notamment le cas de H<sup>1</sup>D1 chez *Arabidopsis thaliana*, un lncARN de 236nt impliqué dans le processus photomorphogénique [Wang *et al.*, 2014]. Les mutants pour ce lncARN présentant un phénotype hypophotomorphogénique en lumière rouge continue, H<sup>1</sup>D1 serait capable d'inhiber l'expression du gène P<sup>1</sup>F3, un régulateur clé de la photomorphogénèse modulant la réponse de la plante à la lumière, en se liant au promoteur de ce dernier [Wang *et al.*, 2014 ; Liu *et al.*, 2015].

## 8. Epigénétique chez les pucerons

### a. *Acyrtosiphon pisum*, un modèle d'étude pour la génomique des espèces aphidiennes

Le puceron du pois, *Acyrtosiphon pisum*, a été la première espèce de pucerons pour laquelle un grand nombre de ressources génomiques ont été développées, par l'International Aphid Genomics Consortium ("AGC). *A. pisum* présente un génome de 540 Mb, des individus aisément manipulables compte tenu de leur taille relativement importante, ainsi que de nombreux traits d'histoire de vie typiques des pucerons comme les polyphénismes de reproduction et de dispersion, une diversité de bactéries endo-symbiotiques et une vection de phytovirus circulants et non-circulants. Depuis 2006, un grand nombre de ressources génomiques, génétiques et bioinformatiques ont ainsi été développées pour cet hémiptère [International Aphid Genomics Consortium, 2010 ; Tagu *et al.*, 2014], avec notamment le séquençage et l'assemblage de son génome en plus de 20 000 scaffolds. Grâce à ces outils, il a été mené plusieurs études portant sur des analyses transcriptomiques et épigénétiques chez *A. pisum*.

### b. Epigénétique chez les pucerons

Les études épigénétiques menées chez les pucerons restent aujourd'hui peu nombreuses. Le génome de *Acyrtosiphon pisum* présente deux homologues des méthyltransferases DNMT1 (*dnmt1a* et *dnmt1b*), une DNMT2 et deux DNMT3 (*dnmt3a* et *dnmt3b*) retrouvées chez les vertébrés [International Aphid Genomics Consortium, 2010]. Ceci suggère que cette espèce comporte un système fonctionnel de méthylation de l'ADN, bien que leur génome ne soit que très peu méthylé (0.69% du génome au total) [Walsh *et al.*, 2010]. Une autre étude a quant à elle permis l'identification de protéines participant au remodelage de la chromatine, toujours chez *Acyrtosiphon pisum*, telles que des histones désacétylases, et des histones méthyltransferases (par identification de histones H3K4 et H3K9 méthylés), démontrant ainsi la présence d'une machinerie de régulation transcriptionnelle basée sur des marques histones [Rider Jr *et al.*, 2010]. Des données de cytogénétique ([Mandrioli et Borsatti, 2007] concernant l'analyse des marqueurs épigénétiques de l'hétérochromatine de *Acyrtosiphon pisum* ont quant à elles suggérées que l'hétérochromatine de cet organisme comporte un très faible taux de méthylation et que la majorité des méthylations se trouveraient au niveau de régions euchromatiniennes, et donc activement transcrites.

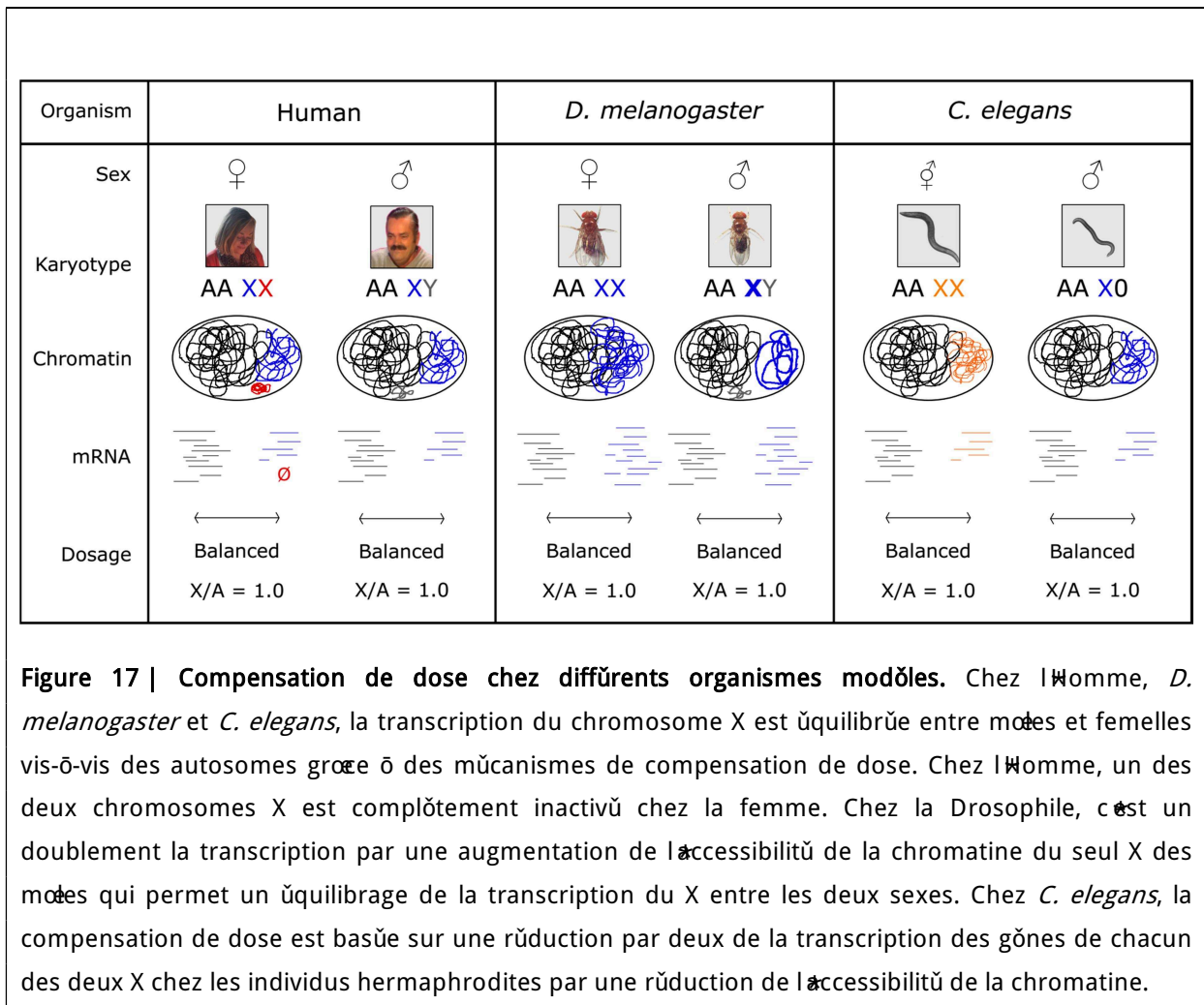
Néanmoins aucune étude visant à étudier des modifications post-traductionnelles des histones ou l'organisation de la chromatine à l'échelle du génome en relation avec le polyphénisme de reproduction ou de la potentielle compensation de dose n'a à ce jour été publiée concernant les pucerons. Chez d'autres organismes, les mécanismes épigénétiques impliqués dans ces deux processus biologiques ont été finement caractérisés, et nous allons à présent les détailler en commençant par la compensation de dose chez l'Homme, *D. melanogaster* et *C. elegans*

## C. Chromosome X et compensation de dose

### 1. Définition de la compensation de dose

Comme décrit précédemment, certaines espèces animales présentent un caryotype différent entre mâles et femelles avec une ou deux copies d'un chromosome sexuel selon le sexe, le chromosome sexuel unique pouvant être accompagné d'un autre chromosome généralement d'origine [Bull et Charnov, 1977 ; Ellegren, 2011]. Chez les espèces hétérogamétiques femelles, telles que les oiseaux ou les serpents, les mâles possèdent deux chromosomes Z tandis que les femelles possèdent un seul chromosome Z accompagné ou non d'un chromosome W (systèmes ZZ/Z0, ZZ/ZW) [Marshall Graves et Shetty, 2001]. Chez les espèces hétérogamétiques mâles comme les mammifères, de nombreux insectes ou encore le nématode *Caenorhabditis elegans*, les femelles possèdent deux X et les mâles un seul, accompagné ou non d'un chromosome Y (systèmes XX/X0, XX/XY). Ces hétérogaméties ou la présence d'un chromosome sexuel mâle ou femelle spécifique (W ou Y) sont à l'origine du déterminisme du sexe chez de nombreuses espèces, tandis que d'autres se différencient en mâle ou en femelles par des facteurs environnementaux comme chez certains poissons par exemple [Devlin et Nagahama, 2002]. Quoi qu'il en soit, ces hétérogaméties génèrent un déséquilibre du nombre d'allèles associés au chromosome sexuel en simple ou double copie, générant éventuellement un déséquilibre de transcription des gènes portés par ce chromosome sexuel [Charlesworth, 1996]. Ces déséquilibres peuvent générer des dysfonctionnements cellulaires importants voire délétères, notamment pour les gènes du chromosome sexuel codant pour des protéines en interaction avec d'autres protéines codées par des gènes portés par les autosomes [Charlesworth, 1996]. Certaines espèces hétérogamétiques ont donc acquis au cours de l'évolution des mécanismes dits de compensation de dose, permettant de rétablir un équilibre de la transcription du chromosome sexuel X ou Z entre sexes par rapport aux autosomes [Muller, 1932 ; Lyon, 1961 ; Lucchesi, 1993].

La compensation de dose peut être complète si elle régule la transcription d'une majeure partie des gènes associés au X ou au Z, ou elle peut être partielle lorsqu'elle régule la transcription uniquement d'un certain lot de gènes, sans doute critiques afin d'éviter des dysfonctionnements cellulaires. Trois mécanismes de compensation de dose ont été étudiés en profondeur à ce jour, et tous impliquent des mécanismes épigénétiques (Figure 17).



## 2. Les mécanismes épigénétiques participant à la compensation de dose chez les organismes modèles

### a. Compensation de dose chez les mammifères

Chez les mammifères, la compensation de dose s'effectue en grande partie par une répression globale de la transcription chez les femelles, mais également par une augmentation de la transcription de quelques gènes chez les mâles [Brockdorff et Turner, 2015]. Plusieurs études transcriptomiques [Nguyen et Distèche, 2006 ; Lin *et al.*, 2007 ; Deng *et al.*, 2011] ont en effet montré que les mâles comporteraient une augmentation de l'expression de plusieurs gènes de leur unique chromosome X, ce dernier comportant un ratio d'expression X : Autosomes supérieur à 0.5, et ce, de manière réversible. Cette augmentation n'étant pas globale, il est fait l'hypothèse que seuls les gènes les plus sensibles à une différence de dose (pour laquelle la non-compensation est par exemple létale) feraient l'objet de cette augmentation de la transcription [Pessia *et al.*, 2012]. La majeure partie de la compensation de dose chez les mammifères s'effectue donc chez les femelles, puisque la plupart des cellules de ces dernières comportent un corps de Barr, correspondant à une condensation complète de la chromatine d'un des deux chromosomes X [Brockdorff et Turner, 2015]. Ce mécanisme permet de n'avoir qu'un seul chromosome X actif que ce soit dans les cellules mâles ou femelles. Cette spectaculaire condensation de la chromatine du chromosome X maternel ou paternel chez les femelles est d'ailleurs reliée à certains phénotypes chez les mammifères, notamment les chattes à la robe tachetée de tortue ou calicot. Chez le chat, le chromosome X est en effet porteur du gène responsable de la pigmentation du pelage, alors que les cellules somatiques des femelles ne comportent qu'un seul X actif, maternel ou paternel, choisi de manière aléatoire selon les cellules. Ainsi, en fonction que le X actif soit maternel ou paternel et en fonction des allèles portés par chacun de ces chromosomes X, la couleur du pelage peut être différente d'une cellule de l'épiderme à l'autre, et ainsi générer une robe spécifique aux femelles. Les mâles ne présentant qu'un seul chromosome X dans leur génome, soit maternel soit paternel, ils présentent un pelage uni du même type que leur mère ou leur père [Kalantry, 2011].

La condensation de la chromatine à l'origine de l'inactivation du chromosome X (XC) s'effectue grâce au X inactivation Center (Xic), un locus comportant les gènes codant pour les ARN non-codants Xist [Brown *et al.*, 1991, 1992 ; Brockdorff *et al.*, 1992], Tsix [Lee *et al.*, 1999], Jpx [Tian *et al.*, 2010] et Ftx [Chureau *et al.*, 2011], tous impliqués dans l'inactivation du chromosome X, ainsi que des sites de fixation de protéines régulatrices. Le Xist est un lncARN transcrit tout d'abord faiblement par les deux X puis majoritairement par le X inactif. Le Xist s'enroule tout autour de la chromatine du chromosome X à inactiver, puis recrute les complexes protéiques Polycomb Repressor Complex 1 et 2 (PRC1 et PRC2), responsables d'un enrichissement global de la marque histone H3K27me3. PRC1 va ensuite permettre de stabiliser et maintenir l'inactivation du X par enrichissement de la mono-ubiquitination de la Lysine 119 de l'histone H2A (H2AK119Ub1) [Silva *et al.*, 2003 ; de Napoles *et al.*, 2004 ; Grossniklaus et Paro, 2014]. La chromatine du chromosome X inactif vis-à-vis du chromosome X actif

présente des niveaux élevés de méthylation de l'ADN, de méthylation de la Lysine 9 et de triméthylation de la Lysine 27 des histones H3, de faibles niveaux d'acétylation des queues des histones, ainsi que de faibles enrichissements de la marque H3K4me3 [O'Neill *et al.*, 2008], et un enrichissement de l'histone variant macroH2A [Costanzi et Pehrson, 1998 ; Csankovszki *et al.*, 1999].

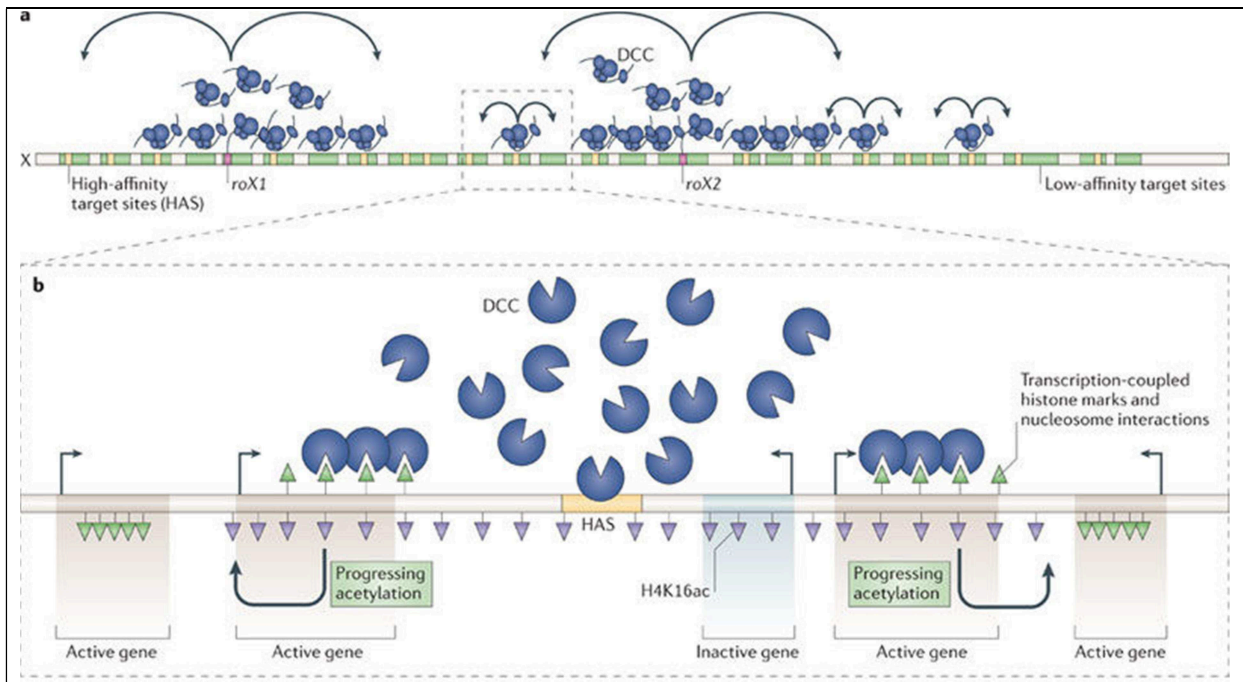
#### b. Compensation de dose chez *C. elegans*

Le nématode *C. elegans* comporte des individus mâles XO et des individus hermaphrodites XX. La compensation de dose s'effectue chez cette espèce de manière exclusive chez les hermaphrodites par une division globale par deux de la transcription de leurs deux chromosomes X [Meyer, 2005]. Ce mécanisme de compensation implique un Dosage Compensation Complexe (DCC), ressemblant en de nombreux points au Condensin Complexe responsable de la compaction de la chromatine et de la ségrégation des chromosomes lors de la méiose et de la mitose chez *C. elegans* [Lau *et al.*, 2014]. Une des protéines essentielles du DCC est sex determination and dosage compensation defective 2 (SDC-2), responsable : i) de la reconnaissance de manière spécifique des deux chromosomes X des hermaphrodites pour l'assemblage du DCC, ii) de la détermination du sexe en agissant de concert avec SDC-1 et SDC-3, plaçant de ce fait SDC-2 au cœur du programme génétique des individus hermaphrodites. L'expression de *sdc-2* est spécifique des individus hermaphrodites et son expression ectopique chez des individus XO déclenche l'assemblage du DCC sur leur unique chromosome X, entraînant la mort de ces individus [Meyer, 2005]. SDC-2 est ainsi une protéine hermaphrodite spécifique nécessaire et suffisante à l'activation de la compensation de dose. Cette dernière est également basée sur la régulation par le DCC de l'activité des protéines SET-1 et SET-4 qui contrôlent ensemble l'enrichissement de H4K20me1 le long des deux chromosomes X des individus hermaphrodites. Le DCC régule également l'activité de S<sup>tr</sup>-2.1, dupliquant ainsi les chromosomes X de la marque active H4K16ac [Lau *et al.*, 2014]. Ces deux phénomènes contribuent à réduire globalement l'expression des gènes portés par les deux chromosomes X des hermaphrodites afin qu'ils atteignent des niveaux de transcription équivalents au seul X des mâles.

#### c. Compensation de dose chez la Drosophile

Chez *Drosophila melanogaster*, la compensation de dose s'effectue exclusivement chez les individus mâles et implique différents éléments régulateurs [Conrad et Akhtar, 2012] : le complexe male-specific lethal (MSL, également appelé DCC bien que très différent du DCC de *C. elegans*), et les lncARN roX1 et roX2. Le complexe MSL est composé des protéines MSL1, MSL2, MSL3, maleless (MLE) et males-absent-on-the-first (MOF) [Hamada *et al.*, 2005 ; Straub *et al.*, 2005]. L'assemblage du complexe MSL est orchestré par la protéine MSL2, retrouvée uniquement chez les mâles [Conrad et Akhtar, 2012]. En effet, chez les femelles, la traduction de l'ARNm *msl2* est inhibée par la fixation de sex lethal (SXL), qui n'est pas correctement splicée chez les mâles, et donc pas exprimée [Sakamoto *et al.*, 1992 ; Kelley *et al.*, 1997 ; Bashaw et Baker, 1997]. MSL2 va d'une part stabiliser MSL1 qui régule quant à elle l'intégration de MSL3 et MOF [Morales *et al.*, 2004 ; Hallacli et Akhtar, 2009 ; Kadlec *et al.*, 2011], et d'autre part permettre la transcription des lncARN roX1 et roX2 [Rattner et Meller, 2004 ; Lee *et al.*, 2004 ; Li *et al.*, 2008 : 1 ; Aratani *et al.*, 2008], intégrés par la suite au sein du complexe par MLE [Meller *et al.*, 2000]. L'expression ectopique chez les femelles de MSL2 active l'assemblage complet du complexe MSL sur

les deux chromosomes X des femelles, ayant pour conséquence un phénotype létal ; ceci démontre bien le rôle majeur de MSL2 au sein du mécanisme de compensation de dose de la Drosophile [Kelley *et al.*, 1995]. Une fois assemblé, le complexe MSL va reconnaître des sites de haute affinité (HAS) distribués le long du chromosome X (Figure 18), puis se disséminer autour de ces HAS en ciblant des sites de faibles affinité correspondant notamment à des gènes [Palmer *et al.*, 1994 ; Lyman *et al.*, 1997 ; Straub *et al.*, 2008 ; Alekseyenko *et al.*, 2008 ; Conrad et Akhtar, 2012]. Ce mécanisme est favorisé au niveau du chromosome X par la transcription des lncARN roX1 et roX2, mais peut également avoir lieu au niveau des autosomes que le complexe MSL peut cibler dans certaines conditions. Au niveau des sites de faible affinité, le complexe MSL va modifier les queues des histones en y déposant la marque H4K16ac grâce à la protéine MOF [Hallacli et Akhtar, 2009]. Cette marque active est globalement corrélée avec une augmentation de l'accessibilité de la chromatine aux facteurs de transcription, dont la RNAP ". Ce mécanisme fin va ainsi permettre une acétylation globale de la chromatine du seul X des mâles, conduisant à un doublement de la transcription du chromosome X, permettant ainsi d'équilibrer l'expression du X vis-à-vis des autosomes par rapport aux femelles [Conrad et Akhtar, 2012]. Le mécanisme permettant le doublement précis de la transcription du seul chromosome X des mâles reste cependant mal compris.



**Figure 18 | Modèle de dispersion du complexe MSL au niveau du chromosome X des individus mâles dans le cadre de la compensation de dose de *D. melanogaster*.** L'identification du chromosome X des mâles par le complexe de compensation de dose (DCC) est la première étape du processus de compensation de dose. A) Le DCC devient entièrement compétent pour le ciblage de la chromatine suite à l'incorporation des lncARN roX au niveau de leurs sites de synthèse sur le chromosome X. Le DCC atteint ensuite ses sites cibles finaux par diffusion au sein de la chromatine. Ces derniers correspondent à des sites de haute affinité (HAS), attirant le DCC vers de multiples loci à travers le chromosome X à partir d'où il se propage sur ses sites de faible affinité. B) Il a été proposé que l'attraction physique du DCC par un HAS entraînerait une augmentation locale de la concentration en DCC. Ce déséquilibre biochimique local déclencherait alors la fixation du DCC à des sites de faible affinité, où le DCC générerait des modifications post-traductionnelles d'histones, à savoir H4K16ac, permettant une augmentation de la transcription de gènes actifs. Comme le DCC est capable de modifier les histones pendant des interactions transitoires dans l'espace tridimensionnel de la chromatine, la marque H4K16ac peut se répartir sur plusieurs kilobases autour des sites de liaison du DCC. L'exposition de ces sites au DCC grâce à la structure tridimensionnelle de la chromatine dans le noyau pourrait constituer un mécanisme de diffusion alternatif du DCC. D'après [Conrad et Akhtar, 2012].

### 3. Étude de la compensation de dose chez les insectes non-drosophilides

Outre les organismes modèles précités chez qui la compensation de dose a été caractérisée à des niveaux très fins, il a été montré que d'autres organismes, non-modèles ou non-modèles, présentent une compensation de dose au niveau transcriptionnel (aucune étude épigénétique n'ayant été menée chez ces organismes dans le cadre de la compensation de dose). La caractérisation au niveau transcriptionnel de la compensation de dose est principalement effectuée par la comparaison des ratios d'expression de l'ensemble des gènes portés par le (les) chromosome(s) X ou Z sur l'ensemble des gènes portés par les autosomes, et ce au sein de chaque sexe. Il est également possible de comparer la distribution de l'expression des gènes portés par le X (ou le Z) vis-à-vis des gènes portés par les autosomes [Deng *et al.*, 2011] en filtrant au préalable les gènes les moins exprimés ou tissus spécifiques [Deng *et al.*, 2011]. Ce type d'expérimentation a été mené notamment chez *Manduca sexta* [Smith *et al.*, 2014], *Anopheles stephensi* [Jiang *et al.*, 2015] et *Anopheles gambiae* [Hahn et Lanzaro, 2005 ; Baker *et al.*, 2011] : des compensations de dosage complètes ont été mises en évidence avec des ratios X:AA = 1 chez les mâles et les femelles, et des différences non-significatives de distribution d'expression des gènes liés aux autosomes et au X. Chez *Tribolium castaneum*, une compensation de dose semble être présente chez les mâles, présentant un ratio d'expression X:AA égal à 1. Cependant le chromosome X de cette espèce étant enrichie en gènes femelle-biaisés, les femelles présentent un ratio X:AA > 1, s'éloignant ainsi de la compensation de dose canonique [Prince *et al.*, 2010]. Certains insectes du genre *Strepsitera* ont quant à eux montré une compensation de dose partielle (ciblant certains gènes) [Mahajan et Bachtrog, 2015], tandis que *Teleopsis dalmanii* ne présente pas de compensation de dose (X:AA = 0.5 chez les mâles, XX:AA= 1 chez les femelles) [Wilkinson *et al.*, 2013]. Chez le puceron du pois *Acyrtosiphon pisum*, la compensation de dose a été étudiée au niveau transcriptomique [Jaquiury *et al.*, 2013 ; Pal et Vicoso, 2015], et fait l'objet d'un chapitre dédié au sein de ce manuscrit (V. Chapitre 1).

Chez la plupart de ces organismes non-modèles, la compensation de dose a été étudiée par l'intermédiaire de RNA-seq d'individus mâles et femelles entiers ou à des stades larvaires précoces, et donc sexuellement immatures. Cependant, l'utilisation de ce type de matériel biologique peut induire des biais importants dans l'évaluation et la caractérisation de la compensation de dose au niveau du chromosome X entre les deux sexes, ce mécanisme pouvant être tissu spécifique [Deng *et al.*, 2011]. Ainsi, il a été suggéré qu'afin de tester de manière précise la compensation de dose chez un organisme, il est important de réaliser des analyses d'expression différentielle de gènes au sein de plusieurs tissus afin de déterminer quels gènes sont tissu spécifiques, et éliminer ces gènes lors du calcul des ratios X:A. Cette filtration concerne notamment les gènes spécifiques des gonades et des têtes, ces tissus présentant généralement des profils de compensation différents comparativement à d'autres tissus considérés neutres (membres, muscles, organes vitaux) [Deng *et al.*, 2011]. En outre, une filtration graduelle des gènes selon leur expression avant le calcul des ratios X:A, au sein de chaque tissu, permet une détermination précise de la compensation de dose chez les mâles et les femelles.

Après avoir détaillé l'implication des mécanismes épigénétiques au sein de la compensation de dose, intéressons-nous à présent au rôle de ces mécanismes dans le cadre de la plasticité phénotypique, et plus particulièrement dans la régulation du polyphénisme chez les insectes.

## D. Plasticité phénotypique, polyphénismes et mécanismes épigénétiques chez les insectes

Au cours de ces dernières années, des avancées ont été réalisées concernant la compréhension du rôle des mécanismes épigénétiques dans le cadre de divers polyphénismes existant chez les insectes. Les exemples de polyphénisme les mieux caractérisés chez ces derniers du point de vue des mécanismes moléculaires sous-jacents correspondent à ce jour au polyphénisme de caste chez les abeilles, par exemple *Apis mellifera* [Evans et Wheeler, 1999], et chez les fourmis telles que *Camponotus floridanus* [Bonasio, 2014], ainsi qu'au polyphénisme de phase chez les criquets, permettant aux individus d'adopter un comportement solitaire ou grégaire en fonction de stimuli environnementaux [Ernst et al., 2015]. Chacun de ces trois polyphénismes implique des mécanismes épigénétiques qui vont maintenant être détaillés.

### 1. Le polyphénisme de castes chez l'abeille

Le polyphénisme de castes chez *Apis mellifera* consiste principalement en une différenciation entre la caste des ouvrières, accomplissant la plupart des tâches nécessaires au fonctionnement et à la survie de la ruche, et la caste de la reine, unique individu capable de se reproduire au sein de la ruche et comportant un phénotype extrêmement différent des ouvrières (taille plus importante, appareil reproducteur fonctionnel, comportement différent vis-à-vis des prédateurs et des congénères) [Page et al., 2006] (Figure 1). Ce polyphénisme de castes est déclenché par une nutrition spécifique des larves afin qu'elles deviennent reines : si une larve est nourrie à base de gelée royale lors de son développement, alors des programmes génétiques spécifiques seront mis en place afin d'orienter son développement en reine [Wilson, 1971 ; Welch et Lister, 2014]. D'un point de vue des mécanismes épigénétiques mis en place dans ce polyphénisme, l'abeille constitue sans doute l'exemple le mieux caractérisé à ce jour concernant l'implication de la méthylation de l'ADN dans un polyphénisme chez les insectes [Welch et Lister, 2014]. La première preuve que la méthylation de l'ADN joue un rôle central dans la détermination des castes des abeilles a été obtenue par ARN interférant (RNAi) ciblant le gène *Dnmt3*. Cette expérience a entraîné une modification du devenir des larves, puisque un quart des larves non-traitées se sont développées en reine, tandis que trois-quarts des larves présentant un gène *Dnmt3* silencieux sont devenues des reines [Kucharski et al., 2008]. Une autre étude plus récente a montré que l'expression de *Dnmt3* était significativement plus faible au cours du développement de larves alimentées expérimentalement avec de la gelée royale [Shi et al., 2011]. Le profilage du méthylome de larves nourries avec de la gelée royale ou commune pendant 4 jours après éclosion a révélé que dans 80% des régions différenciellement méthylées identifiées, les larves à destinée

ouvrière comportaient des niveaux de méthylation plus élevés que les larves destinées de reine [Foret *et al.*, 2012]. Ceci suggère que la méthylation d'ADN *de novo* régulée par *Dnmt3* serait cruciale pour la définition de la trajectoire développementale des larves chez *A. mellifera*, sans doute par réduction de la transcription de nombreux gènes chez les larves destinées ouvrière. Enfin, des méthylations différentielles ont été identifiées entre le cerveau des reines et des ouvrières au niveau de 550 gènes, dont une protéine membranaire GB18602, un gène codant une protéine transmembranaire cadhérine GB18207 et TAP42 impliqué dans la signalisation de type 2 Target of Rapamycin P (TOR) responsable de la réaction des cellules vis-à-vis de la disponibilité en énergie et de l'hypoxie. Cela suggère l'implication de la méthylation des cytosines dans la définition du comportement de ces deux castes [Lyko *et al.*, 2010].

Les comparaisons de séquences des gènes de *A. mellifera* avec leurs homologues chez les espèces modèles (Humain, *C. elegans*, *A. thaliana*, *D. melanogaster*, *Escherichia coli*) indiquent une plus grande conservation de séquence des gènes présentant un corps méthylé par rapport au gènes comportant un corps non-méthylé, suggérant ainsi que la régulation de ces loci par des modifications épigénétiques serait fortement conservée par l'évolution [Lyko *et al.*, 2010]. Welch et Lister ont émis l'hypothèse que l'expression de *Dnmt3* serait inhibée par la prise de gelée royale, limitant alors le potentiel de la méthylation *de novo* de l'ADN chez les larves destinées à la caste royale, tandis que l'accumulation active de méthylations *de novo* de l'ADN au niveau de certains gènes conservés par l'évolution permettrait l'orientation du développement vers la caste ouvrière [Welch et Lister, 2014]. En outre, il existe un lien fort entre la méthylation du corps des gènes et la formation de transcrits alternatifs en relation avec le polyphénisme de caste chez les abeilles. "Il y a en effet un enrichissement du complexe spliceosomal et de facteurs de splicing au niveau des gènes méthylés, ainsi qu'une formation accrue de transcrits alternatifs au niveau de ces gènes." Il est ainsi suggéré que les régions méthylées dans le corps des gènes pourraient jouer un rôle dans la sélection des exons inclus dans les transcrits matures [Foret *et al.*, 2012].

En outre, plusieurs modifications post-traductionnelles d'histones différentielles entre les deux castes ont été caractérisées et ne semblent pas être causées directement par l'ingestion de gelée royale, pourtant à l'origine de la modification de la trajectoire de développement des larves [Wilson, 1971 ; Dickman *et al.*, 2013]. La gelée royale est une substance complexe contenant des protéines, des lipides, des sucres, des vitamines, des minéraux et divers composés bioactifs [Viuda-Martos *et al.*, 2008]. L'un des principaux constituants de la gelée royale est un inhibiteur d'histone désacétylase (HDAC inhibitor) : l'acide (E)-10-hydroxy-2-décénoïque (10HDA). Comme décrit précédemment, les histones acétylées permettent une conformation plus accessible de la chromatine, en général associée à des niveaux plus élevés d'expression génique [Strahl et Allis, 2000]. Il a été montré que le 10HDA peut réactiver des gènes silencieux par son activité inhibitrice d'HDAC dans certains modèles de cellules humaines [Spannhoff *et al.*, 2011]. Cependant, il n'a toujours pas été démontré chez l'abeille que l'apport alimentaire d'inhibiteurs d'HDAC peut entraîner des changements de l'épigénome.

## 2. Le polyphénisme de castes chez les fourmis

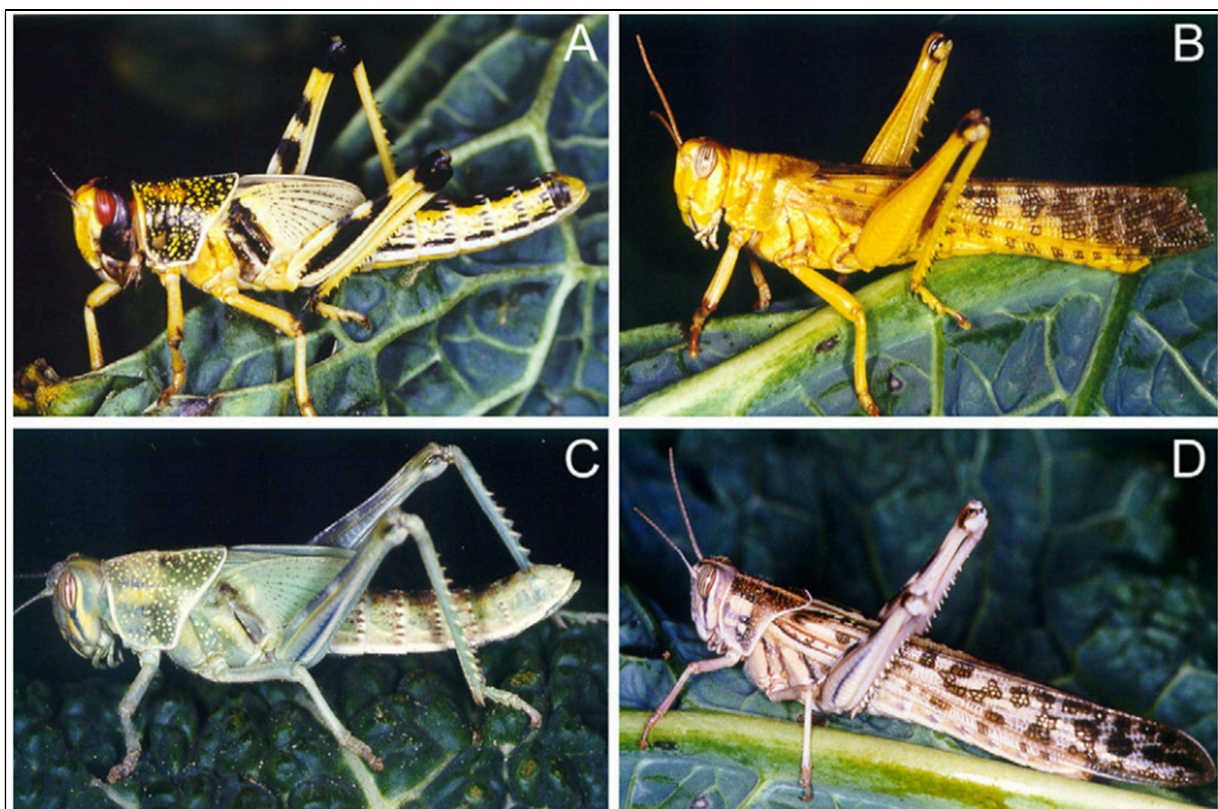
Les fourmis présentent également un polyphénisme de castes orientant le développement des larves en ouvrières, en guerrières ou en reine. En fonction des espèces, ce polyphénisme peut être déterminé par la nutrition avec de la gelée royale, comme chez *Camponotus floridanus* et *Harpegnathos saltator* [Bonasio *et al.*, 2012], ou par un déterminisme génétique comme chez certaines lignées de *Pogonomyrmex barbatus* [Volny et Gordon, 2002]. Chez *Camponotus floridanus*, un ensemble de gènes différenciellement méthylés entre reines et travailleuses a été identifié par comparaison de leurs méthylomes [Bonasio *et al.*, 2012]. Certains de ces gènes ont été identifiés comme conservés entre *C. floridanus* (sous-famille Formicinae) et *Harpegnathos saltator* (sous-famille Ponerinae) [Bonasio *et al.*, 2012]. Cette observation suggère que la régulation épigénétique des traits spécifiques aux castes pourrait donc être similaire entre ces deux familles. Parmi ces gènes conservés, certains comportent des annotations fonctionnelles en relation avec la reproduction, suggérant ainsi que le polyphénisme de caste, ici relié à la reproduction, pourrait être sous contrôle épigénétique, et plus particulièrement régulé par la méthylation de l'ADN [Bonasio, 2014]. En accord avec cette hypothèse, il a été identifié des CpG méthylés de façon moins fréquente chez des lignées comportant des castes déterminées de manière génétique par rapport à des lignées comportant des castes déterminées de manière environnementale chez *Pogonomyrmex barbatus* [Volny et Gordon, 2002] [Smith *et al.*, 2012].

Outre la méthylation de l'ADN, les modifications post-traductionnelles des histones ont fait l'objet d'études approfondies chez *Camponotus floridanus*, qui est actuellement l'exemple le mieux caractérisé de l'implication de marques histones en lien avec un polyphénisme chez les insectes [Simola *et al.*, 2013a]. Des études de ChIP-seq et/ou de RNA-seq ont été menées chez *C. floridanus* et *H. saltator* afin d'identifier les marques histones différenciellement enrichies entre les différentes castes de ces deux espèces. Il a d'abord été identifié que les différentes familles d'enzymes permettant la modification des histones sont largement conservées entre *C. floridanus* et *H. saltator* [Bonasio *et al.*, 2010]. Chez cette dernière, il a été montré que des gènes codant pour des histones méthyltransférases de la famille des SMYD (contenant des domaines SET et MYND) possèdent des profils d'expression caste-spécifiques [Bonasio *et al.*, 2010]. Cette famille d'enzymes jouerait ainsi un rôle important dans la détermination des castes. Chez *C. floridanus*, il a été montré que des changements dans le paysage épigénétique au niveau de certains gènes codant permettraient de discriminer deux castes d'ouvrières, mineures et majeures, chacune ayant son propre comportement d'exploration et de recherche de ressources. La marque histone H3K27ac a notamment été identifiée comme la marque la plus différenciellement enrichie entre ces deux castes [Simola *et al.*, 2013b]. De plus, le gène codant pour un homologue de CBP chez *C. floridanus*, c'est-à-dire la protéine responsable du dépôt de H3K27ac chez *D. melanogaster* [Spannhoff *et al.*, 2011], a été identifié comme différenciellement exprimée entre ces deux castes.

### 3. Le polyphénisme de phase chez le criquet

Les criquets présentent un polyphénisme de phase, c'est-à-dire la capacité d'un individu à adopter un comportement solitaire ou grégaire en fonction de son environnement. Ce polyphénisme est déclenché par le contact d'un individu avec son environnement ou avec ses congénères au niveau de ses pattes arrières [Ellis, 1959 ; Simpson *et al.*, 2001]. Les criquets ainsi stimulés vont adopter un comportement grégaire et leur descendance arborera des couleurs vives permettant d'attirer d'autres individus grégaires [Uvarov *et al.*, 1966 ; Pener et Simpson, 2009] (Figure 19), mais également des individus solitaires, occasionnant alors des contacts physiques entre individus et permettant donc l'adoption d'un comportement grégaire par des individus solitaires [Ernst *et al.*, 2015]. Lors de la raréfaction de la nourriture, les surfaces vont se réduire, les criquets seront davantage concentrés, de nombreux individus vont entrer en contact entre eux et le polyphénisme de caste va provoquer la formation de nuages entiers de criquets, favorisant leur déplacement vers de nouvelles ressources, et capables de réduire à nu des hectares de cultures agronomiques, le criquet étant une espèce phytophage.

La méthylation de l'ADN chez le criquet a commencé à être étudiée par Wyatt en 1951, qui a montré que 0,96% des cytosines était méthylées au sein du génome de *Locusta migratoria* en utilisant une hydrolyse quantitative des acides nucléiques [Wyatt, 1951]. Ces résultats ont été vérifiés plus récemment par un séquençage au bisulfite où 1.3% des cytosines totales ont été identifiées comme méthylées chez *L. migratoria* [Robinson *et al.*, 2011], et 1.3% à 1.9% chez *Schistocerca gregaria* en fonction des tissus [Boerjan *et al.*, 2011]. Chez ces deux espèces de criquets, et à l'instar de nombreux insectes, des méthylcytosines (5mc) sont retrouvées dans le corps des gonées et au niveau des éléments reproductifs [Robinson *et al.*, 2011]. Une analyse comparative du méthylome de tissus cérébraux d'individus solitaires et grégaires chez *L. migratoria* a montré que les régions codantes comportaient moins de méthylation des CpG que le reste du génome [Wang et Kang, 2014]. De plus, cette même étude a montré que 90 gonées sont différenciellement méthylées entre individus solitaires et grégaires. En analysant la fonction de ces gonées, il s'avère que des processus tels que le développement du cytosquelette, la plasticité synaptique et le contrôle de la dynamique des microtubules joueraient un rôle dans la transition d'une phase à l'autre au niveau du cerveau de *L. migratoria* [Ernst *et al.*, 2015]. Les modifications post-traductionnelles des histones pourraient également être impliquées puisque l'on a été montré chez *S. gregaria* que les tissus cérébraux des individus grégaires comportent des histones H3 en moyenne plus phosphorylées que celles des individus solitaires [Ernst *et al.*, 2015]. L'ensemble de ces résultats suggère un rôle potentiel des mécanismes épigénétiques dans le contrôle du polyphénisme de phase chez le criquet.



**Figure 19 | La coloration corporelle de *S. gregaria* dépend en grande partie de sa phase et de son stade de développement.** Sont ici représentés des mâles au dernier stade larvaire (A, C) et au stade imago (B, D). Les criquets solitaires (C, D) sont plus grandes et colorés de manière discrète tandis que les criquets grégaires (A, B) présentent des couleurs aposomatiques. D'après [Ernst *et al.*, 2015].

Nous avons donc montré dans cette synthèse bibliographique que divers mécanismes épigénétiques sont impliqués d'une part dans la régulation du polyphénisme chez des insectes tels que les abeilles, les fourmis et les criquets, et d'autre part dans la compensation de dose de l'insecte modèle *D. melanogaster*. Concernant le puceron du pois - l'organisme d'étude de cette thèse - il a été montré que dans le cadre du polyphénisme de reproduction, des gènes impliqués dans des régulations épigénétiques sont régulés au cours de la différenciation des cellules germinales des embryons présentant une future reproduction sexuelle ou asexuelle (histones linker H1, histone H2B.3, *suv(4)-20H1* et *uhrf1*). Dans le cadre de la compensation de dose de *A. pisum*, un profil différentiel d'expression du chromosome X a été identifié entre mâles et femelles, suggérant une potentielle compensation de dose chez *A. pisum*, et donc l'implication éventuelle de mécanismes épigénétiques régulant l'accessibilité de la chromatine dans ce processus. Nous allons maintenant présenter les objectifs et stratégies de cette thèse, puis poursuivre avec une description générale du matériel biologique et des méthodes qui ont été utilisées.

## "" . Objectifs et stratégies

Ma thèse a pour objectif d'étudier chez *A. pisum* l'accessibilité de la chromatine et la transcription des gènes dans le cadre i) de la potentielle compensation de dose pouvant avoir lieu au niveau des gènes portés par le chromosome X des pucerons mâles et des femelles, et ii) du polyphénisme de reproduction, ces deux processus biologiques ayant lieu au cours du cycle annuel de reproduction des pucerons (Figure 20).

Concernant la compensation de dose chez les pucerons, nous allons vérifier s'il existe un différentiel d'expression global au niveau du chromosome X entre mâles et femelles par RNA-seq, ou au contraire si leurs niveaux d'expression sont équivalents chez *A. pisum*. Ces étapes seront réalisées grâce à des données de RNA-seq qui ont été générées par [Jaquiéry *et al.*, 2013] de manière antérieure à ma thèse, ainsi qu'à l'aide d'une nouvelle assignation du chromosome X et des autosomes du génome de *A. pisum* [Jaquiéry *et al.*, 2017]. Grâce à cette nouvelle assignation, nous allons également vérifier s'il est possible d'identifier un profil d'accessibilité de la chromatine différentiel entre mâles et femelles au niveau du chromosome X et des autosomes par Formaldehyde Assisted "solation of Regulatory Elements suivie d'un séquençage haut-débit (FA"RE-seq). Ces données vont principalement être analysées d'un point de vue de l'accessibilité globale de la chromatine entre chromosome X et autosomes chez ces deux sexes. Les résultats obtenus suite à ces expériences font l'objet d'un article publié dans Epigenetics & Chromatin, présenté au sein du Chapitre ".

Concernant le polyphénisme de reproduction, nous avons appliqué les méthodes de FA"RE-seq et RNA-seq aux quatre derniers stades embryonnaires de *A. pisum* (stades 17 à 20), à devenir sexué ou asexué, en fonction d'un traitement hormonal subit par leur mère, selon un protocole utilisé par [Gallot *et al.*, 2012] qui sera détaillé plus amplement en Matériel et Méthodes. Nous allons étudier les programmes génétiques associés aux phénotypes alternatifs lors du basculement du mode de reproduction en comparant les données d'expression et d'accessibilité de la chromatine. Dans un premier temps, nous allons cartographier l'ensemble des régions du génome associées à une accessibilité différentielle de la chromatine entre embryons à devenir sexué et asexué selon leur stade de développement. Ces régions d'intérêt seront ensuite caractérisées afin d'identifier des éléments régulateurs putatifs (enhancers, silencers) qui seront mis en relation avec des gènes dont l'expression est régulée et dont nous analyserons les fonctions putatives. Les résultats seront présentés dans un second chapitre sous forme d'un manuscrit.

Un synopsis général concernant les expérimentations biologiques et les expérimentations bioinformatiques utilisées lors de ces travaux va maintenant être décrit.

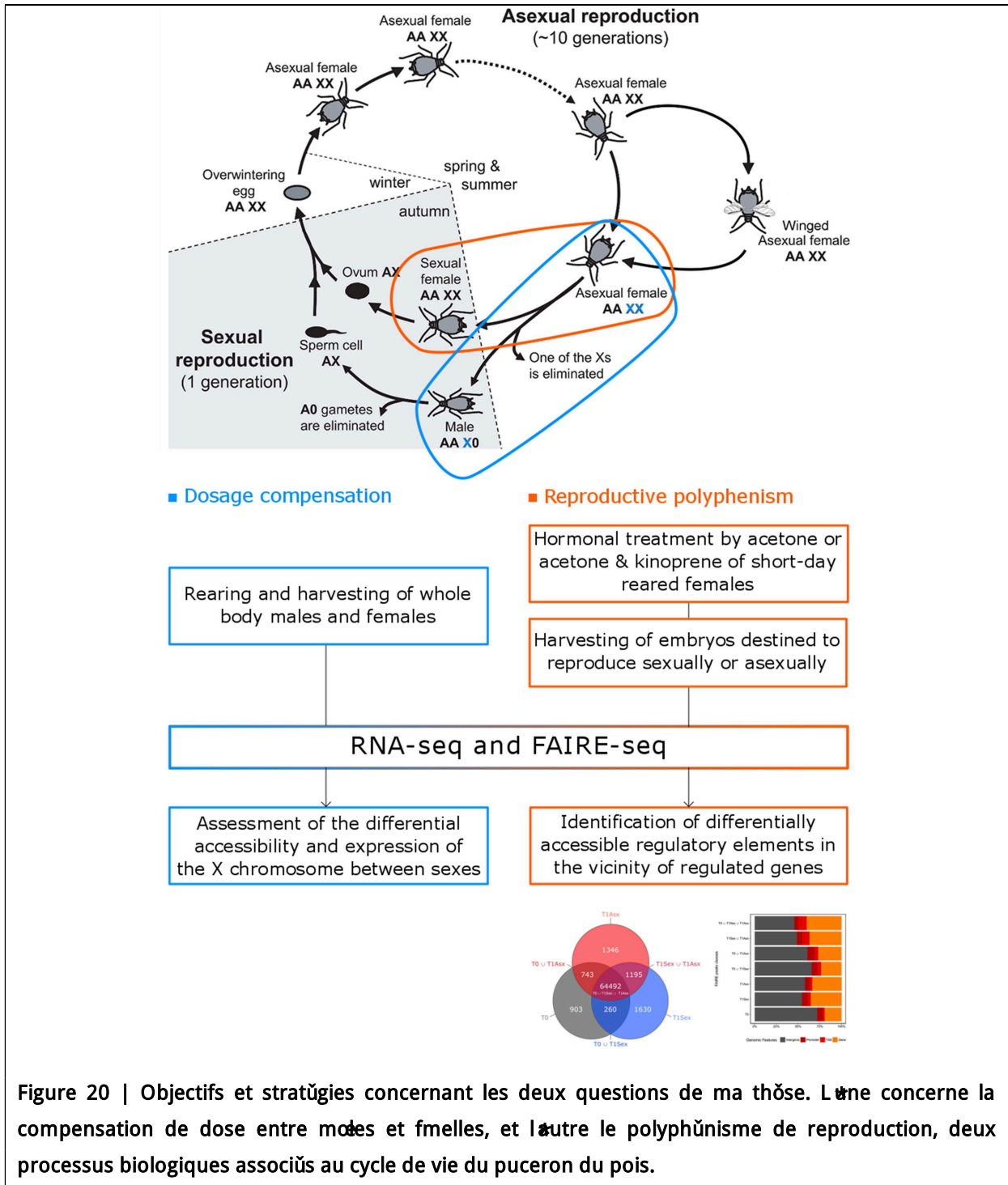


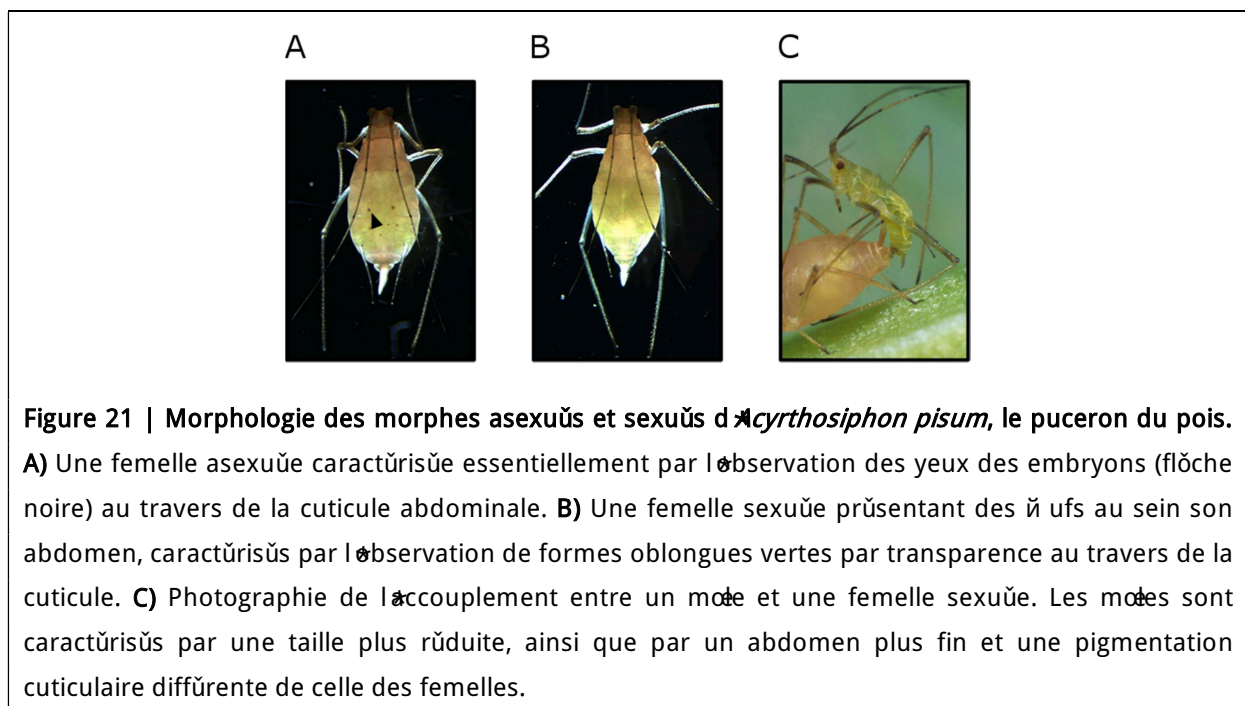
Figure 20 | Objectifs et stratégies concernant les deux questions de ma thèse. L'une concerne la compensation de dose entre mâles et femelles, et l'autre le polyphénisme de reproduction, deux processus biologiques associés au cycle de vie du puceron du pois.

## V. Matériel et Méthodes

### A. Acquisition du matériel biologique

#### 1. Compensation de dose

Des individus entiers mâles et femelles asexués du clone LSR1 ont été utilisés pour l'extraction FA<sup>RE</sup>. Des femelles parthénogénétiques sont initialement élevées sur *Vicia faba* à faible densité (moins de cinq individus par plante) pendant au moins deux générations afin d'empêcher la production de morphes ailés. La parthénogenèse des femelles asexuées est maintenue sous une photopériode de 16h et à une température de 18, C. Après 3 générations à faible densité, 20 femelles asexuées adultes (Figure 21) ont ensuite été directement congelées dans de l'azote liquide pour l'extraction FA<sup>RE</sup>. La production d'individus mâles nécessite un protocole particulier. Des femelles de stade L3 sont ainsi transférées de conditions de photopériode longue en condition de jours courts (12h de jour et 12h de nuit, toujours à 18, C), un tel régime photopériodique permettant l'induction de morphes sexués 2 générations plus tard [Jaquiéry *et al.*, 2013]. En effet, une fois devenues adultes, quelques-uns des descendants de ces femelles asexuées sont conservés au stade L1 (ce sont les individus dits <sup>2</sup> sexupares P, à savoir des individus parthénogénétiques qui produisent des individus sexués dans leur descendance) jusqu'à ce qu'ils atteignent leur tour le stade adulte. Ces derniers vont ensuite produire successivement dans leur descendance des femelles sexuées, des mâles puis quelques femelles asexuées. Les individus mâles sont ensuite spécifiquement conservés une fois devenus adultes (Figure 21) puis congelés dans l'azote liquide par lots de 100 pour la réalisation des extractions FA<sup>RE</sup>. Il est à noter que des extractions FA<sup>RE</sup> ont également été réalisées sur des femelles sexuées, mais les résultats n'ont pas été concluants (discutés en Chapitre 4).



**Figure 21 | Morphologie des morphes asexués et sexués d'*Acyrtosiphon pisum*, le puceron du pois.** **A)** Une femelle asexuée caractérisée essentiellement par l'observation des yeux des embryons (flèche noire) au travers de la cuticule abdominale. **B)** Une femelle sexuée présentant des œufs au sein son abdomen, caractérisés par l'observation de formes oblongues vertes par transparence au travers de la cuticule. **C)** Photographie de l'accouplement entre un mâle et une femelle sexuée. Les mâles sont caractérisés par une taille plus réduite, ainsi que par un abdomen plus fin et une pigmentation cuticulaire différente de celle des femelles.

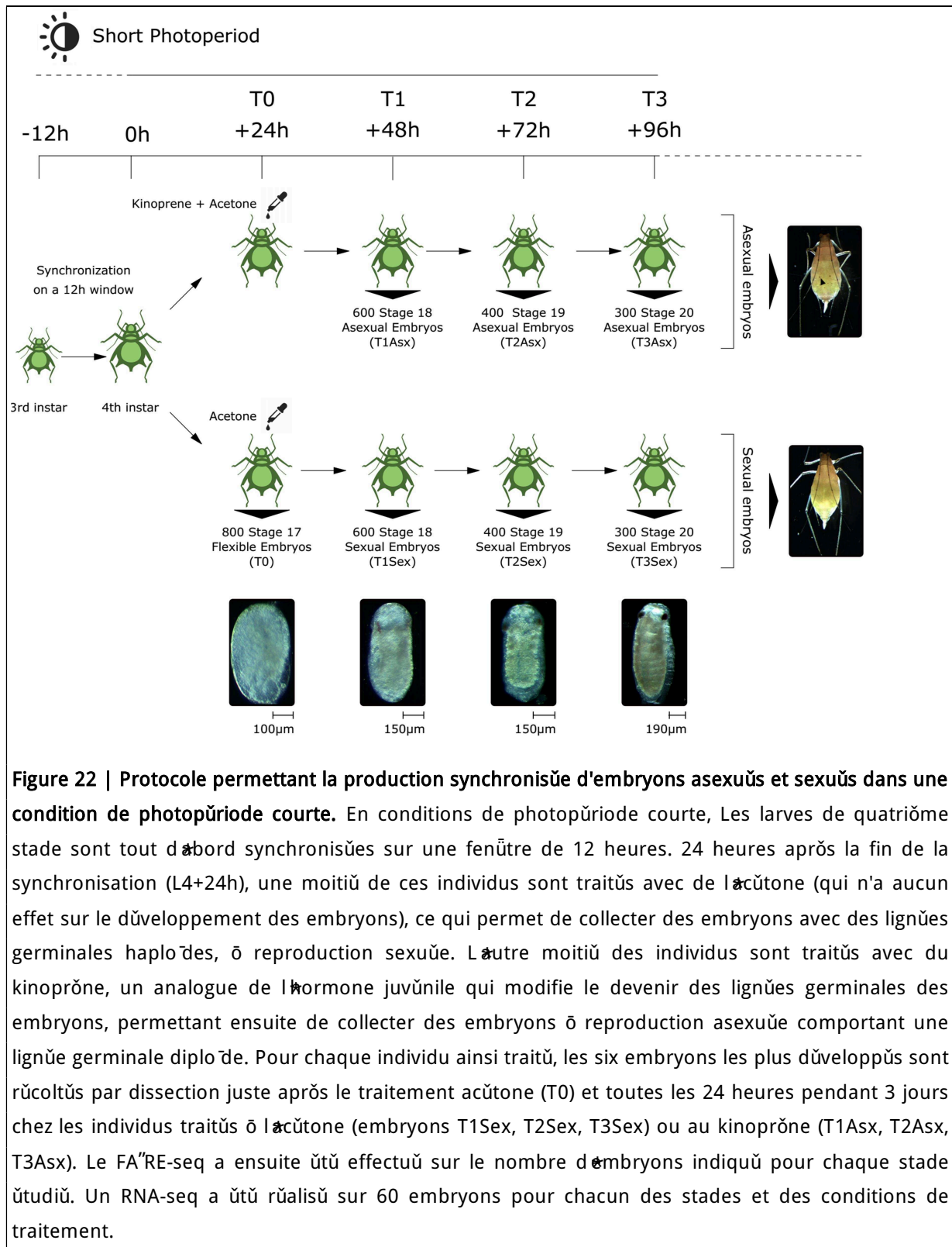
## 2. Polyphénisme de reproduction

Pour étudier le polyphénisme de reproduction chez *A. pisum*, un protocole similaire à celui décrit par [Gallot *et al.*, 2012] a été utilisé, et consiste à récupérer par dissection les embryons provenant de femelles parthénogénétiques placées en conditions de photopériode courte suivant deux modalités de traitement hormonal (avec ou sans kinoprène, un analogue de l'hormone juvénile) [Gallot *et al.*, 2012], permettant *in-fine* d'isoler des embryons synchrones de femelles à destinée reproductive asexuée ou sexuée (Figure 22).

Des femelles parthénogénétiques du clone LSR1 de *A. pisum* sont initialement élevées sur *Vicia fabae* à 18°C sous une photopériode de 16 heures de jour (8h de nuit) et à faible densité (5 individus par plante) pendant au moins deux générations pour empêcher la production de morphes ailés. Environ 200 femelles parthénogénétiques de 3<sup>ème</sup> stade larvaire (5 par plante) sont ensuite transférées en régime de photopériode courte (12h de jour, 12h de nuit toujours à 18°C), qui est la condition nécessaire à l'induction d'individus sexués 2 générations plus tard. Une fois devenus adultes, une partie de leurs descendants est conservée au stade L1. Ces individus dits <sup>2</sup> sexupares produiront ensuite une fois devenus adultes successivement des femelles sexuées (40 1ers produits), une trentaine de mâles et enfin quelques femelles asexuées. Dans le cadre de notre protocole, ces individus L1 (8 par femelle adulte) sont récupérés sur une fenêtre de temps réduite (environ 6h) afin de synchroniser leur développement pour la suite de l'expérimentation. Environ 1600 individus de stade L1 sont ainsi conservés (5 par plante). Une fois que ces individus atteignent le stade L3, une deuxième phase de synchronisation (fenêtre de 12h) est réalisée au moment de la mue du stade L3 vers le stade L4, afin de ne conserver que des femelles sexupares de même âge et dont les embryons les plus avancés sont exactement au stade 16 de leur développement. 1000 individus synchronisés de stade L4 sont ainsi utilisés (5 par plante) pour la phase de traitement hormonal.

En effet, 24 heures après cette synchronisation, ces femelles sexupares synchrones sont séparées en deux groupes de 500 individus, chacun subissant un traitement différent. Le premier groupe est traité par application ectopique sur leur abdomen de 50  $\mu$ g d'acétone. Celui-ci n'ayant aucun effet particulier, les embryons de ces femelles sexupares ainsi traités continuent de se différencier en embryons à destinée reproductive sexuée comportant donc des lignées germinales haploïdes se différenciant en mâles non-fertilisables à l'âge adulte. Le second groupe de femelles sexupares est quant à lui traité par application ectopique sur leur abdomen de 400ng de kinoprène dilués dans 50  $\mu$ g d'acétone. Le kinoprène est un analogue de l'hormone juvénile ayant pour effet de modifier les programmes génétiques exprimés dans les cellules germinales des embryons de ces femelles sexupares. Ces embryons vont alors se destiner à une reproduction asexuée et produire des lignées germinales diploïdes se différenciant en embryons à l'âge adulte. Il est important de rappeler que toutes ces manipulations s'effectuent sous un régime de photopériode courte (12h de jour, 12h de nuit). Ce protocole basé sur [Corbitt et Hardie, 1985], puis modifié par [Gallot *et al.*, 2012] permet ainsi la production synchronisée d'embryons à devenir sexuels ou à devenir asexuels sous une condition de photopériode unique. Suite à ce traitement, des embryons sont collectés par dissection toutes les 24h comme décrit en figure 22. Il a été préalablement établi au laboratoire [Gallot *et al.*, 2012] qu'au sein

des abdomens d'individus sexupares, le dernier stade embryonnaire capable de répondre au traitement au kinoprene et de inverser le devenir de sa lignée germinale est le stade 17, ce stade de développement précis étant caractérisé de 2 flexible. Les 3 stades suivants (18, 19 et 20, qui est le dernier stade embryonnaire avant la naissance d'individus au premier stade larvaire) sont développementalement déterminés (non-flexibles donc) pour être sexués (et acétone) ou asexués (kinoprene). Les embryons les plus développés chez les femelles sexupares L4 + 24h sont ainsi au stade 17 (T0) tandis que les embryons les plus développés des individus L4 + 48 h sont au stade 18 (T1Sex ou T1Asex suivant le type de traitement). 24h plus tard, ces femelles sexupares font ensuite leur dernière mue deviennent adulte et leurs embryons les plus développés sont au stade 19 (T2Sex ou T2Asex). A nouveau 24h plus tard, les embryons les plus développés de ces femelles adultes sont au stade 20 (T3Sex ou T3Asex). Ainsi, par dissection des sexupares L4 ou adultes toutes les 24h après les traitements initiaux, il est possible de récupérer des embryons de stades déterminés, synchronisés, sexués ou asexués. Pendant les 4 jours de l'expérimentation, les femelles sexupares (L4 ou adulte) sont disséquées et les 6 embryons les plus développés ont été collectés. Les quantités minimales de matériel biologique nécessaires à la réalisation des extractions FA'RE ont été déterminées au préalable, aboutissant à la collecte de 800 embryons de stade 17, 600 embryons de stade 18, 400 embryons du stade 19 et 300 embryons du stade 20 et ce pour les 2 conditions de traitement. L'effectif est ainsi variable selon les différents stades afin de récupérer un poids frais de matériel équivalent pour chaque échantillon. Les extractions d'ARN nécessitent moins de matériel biologique puisque 60 embryons par stade sont suffisants. Les embryons ainsi collectés ont été congelés dans de l'azote liquide et conservés à -80°C. Cette procédure a été répétée au total 9 fois au cours de ma thèse afin de générer autant de répétitions biologiques de la cinétique entière et de ne séquencer *in fine* que les 3 réplicats biologiques les mieux corrélés sur la base de leur ratio FA'RE (détaillé ci-dessous). Pour chaque expérience et traitement, quelques femelles sexupares adultes traitées avec de l'acétone ou de l'acétone et du kinoprene ont été gardées vivantes afin de vérifier par analyse morphologique de leur descendance que le traitement a bien induit les phénotypes attendus (femelles sexuées ou femelles asexuées respectivement).



## B. Formaldehyde Assisted Isolation of Regulatory Elements suivi d'un séquençage haut-débit (FAIRE-seq)

### 1. Protocole utilisé

Le FAIRE est un protocole permettant l'extraction des régions chromatiniques accessibles du génome. Dans le cadre du protocole utilisé au cours de ma thèse, les tissus congelés sont tout d'abord pulvérisés dans de l'azote liquide, puis subissent une lyse cellulaire et un broyage fin des tissus. Ces étapes sont suivies d'une fixation au formaldéhyde créant des liaisons covalentes entre ADN et protéines par la formation de ponts méthylène. Cette fixation est ensuite stoppée par ajout de glycine, et la chromatine ainsi fixée est fragmentée par sonication. Une partie de cette chromatine soniquée est utilisée pour l'extraction d'ADN FAIRE. Dans ce cas, la chromatine soniquée est directement extraite à l'aide de phénol-chloroforme, permettant de séparer l'ADN libre dépourvu de nucléosomes (phase aqueuse) de l'ADN fermé, constitué d'ADN et de protéines (phase organique). L'ADN libre ainsi obtenu correspond à l'ADN FAIRE. L'autre partie de l'ADN soniqué est utilisée pour l'extraction de l'ADN Contrôle, celui-ci correspondant à l'ensemble de l'ADN génomique (fermé et ouvert). Cet ADN Contrôle est obtenu par une annulation de la fixation au formaldéhyde, notamment par une dégradation des protéines à l'aide de protéinase K, puis par une extraction au phénol-chloroforme où l'ensemble de l'ADN génomique est extrait (Figure 23).

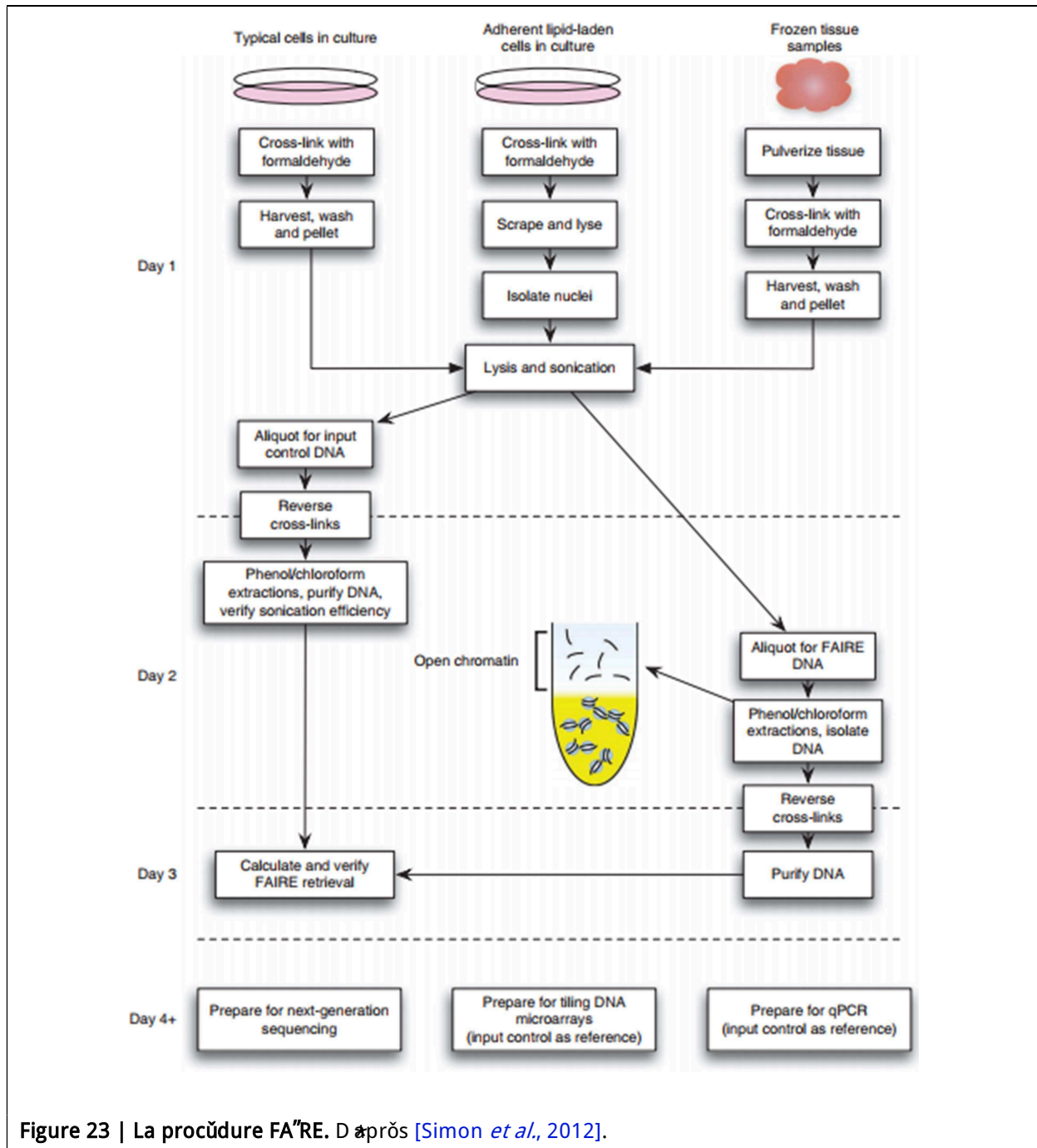


Figure 23 | La procűdure FAIRE. D'aprűs [Simon *et al.*, 2012].

Les extractions FA"RE ont űtű rűalisűes dans les deux volets de cette thűse selon le protocole dűcrit pour des tissus congelűs par [Simon *et al.*, 2012] avec certains paraműtres optimisűs au sein du laboratoire pour notre modűle. Les tissus ont d'abord űtű broyűs ō l'aide d'un Biospec Bio-pulverizer et ont ensuite űtű fixűs par addition de 3% de formaldűhyde (Thermo Scientific Pierce) pendant 8 min. La fixation est arrűtűe par addition de glycine ō 125 mM. Suite au rinűge et ō la suspension des culots, les tissus sont broyűs avec un Tissue Lyser Qiagen en utilisant des billes műtalliques Qiagen, pendant cinq cycles de 5 minutes, chacun űtant interrompu pendant 2 min. Les űtapes de sonication sont effectuűes en utilisant un Bioruptor Plus Diagenode pendant 12 cycles de 30 secondes, chacun űtant interrompu pendant 30 secondes. Les űtapes d'extraction au phűnol chloroforme sont effectuűes en utilisant du phűnol-chloroforme Sigma. Suite ō son extraction, l'ADN de FA"RE est purifiű en utilisant un kit Ch" P DNA Clean & Concentrator (Zymo). Enfin, l'ADN FA"RE et Contrűle sont quantifiűs en utilisant un Quantus de Promega avec le kit Quantifluor dsDNA. La reproductibilitű des rűplicats a űtű űvaluűe avant le sűquenĳage en calculant le rapport FA"RE / Control [Simon *et al.*, 2012] selon la formule suivante :

$$\text{Ratio FA"RE} = \frac{\text{Signal FA"RE}}{\text{Signal Control}}$$

## 2. Műthodes d'analyse bioinformatique de donnűes FA"RE-seq

### a. Construction des banques et dispositif de sűquenĳage

Un protocole de sűquenĳage et de construction des banques similaire a űtű utilisű pour les expűrimentations des deux chapitres de cette prűsente thűse (compensation de dose et polyphűnisme de reproduction), en utilisant un kit de prűparation des banques TruSeq Ch" P "llumina, ainsi qu'une sűlection de la taille des fragments par le kit AMPure XP Beads Beckman Coulter. Concernant la compensation de dose, les banques FA"RE-seq d'individus műles ou femelles asexuűes ont űtű sűquencűes en paired-end avec des lectures de 100 nuclűotides chacune. L'ensemble des banques FA"RE et Contrűles, c'est-ű-dire 3 banques FA"RE műles, 3 banques FA"RE femelles asexuűes, 1 banque Contrűle műles et une banque Contrűle femelles asexuűes, ont űtű sűquencűes sur une műme piste d'une cellule ō flux au sein d'un appareil Hi-seq 2500 (Figure 24). Chaque piste d'une telle cellule ō flux produisant en thűorie environ 360 millions de lectures, ce protocole aurait permis la production d'environ 45 millions de lectures pour chacune des 8 banques. Concernant le FA"RE-seq de la partie polyphűnisme de reproduction, 28 banques ont űtű constituűes comprenant pour chacune des 7 conditions (embryons sexuűs et asexuűs de stade 18, 19, 20 et embryons indifűrenciűs de stade 17) 3 rűplicats biologiques FA"RE par stade embryonnaire et 1 pool űquimolaire d'ADN Contrűle par stade embryonnaire correspondant aux 3 rűplicats biologiques FA"RE. En plus de l'expűrimentation FA"RE-seq, un sűquenĳage d'ARNm a űtű effectuű, avec une banque gűnűrűe pour chacune des 7 conditions. L'ensemble des banques FA"RE-seq ont űtű rűparties sur 4 pistes de sűquenĳage, tandis que les banques RNA-seq ont űtű rűparties sur une piste de sűquenĳage afin de rűaliser un sűquenĳage paired-end 2x100 nuclűotides ō l'aide d'un appareil Hi-seq 2500 (Figure 24). Chaque piste produisant

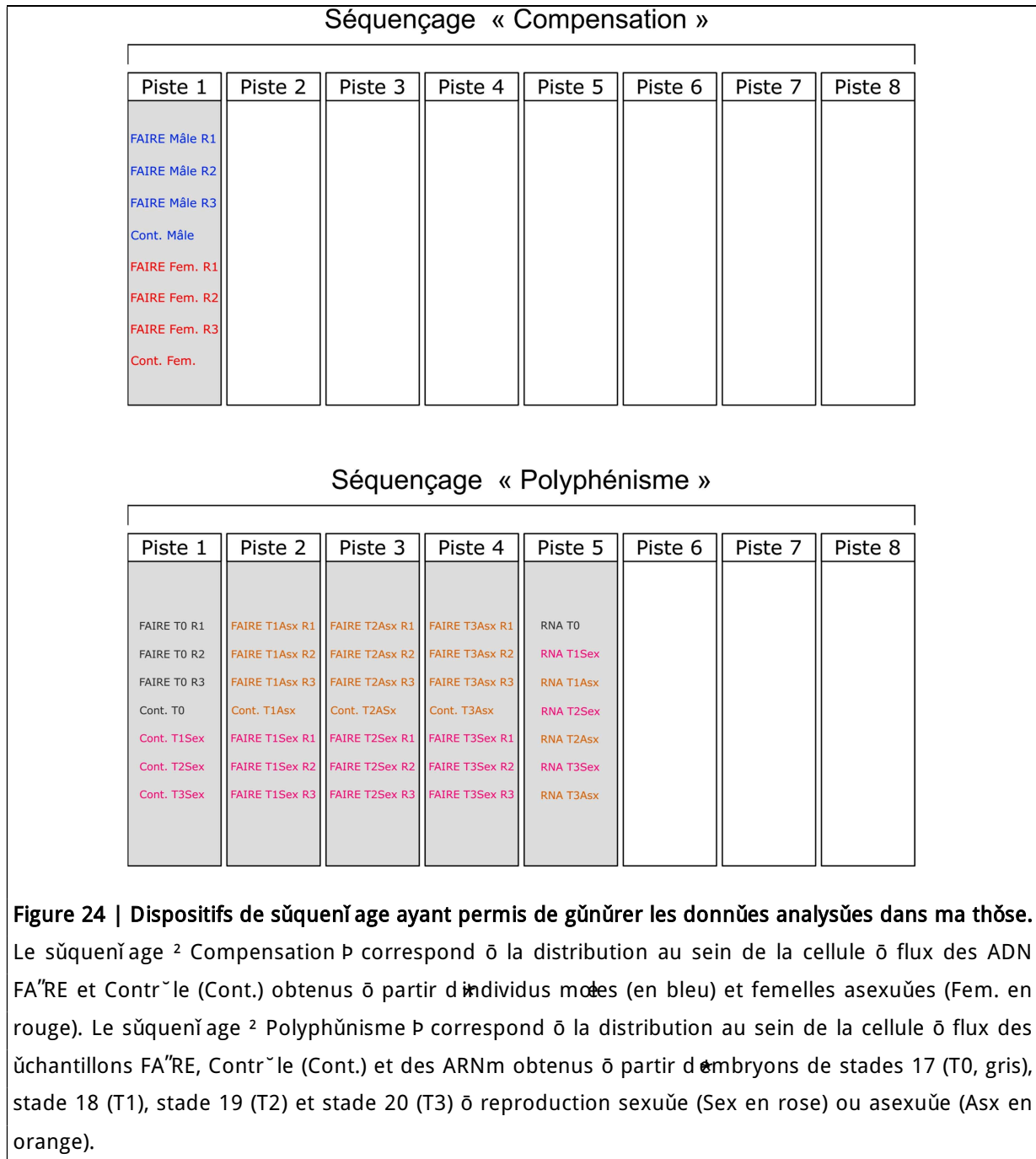
en moyenne 360 millions de lectures, ce protocole de sūquenĳage prūsent une profondeur thūorique de 51.4 millions de reads par ũchantillon.

#### b. Contrūle qualitū des banques et du sūquenĳage

Suite au sūquenĳage haut-dūbit des ADN FA"RE et Contrūle, un contrūle qualitū du sūquenĳage est effectuū avec FastQC (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Le mapping des banques FA"RE sur le gūnome du puceron du pois est rūalisū avec bowtie2 [Langmead et Salzberg, 2012], un programme d'alignement de fragments d'ADN courts single-end ou paired-end, donnant ũgalement le nombre de fragments alignūs sur le gūnome. SAMtools [Li *et al.*, 2009] est ensuite utilisū afin de garder uniquement les fragments d'ADN ne s'alignant qu'une fois dans le gūnome, ainsi que les fragments d'ADN avec une qualitū d'alignement (MAPQ, un score dūterminū par bowtie2) supūrieur ō 30. Toujours grāce ō SAMtools, les fragments alignūs sont ensuite triūs par ordre de position gūnomique (tri alphabūtique sur le nom des scaffolds et numūrique sur la position de dūpart du fragment alignū), puis stockūs dans un format BAM compressū. Les fragments ainsi archivūs sont ensuite utilisūs par MACS2 [Zhang *et al.*, 2008b] qui permettra de rūaliser l'appel des rūgions comportant une accumulation de fragments d'ADN FA"RE comparativement aux fragments d'ADN Contrūle.

#### c. Appel des pics

Cette ũtape d'appel de pics est rūalisūe sur diffūrents types de rūplicas : i) les rūplicas biologiques originaux, ii) les pseudo-rūplicas originaux, iii) les rūplicas originaux <sup>2</sup> poolūs P, iv) les pseudo-rūplicas originaux <sup>2</sup> poolūs P. Les pseudo-rūplicas sont obtenus par un ũchantillonnage alūatoire des fragments d'ADN FA"RE sūparūs en deux parts ũgales. Les rūplicats <sup>2</sup> poolūs P sont obtenus par une addition de l'ensemble des fragments d'ADN FA"RE contenus dans chacun des rūplicas originaux. Une fois les appels de pics appliquūs sur l'ensemble de ces rūplicas, "DR [Kundaje, 2012] est utilisū afin de classer les pics du plus rūpūtable au moins rūpūtable selon deux scores, dont l'un est calculū en comparant l'appel de pics des rūplicas originaux deux ō deux, et l'autre ũtant calculū en comparant l'appel de pics des pseudo-rūplicas <sup>2</sup> poolūs P et des pseudo-rūplicas originaux. Cette ũtape permet ensuite de ne conserver qu'un certain nombre de pics rūpūtables au sein des rūplicas originaux <sup>2</sup> poolūs P, selon ces comparaisons. Les rūgions accessibles au sein de chacune des conditions ũtudiūes sont ainsi dūterminūes.



d. Calcul du signal FA<sup>2</sup>RE

Un signal FA<sup>2</sup>RE normalisé par la profondeur de séquençage et par l'ADN contrôle est enfin calculé à l'aide de deepTools [Ramírez *et al.*, 2014, 2016] sur des fenêtres de 10 nucléotides, en utilisant les données de réplicas originaux <sup>2</sup> poolés. La normalisation par la profondeur de séquençage est réalisée entre les ADN FA<sup>2</sup>RE et Contrôle en ramenant l'échantillon ayant la plus grande profondeur de séquençage au niveau de l'échantillon ayant la plus faible profondeur de séquençage des deux. Ceci est effectué par le calcul d'un coefficient multiplicatif inférieur à 1 ; ce coefficient est défini par le ratio du nombre total de fragments d'ADN FA<sup>2</sup>RE et du nombre total de fragments d'ADN Contrôle. Les ADN FA<sup>2</sup>RE étant poolés, leur profondeur de séquençage est systématiquement plus importante que celle des ADN Contrôle, étant alors de manière systématique le numérateur dans le calcul de ratio précédemment décrit, permettant ainsi la comparaison du signal FA<sup>2</sup>RE entre plusieurs conditions. La normalisation par l'ADN Contrôle est effectuée en calculant tout d'abord la valeur moyenne ou maximale de la couverture FA<sup>2</sup>RE et Contrôle au sein de chacune des fenêtres de 10 nucléotides disséminées le long du génome, puis par la division de la couverture FA<sup>2</sup>RE par la couverture Contrôle dans chacune de ces fenêtres. Cette étape est réalisée après la normalisation par la profondeur de séquençage. Ce signal FA<sup>2</sup>RE permet ensuite de définir avec précision et de manière non-biaisée le niveau d'accessibilité de la chromatine le long du génome, puis autour de points d'intérêt, tels que le sommet des pics précédemment identifiés ou au niveau des sites d'initiation de la transcription. Ce signal FA<sup>2</sup>RE permet par exemple de corréler ouverture et expression en représentant le signal FA<sup>2</sup>RE sus-décrit au niveau des TSS des gènes, classés du plus exprimé au moins exprimé. Il s'avère généralement que plus un gène est exprimé, plus son TSS sera accessible, reflétant l'accessibilité accrue du site d'initiation de la transcription d'un gène exprimé pour la fixation de la RNAP III au sein de l'ensemble des cellules dont l'ADN FA<sup>2</sup>RE a été extraite. D'autres analyses plus spécifiques peuvent être réalisées et seront décrites au sein des articles correspondants, et notamment l'obtention, le mapping et l'analyse différentielle des données RNA-seq.

# V. Chapitre " : Compensation de dose et accessibilit  sexe-sp cifique du chromosome X chez *Acyrtosiphon pisum*

## A. "ntroduction

### 1. Contexte

Les esp ces h t rogam tiques males pr sentent deux chromosomes XX chez les femelles, et un seul chromosome X chez les males, accompagn  ou non d'un chromosome Y. Ce d s quilibre du nombre d'all les associ s au chromosome X vis- -vis des autosomes entre males et femelles peut  tre   l'origine de divers dysfonctionnements cellulaires, comme le cancer du sein ou des ovaires chez l'humain [Kristiansen *et al.*, 2002]. Certains organismes h t rogam tiques ont ainsi acquis au cours de l' volution des m canismes de compensation de dose, permettant d' quilibrer la transcription du ou des chromosomes X entre males et femelles vis- -vis des autosomes, en assurant un ratio d'expression global X / Autosomes de 1 chez les deux sexes. Des m canismes  pig n tiques ont  t  montr s comme impliqu s dans la compensation de dose chez les mammif res, *C. elegans* et *D. melanogaster*. Chez ces organismes, les m canismes  pig n tiques sous-jacents impliquent tous une accessibilit  diff rentielle de la chromatine du chromosome X entre males et femelles, par des enrichissements ou des d pl tions de diverses marques histones.

Le puceron du pois est un h mipt re h t rogam tique male de type XX / XO. Des  tudes transcriptomiques pr liminaires ont montr  une expression moyenne de ces g nes globalement sup rieure au niveau de l'unique chromosome X des males comparativement aux deux chromosomes X des femelles, sexu es ou asexu es. Cependant, les ratios d'expression X / Autosomes n'ont pas  t  calcul s, et aucune  tude n'a pour le moment analys  le niveau d'accessibilit  de la chromatine ou les enrichissements de marques histones du chromosome X entre sexes et comparativement aux autosomes chez les pucerons.

## 2. Rappel des objectifs et stratégie

Au cours de ma thèse, je me suis ainsi intéressé à estimer la présence ou non d'un mécanisme putatif de compensation de dose chez *A. pisum*, d'une part d'un point de vue transcriptionnel, et d'autre part d'un point de vue de l'accessibilité de la chromatine du chromosome X. J'ai donc dans un premier temps caractérisé finement le profil d'expression du chromosome X et des autosomes d'individus mâles, femelles sexuées et femelles asexuées grâce à des données RNA-seq d'individus entiers, notamment en modulant les critères de filtration d'expression des gènes ainsi qu'en comparant entre sexes et entre chromosomes les profils d'expression des gènes sexe-biaisés et non-biaisés. Dans un second temps, par réalisation pour la première fois d'un FA<sup>3</sup>RE-seq chez *A. pisum* à partir d'individus entiers mâles et femelles asexués, les niveaux d'accessibilité globaux de la chromatine du chromosome X et des autosomes au sein de chaque sexe et entre sexes ont été comparés. Les résultats de ces expériences font l'objet d'un article publié dans *Epigenetics and Chromatin* en Juin 2017, présenté et discuté ci-dessous.

## B. Premier article

RESEARCH

Open Access



# Dosage compensation and sex-specific epigenetic landscape of the X chromosome in the pea aphid

Gautier Richard<sup>1\*</sup> , Fabrice Legeai<sup>2,3</sup>, Nathalie Prunier-Leterme<sup>1</sup>, Anthony Bretaudeau<sup>2,4</sup>, Denis Tagu<sup>1</sup>, Julie Jaquiéry<sup>5\*†</sup> and Gaël Le Trionnaire<sup>1\*†</sup>

## Abstract

**Background:** Heterogametic species display a differential number of sex chromosomes resulting in imbalanced transcription levels for these chromosomes between males and females. To correct this disequilibrium, dosage compensation mechanisms involving gene expression and chromatin accessibility regulations have emerged throughout evolution. In insects, these mechanisms have been extensively characterized only in *Drosophila* but not in insects of agronomical importance. Aphids are indeed major pests of a wide range of crops. Their remarkable ability to switch from asexual to sexual reproduction during their life cycle largely explains the economic losses they can cause. As heterogametic insects, male aphids are X0, while females (asexual and sexual) are XX.

**Results:** Here, we analyzed transcriptomic and open chromatin data obtained from whole male and female individuals to evaluate the putative existence of a dosage compensation mechanism involving differential chromatin accessibility of the pea aphid's X chromosome. Transcriptomic analyses first showed X/AA and XX/AA expression ratios for expressed genes close to 1 in males and females, respectively, suggesting dosage compensation in the pea aphid. Analyses of open chromatin data obtained by Formaldehyde-Assisted Isolation of Regulatory Elements (FAIRE-seq) revealed a X chromosome chromatin accessibility globally and significantly higher in males than in females, while autosomes' chromatin accessibility is similar between sexes. Moreover, chromatin environment of X-linked genes displaying similar expression levels in males and females—and thus likely to be compensated—is significantly more accessible in males.

**Conclusions:** Our results suggest the existence of an underlying epigenetic mechanism enhancing the X chromosome chromatin accessibility in males to allow X-linked gene dose correction between sexes in the pea aphid, similar to *Drosophila*. Our study gives new evidence into the comprehension of dosage compensation in link with chromatin biology in insects and newly in a major crop pest, taking benefits from both transcriptomic and open chromatin data.

**Keywords:** X chromosome, Dosage compensation, Transcriptomics, Open chromatin, Non-model organism, Pea aphid, *Acyrtosiphon pisum*, Formaldehyde-Assisted Isolation of Regulatory Elements (FAIRE)

\*Correspondence: gautier.richard@inra.fr; julie.jaquier@univ-rennes1.fr; gael.le-trionnaire@inra.fr

†Julie Jaquiéry and Gaël Le Trionnaire participated equally

<sup>1</sup> EGI, UMR 1349, INRA, Institut de Génétique, Environnement et Protection des Plantes (IGEPP), Domaine de la Motte, BP 35327, Le Rheu, France

<sup>5</sup> CNRS, UMR 6553, EcoBio, University of Rennes 1, 35042 Rennes, France

Full list of author information is available at the end of the article



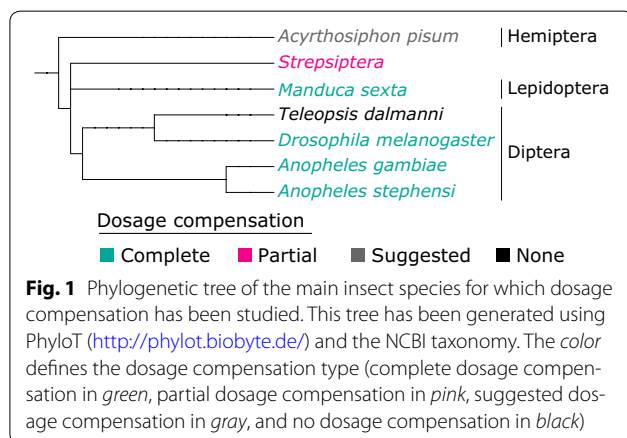
## Background

The sex of individuals relies in many organisms upon morphologically differentiated sex chromosomes. Those chromosomes are referred to as X, Y, Z and W chromosomes depending on which sex is heterogametic. In organisms with male heterogamety, females encompass two X chromosomes, while males possess only one X, accompanied or not by a Y chromosome (XX/XY or XX/XO systems) [1]. In these organisms showing a degenerated or an absence of Y chromosome, there are an imbalanced number of X-linked alleles between males and females that might induce differential transcription levels for those genes between sexes [2]. Such a disequilibrium needs to be corrected, especially for X-linked genes that interact with autosomal genes, since reduced gene dose in the heterogametic sex might have deleterious phenotypic consequences [2–5]. Dosage compensation mechanisms are thought to have evolved to correct such disequilibrium. These mechanisms tend to generate equilibrated X-linked and autosomal transcript levels, often resulting in XX/AA and X/AA expression ratios equal to 1 in both the homogametic (XX) and the heterogametic (XY or XO) sexes, and consequently of XX/X ratios also equal to 1, as described in several model organisms such as Eutherian mammals [2, 6–9], *Caenorhabditis elegans* [9–11] and the insect model *Drosophila melanogaster* [12–14]. Recently, complete dosage compensation in other male heterogametic insect species (Fig. 1)—namely *Anopheles stephensi* [15], *Anopheles gambiae* [16] and *Manduca sexta* [17]—has been demonstrated using transcriptomic data. A partial dosage compensation was found in *Strepsiptera* [18], and no compensation was detectable in *Teleopsis dalmanni* [19].

Various mechanisms have evolved to counteract the deleterious effects of different doses of X-linked genes in males and females. These mechanisms are all based on the modulation of chromatin accessibility of the X

chromosome(s) in one sex in order to ensure that both sex's somatic—and sometimes germ cells [20]—show similar transcription levels for X-linked genes [21]. In mammals, dosage compensation mechanisms take place in the female where one of the two X chromosomes—either the maternal or paternal one—is entirely inactivated, thus balancing the X/A transcription level in males and females. This transcriptional regulation is achieved by the progressive depletion of active marks such as H3K4me1 and H3K9ac [22] and the enrichment of histone macroH2A1 [23] on the female's inactive X chromosome. Moreover, the histone repressive mark H3K27me3 is enriched on this inactive chromosome through the action of the X-inactive specific transcript (Xist), a long noncoding RNA (lncRNA) [24]. In *C. elegans* (a XX/XO system), dosage compensation also takes place in the homogametic sex, where the two X chromosomes display halved transcription levels [25]. Reduced transcription is allowed by an enrichment of the repressive H4K20me1 histone mark and a depletion of the active H4K16ac histone mark [26] on the two hermaphrodites X chromosomes through the action of the dosage compensation complex (DCC) resulting in a global reduction in RNA Pol II recruitment [27]. On the other hand, in *D. melanogaster* (XX/XY system) the transcription of the single males' X chromosome is doubled by an overall increase in the chromatin accessibility by the DCC [28, 29]. This complex is composed of five proteins: male-specific lethal 1, 2 and 3 (MSL1, MSL2 and MSL3), males absent on the first (MOF), maleless (MLE) and two lncRNAs (*roX1* and *roX2*) [29]. The DCC binds to the males X chromosome at several genomic locations where it modifies histones chemistry with the active histone mark H4K16ac [30, 31] that enhances the chromatin accessibility of the X in males compared to females, hence allowing an increased transcription of X-linked genes in males.

Aphids are major crop pests that can cause severe damages on a wide range of crops. Their success as pests is largely explained by their remarkable adaptive potential to their environment, and especially the phenotypic plasticity they display during their annual life cycle where they can switch from clonal to sexual reproduction [32, 33] in response to environmental cues. Viviparous parthenogenetic females reproduce clonally from spring to summer and, in response to photoperiod shortening at fall arrival, can generate sexual females and males that will mate and produce overwintering eggs [34–36]. The pea aphid *Acyrtosiphon pisum*—for which the genome is sequenced and partially assembled [37]—is a male heterogametic species with XO males and XX females. A recent transcriptomic analysis of the different pea aphid sexual morphs by Jaquière et al. [34] revealed a general trend where X-linked genes are on average more transcribed



in males than in females. Nevertheless, the global expression of X-linked and autosomal genes and consequently XX/AA, X/AA and XX/X ratios were not assessed at that time. Here, our study first aimed at addressing this point taking benefits from a much larger set of genes—being generated overtime [38]—assigned as X-linked or autosomal. We report that expression profiles of expressed X-linked and autosomal genes yield X/AA and XX/AA ratios close to 1, suggesting a chromosome-wide regulation of the expression of X-linked genes, and consequently a potential differential X chromosome chromatin accessibility between sexes. To test this hypothesis, we applied the FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) methodology to extract and then sequence the open chromatin from whole male and parthenogenetic female (here after termed as “females”) individuals [39, 40]. This open chromatin mainly corresponds to nucleosome-depleted regions (NDR) likely to be more accessible to transcriptional (RNA polymerase II) or regulatory elements (mainly enhancers or insulators). Statistical analyses revealed that the transcription start sites (TSSs) from expressed genes were significantly more accessible on the X chromosome in males compared to females. More specifically, X-linked genes expressed at the same level in males and females and, thus likely to be compensated, have significantly higher chromatin accessibility in males than in females, notably in the TSS. These results suggest that a potential global regulation of chromatin accessibility might occur on the X chromosome of aphids to compensate for the gene dose in males.

## Results

### Expression of X-linked and autosomal genes in males and females

A new assignation of the genomic scaffolds of the *A. pisum* genome to autosomes and X chromosome has recently been performed [38]. We thus reanalyzed previously published *A. pisum* RNA-seq data from whole-individual males and parthenogenetic females (and to a lesser extent, sexual females, since FAIRE has been performed on asexual females and males; see the next result part concerning the FAIRE for a more in-depth explanation) [34] in regard to this new gene assignation to characterize the expression level of the 19,232 and 13,711 genes located on the X and autosomes, respectively. X/A expression ratios were calculated at different expression levels thresholds [15–17] (Table 1). When all genes are taken into account, XX/AA and X/AA ratios close to 0 are observed since over 80% of the X-linked genes are weakly or not expressed in both males and asexual females. When increasing the minimum expression levels thresholds, males X/AA ratio increases and approaches 1

when genes with a mean male and female RPKM (considered as mean RPKM hereafter) superior to 2 are considered. Above this expression level threshold, X-linked and autosomal genes are evenly expressed in males (Wilcoxon rank sum tests with  $p > 0.05$ ). On the other hand, asexual females XX/AA ratios are in all cases lower than 1 (ranging from 0.45 to 0.69,  $p < 10^{-16}$  in all cases), suggesting that females autosomal genes are in average more expressed than X-linked genes.

Similar results were obtained using an increasing minimum transcripts per million (TPM) filter separately in sexual females, asexual females and males, ranging from a minimum TPM of 1 to 100 (Fig. 2a–d). The X/AA ratio of males quickly goes up to 0.8, when very lowly expressed genes are filtered on both autosomes and the X chromosome (Fig. 2a). Dosage compensation for males, i.e., when their X/AA ratio is between 0.95 and 1.05 for Wilcoxon rank sum tests with  $p > 0.05$ , is attained with a minimum TPM threshold of 34–74 (overshadowed in gray, Fig. 2a, b). Both asexual and sexual females share a very similar expression pattern with a maximum XX/AA expression ratio of 0.8 (Fig. 2a) and with Wilcoxon rank sum tests that are always significant ( $p < 0.05$ ), meaning that the distribution of expression of X-linked and autosomal genes is significantly different, despite the two X chromosomes of females (Fig. 2b).

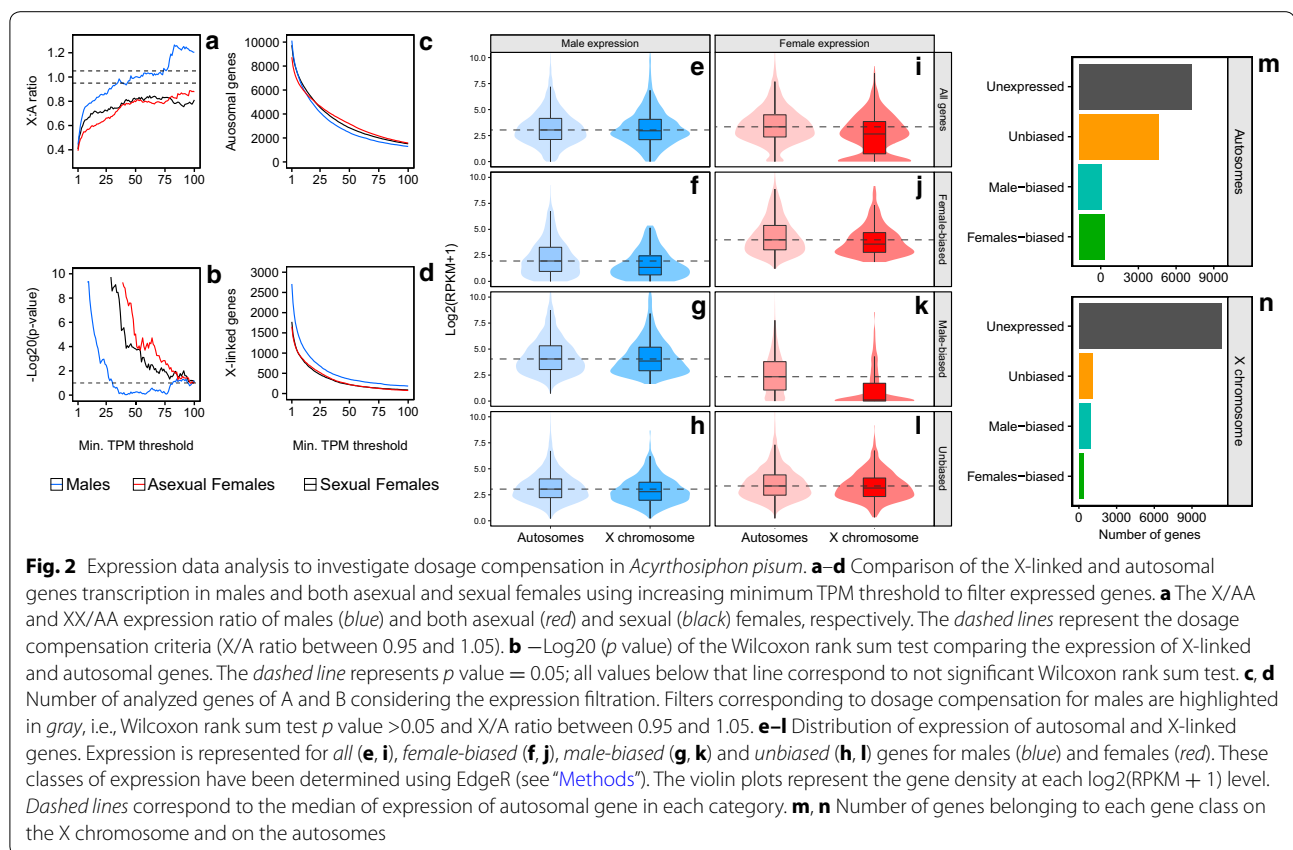
To further investigate the uneven expression of X-linked and autosomal genes in the females, we classified genes (with mean RPKM  $>2$  in males or asexual females) into four non-exclusive classes (*all*, *female-biased*, *male-biased* and *unbiased* genes) based on differential expression analysis (see “Methods,” Fig. 2) between males and asexual females (referred to as females hereafter), since FAIRE has been performed on these two sexes. When considering *all* genes (Fig. 2i), violin plots for females display a large group of lowly expressed genes on the X chromosome that pulls down the females’ median expression of X-linked genes. This lowly expressed gene group on the females X corresponds to *male-biased* genes (Fig. 2k), and when those are not taken into account, such as in the *unbiased* gene class (Fig. 2l), the XX/AA ratio is approaching 1 (ranging from 0.79 to 0.92 for genes with a mean RPKM  $>1-4$ ). This suggests an almost balanced transcription in the females of the X-linked and autosomal *unbiased* genes, although the difference in expression between chromosomes is still significant (Wilcoxon rank sum test,  $p$  ranging from  $7.10^{-7}$  to 0.03 for genes with a mean RPKM  $>1-4$ ). Such group of lowly expressed genes is absent in males (Fig. 2e–h), thus leading to an equal expression of X-linked and autosomal genes for *all* (Fig. 2e) and *unbiased* (Fig. 2h) gene classes. Moreover, the number of each gene class on autosomes and on the X (Fig. 2m, n,

**Table 1** Ratio of expression between X chromosome(s) and autosomes for males and females using minimum RPKM threshold

Genes taken into account	Number of genes retained		Female		Male	
	X	A	XX/AA ratio	<i>p</i> value*	X/AA ratio	<i>p</i> value*
All genes	13,708	19,230	0.00	<2.20E−16	0.00	<2.20E−16
Genes with RPKM >1	1886	9087	0.45	<2.20E−16	0.83	1.16E−04
Genes with RPKM >2	1484	7871	0.58	<2.20E−16	0.94	0.372
Genes with RPKM >3	1260	7010	0.62	<2.20E−16	0.99	0.891
Genes with RPKM >4	1069	6275	0.69	<2.20E−16	0.91	0.119

All genes or different mean RPKM cutoffs have been considered to filter expressed genes

\* *p* values were calculated using Wilcoxon rank sum tests to compare X-linked and autosomal genes expression for a given morph



respectively) reveals that *male-biased* genes are almost three times more represented than *female-biased* genes on the X while even numbers are observed on the autosomes for these two classes. This further explains the fact that the expression of the X chromosome in females is driven down compared to autosomes, and thus the 0.8 XX/AA ratio. When taking the results of both males and females into account, our analyses show that the XX/AA and X/AA ratios are approaching 1, especially for

*unbiased* genes (Fig. 2h, l). Since males carry only one X per cell, and females two, these results suggest a global readjustment of X-linked genes transcription in males or females.

#### Raw FAIRE-seq data analysis

To compare chromatin accessibility in males and females, we used the Formaldehyde-Assisted Isolation of Regulatory Elements (FAIRE) procedure to extract the DNA associated

with NDR from whole individuals and for three pools of males and parthenogenetic females [39, 40], as well as a pool of control DNA for each sex. FAIRE experiments on sexual females were unsuccessful, which might be related to the important quantity of yolk contained in the eggs, interfering with the FAIRE procedure. We thus focused on asexual females and males for the FAIRE experiment. An average FAIRE ratio (measured as in 36) of 1.43% ( $\pm 0.54$ ) and 2.96% ( $\pm 0.63$ ) was obtained for males and asexual females, respectively (referred to as females hereafter). After sequencing the six FAIRE and the two Control DNA samples, FAIRE and Control Reads were mapped onto pea aphid genome and filtered to only conserve uniquely mapped reads.

The coverage of autosomes and X chromosomes was assessed in the control DNA libraries for both males and females (Additional file 1) yielding 14.3 million and 28.0 million 100-bp paired-end uniquely mapped reads for females and males, respectively. The female control library showed an equal coverage for autosomes and X chromosome which is expected since the female is diploid at both X chromosome and autosomes. In males—that are diploid for autosomes and haploid for the X chromosome—autosomes display an expected twofold higher coverage than the haploid X chromosome in males.

The reproducible FAIRE biological replicates for males and females were pooled according to their reproducibility using the MACS2 peakcaller [41] followed by irreproducible discovery rate (IDR) [42, 43] peaks ranking analysis. For males, these algorithms resulted in the discrimination of one of the three male replicates as lowly correlated with the other two. After pooling the relevant FAIRE replicates, three female and two male libraries were then conserved which ended up with 20.5 million and 23.6 million 100-bp paired-end reads mapping to unique positions in the genome for female and male pools, respectively. MACS2 [41] and IDR [42] also allowed us to retain a set of reproducible FAIRE peaks across the biological replicates for each sex: 8143 FAIRE peaks for males and 6369 FAIRE peaks for females were then identified; 39% of these are overlapping between males and females and are thus non-sex-specific (Fig. 3a). In order to assess the level of correlation between the control and retained FAIRE replicates, we concatenated these peaks coordinates into a set of 10,433 FAIRE peaks using BEDTools [44] and performed a Pearson correlation and a hierarchical clustering using deepTools2 [45] (Fig. 3b). Control libraries are alike (Pearson's  $R^2$  of 0.83) and segregate together. Female and male FAIRE libraries are separated in the hierarchical clustering, and Pearson's  $R^2$  values within each sex are high: 0.88 for males and more than 0.95 for female libraries comparisons.

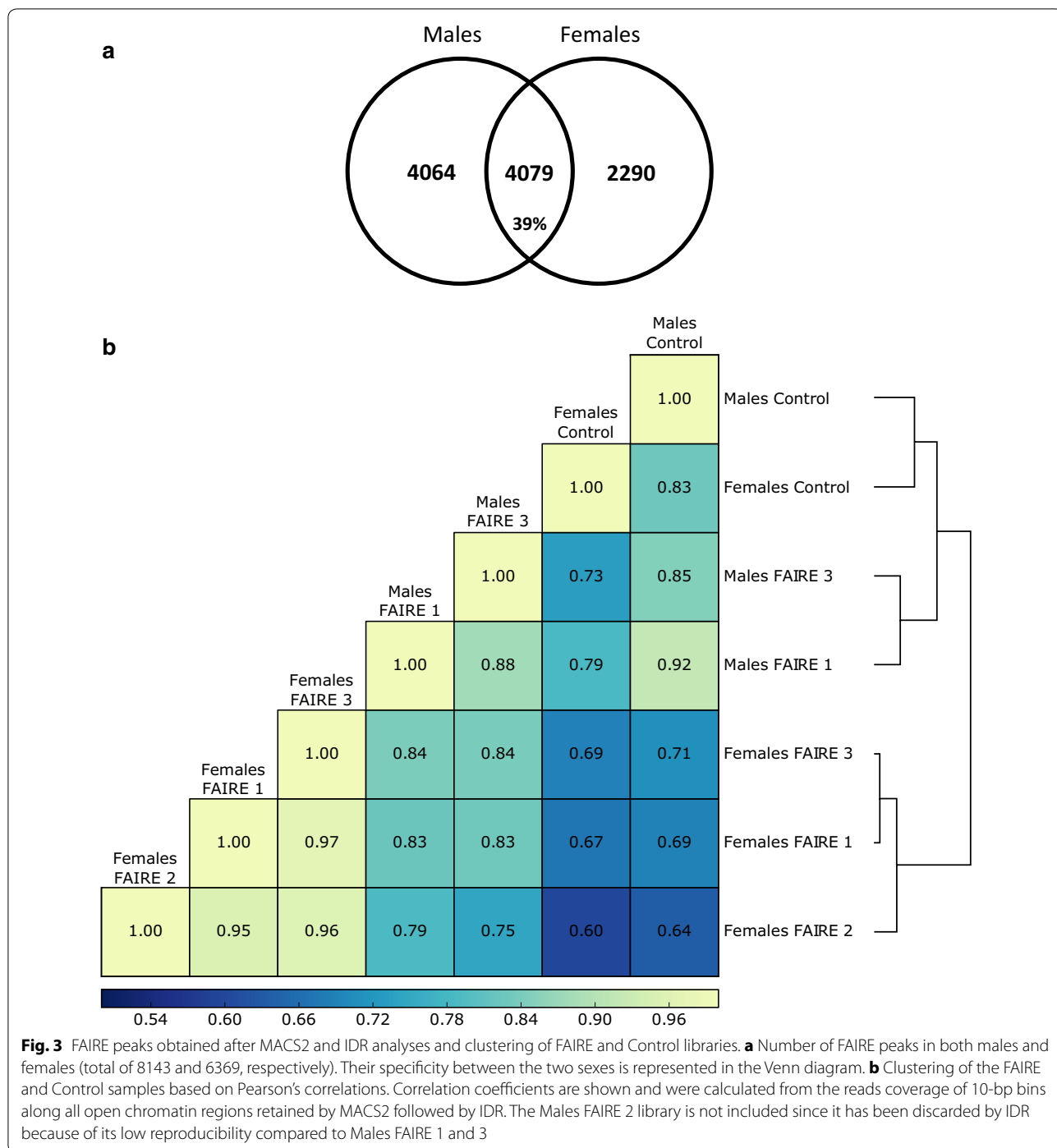
To describe the overall distribution of the identified FAIRE peaks across the genome, we calculated the

abundance of FAIRE peaks overlapping with different genomic features (promoters, TSSs, UTRs, exons, introns and intergenic regions) (Fig. 4). 5'UTRs and TSSs are characterized by high FAIRE peaks densities (Fig. 4a): Although these genomic features represent, respectively, only 1.5 and 0.8% of the genome (Fig. 4b), they contain a large proportion of FAIRE peaks compared with the other genomic features that represent larger parts of the genome. The FAIRE peaks density is thus much lower in all other genomic features (Fig. 4a). Examples of genomic regions displaying some of these open chromatin peaks specific of each sex or in common between sexes near sex-specific or housekeeping genes are observable in Additional file 2.

### Open chromatin signal in males and females

To assess the pea aphid's chromatin opening profile, we calculated the FAIRE signal at the genome scale by dividing the normalized FAIRE coverage by the normalized Control coverage on 10-bp windows (hereafter named bins, as described by deepTools [45]). In Fig. 5a, the FAIRE signal is represented for each gene of the pea aphid genome and RNA-seq data have been used to rank genes according to their level of expression in males or in females. The strongest FAIRE signals are observed in regions upstream the TSS. The more a gene is expressed, the more its FAIRE signal upstream the TSS is high, indicating a positive correlation between gene expression and chromatin accessibility. The heatmaps representing the FAIRE signal along X-linked genes suggest a stronger FAIRE enrichment in males than in females, in comparison with the autosomes that share similar profiles between sexes (Fig. 5a).

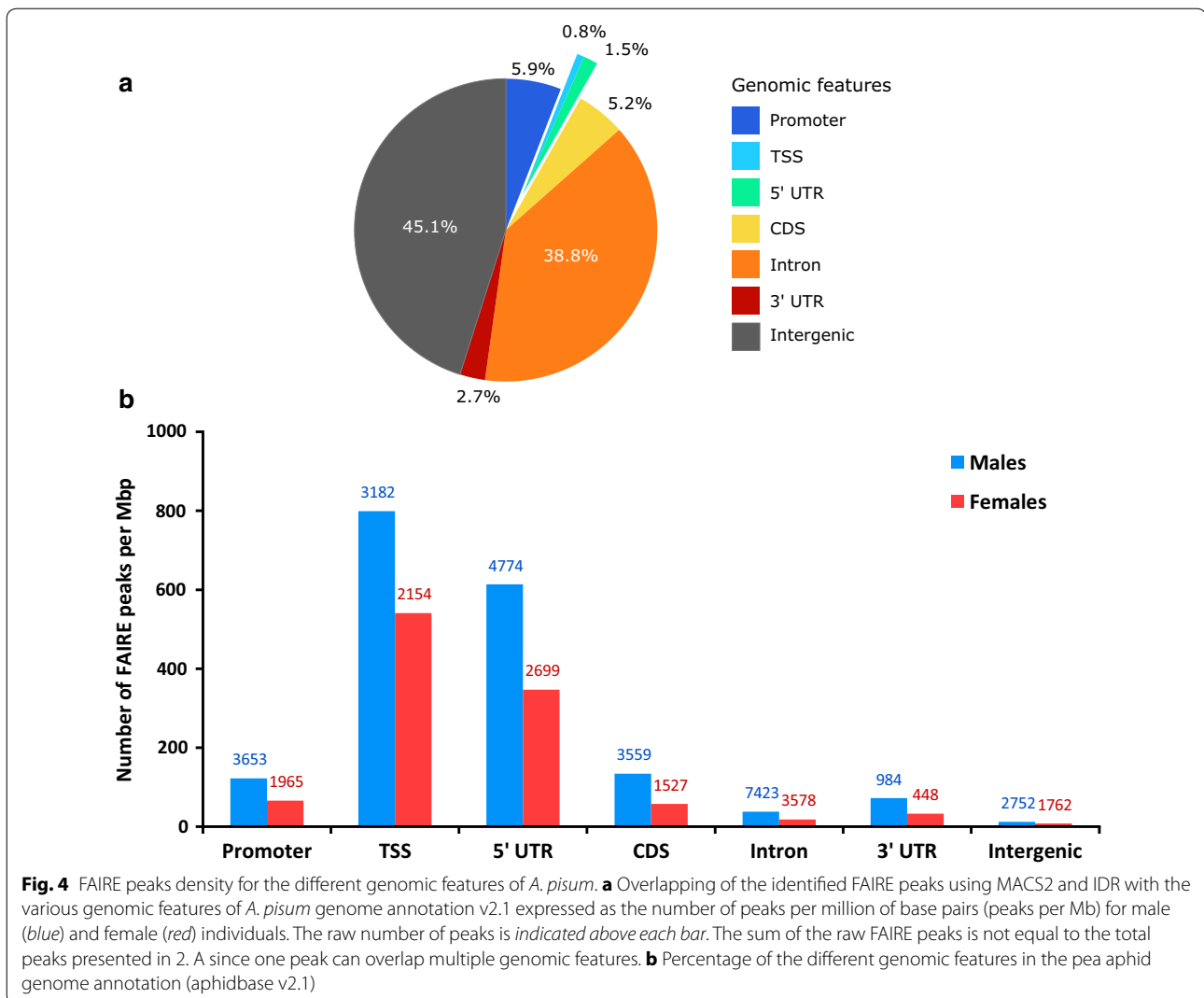
We then calculated the mean genic FAIRE signal and 99% confidence interval (CI) for both males and females on autosomal and X-linked genes (Fig. 5b). According to the previous observation, autosomal genes share similar accessibility levels in males and females, with a slightly higher accessibility near the TSS in males than in females based on 99% CI ( $2.94 \pm 0.13$  in males and  $2.59 \pm 0.09$  in females on the most accessible bin). Contrastingly, the mean FAIRE signal along and around X-linked genes is almost two times higher in males than in females, especially near the TSS ( $2.40 \pm 0.18$  for males and  $1.45 \pm 0.07$  for females on the most accessible bin). In order to verify whether the FAIRE signal normalization on Control DNA could introduce a bias for the X chromosome, we calculated the mean FAIRE coverage normalized by sequencing depth around all autosomal and X-linked genes in males and females (Additional file 3). It appears that genes on the autosomes and on the X chromosome are equally accessible between males and females, despite the single X in males compared to the two X



chromosomes of females. This profile does not invalidate the results observed using input normalized FAIRE signal, since the single X chromosome of males is still more accessible than each female X chromosome, and is thus differentially accessible.

We then investigated the FAIRE signal within intergenic regions in males and females. The assessment of the

chromatin accessibility in these intergenic regions was performed around the summit of the intergenic FAIRE peaks retained by IDR. FAIRE peaks that do not overlap any annotation feature were thus extracted and defined as intergenic FAIRE peaks. The mean FAIRE signal and 99% CI were calculated 1000 bp around the summits of those intergenic FAIRE peaks for males and females (Fig. 6). On

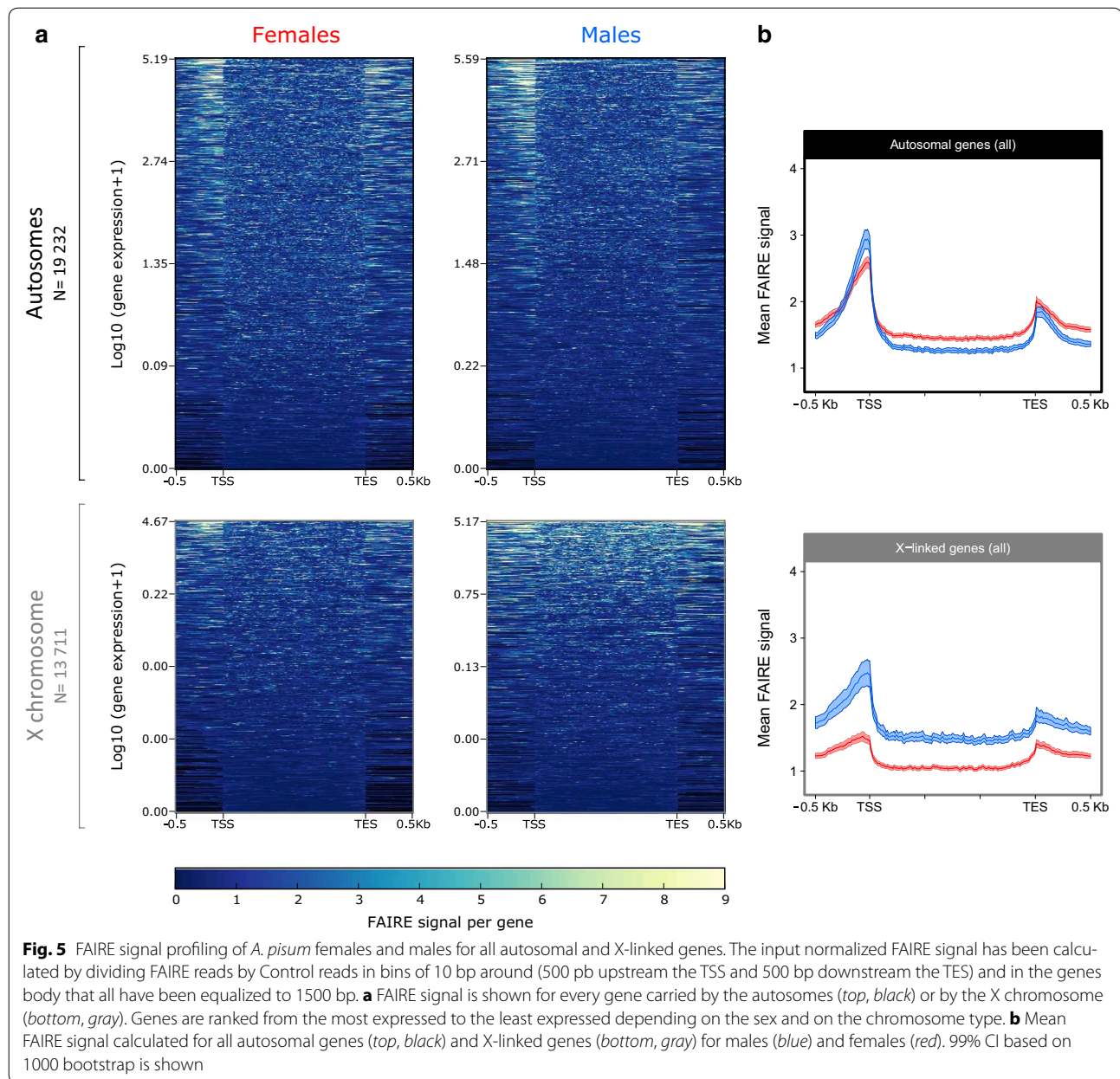


autosomal intergenic FAIRE peaks, no significant differences of chromatin accessibility were identified between males and females. Contrastingly, for the X chromosome, intergenic FAIRE peaks in males are significantly more accessible than in females, especially near the peaks summits (mean FAIRE signal of  $14.90 \pm 2.46$  for males and  $9.35 \pm 0.82$  for females on the most accessible bin). The results for intergenic regions are thus similar to those observed for genic regions and demonstrate a comparable accessibility of the autosomes in the two sexes and an increased accessibility of the single X in males compared to the two X chromosomes in females.

#### Link between genes expression and chromatin accessibility

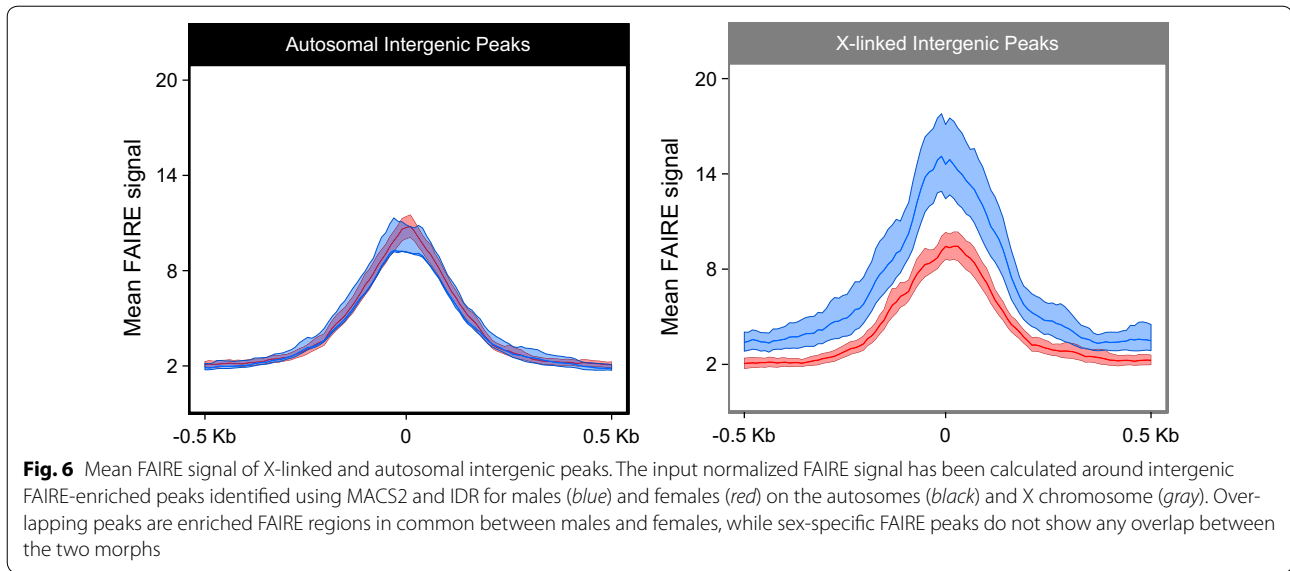
To explore the link between gene expression and chromatin accessibility, and especially around the TSSs, we grouped genes into four classes depending on their

expression patterns in males and females and studied their FAIRE signal profile considering both sexes and chromosome type (Fig. 7): *unexpressed* (Fig. 7a, e) (46.8% on the autosomes, 82.7% on the X), *male-biased* (Fig. 7b, f) (9.6% on the autosomes, 6.8% on the X), *female-biased* (Fig. 7c, g) (10.6% on the autosomes, 2.6% on the X) and *unbiased* genes (Fig. 7d, h) (33.0% on the autosomes, 7.9% on the X). These classes correspond to the ones defined during the RNA-seq experiment (Fig. 2). On the autosomes, *unexpressed*, *male-biased* and *female-biased* gene classes share a similar chromatin accessibility profile in males and females (Fig. 7a–c), and only *unbiased* genes (Fig. 7d) are slightly more accessible in the males TSS compared to females ( $4.41 \pm 0.32$  in males and  $3.68 \pm 0.22$  in females on the most accessible bin). Contrastingly, on the X chromosome (Fig. 7e–h), two main profiles can be distinguished. The first group is comprised



of *unexpressed* and *male-biased* genes (Fig. 7e, f) that display comparable—yet significantly different—FAIRE signal in males and in females:  $1.56 \pm 0.09$  for males and  $1.17 \pm 0.05$  for females regarding *unexpressed* genes and  $3.27 \pm 0.72$  for males and  $2.05 \pm 0.33$  for females concerning the *male-biased* genes. The second group is composed of *female-biased* and *unbiased* genes (Fig. 7g, h) that display a stronger chromatin accessibility difference between sexes. Mean FAIRE signals of  $10.72 \pm 3.06$  in males and  $4.41 \pm 0.84$  in females are observed for *female-biased genes* and those of  $8.80 \pm 1.55$  in males and  $3.95 \pm 0.52$  in females are observed for *unbiased*

*genes*. These results suggest that *unbiased* and *female-biased* genes participate in majority to the observed global enhanced accessibility of the X chromosome in males compared to females (Fig. 5), even if they represent only 10% of X-linked genes. Interestingly, we can also observe that in females for *unbiased* (Additional file 4H) and *female-biased* (Additional file 4G) gene classes, chromatin accessibility is similar between the X chromosome and autosomes. On the contrary in males and for the same gene classes, chromatin accessibility is higher on the X chromosome than on autosomes (Additional file 4C, D). This suggests that autosomal genes in males



and females and X chromosome in females share similar chromatin patterns and that the differential X chromosome chromatin pattern between sexes is explained by an enhancement of chromatin accessibility in males rather than a reduced accessibility in females.

## Discussion

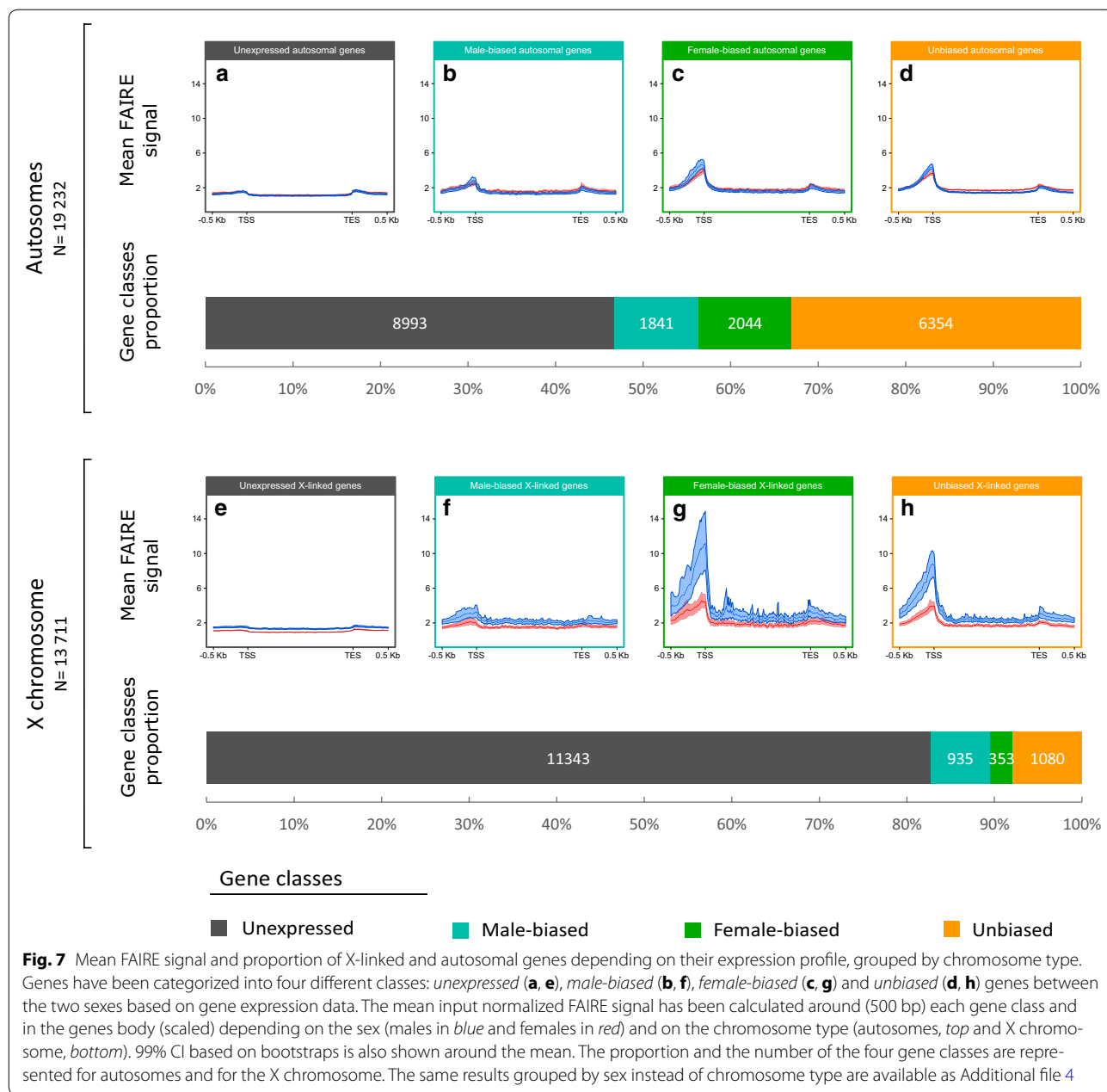
In this study, we first reanalyzed whole-body RNA-seq data of pea aphid males and females and evidenced a potential dosage compensation mechanism. We then generated an overview of the open chromatin structure (nucleosome-depleted regions) of whole-body pea aphid males and females using the Formaldehyde-Assisted Isolation of Regulatory Elements. This first genome-wide epigenetic study in aphids demonstrated an enhanced chromatin accessibility of the males' single X chromosome compared to the two X of the females.

### Global transcriptomic profiles suggest potential dosage compensation in pea aphid males

Using a total of 3712 genes assigned to autosomes or the X chromosome, Jaquiéry et al. [34] outlined that in the pea aphid, autosomal genes display similar transcription levels between males and females while X-linked genes are slightly more expressed in males than in females. Based on the same data, Pal and Vicoso [46] outlined different results as they found that males and females share similar expression levels on both X and autosomes. These contradictory results can, however, be explained by a misassignment of genes to chromosomes due to a widespread scaffold misassembly in the pea aphid genome [38]. Here, we used a new assignment of scaffolds to chromosomes [38] to analyze X-linked and autosomal gene expression patterns in females and males.

In males, we showed X/AA expression ratios almost equal to 1 when taking into account consistently expressed genes (mean RPKM >2). Similar expression patterns in organisms showing complete dosage compensation have been demonstrated such as in *Manduca sexta* heads [17] where the Z/AA ratio is approaching 1 when lowly expressed genes are not considered. These results suggest a potential dosage compensation mechanism in the male pea aphids.

Contrastingly, the females XX/AA expression ratios are inferior to 1 which is not a common feature within male heterogametic species with complete dosage compensation. Nevertheless, such an unusual ratio can also be observed in the flour beetle (*Tribolium castaneum*). In this species, females display a XX/AA ratio superior to 1 which is explained by the fact that females show a global X-linked genes overexpression (supported by an enrichment of female-biased genes on the X), while males show a X/AA ratio equal to 1, ending up with a XX/X ratio different from 1 [47]. In the pea aphid, the X chromosome is predicted to be masculinized [34], and we indeed observed that X-linked *male-biased* genes are almost three times more represented than X-linked *female-biased* genes and that most of the X-linked *male-biased* genes are male specific (i.e., expressed in males but not at all in females, as shown in Fig. 2k). As a result, XX/AA female ratio and female-to-male XX/X ratio are inferior to 1. It has to be noted that mRNA extractions were performed on whole individuals and thus contain mRNAs from both somatic and germ cells. The fact that most X-linked *male-biased* genes are almost male specific suggests that these genes could be testis specific or more generally involved in male phenotypic traits. GO terms enrichment analyses on *female-biased*, *male-biased* and



*unbiased* gene classes on the X and autosomes revealed that only *male-biased* and *female-biased* genes on the X display reproductive-related functions. *Male-biased* X-linked genes display an enrichment of biological processes related to male-specific reproduction such as cilium movement, microtubule-based movement, sperm competition and multi-organism reproductive process (Additional file 5). Because the pea aphid's spermatozooids are flagellated [48], such functions are likely to correspond to testis-specific genes, thus reinforcing the hypothesis of their absence of expression in females

because of their male germ cells specificity. The unusual females XX/AA ratio (<1) for a male heterogametic species can thus be explained and is compatible with complete dosage compensation in the male pea aphids.

In mammals and in *Drosophila* [20, 49], dosage compensation mechanism activity on the X chromosome can be different between germ cells and somatic cells. Mammals do not show any dosage compensation in the germ cells where the X chromosome inactivation in females is withdrawn in developing primordial germ cells, allowing X-linked genes expression in female gametes [20].

Contrastingly, *D. melanogaster* germ cells compensate for gene dose in germlines with, however, distinct mechanisms than the well-known somatic cells dosage compensation [49]. These studies suggest that dosage compensation does not always occur in germ cells. Since the RNA-seq data analyzed in our study comprise mRNAs from germ cells and somatic cells, we cannot conclude whether the potential dosage compensation in the male pea aphids occurs only in somatic cells or in both somatic and germ cells.

#### **Dosage compensation in the pea aphid might be achieved by an enhanced chromatin accessibility of the X chromosome in males**

Dosage compensation is often controlled by epigenetic mechanisms mediating the X or Z chromosome chromatin accessibility. Since our RNA-seq analyses suggest potential complete dosage compensation in the male pea aphids, we aimed at assessing the global chromatin accessibility of the autosomes and the X chromosome(s) in both males and females using FAIRE-seq. The FAIRE methodology allows the identification of open chromatin (nucleosome-depleted) regions. It has been performed for the first time by Giresi et al. [39] and since then has been used in various organisms and tissues [39, 40, 50]. Methods such as assay for transposase-accessible chromatin (ATAC) [51–53] are becoming alternatives to the FAIRE overtime; however, the latter remains a reliable way to assess open chromatin regions at the genome scale when coupled to high-throughput sequencing.

To validate the FAIRE-seq approach used here, we checked the presence of a high FAIRE peaks frequency in TSS and 5'UTR regions, as expected since TSSs are known as the most accessible regions in the genome in order to allow the binding of the RNA polymerase II and thus the transcription of the gene [40, 54–56]. We also observed a positive correlation between genes expression and chromatin opening upstream their TSS, which corresponds to a typical profile of a successful FAIRE experiment [40].

With our FAIRE-seq approach, we then showed a similar mean chromatin accessibility of autosomes in males and females over all genes (Fig. 5b; Additional file 3) and also when genes are classified according to their expression patterns in males and females (Fig. 7; Additional file 4). Similar chromatin accessibility in males and females is also observed within open intergenic regions. These results suggest that the autosomes of the pea aphid are equally accessible in males and females.

Contrastingly, the FAIRE signal analysis over *all* genes (Fig. 5b) revealed that X-linked genes are more accessible in males than in females, or evenly accessible despite the single X of males using sequencing depth normalization

alone (Additional file 3). *Female-biased* and *unbiased* X-linked genes (Fig. 7g, h) display the most important difference of chromatin accessibility between males and females and thus contribute at most to the global male-enhanced chromatin accessibility of the X chromosome, suggesting that the dosage compensation in males might be partial. Since *unbiased* genes have (by construction) similar expression pattern between males and females, X-linked *unbiased* genes correspond to potentially compensated genes. A higher chromatin accessibility of these genes in males than in females thus supports the hypothesis of potential epigenetic mechanism underlying dosage compensation in males. Surprisingly, X-linked *female-biased* genes chromatin accessibility is comparatively higher in males than in females (Fig. 7g). This cannot be used to reject this hypothesis since in *D. melanogaster* males, some female-biased genes also show an enrichment of the active histone mark H4K16ac mediated by the DCC, thus resulting in an enhanced chromatin accessibility for these genes [57]. Additionally, X-linked intergenic FAIRE peaks, which represent only a very small proportion of intergenic regions, are significantly more accessible in males than in females. Altogether, these data suggest a global regulation of the X chromosome chromatin accessibility in order to compensate for gene dose in males.

Interestingly, the chromatin accessibility of X-linked *male-biased* genes is only slightly enhanced in males, despite their high level of expression in that morph. As suggested by our GO analyses, these genes could correspond in majority to testis-specific genes (hence including germ cells-specific genes). The relatively low chromatin accessibility of these genes could be explained on the one hand by the fact that FAIRE is a population assay and that male-specific cells make up only a small percentage of the cell population in the analyzed whole-body male individuals. The FAIRE signal could thus be low for the particular class of *male-biased* genes, hence resulting in a low difference in FAIRE signal between males and females for that gene class. On the other hand, these results could suggest that the epigenetic mechanism underlying the male-enhanced X chromosome chromatin accessibility does not take place in male germ cells, like in the female mammals for example [20]. *Female-biased* and *unbiased* genes might then be part of a given X chromosome territory compensated for the lack of a second X chromosome in the males' genome, while the *male-biased* and *unexpressed* genes might be part of another X chromosome territory where the chromatin accessibility is not submitted to a given mechanism underlying dosage compensation. Such hypothesis would resemble what has been identified in the *Drosophila* where only 75% of the X chromosome territory is regulated by the DCC [29].

The high chromatin accessibility of the male single X chromosome might be the consequence of an underlying epigenetic mechanism taking place either in male (enhanced chromatin accessibility) or in female individuals (reduced chromatin accessibility). When taking *unbiased* and *female-biased* gene classes into account (Fig. 7c, d, g, h), we can observe that autosomal genes in males and females as well as X-linked genes in females are equally accessible, whereas X-linked genes are more accessible only in males. The differential X chromosome chromatin accessibility observed between sexes can thus be explained by an augmentation of such accessibility in males rather than a reduction in females. This suggests the existence in pea aphids males of an epigenetic mechanism promoting the enhancement of the X chromosome chromatin accessibility associated with dosage compensation. These epigenetic patterns resemble the *Drosophila* model where an overall increase in chromatin accessibility is observed on the males X chromosome through a male-specific H4K16ac histone posttranslational modification enrichment mediated by the DCC in the gene bodies (H4K16ac being enriched in the genes' TSS and promoter by the NSL complex) [28, 57]. Interestingly, the pea aphid genome contains protein coding for genes homologous to the five proteins composing the *D. melanogaster* DCC (Additional files 6, 7). The DCC is also constituted by two additional *roX* lncRNAs that were not found in the pea aphid genome, but it is noteworthy that lncRNA sequences are rarely conserved between organisms [58]. Interestingly, homologs identities and similarities between *D. melanogaster* and *A. pisum* DCC proteins are found particularly in functional domains. The most conserved proteic domains between the *Drosophila* and the pea aphid are MOF's MOZ-SAS domain and all MLE's proteic domains, namely dsrm, DEAD, helicase C, HA2 and OB NTP BIND (Additional file 7). MOZ-SAS domain from *Drosophila*'s MOF is involved in the acetylation of lysine 16 of histone 4 (H4K16ac) and is thus responsible for the enhanced chromatin accessibility along the *Drosophila* males X chromosome. The high proteic domain conservation observed between the pea aphid and the *Drosophila* could lead to conserved functions of the MOF and MLE proteins in these two organisms. However, functional analyses are required to validate this hypothesis. Moreover, some of the proteins composing the *Drosophila*'s DCC, including MOF and MLE, has been identified as conserved in mammals without, however, playing a role in dosage compensation [29, 59–63]. The pea aphid could then, independently of a DCC complex, display a different epigenetic mechanism supporting dosage compensation by globally enhancing chromatin accessibility of the single X chromosome in males, an alternative that cannot be ruled out taking into account the rather weak conservation of MSL2 protein sequence between *Drosophila* and the pea

aphid (Additional file 7), particularly in the RING domain which plays a role in the DCC assembly [64].

## Conclusions

This study gives a first insight into *A. pisum* chromatin accessibility patterns in relation to a possible dosage compensation mechanism. The males single X chromosome is globally more accessible than the two female X chromosomes. More importantly, X-linked genes showing similar expression levels between females and males and that could potentially be compensated in the latter are more accessible in males. Further experiments—especially chromatin immunoprecipitation followed by high-throughput sequencing targeting specific histone marks—must be conducted in order to characterize the underlying epigenetic mechanism involved in this overall enhanced chromatin accessibility of the males single X chromosome.

## Methods

### Aphids rearing

*Acyrtosiphon pisum* individuals from the clone LSR1 (the reference clone that was used for genome sequencing [37]) were reared on broad bean *Vicia faba* at low density (less than five individuals per plant) to prevent the production of winged morphs for at least two generations. Parthenogenesis was maintained under a 16-h photoperiod and a temperature of 18 °C. At the third generation, 20 asexual females were then directly frozen into liquid nitrogen for FAIRE extraction. The production of male individuals was initiated by transferring larvae from a 16-h to a 12-h photoperiod at the same temperature of 18 °C [34]. Two generations later, males were produced. A total of 100 adult males were then directly frozen into liquid nitrogen for FAIRE extraction. No RNA extraction was performed since already published RNA-seq data were used [34].

### RNA high-throughput sequencing

We reanalyzed six RNA-seq libraries used by Jaquiéry et al. [34]. Briefly, these six libraries correspond to three male libraries and three parthenogenetic female libraries of clone LSR1. Details regarding aphid rearing, RNA extraction, libraries preparation and sequencing are provided in Jaquiéry et al. [34]. Libraries were mapped on the version 2 of the pea aphid genome assembly using TopHat2 (RRID:SCR\_013035) default parameters [65]. The number of reads covering each CDS of the gene prediction v2.1 was then counted using HTSeq-count (RRID:SCR\_011867) [66] with the following parameters: `-m intersection-strict -s no -t exon`. The numbers of mapped reads per library ranged from 7.3 to 20 million, with an average over libraries of 14.3 million reads. Raw

read counts were normalized in R (RRID:SCR\_001905) [67] with edgeR (RRID:SCR\_012802) [68] package by sequencing depth using the TMM method and by genes length (RPKM and TPM calculation). Genes were filtered using increasing minimum TPM threshold ranging from 1 to 100 using a step of 1 in order to remove the less expressed genes. Within each TPM filtration step, X/A ratio in males and females, Wilcoxon rank sum test  $p$  value and the number of retained X-linked and autosomal genes have been calculated. Differential expression between males and parthenogenetic females for each gene was tested with edgeR, considering the different libraries for each morph as replicates using genewise exact tests for differences in the means between two groups of negative-binomially distributed counts, based on normalized read counts. Genes have then been grouped into four classes: *unexpressed*, *male-biased*, *female-biased* and *unbiased* genes. *Unexpressed* genes have been determined as such if they displayed less than 1 count per million in at least three libraries as in [69]. *Male-biased* and *female-biased* genes have been determined as such using a FDR <0.05. *Unbiased* genes comprise differentially expressed genes with a FDR >0.05 and genes non-differentially expressed.

#### Formaldehyde-Assisted Isolation of Regulation Elements, sequencing and bioinformatics analyses

The FAIRE extraction of three frozen pools of male and parthenogenetic female individuals was performed following the protocol of frozen tissues proposed by Simon et al. [40] with the following parameters. Tissues were first ground using a Biospec Bio-pulverizer and then fixed by the addition of 3% of Thermo Scientific Pierce formaldehyde during 8 min. The fixation was then stopped by the addition of glycine at 125 mM. Subsequently to the pellet rinsing and resuspension, the tissues were ground with a Tissue Lyser, using Qiagen metallic beads, during five cycles of 5 min, each being interrupted during 2 min. The sonication steps were performed using a Bioruptor Plus during 12 cycles of 30 s, each being interrupted during 30 s. Phenol chloroform extraction steps were performed using Sigma phenol chloroform. Subsequently to its extraction, the FAIRE DNA was purified the using ZYMO CHIP DNA Clean & Concentrator. Finally, FAIRE and Control DNA were quantified using a Quantus with the Quantifluor dsDNA kit. The reproducibility of the replicates was assessed before sequencing by calculating the FAIRE/Control ratio described in Simon et al. [40].

FAIRE and Control DNA were sequenced using the Illumina HiSeq 2000 instrument and CHIP TrueSeq kit. Three FAIRE samples for each morph were sequenced, while only a pool of the three samples from the same morph individuals was sequenced for control DNA. The eight different samples (six FAIRE and two Control samples)

were 100-pb paired-end sequenced on a single lane. The raw sequenced data for these eight samples are available online at the NCBI under the BioProject accession number PRJNA348188. The reads were mapped using bowtie2 with default parameters [70, 71]. Only uniquely mapped reads with a mapping quality over or equal to 30 were kept using SAMtools (RRID:SCR\_002105) [72], following the IDR recommendations [42, 43]. For peak calling, we analyzed separately the reads mapping on the X and on autosomes to avoid bias related to the difference of coverage of the X chromosome (haploid in males) and autosomes (diploid in males and females). In addition, Jaquiéry et al. [38] showed that over 50% of scaffolds greater than 150 kb are chimeras of X and autosomes, which could induce artificial peaks at breakpoints. Following IDR recommendations, MACS2 (RRID:SCR\_013291) [41] was used to perform the peak calling with the following parameters: `-f BAM -nomodel -extsize 200 -slocal 1000 -llocal 10,000 -p 0.05`. The `-gsize` parameter (which corresponds to the total size of the genome analyzed) has been adapted depending on the size of chromosome types (166,018,072 bp for the X and 340,086,956 bp for autosomes). IDR analyses were then performed as suggested in [43] using a threshold of 0.04 for original replicates, of 0.04 for self-consistency replicates and of 0.01 for pooled pseudoreplicates. Once the final peak set for males and females was generated, HOMER suite (RRID:SCR\_010881) [73] was used to calculate Venn Diagrams of males and females FAIRE peaks specificity. Segtools (RRID:SCR\_004394) [74] with the aphidbase (RRID:SCR\_001765) version 2.1 annotation of the pea aphid genome were used to assign the FAIRE peaks to various genomic features such as promoters (1300 bp upstream TSS regions), TSS (defined as 200 bp upstream the first base of 5'UTRs), 5'UTR, coding regions, introns, 3'UTR and intergenic regions. DeepTools2 [45] was used to calculate and represent the FAIRE signal as the ratio between pooled FAIRE reads over Control reads on bins of 10 pb on whole genome. FAIRE signal data were first retrieved along every gene and represented in heatmaps with deepTools2 [45]. The mean FAIRE signal was calculated using R and represented using the ggplot2 package [67, 75]. In order to estimate a confidence interval for each mean FAIRE signal calculated, bootstrap was done using a custom R script that performed 1000 random resampling of the genes by taking into account the number of genes in each resampled set. 99% confidence interval, which depends on gene number and FAIRE signal values, was then calculated and represented using the ggplot2 [75] package in the R software [67].

#### Drosophila BLAST and GO terms enrichment of molecular processes in the pea aphid

Since *A. pisum* gene functions are poorly characterized, the most efficient way to perform GO terms enrichment

in this organism was to compare unique *D. melanogaster* homologs of a given list of pea aphid genes of interest against all the unique *D. melanogaster* homologs found in the pea aphid annotation (aphidbase v2.1). In order to find *D. melanogaster* homologs in the pea aphid, BLASTp (RRID:SCR\_001010) [76] analyses were performed using BLAST+ [77] with default parameters. GO terms enrichment of biological processes has then been done using GOzilla (RRID:SCR\_006848) with default parameters [78, 79] using unique *D. melanogaster*'s homologs coming from a gene list of interest as "target set" and all the unique *D. melanogaster* homologs found in the pea aphid annotation v2.1 as "background set".

## Additional files

**Additional file 1.** FAIRE and Control libraries coverages. Number of reads and coverage of X chromosome (X) and autosomes (A) for male and female control libraries.

**Additional file 2.** Genome browser view of remarkable autosomal and X-linked regions displaying sex-specific and non-specific FAIRE-seq and RNA-seq signal. **A, D:** female-specific regions around the genes ACYPI003071 (uncharacterized protein) and ACYPI001644 (cuticular protein 44). **B, E:** male-specific regions around the genes ACYPI080359 (uncharacterized protein) and ACYPI081672 (uncharacterized protein). **C, F:** regions in common between males and females for the genes ACYPI000061 (ATP synthase subunit beta) and ACYPI006656 (molybdate-anion transporter). The RNA-seq and FAIRE-seq signals have been made equal between males and females for each region.

**Additional file 3.** Mean FAIRE coverage calculated for all autosomal genes (left, black) and X-linked genes (right, gray) for males (blue) and females (red). 99% CI based on 1000 bootstrap is shown around the mean. The FAIRE coverage has been normalized by read-depth in order to allow the comparison between males and females data.

**Additional file 4.** Mean FAIRE signal and proportion of X-linked and autosomal genes depending on their expression profile, grouped by sex. Genes have been categorized in four different classes: *unexpressed* (**A, E**), *male-biased* (**B, F**), *female-biased* (**C, G**) and *unbiased* (**D, H**) genes between the two sexes based on gene expression data. The mean FAIRE signal has been calculated around each gene class (500 bp) and in their gene body (scaled), depending on the chromosome type (autosomes in dark color, X chromosome in light color) and depending on the sex (males, top in blue and females, bottom in red). 99% CI based on bootstraps is also shown around the mean.

**Additional file 5.** GO enrichment of biological processes for autosomal and X-linked, male- and female-biased genes.

**Additional file 6.** Protein sequence comparison of the *Drosophila*'s DCC in the pea aphid. Clustal multiple sequence alignment by MUSCLE (3.8) of *A. pisum*'s homologs of *D. melanogaster* DCC. Proteic domains of the *Drosophila* are represented in full color boxes above the alignments. *A. pisum* gene names found by BLASTp are indicated in parenthesis.

**Additional file 7.** Protein conservation of the *Drosophila*'s DCC in the pea aphid. Proteic domains percentage of identity between *A. pisum* homologs of the five proteins composing the DCC of *D. melanogaster*.

## Abbreviations

H3K4me1: monomethylated lysine 4 of histone 3; H3K9ac: acetylated lysine 9 of histone 3; macroH2A1: histone variant macroH2A1; H3K27me3: trimethylated lysine 27 of histone 3; lncRNA: long noncoding RNA; H4K20me1: mono-methylated lysine 20 of histone 4; H4K16ac: acetylated lysine 16 of histone

4; DCC: dosage compensation complex; RNA Pol II: RNA polymerase II; MSL1, MSL2 and MSL3: male-specific lethal 1, 2 and 3; MOF: males absent on the first; MLE: maleless; *roX1*, *roX2*: RNA on X chromosome 1 or 2; FAIRE: Formaldehyde-Assisted Isolation of Regulatory Elements; NDR: nucleosome-depleted regions; TSS: transcription start site; RPKM: reads per kilobase of transcript per million mapped reads; TPM: transcripts per million; IDR: irreproducible discovery rate; MACS2: model-based analysis for ChIP-seq; 5'/3'UTR: five or three prime untranslated region; CI: confidence interval; GO: gene ontology; ATAC: assay for transposase-accessible chromatin; MOZ-SAS: MOZ is a monocytic leukemia Zn\_finger protein and the SAS protein from *Saccharomyces cerevisiae*; FDR: false discovery rate; ChIP: chromatin immunoprecipitation; HOMER: Hypergeometric Optimization of Motif Enrichment; BLAST: basic local alignment search tool; NCBI: National Center for Biotechnology Information; MUSCLE: Multiple Sequence Comparison by Log-Expectation.

## Authors' contributions

GR managed the aphids rearing for the FAIRE-seq part of the current study, performed all the FAIRE-seq and RNA-seq bioinformatics analyses with the help of FL, all the statistical analyses with the help of JJ, supervised and performed all the graphical representations and data visualization and fully drafted the manuscript of the current study. FL and AB helped in providing computing resources, analysis tools and software implementation and FL designed some custom scripts. NPL developed and performed the FAIRE methodology to generate the samples used in the current study. GLT and DT acquired the financial support for the project leading to this study. GLT conceptualized the project leading to this study with the help of JJ. GLT mentored GR and supervised the research activity leading to the current study with the help of JJ. GLT and JJ made the major manuscript corrections, verifications and validation with the participation of DT. All authors read and approved the final manuscript.

## Author details

<sup>1</sup> EGI, UMR 1349, INRA, Institut de Génétique, Environnement et Protection des Plantes (IGEPP), Domaine de la Motte, BP 35327, Le Rheu, France. <sup>2</sup> BIPAA, UMR 1349, INRA, Institut de Génétique, Environnement et Protection des Plantes (IGEPP), Campus Beaulieu, Rennes, France. <sup>3</sup> Genscale, INRIA, IRISA, Campus Beaulieu, Rennes, France. <sup>4</sup> Genouest, INRIA, IRISA, Campus Beaulieu, Rennes, France. <sup>5</sup> CNRS, UMR 6553, EcoBio, University of Rennes 1, 35042 Rennes, France.

## Acknowledgements

We acknowledge Bernard Chaubet and Evelyne Turpeau for the photography of mating *A. pisum* males and females used as the personal cover of the current study.

## Competing interests

The authors declare that they have no competing interests.

## Availability of data and materials

The datasets generated during the current study are available from the NCBI SRA database with the following Accession Numbers: RNA-seq PRJNA209321, SRR071347, SRR073573 [34]; FAIRE-seq PRJNA348188. All the scripts used to analyze these datasets in the current study are available from the corresponding author on reasonable request.

## Funding

Gautier Richard's PhD is funded by a Young Scientist Contract from INRA, The Division for Plant Health and Environment (SPE) Department. This project was funded by The French National Research Agency Project MiRNAdapt (ANR-11-BSV6-01701) and by a French National Institute for Agricultural Research (INRA), The Division for Plant Health and Environment (SPE) Department Project Grant.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 3 February 2017 Accepted: 5 June 2017

Published online: 15 June 2017

## References

- Ellegren H. Sex-chromosome evolution: recent progress and the influence of male and female heterogamety. *Nat Rev Genet.* 2011;12:157–66.
- Mank JE, Hosken DJ, Wedell N. Some inconvenient truths about sex chromosome dosage compensation and the potential role of sexual conflict. *Evolution.* 2011;65:2133–44.
- Ohno S. Sex chromosomes and sex-linked genes. 1st ed. Berlin: Springer; 1967.
- Charlesworth B. Model for evolution of Y chromosomes and dosage compensation. *Proc Natl Acad Sci USA.* 1978;75:5618–22.
- Marin I, Siegal ML, Baker BS. The evolution of dosage-compensation mechanisms. *BioEssays.* 2000;22:1106–14.
- Julien P, Brawand D, Soumillon M, Necsulea A, Liechti A, Schütz F, et al. Mechanisms and evolutionary patterns of mammalian and avian dosage compensation. *PLoS Biol.* 2012. doi:10.1371/journal.pbio.1001328.
- Pessia E, Makino T, Bailly-Bechet M, McLysaght A, Marais GAB. Mammalian X chromosome inactivation evolved as a dosage-compensation mechanism for dosage-sensitive genes on the X chromosome. *Proc Natl Acad Sci.* 2012;109:5346–51.
- Lin F, Xing K, Zhang J, He X. Expression reduction in mammalian X chromosome evolution refutes Ohno's hypothesis of dosage compensation. *Proc Natl Acad Sci.* 2012;109:11752–7.
- Deng X, Hiatt JB, Nguyen DK, Ercan S, Sturgill D, Hillier LW, et al. Evidence for compensatory upregulation of expressed X-linked genes in mammals, *Caenorhabditis elegans* and *Drosophila melanogaster*. *Nat Genet.* 2011;43:1179–85.
- Kramer M, Rao P, Ercan S. Untangling the contributions of sex-specific gene regulation and X-chromosome dosage to sex-biased gene expression in *Caenorhabditis elegans*. *Genetics.* 2016;204:355–69.
- Albritton SE, Kranz A-L, Rao P, Kramer M, Dieterich C, Ercan S. Sex-biased gene expression and evolution of the x chromosome in nematodes. *Genetics.* 2014;197:865–83.
- Straub T, Becker PB. Dosage compensation: the beginning and end of generalization. *Nat Rev Genet.* 2007;8:47–57.
- Gelbart ME, Kuroda MI. *Drosophila* dosage compensation: a complex voyage to the X chromosome. *Development.* 2009;136:1399–410.
- Lucchesi JC. Dosage compensation in *Drosophila*. *Annu Rev Genet.* 1973;7:225–37.
- Jiang X, Biedler JK, Qi Y, Hall AB, Tu Z. Complete dosage compensation in *Anopheles stephensi* and the evolution of sex-biased genes in mosquitoes. *Genome Biol Evol.* 2015;7:1914–24.
- Rose G, Krzywinska E, Kim J, Revuelta L, Ferretti L, Krzywinski J. Dosage compensation in the African malaria mosquito *Anopheles gambiae*. *Genome Biol Evol.* 2016;8:411–25.
- Smith G, Chen Y-R, Blissard GW, Briscoe AD. Complete dosage compensation and sex-biased gene expression in the moth *Manduca sexta*. *Genome Biol Evol.* 2014;6:526–37.
- Mahajan S, Bachtrog D. Partial dosage compensation in Strepsiptera, a sister group of beetles. *Genome Biol Evol.* 2015;7:591–600.
- Wilkinson GS, Johns PM, Metheny JD, Baker RH. Sex-biased gene expression during head development in a sexually dimorphic stalk-eyed fly. *PLOS ONE.* 2013. doi:10.1371/journal.pone.0059826.
- Brockdorff N, Turner BM. Dosage compensation in mammals. *Cold Spring Harb Perspect Biol.* 2015. doi:10.1101/cshperspect.a019406.
- Lucchesi JC, Kelly WG, Panning B. Chromatin remodeling in dosage compensation. *Annu Rev Genet.* 2005;39:615–51.
- Payer B, Lee JT. X chromosome dosage compensation: how mammals keep the balance. *Annu Rev Genet.* 2008;42:733–72.
- Costanzi C, Stein P, Worrad DM, Schultz RM, Pehrson JR. Histone macroH2A1 is concentrated in the inactive X chromosome of female preimplantation mouse embryos. *Development.* 2000;127:2283–9.
- Engreitz JM, Pandya-Jones A, McDonel P, Shishkin A, Sirokman K, Surka C, et al. The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. *Science.* 2013. doi:10.1126/science.1237973.
- Ercan S. Mechanisms of X chromosome dosage compensation. *J Genomics.* 2015;3:1–19.
- Wells MB, Snyder MJ, Custer LM, Csankovszki G. *Caenorhabditis elegans* dosage compensation regulates histone H4 chromatin state on X chromosomes. *Mol Cell Biol.* 2012;32:1710–9.
- Kruesi WS, Core LJ, Waters CT, Lis JT, Meyer BJ. Condensin controls recruitment of RNA polymerase II to achieve nematode X-chromosome dosage compensation. *eLife.* 2013. doi:10.7554/eLife.00808.
- Prestel M, Feller C, Straub T, Mitlöhner H, Becker PB. The activation potential of MOF is constrained for dosage compensation. *Mol Cell.* 2010;38:815–26.
- Conrad T, Akhtar A. Dosage compensation in *Drosophila melanogaster*: epigenetic fine-tuning of chromosome-wide transcription. *Nat Rev Genet.* 2012;13:123–34.
- Bell O, Schwaiger M, Oakeley EJ, Lienert F, Beisel C, Stadler MB, et al. Accessibility of the *Drosophila* genome discriminates PcG repression, H4K16 acetylation and replication timing. *Nat Struct Mol Biol.* 2010;17:894–900.
- Shogren-Knaak M, Ishii H, Sun J-M, Pazin MJ, Davie JR, Peterson CL. Histone H4-K16 acetylation controls chromatin structure and protein interactions. *Science.* 2006;311:844–7.
- Le Trionnaire G, Jaubert-Possamai S, Bonhomme J, Gauthier J-P, Guernec G, Le Cam A, et al. Transcriptomic profiling of the reproductive mode switch in the pea aphid in response to natural autumnal photoperiod. *J Insect Physiol.* 2012;58:1517–24.
- Le Trionnaire G, Wucher V, Tagu D. Genome expression control during the photoperiodic response of aphids. *Physiol Entomol.* 2013;38:117–25.
- Jaquière J, Rispe C, Roze D, Legeai F, Le Trionnaire G, Stoeckel S, et al. Masculinization of the X chromosome in the pea aphid. *PLoS Genet.* 2013. doi:10.1371/journal.pgen.1003690.
- Jaquière J, Stoeckel S, Rispe C, Mieuze L, Legeai F, Simon J-C. Accelerated evolution of sex chromosomes in aphids, an XO system. *Mol Biol Evol.* 2012;29:837–47.
- Wilson ACC, Sunnucks P, Hales DF. SHORT PAPER random loss of X chromosome at male determination in an aphid, *Sitobion near fragariae*, detected using an X-linked polymorphic microsatellite marker. *Genet Res.* 1997;69:233–6.
- International Aphid Genomics Consortium. Genome sequence of the pea aphid *Acyrtosiphon pisum*. *PLoS Biol.* 2010;8:e1000313. doi:10.1371/journal.pbio.1000313pmid:20186266.
- Jaquière J, Peccoud J, Ouisse T, Legeai F, Prunier-Leterme N, Gouin A, et al. Disentangling the causes for faster-X evolution in aphids. *bioRxiv.* 2017;125310.
- Giresi PG, Kim J, McDaniell RM, Iyer VR, Lieb JD. FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin. *Genome Res.* 2007;17:877–85.
- Simon JM, Giresi PG, Davis IJ, Lieb JD. Using Formaldehyde-Assisted Isolation of Regulatory Elements (FAIRE) to isolate active regulatory DNA. *Nat Protoc.* 2012;7:256–67.
- Zhang Y, Liu T, Meyer CA, Eeckhoutte J, Johnson DS, Bernstein BE, et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* 2008;9:R137. doi:10.1186/gb-2008-9-9-r137.
- Li Q, Brown JB, Huang H, Bickel PJ. Measuring reproducibility of high-throughput experiments. *Ann Appl Stat.* 2011;5:1752–79.
- Kundaje A. ENCODE: TF ChIP-seq peak calling using the Irreproducibility Discovery Rate (IDR) framework (2012). <https://sites.google.com/site/anshulkundaje/projects/idr>. Accessed 5 Oct 2016.
- Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics.* 2010;26:841–2.
- Ramírez F, Ryan DP, Grünig B, Bhardwaj V, Kilpert F, Richter AS, et al. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* 2016;44:W160–5. doi:10.1093/nar/gkw257.
- Pal A, Vicoso B. The X chromosome of hemipteran insects: conservation, dosage compensation and sex-biased expression. *Genome Biol Evol.* 2015;7:3259–68.
- Prince EG, Kirkland D, Demuth JP. Hyperexpression of the X chromosome in both sexes results in extensive female bias of X-linked genes in the flour beetle. *Genome Biol Evol.* 2010;2:336–46.
- Dagg J. Strategies of sexual reproduction in aphids (2002). <https://ediss.uni-goettingen.de/handle/11858/00-1735-0000-0006-ABF2-B>. Accessed 22 Dec 2016.
- Gupta V, Parisi M, Sturgill D, Nuttall R, Doctolero M, Dudko OK, et al. Global analysis of X-chromosome dosage compensation. *J Biol.* 2006;5:3.

50. Simon JM, Giresi P, Davis IJ, Lieb JD. Addendum: using Formaldehyde-Assisted Isolation of Regulatory Elements (FAIRE) to isolate active regulatory DNA. *Nat Protoc.* 2014;9:501–3.
51. Buenrostro J, Wu B, Chang H, Greenleaf W. ATAC-seq: a method for assaying chromatin accessibility genome-wide. In: Ausubel FM, editor. *Current protocols in molecular biology*, vol. 109. London: Wiley; 2015. p. 21.29.1–21.29.9.
52. Buenrostro JD, Wu B, Litzenburger UM, Ruff D, Gonzales ML, Snyder MP, et al. Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature.* 2015;523:486–90.
53. Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods.* 2013;10:1213–8.
54. Li B, Carey M, Workman JL. The role of chromatin during transcription. *Cell.* 2007;128:707–19.
55. Thomas S, Li X-Y, Sabo PJ, Sandstrom R, Thurman RE, Canfield TK, et al. Dynamic reprogramming of chromatin accessibility during *Drosophila* embryo development. *Genome Biol.* 2011;12:R43.
56. Song L, Zhang Z, Grasfeder LL, Boyle AP, Giresi PG, Lee B-K, et al. Open chromatin defined by DNaseI and FAIRE identifies regulatory elements that shape cell-type identity. *Genome Res.* 2011;21:1757–67.
57. Kharchenko PV, Alekseyenko AA, Schwartz YB, Minoda A, Riddle NC, Ernst J, et al. Comprehensive analysis of the chromatin landscape in *Drosophila melanogaster*. *Nature.* 2011;471:480–5.
58. Johnsson P, Lipovich L, Grandér D, Morris KV. Evolutionary conservation of long non-coding RNAs; sequence, structure, function. *Biochim Biophys Acta.* 2014;1840:1063–71.
59. Hartman TR, Qian S, Bolinger C, Fernandez S, Schoenberg DR, Boris-Lawrie K. RNA helicase A is necessary for translation of selected messenger RNAs. *Nat Struct Mol Biol.* 2006;13:509–16.
60. Robb GB, Rana TM. RNA helicase A interacts with RISC in human cells and functions in RISC loading. *Mol Cell.* 2007;26:523–37.
61. Mendjan S, Taipale M, Kind J, Holz H, Gebhardt P, Schelder M, et al. Nuclear pore components are involved in the transcriptional regulation of dosage compensation in *Drosophila*. *Mol Cell.* 2006;21:811–23.
62. Smith ER, Cayrou C, Huang R, Lane WS, Côté J, Lucchesi JC. A human protein complex homologous to the *Drosophila* MSL complex is responsible for the majority of histone H4 acetylation at lysine 16. *Mol Cell Biol.* 2005;25:9175–88.
63. Taipale M, Rea S, Richter K, Vilar A, Lichter P, Imhof A, et al. hMOF histone acetyltransferase is required for histone H4 lysine 16 acetylation in mammalian cells. *Mol Cell Biol.* 2005;25:6798–810.
64. Copps K, Richman R, Lyman LM, Chang KA, Rampersad-Ammons J, Kuroda MI. Complex formation by the *Drosophila* MSL proteins: role of the MSL2 RING finger in protein complex assembly. *EMBO J.* 1998;17:5409–17.
65. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* 2013;14:R36.
66. Anders S, Pyl PT, Huber W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics.* 2015;31:166–9.
67. R Core Team. R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2014. <http://www.R-project.org/>.
68. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinform Oxf Engl.* 2010;26:139–40.
69. Law CW, Alhamdoosh M, Su S, Smyth GK, Ritchie ME. RNA-seq analysis is easy as 1-2-3 with limma, Glimma and edgeR. *F1000Research.* 2016;5:1408.
70. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 2009;10:R25.
71. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods.* 2012;9:357–9.
72. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. *Bioinform Oxf Engl.* 2009;25:2078–9.
73. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell.* 2010;38:576–89.
74. Buske OJ, Hoffman MM, Pons N, Le Roch KG, Noble WS. Exploratory analysis of genomic segmentations with Segtools. *BMC Bioinform.* 2011;12:415.
75. Wickham H. ggplot2: elegant graphics for data analysis. New York: Springer; 2009. <http://ggplot2.org>.
76. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol.* 1990;215:403–10.
77. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. *BMC Bioinform.* 2009;10:421.
78. Eden E, Lipson D, Yogev S, Yakhini Z. Discovering motifs in ranked lists of DNA sequences. *PLoS Comput Biol.* 2007;3:e39.
79. Eden E, Navon R, Steinfeld I, Lipson D, Yakhini Z. GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinform.* 2009;10:48.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

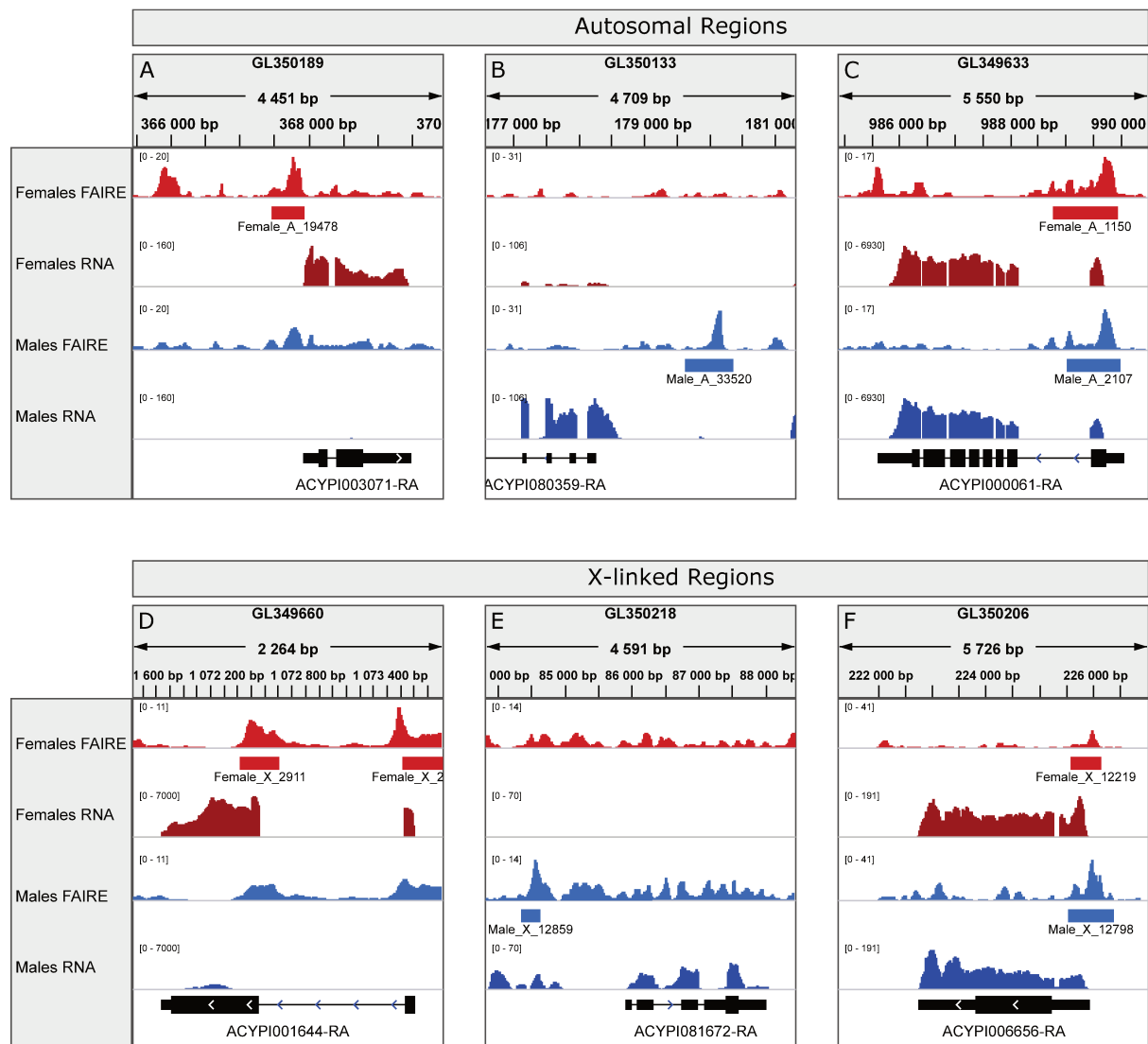
Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)



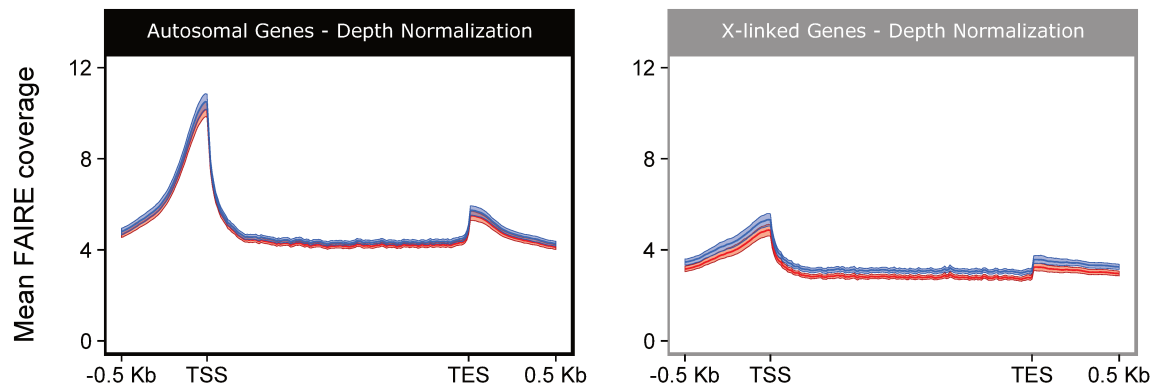
**Additional file 1: FAIRE and Control libraries coverages.** Number of reads and coverage of X chromosome (X) and autosomes (A) for male and female control libraries.

	Female Control (AA XX)		Male Control (AA X0)	
	X	A	X	A
Number of reads	4 906 829	9 454 084	5 304 035	22 708 453
Number of base covered	490 682 900	945 408 400	530 403 500	2 270 845 300
Size	166 018 072	340 086 956	166 018 072	340 086 956
<b>Coverage (expressed in x)</b>	<b>3.0</b>	<b>2.8</b>	<b>3.2</b>	<b>6.0</b>

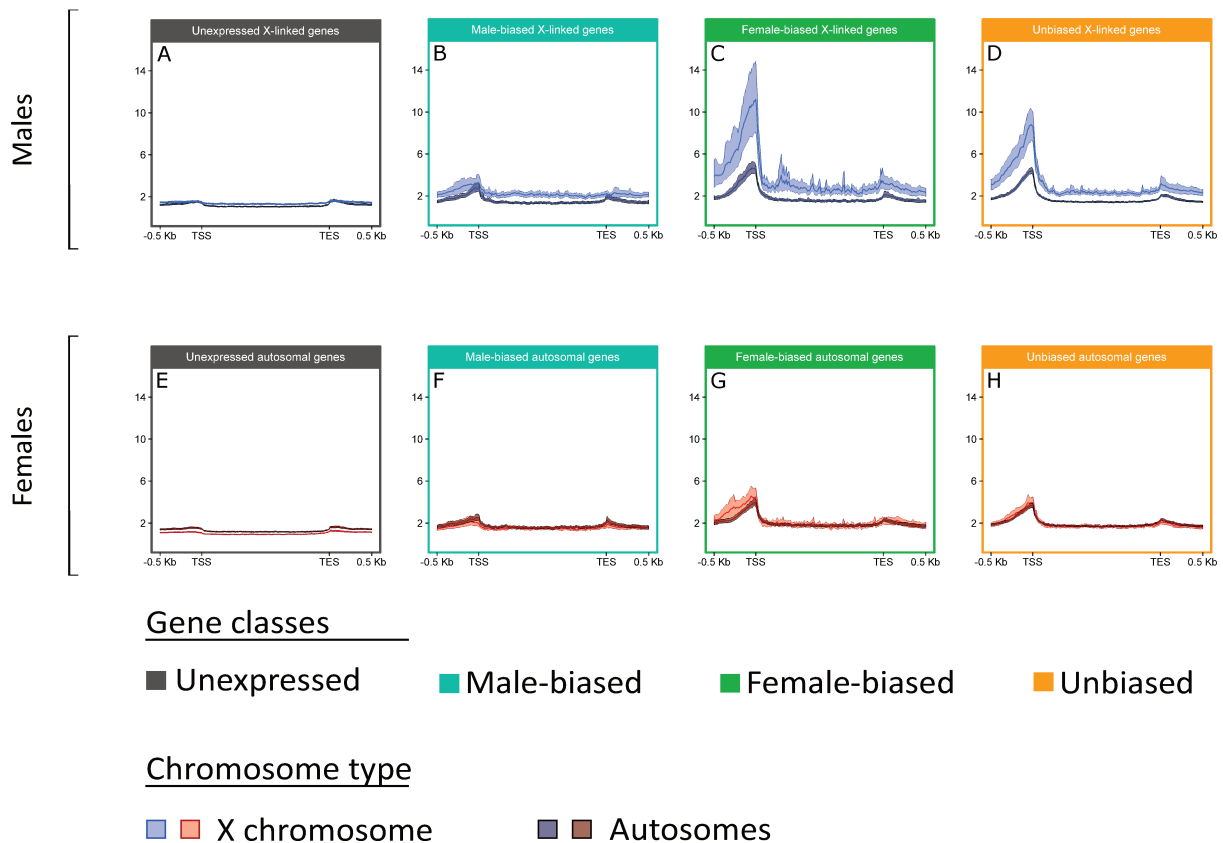
**Additional file 2: Genome browser view of remarkable autosomal and X-linked regions displaying sex-specific and non-specific FAIRE-seq and RNA-seq signal. A, D: female specific regions around the genes ACYPI003071 (uncharacterized protein) and ACYPI001644 (cuticular protein 44). B, E: male specific regions around the genes ACYPI080359 (uncharacterized protein) and ACYPI081672 (uncharacterized protein). C, F: regions in common between males and females for the genes ACYPI000061 (ATP synthase subunit beta) and ACYPI006656 (molybdate-anion transporter). The RNA-seq and FAIRE-seq signals have been made equal between males and females for each region.**



**Additional file 3: Mean FAIRE coverage calculated for all autosomal genes (left, black) and X-linked genes (right, grey) for males (blue) and females (red). 99% CI based on 1000 bootstrap is showed around the mean. The FAIRE coverage has been normalized by read-depth in order to allow the comparison between males and females data.**



**Additional file 4: Mean FAIRE signal and proportion of X-linked and autosomal genes depending on their expression profile, grouped by sex.** Genes have been categorized in four different classes: *unexpressed* (A, E), *male-biased* (B, F), *female-biased* (C, G) and *unbiased* (D, H) genes between the two sexes based on gene expression data. The mean FAIRE signal has been calculated around each gene class (500bp) and in their gene body (scaled), depending on the chromosome type (autosomes in dark color, X chromosome in light color) and depending on the sex (males, top in blue and females, bottom in red. 99% CI based on bootstraps is also showed around the mean.



**Additional file 5: GO enrichment of biological processes for autosomal and X-linked, male and female-biased genes.**

**X-linked male-biased genes GO enrichment.** Enriched Biological Processes GO terms of male-biased X-linked genes with unique *D. melanogaster* homologs (n=235).

GO term	Description	P-value	FDR q-value	Enrichment (N. B. n. b)
GO:0003341	cilium movement	1.82E-12	1.05E-08	19.59 (5349.13.210.10)
GO:0007018	microtubule-based movement	9.01E-07	2.60E-03	4.75 (5349.75.210.14)
GO:0007017	microtubule-based process	5.86E-05	1.13E-01	2.56 (5349.209.210.21)
GO:0006513	protein monoubiquitination	5.97E-05	8.63E-02	25.47 (5349.3.210.3)
GO:0046692	sperm competition	7.36E-05	8.52E-02	14.56 (5349.7.210.4)
GO:0044706	multi-multicellular organism process	1.43E-04	1.38E-01	12.74 (5349.8.210.4)
GO:0045814	negative regulation of gene expression. epigenetic	3.42E-04	2.83E-01	5.09 (5349.35.210.7)
GO:0040029	regulation of gene expression. epigenetic	6.23E-04	4.50E-01	3.40 (5349.75.210.10)
GO:0044703	multi-organism reproductive process	8.91E-04	5.73E-01	8.49 (5349.12.210.4)

**X-linked female-biased genes GO enrichment.** Enriched Biological Processes GO terms of female-biased X-linked genes with unique *D. melanogaster* homologs (n=182).

GO term	Description	P-value	FDR q-value	Enrichment (N. B. n. b)
GO:0009886	post-embryonic animal morphogenesis	3.97E-05	2.29E-01	4.56 (5349.309.38.10)
GO:0048518	positive regulation of biological process	4.02E-05	1.16E-01	2.85 (5349.790.38.16)
GO:0048583	regulation of response to stimulus	9.16E-05	1.77E-01	3.19 (5349.573.38.13)
GO:0009653	anatomical structure morphogenesis	1.22E-04	1.76E-01	2.75 (5349.768.38.15)
GO:0048522	positive regulation of cellular process	1.40E-04	1.61E-01	2.88 (5349.685.38.14)
GO:0010646	regulation of cell communication	1.45E-04	1.40E-01	3.28 (5349.515.38.12)
GO:0023051	regulation of signaling	1.45E-04	1.20E-01	3.28 (5349.515.38.12)
GO:0035120	post-embryonic appendage morphogenesis	2.87E-04	2.07E-01	5.27 (5349.187.38.7)
GO:0035114	imaginal disc-derived appendage morphogenesis	3.27E-04	2.10E-01	5.16 (5349.191.38.7)
GO:0035107	appendage morphogenesis	3.95E-04	2.28E-01	5.00 (5349.197.38.7)
GO:0002682	regulation of immune system process	4.00E-04	2.10E-01	6.03 (5349.140.38.6)
GO:0003002	regionalization	4.46E-04	2.15E-01	4.90 (5349.201.38.7)
GO:0030718	germ-line stem cell population maintenance	5.31E-04	2.36E-01	10.43 (5349.54.38.4)
GO:0009966	regulation of signal transduction	8.96E-04	3.70E-01	3.12 (5349.451.38.10)

**X-linked unbiased genes GO enrichment.** Enriched Biological Processes GO terms of unbiased X-linked genes with unique *D. melanogaster* homologs (n=517).

GO term	Description	P-value	FDR q-value	Enrichment (N. B. n. b)
GO:0007424	open tracheal system development	4.23E-04	1.00E+00	2.44 (5349.81.461.17)
GO:0060541	respiratory system development	4.23E-04	1.00E+00	2.44 (5349.81.461.17)
GO:0098732	macromolecule deacylation	8.14E-04	1.00E+00	5.80 (5349.10.461.5)

**Autosomal male-biased genes GO enrichment.** Enriched Biological Processes GO terms of male-biased autosomal genes with unique *D. melanogaster* homologs (n=598).

GO term	Description	P-value	FDR q-value	Enrichment (N. B. n. b)
GO:0006629	lipid metabolic process	8.50E-08	4.91E-04	2.21 (5349.214.532.47)
GO:0044281	small molecule metabolic process	1.34E-07	3.87E-04	1.85 (5349.374.532.69)
GO:0044710	single-organism metabolic process	1.83E-07	3.52E-04	1.54 (5349.776.532.119)
GO:0055085	transmembrane transport	7.74E-07	1.12E-03	2.14 (5349.202.532.43)
GO:0044712	single-organism catabolic process	2.96E-05	3.42E-02	2.41 (5349.100.532.24)
GO:0042445	hormone metabolic process	4.14E-05	3.99E-02	3.66 (5349.33.532.12)
GO:0055114	oxidation-reduction process	4.20E-05	3.47E-02	1.86 (5349.227.532.42)
GO:0006091	generation of precursor metabolites and energy	1.20E-04	8.67E-02	2.99 (5349.47.532.14)
GO:0006820	anion transport	1.49E-04	9.58E-02	3.26 (5349.37.532.12)
GO:0015698	inorganic anion transport	2.03E-04	1.17E-01	3.93 (5349.23.532.9)
GO:0044255	cellular lipid metabolic process	2.36E-04	1.24E-01	1.95 (5349.155.532.30)
GO:0006082	organic acid metabolic process	3.93E-04	1.90E-01	1.77 (5349.204.532.36)
GO:0044262	cellular carbohydrate metabolic process	4.15E-04	1.85E-01	3.35 (5349.30.532.10)
GO:0006865	amino acid transport	4.22E-04	1.74E-01	3.62 (5349.25.532.9)
GO:0030239	myofibril assembly	5.15E-04	1.99E-01	5.03 (5349.12.532.6)
GO:0006099	tricarboxylic acid cycle	5.72E-04	2.07E-01	3.83 (5349.21.532.8)
GO:0044282	small molecule catabolic process	5.92E-04	2.01E-01	2.61 (5349.54.532.14)
GO:0015849	organic acid transport	8.08E-04	2.60E-01	3.35 (5349.27.532.9)
GO:0006979	response to oxidative stress	8.41E-04	2.56E-01	2.43 (5349.62.532.15)
GO:0032787	monocarboxylic acid metabolic process	8.64E-04	2.50E-01	2.28 (5349.75.532.17)
GO:0043648	dicarboxylic acid metabolic process	8.77E-04	2.41E-01	4.64 (5349.13.532.6)
GO:0006121	mitochondrial electron transport. succinate to ubiquinone	9.79E-04	2.57E-01	10.05 (5349.3.532.3)

**Autosomal female-biased genes GO enrichment.** Enriched Biological Processes GO terms of female-biased autosomal genes with unique *D. melanogaster* homologs (n=1107).

GO term	Description	P-value	FDR q-value	Enrichment (N. B. n. b)
GO:0002181	cytoplasmic translation	2.75E-32	1.59E-28	4.31 (5349.74.1023.61)
GO:0006807	nitrogen compound metabolic process	1.69E-20	4.89E-17	1.52 (5349.1144.1023.332)
GO:0034641	cellular nitrogen compound metabolic process	1.80E-20	3.46E-17	1.56 (5349.1020.1023.304)
GO:0022402	cell cycle process	4.88E-17	7.05E-14	1.89 (5349.401.1023.145)
GO:0090304	nucleic acid metabolic process	1.85E-15	2.14E-12	1.62 (5349.672.1023.208)
GO:0051298	centrosome duplication	2.61E-13	2.51E-10	3.56 (5349.47.1023.32)
GO:0009059	macromolecule biosynthetic process	1.59E-12	1.32E-09	1.74 (5349.399.1023.133)
GO:0031023	microtubule organizing center organization	2.18E-12	1.58E-09	2.73 (5349.88.1023.46)
GO:0034645	cellular macromolecule biosynthetic process	2.69E-12	1.73E-09	1.79 (5349.354.1023.121)
GO:0006139	nucleobase-containing compound metabolic process	4.14E-12	2.39E-09	1.48 (5349.789.1023.224)
GO:1901360	organic cyclic compound metabolic process	4.43E-12	2.33E-09	1.45 (5349.883.1023.245)
GO:0006412	translation	7.71E-12	3.71E-09	2.16 (5349.172.1023.71)
GO:0051297	centrosome organization	1.03E-11	4.58E-09	2.71 (5349.85.1023.44)
GO:0006396	RNA processing	1.16E-11	4.79E-09	1.83 (5349.306.1023.107)
GO:0043043	peptide biosynthetic process	1.26E-11	4.87E-09	2.13 (5349.177.1023.72)
GO:0032502	developmental process	1.28E-11	4.61E-09	1.27 (5349.1787.1023.434)
GO:0016071	mRNA metabolic process	1.30E-11	4.42E-09	1.97 (5349.231.1023.87)
GO:1903047	mitotic cell cycle process	1.38E-11	4.44E-09	1.90 (5349.261.1023.95)
GO:0006725	cellular aromatic compound metabolic process	1.47E-11	4.46E-09	1.45 (5349.856.1023.237)
GO:0016070	RNA metabolic process	1.68E-11	4.87E-09	1.60 (5349.522.1023.160)
GO:0046483	heterocycle metabolic process	2.30E-11	6.32E-09	1.45 (5349.828.1023.230)
GO:0043604	amide biosynthetic process	2.39E-11	6.28E-09	2.10 (5349.179.1023.72)

**Autosomal unbiased genes GO enrichment.** Enriched Biological Processes GO terms of unbiased autosomal genes with unique *D. melanogaster* homologs (n=2805).

GO term	Description	P-value	FDR q-value	Enrichment (N. B. n. b)
GO:0051179	localization	7.61E-07	4.40E-03	1.14 (5349.1063.2508.569)
GO:0006810	transport	1.66E-06	4.80E-03	1.15 (5349.910.2508.491)
GO:0051234	establishment of localization	7.24E-06	1.40E-02	1.14 (5349.945.2508.504)
GO:0006898	receptor-mediated endocytosis	1.94E-05	2.80E-02	1.73 (5349.37.2508.30)
GO:0007264	small GTPase mediated signal transduction	6.54E-05	7.57E-02	1.44 (5349.89.2508.60)
GO:0072583	clathrin-dependent endocytosis	6.71E-05	6.47E-02	1.73 (5349.32.2508.26)
GO:0035556	intracellular signal transduction	1.08E-04	8.90E-02	1.26 (5349.221.2508.131)
GO:0016192	vesicle-mediated transport	1.39E-04	1.01E-01	1.20 (5349.355.2508.200)
GO:0048488	synaptic vesicle endocytosis	2.05E-04	1.32E-01	1.71 (5349.30.2508.24)
GO:0099643	signal release from synapse	3.78E-04	2.19E-01	1.40 (5349.82.2508.54)
GO:0001505	regulation of neurotransmitter levels	3.85E-04	2.02E-01	1.38 (5349.91.2508.59)
GO:0023061	signal release	4.17E-04	2.01E-01	1.40 (5349.84.2508.55)
GO:0099003	vesicle-mediated transport in synapse	4.35E-04	1.93E-01	1.55 (5349.44.2508.32)
GO:0006464	cellular protein modification process	5.41E-04	2.24E-01	1.14 (5349.601.2508.320)
GO:0036211	protein modification process	5.41E-04	2.09E-01	1.14 (5349.601.2508.320)
GO:0099531	presynaptic process involved in chemical synaptic transmission	6.48E-04	2.34E-01	1.38 (5349.85.2508.55)
GO:0070647	protein modification by small protein conjugation or removal	7.97E-04	2.71E-01	1.29 (5349.139.2508.84)
GO:0007269	neurotransmitter secretion	9.15E-04	2.94E-01	1.39 (5349.75.2508.49)

**Additional file 6: Protein sequence comparison of the Drosophila's DCC in the pea aphid.** CLUSTAL multiple sequence alignment by MUSCLE (3.8) of *A. pisum*'s homologs of *D. melanogaster* DCC. Proteic domains of the Drosophila are represented in full color boxes above the alignments. *A. pisum* gene names found by BLASTp are indicated in parenthesis.

### MSL1 (XM\_003243415.1)

```

MSL1_DROME      MDKRFKWPFKRANYLESYPHIPSRGRQRNLHGHPNQTOHLHQHPGKIYERQQYGNRG
MSL1_ACYPI      MARR-----RNSRGVGRA
                *  .*                               *:  *  **
MSL1_DROME      GHGGGNNNYRKLHLSLPAEHGGGAMAPPSSGGTVCAGADMVKLISENNNLRMVMLNLNL
MSL1_ACYPI      -----
MSL1_DROME      MQEQTDSIAAKDKELDDQSAKMSVVKAQNEELKQAVAQLEAANQELCKQLRRKNQRRNDN
MSL1_ACYPI      -----SAMSKEKDEK-----
                :*  .**  *::
MSL1_DROME      DDDDDPPLPPAAPQQKLIRCHAETQTVFREREQGTQTIDAQPQFANALPRGINMKESPA
MSL1_ACYPI      -----
MSL1_DROME      LDHHAGAVTNQPASKRSESKGRGEFNGKKVSTFILQRMNQDFEHHIHEQTEVAEEHEMEA
MSL1_ACYPI      -----IL-----
                **
MSL1_DROME      HKEQISQEEQDLVAEEDHLHMQEVHTEEVVGGDIFHDALESIEMEVVTEELVDMEEHGQS
MSL1_ACYPI      ---QLTNEVEYLKQKSDVLSFTQVSS-----
                *:::*  : *  :.*  *  : : *  :
MSL1_DROME      VDANGHIEEDDEDEDEDDENSDKDDDSEDDYPWMHSDADVNARTEELWQNQNYLLELD
MSL1_ACYPI      -----
MSL1_DROME      PTEEKTCAPSAHSTPNHQKSSQAEIRKEGNQRNRITEKLLQLKPEPMVDALEAPILPKW
MSL1_ACYPI      -----TCSS-----
                **:.
MSL1_DROME      VAFKKKDKHEHESVPESPEVPKQQPHQEDAIVDHNAIKNQLEVPKPDLPKQPKDEQRQD
MSL1_ACYPI      -----GIGDLSLAEKQYQOEHQHIKPK-----
                .*  *  .  ::*  :  :  :***
MSL1_DROME      QLDVRVEPQEDVRKVQKETLKRQPEDAPKHLPKAVAPKVTKTSSRESTLPKANTADIKD
MSL1_ACYPI      -----KRTIRQAATKQPNPDP-----LESNQTKRRRTSIPQTKSAKLKN
                ...  .::  ..**:*  *  :  :.*  *  :::*::*:.*:
MSL1_DROME      APAQKVIANHQSTKT---QTDPVKTQRLQVKIRQYEMHPDMRTGSSAPSDIRKQKNVDPV
MSL1_ACYPI      TEELSLITADDAYYTCLGSYDPIKS-----NPV
                :  .:*:  ::  *  .**:*:  :**
MSL1_DROME      STPETKTIKSKSMLVNDKKTTSSETSQSPDQEIDVETVRRKLAEHLKKEKLSQSHSSQVTL
MSL1_ACYPI      SKNTKRNI-----
                *  .  ...*
MSL1_DROME      KKIRERVATNLIYPPPSAPVSSTTITPAPTPTSTPTPGSTPQHAVTSSMDQEISAAKSKS
MSL1_ACYPI      -----
MSL1_DROME      KAAEQIATPLTPQSNSSVSSTTSTIRKTLNNCSPHTYSKATARSGLQSRFRATATFPYST
MSL1_ACYPI      -----

```

MSL1\_DROME RTWEDQEFHCDNEFFLEEADPELLADNPSLEIPKWRDVPVPPSSDKIDTELLSDATFERRH  
MSL1\_ACYPI -----EIPNYRLKKYSCRNSAKGIELMSDAIFQKRH  
\*\*\*:\* . . . \*\*:\* \*\* \*:\* \*\*

PEHE  
MSL1\_DROME QKYVKDEVDRKCRDARYMKEQIRLEQLRMRNQDEVIVALDPLRA--STFYPLPEDIEAI  
MSL1\_ACYPI FKHERSEQICKRRDLRQIRELIRVENLRKGRNKHHSQKSNDDNKEELTLLPSPSNIKSI  
\* : ..\* \* \* \* \* :.\* \*\*:\* \*\* \* : \* . : \* : \* \* . : \* \*

MSL1\_DROME QFVNEVTVQAFGENVVNMEARDDFGVPWVDAIEAPTSIARSKALAEPVATLASKKIPTTA  
MSL1\_ACYPI EITNKLVPVSAFGINLYRLTC-----  
:.\*:.\*.\* \*\* \* : . : .

MSL1\_DROME AEARHQENHSSYVFPKRRKRQKNR  
MSL1\_ACYPI -----

### MSL2 (ACYPI49193)

MSL2\_DROME MAQTAYLVKTRIAMSASNLKRRVEELNSGLGELRQLLSCVCCQLL--VDPYSPKGRK  
MSL2\_ACYPI MALIEYYLSTSQIVLQAKSTDPLTWQKLHSLPLLEHLSCQVCHKLVDRNLNPNY-----  
\*\* \* \* : . \* . . :.\* \* \* \* : \* : \* \* \* \* : \* : \* \*

zf-RING10  
MSL2\_DROME CQHNVCRLCLRGKKHLFPSTQCEGCSDFKTYEENRMMAAQLLCYKTLVHLLHSALFGE  
MSL2\_ACYPI LDGYACTVCVNHQITENPS-----SAIIQCYKKLCAYIHSSPLYKV  
: . \* : \* : : \* \* : \* : \* \* . \* : : \* . \* :

MSL2\_DROME LAGMRPQVARELVPRIKLPPKTTQEFIREGSNISDTFDIFLPQPDLPFLKDMPTSLPAET  
MSL2\_ACYPI MC-----TRVEDKRLVELVTE-----  
: . : \* \* \* : \* \*

MSL2\_DROME PPTSAVTTPPELPHYDHLNINSDIEAEAAATAEQGHFSPLPLPTGSRMGMLSHAGQIVIAI  
MSL2\_ACYPI -----VISL  
\*\* :

MSL2\_DROME ESSESGFMDQAWTDQVDSGTVSVSKYTNNGNFAVSVMPTSATTKFDPQELQIGQVVQ  
MSL2\_ACYPI PRSINGY-----SGMVN-----GTVIK  
\* . \* : \* \* \* . \* \* :

MSL2\_DROME MADSTQLAVLAAVEETVETSTQLTVLSTTVEETVETSTQLEVLTSAEEPNEISDQLANLQ  
MSL2\_ACYPI QEKK-----EKDIEDDSTLMVNGTMNKQEI-----  
.. \* : \* \* . : \* \* \* \* : \* : \*

MSL2\_DROME VEESDEALVEETVEEAEGTSSIPSEVVAEHMEEDQHLVDVHTSQSPTQTEMEEAVEEHVATK  
MSL2\_ACYPI ----EEDIEDDSTLMVNGTMNKQEIKEEDIEDDFTLMVNGIMNKQEIKEENSKDDI----  
\* : : \* : . : \* \* . \* : \* : \* \* \* \* : . : : \* : : :

MSL2\_DROME TQLGHVQTELQDAESLQKDFEDAKAAAEBAKEKEKDLHAISAELQKEDSDEPTLKRKRTR  
MSL2\_ACYPI TLISSDIQLPTVNETPEDYINYQIFDVQSQ-----  
\* . . : \* . . : \* : : : : :

MSL2\_DROME TLKASQAAKIEPVPSEVKTQVSGKALRRIRGKDKEEKVPPKPKRCRCSGSSNTLTT  
MSL2\_ACYPI -----KHKIQN-QSVFNIPKKNKNER---RWGCRCGNATTPGKLT  
\* \* \* . : . : . \* \* \* . \* \* \* : : \* \*

MSL2-CXC  
MSL2\_DROME CRNSRCPCYKSYNSCAGCHCVCKNPHKEDYVESDEDDLEDFFEMPKDVPPEPMTQSEEPV  
MSL2\_ACYPI CFGQRCPCYTEQKPCDQCKCRGCRNPRQK---RSNDDNLE-----  
\* . . \* \* \* . . : \* \* \* \* . : \* : \* \* \* \* :

MSL2\_DROME VAEPRQEENSMAPPDSSAPISLVPLNNLQOSQHPLVLVQNEKGEYQGFNIFQGSKPLDPV  
MSL2\_ACYPI IDQIRRK-----PVTLELVSSLKPSH-----NNSNSLFSAYTMHD-----  
: : \*.: \*:\* :..\*:\* \*.\* :\*... :...:..:

MSL2\_DROME TVGFTIRVQLQHTDGFGLPQYAYIMPTIDPPNPPAPSLSPPPPPAPDREVIEPPAKKFR  
MSL2\_ACYPI MLQFS-----SHSQQFD-----  
: \* : .\*: : \*

MSL2\_DROME TSTRTRRGRANFSALDTVDELVSGGSRNSAAGDRSSATDNAHSLFEEIMSGSDDL  
MSL2\_ACYPI -----QGGTEDN-----LGHVNSIFHSP-----  
.\*\*: .\* :...:\*\*\*

### MSL3 (ACYPI000966)

MSL3\_DROME MTELRDET-PLFHKGELVLCYEPDKSKARVLYTSKVLNVFERRNEHGLRFYEYKIHFQGW  
MSL3\_ACYPI MAAKVLTKSKFEEGEKVLVLCYEPDPAKTKVLYDSKVLRVVPEKDEQGRKFKFLIHFQGW  
\*: . \* . \* :\*\* \*\*\*\*\* :\*:.\*\*\* \*\*\*,\* . :\*: \* . \* : \*\*\*\*\*

MSL3\_DROME RPSYDRCVRATVLLKDTENRQLQRELAEAAKLQIRGDYSYKGTDPKPSAKKKRGGKAAH  
MSL3\_ACYPI NSTWDRFVTDEFILKDTENRKLQKELAEAAQLTPGGNL-YRKKRKRRAVKLE---PKPL  
..: \*\* \* . :\*\*\*\*\*:\*\*\* \*\*\*,\* \* : \* . \* . \* : \* :

MSL3\_DROME VEEPIVVPMDTGHLEAEHEMAPT PRAAGNRTRDNSGGKRKEKPPSGDGRKGNRGRQTET  
MSL3\_ACYPI VIEPAVVTIDDKSM-VENENVPVP-----  
\* \*\* \*\* .:\* : .\*: \* .\*.\*

MSL3\_DROME FYNNAINDVSVYNHVPQEDRIMMRVSERLRELIYDRNMIKVLGKQHALPARVPIVTIME  
MSL3\_ACYPI -----IDTI-----LLPKRRLPDLEFPDNLKFHTGYNCYLVHEKNTLVQLPCQPNVVTLE  
\* : : \* : : ..:.\* \* : : : .. \*\*.. :\*\*\*:\*

MSL3\_DROME NFVKQQAVELAISIKQDSSRARNTQSRNARMEREYDRVMSTVCMLKEVVDGLRIYFEFHV  
MSL3\_ACYPI SYLRYLARNNFSDNKTTKKKRQPEVLDKQLEKRY-----IICV--EVL DGLRICFNTFL  
... \* : . \* ... : :\*. \* :\* : \*\*.\*\*\*\* \* : :

MSL3\_DROME DDHLLYTEEKE---YVHNYLTDDNMRNCSLILNKSYEYINPSGDTELIGLDGTPVVEGSG  
MSL3\_ACYPI FRKLLVNEDEQAQYYEALKVTLQPPVNNIIPQNGEQDYDNVPNDEESESHDNKNNVKAKR  
:\*\* .\*: : \* : \* : \* . \* \* ..\* \* . \*.. \* : ..

MSL3\_DROME DTNGQIGVINIGGPEYEQQLQKCLLYIVTAS--GKNTAQAYERTSPYTAAYKLPVEMRGF  
MSL3\_ACYPI DNTRS-----STSQYASSQOTCSDCSLTKSQCSKIVKDTFEKARD-----EY  
\*.. . . :\* .. \*.\* :\* \* . \* . :\*: : :

MSL3\_DROME LNETFKWRLLS---AESPEKSMVFGAPHLVRLMIKMPMFLNASPISNKKLEDLLPHLD  
MSL3\_ACYPI VSKADSWKAVPDSAYDEEIKQPAVVYGVYHLLRLLLENLPKILANTEVDGEKLSIVYLYSN  
: : : .\*. : . \* : :\*:\*. \*\*:\* : \* \* : : : \*\* . : : :

MSL3\_DROME AFINYLENHREWFDRNFVNSTALPQEDLQRELLDSL DGLIAA  
MSL3\_ACYPI GLLKYLSTQTYLFGMQYVKNEMEDAVKSSAPIRNRHNRHKN  
: : : \*\*.. : \* . : \* : . . : : :

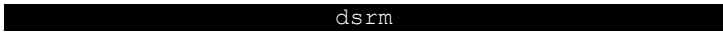



# MLE (ACYPI003650)

MLE\_DROME -----  
MLE\_ACIPI MANPKAYLHEWCAKNNLELQFETKQAGSNCKPRLFICEVTVTGHNYIGVGNSTNKKLAQA

MLE\_DROME -----  
MLE\_ACIPI NASKDYLLYLTK EGLISIDTYFLPSNCISGPASNGLMVNASQQLMNVVPPKSEYEEETS

MLE\_DROME -----  
MLE\_ACIPI ILGEAYRPNKTNFQLNDFLNDENNVEEA EGLNLHAAVCGDWTIDNAKSHLNQFIQSNKL

MLE\_DROME -----  
MLE\_ACIPI  dsrm  
-----MDIKSFLYQFCAKSQIEPKFDIRQTGPKNRQRFLCEVRVEPTYIGVGNSTN  
KNIDYKYS CNPKTFLHEWCEKNNLEPQFKTKQAGSNRKRPFICEVITGHDIIGIGNSAN  
: \*:\*:\*:\*:\* \*:\*:\*:\*:\* .:\*:\*:\*:\* \*:\*:\*:\*:\* : : \*:\*:\*:\*:\*


MLE\_DROME -----  
MLE\_ACIPI   
KKDAEKNACRDFVNYLVRVGLNTNDVPADAGASGGGPRRTGLEGAGMAGGSGQQKRVFDG  
IKYAQENASKDFLLYLTRKGLVSTDSLASNC-ISGPASNDLMVNASQKLMNVVPPKSVYQD  
\* \*:\*:\*:\*:\*:\* \*:\* \* \* :\*:\*:\*:\*:\* \* \* . . . : . \* . \* \* : .

MLE\_DROME -----  
MLE\_ACIPI QSGPQDLGEAYRPLNHDGGDGNRYSVIDRIQEQRDMNEAEAFDVNAAIHGNWTIENAKE  
HETPKMLGESYRPV-----GKANFQFNDFLNDENNVEEAHLVDNAAVHGGWTIDNAKS  
: . \* : \*:\*:\*:\*:\* : \* . . . \* : \*:\*:\*:\*:\* \*:\*:\*:\*:\* \*:\*:\*:\*:\* .

MLE\_DROME -----  
MLE\_ACIPI RLNIYKQTNNIRD-DYKYTPVGPEHARSFLAELSIYVPALNRTVTARESGSNKKSASKSC  
RLNQFIQSNKLNKNTDYKYSFIG----KSHVAELSIYIAKLRNVTARESGLNKRSASTSC  
\*\*\* : \*:\*:\*:\*:\* : \*:\*:\*:\*:\* : \* . \*:\*:\*:\*:\* . \* . \*:\*:\*:\*:\* \* \* \* \* \* \*


MLE\_DROME -----  
MLE\_ACIPI ALSLVRQLFHLNVIEPFGSLTKKKKDEQLKPYPVKLSPNLINKIDEVIKGLDLPVVNPRN  
ALSFIRQLYHLGVIEPFGSLKKSILNTVKPYEVNVDPLILNKVYNILOEFNIEPVIKND  
\*\*\* : \*:\*:\*:\*:\* \* \* \* \* \* \* \* \* \* \* : : \* \* \* \* \* \* \* \* \* \* : : : : \* . .


MLE\_DROME -----  
MLE\_ACIPI IKIELDGPPIPLIVNLSRIDSSQODGEKROESSVIPWAPPQANWNTWHACNIDEGELATT  
TIIPEP-TTISLISNQLAEFKELKPIKL--SRAILWSPPQPNWNPWLAINIDEGPLATA  
\* \* : . \* . \* \* \* \* . : . . \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \*

MLE\_DROME -----  
MLE\_ACIPI   
SIDDLSDMYERSLRDRQNDNEYRQLEFREKLPAAAMRSEILTAINDPVVIIRGNTGC  
TLDQLSEELAKDYTKKLNSSKGFKESIEMRSQLPVFNKNEILDIIRKNSIVIIQGSGTGC  
:\*:\*:\*:\* : . . . : : : \*:\*:\*:\*:\* : . \* \* \* \* \* \* \* \* \* \* \* \* \*

MLE\_DROME -----  
MLE\_ACIPI  DEAD  
GKTTQIAQYILDDYICSGQGGYANIYVTQPRRISAI SVAERVARERCEQLGDTVGYSVRF  
GKTTQVCQFILDEYLNKNDQAYCNIICTQPRKISAI SADRVAFERKEDIGLSVGYSVRF  
\* \* \* \* \* : . \* : \* \* \* \* \* : . \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \*

MLE\_DROME -----  
MLE\_ACIPI   
ESVFRPYGAILFCTVGVLLRKLKLEAGLRGVSHIIVDEIHERDVNSDFLLVILRDMVDYYP  
DSIFPRSTGAILFCTVGVLLRKLKLENGMRGISHVIVDEIHERGVNSDFLMVVLKDMVYNYYP  
:\*:\*:\*\* . \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \*

MLE\_DROME -----  
MLE\_ACIPI   
DLHVILMSATIDTTKFSKYFGICPVLEVPGRAFPVQQFFLEDIIQMTDFVPSAESRRKRK  
DLRVIFMSATINTAMFSKYFNCCPVIDIKGRCPVKEYFLEDIVQVLNYQPTDIKRRMN  
\* \* . \* \* : \* \* \* \* \* \* \* \* \* \* : : \* \* : \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \*

MLE\_DROME -----  
MLE\_ACIPI   
EVEDEEQLLSEDKDEAEINYNKVCEDKYSQKTRNAMAMLSSESDVSFELLEALLMHIKSK-  
KDYDDEYVI--DAQDHEENCNLLVSDDYPPEVKSVALISEKYVDFEIEALLTHIEIKM  
: \* . \* : : \* : : \* \* \* : . \* . \* . . . : \* : \* \* \* . \* \* \* : \* \* \* \* \* \* \* \*



**Additional file 7: Protein conservation of the Drosophila's DCC in the pea aphid.**

Proteic domains percentage of identity between *A. pisum* homologs of the five proteins composing the DCC in *D. melanogaster*.

<b><i>D. melanogaster</i> protein name</b>	<b><i>A. pisum</i> homologs</b>	<b>Proteic domains</b>	<b><i>A. pisum and D. melanogaster</i> domain identity</b>
MSL1_DROME	XM_003243415.1	PEHE	0.28
MSL2_DROME	ACYPI49193	zf-RING10	0.25
		MSL2-CXC	0.32
MSL3_DROME	ACYPI000966	MRG	0.22
MOF_DROME	ACYPI002102	MOZ-SAS	0.67
MLE_DROME	ACYPI003650	dsrm	0.45
		DEAD	0.68
		Helicase_C	0.69
		HA2	0.43
		OB NTP BIND	0.57

## C. Discussion

Grae aux analyses d crites dans l'article ci-dessus, d'une part nous avons identifi  un ratio d'expression X/A  gal   1 chez les m es, sugg rant une possible compensation de dose au niveau transcriptomique sur le X de ces derniers. D'autre part, nous avons identifi  une accessibilit  accrue de la chromatine de l'unique chromosome X des m es de mani re globale   l' chelle de l'ensemble des g nes port s par ce chromosome, et de mani re plus locale, notamment au niveau des g nes non-sexe biais s, et donc potentiellement compens s. Nous allons   pr sent discuter plusieurs points sp cifiques pr sent s dans l'article,   savoir l'absence de donn es FA"RE-seq chez les femelles sexu es, le challenge bioinformatique qu'a constitu  les analyses de ces donn es compte tenu de leur qualit  et enfin des r sultats suppl mentaires non publi s concernant l'identification de lncARN m e-biais s, notamment sur le chromosome X.

### 1. Absence de donn es de FA"RE-seq chez les femelles sexu es

Comme d crit au sein de l'article, les donn es d'expression ont  t  g n r es   partir de m es, de femelles asexu es et de femelles sexu es, tandis que les donn es FA"RE-seq d'accessibilit  de la chromatine n'ont  t  analys es qu'au niveau de pools d'individus m es et de femelles asexu es. Ceci s'explique principalement par la difficult  d'obtention d'un ratio FA"RE comparable entre femelles sexu es, m es et femelles asexu es (Tableau " ). Comme d crit en mat riel et m thode de ma th se, le ratio FA"RE correspond   une comparaison du rendement en ADN FA"RE et du rendement en ADN Contr le, et est calcul  par la formule suivante pour chaque exp rimentation FA"RE :

$$\text{Ratio FA"RE} = \frac{\text{Rendement en ADN FA"RE}}{\text{Rendement en ADN Contr le}}$$

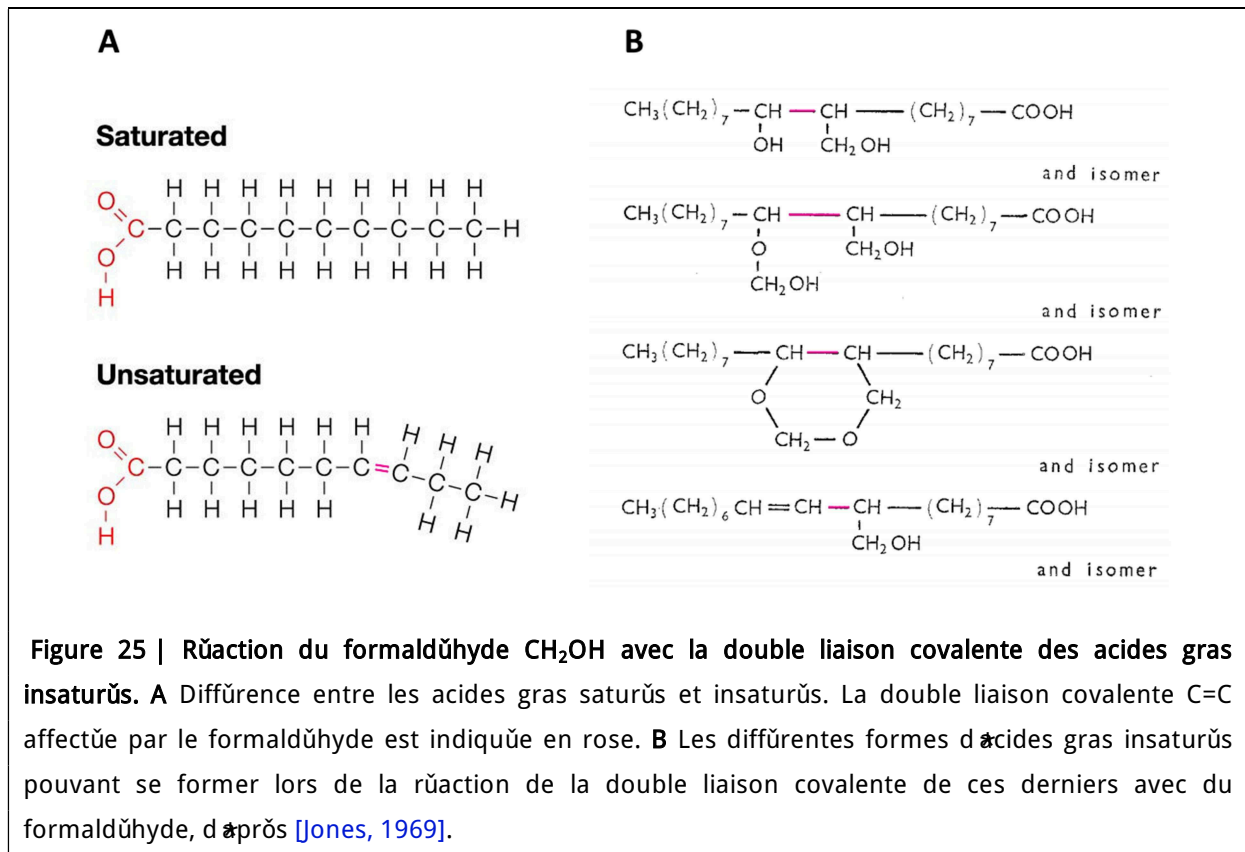
**Tableau " . R sultats d'extractions FA"RE   partir de 30 femelles asexu es, 40 femelles sexu es et 100 m es. Les concentrations correspondent   une moyenne relev e   partir de 3 r plicas techniques.**

;	ADN FA"RE	ADN Contr�le	Ratio FA"RE	ADN FA"RE	ADN Contr�le	Ratio FA"RE
C	1.5	1.5	1.0	1.5	1.5	1.0
C	1.5	1.5	1.0	1.5	1.5	1.0
C	1.5	1.5	1.0	1.5	1.5	1.0
C	1.5	1.5	1.0	1.5	1.5	1.0
a	1.5	1.5	1.0	1.5	1.5	1.0
a	1.5	1.5	1.0	1.5	1.5	1.0
a	1.5	1.5	1.0	1.5	1.5	1.0
a	1.5	1.5	1.0	1.5	1.5	1.0
C	1.5	1.5	1.0	1.5	1.5	1.0
C	1.5	1.5	1.0	1.5	1.5	1.0
C	1.5	1.5	1.0	1.5	1.5	1.0
C	1.5	1.5	1.0	1.5	1.5	1.0

Par l'analyse du Tableau 1, nous observons d'une part des ratios FA/RE plus  lev s chez les femelles sexu es en comparaisons avec les autres morphes. D'autre part, l' cart-type des ratios FA/RE mesur  entre r plicas biologiques est environ 4 fois sup rieur chez les femelles sexu es, comparativement aux m les et aux femelles asexu es. Ces ratios FA/RE plus  lev s et instables t moignent sans doute d'une interaction entre l'une des  tapes du FA/RE avec les quantit s importantes de lipides contenus dans les  ufs des femelles sexu es. En g n ral, les  ufs des insectes, comme ceux des moustiques par exemple, sont constitu s   plus de 70% de lipides gras insatur s [Ziegler, 2013], ce qui pourrait  tre le cas des  ufs non-fertilis s que comportent les femelles ovipares dans leur abdomen chez *A. pisum*. La fixation au formald hyde de l'ADN avec les prot ines histones est sans-doute l' tape la plus   m me d'entrer en conflit avec de grandes quantit s de lipides. On sait que le formald hyde r agit chimiquement avec la double liaison covalente des lipides insatur s (Figure 25). Cette r action chimique g n re un complexe comportant un groupement carbonyle en provenance du formald hyde, le rendant de fait moins disponible pour la formation de liaisons covalentes ADN-prot ine et prot ine-prot ine et r duisant ainsi potentiellement l'efficacit  de la fixation de la chromatine [Wolman et Greco, 1952 ; Jones, 1969]. Ceci expliquerait le niveau  lev  des ratios FA/RE observ s chez les femelles sexu es, correspondant   une sous-fixation de la chromatine du fait de l'interaction de la graisse contenu dans leurs  ufs non-fertilis s avec le formald hyde.

La sous-fixation au formald hyde de la chromatine est responsable d'un ratio signal / bruit final, apr s s quen age, g n ralement tr s bas, rendant difficile les analyses bioinformatiques des donn es g n r es. Ainsi, nous avons pris la d cision de ne pas s quencer ces  chantillons FA/RE obtenus   partir de femelles sexu es. Afin d'obtenir des  chantillons de suffisamment bonne qualit , il aurait fallu r aliser de nouvelles extractions FA/RE en utilisant un protocole comportant un temps de fixation ou une concentration en formald hyde plus  lev s. Il aurait  galement  t  possible de r aliser une pr cipitation pr alable des lipides contenus dans les  ufs apr s pulv risation et broyage des tissus par centrifugation, comme cela est effectu  dans la pr paration des  ufs de *X nope* pour l' tude de la r plication de l'ADN et des prot ines associ es   la chromatine [Gillespie *et al.*, 2012].

Cependant les femelles sexu es constituent un mat riel biologique difficile   g n rer, et ces r sultats ayant  t  obtenus au d but de la mise au point de la m thode FA/RE au sein du laboratoire, l' quipe manquait de temps et de recul sur la m thode pour prendre les d cisions ad quates afin de g n rer de nouveaux  chantillons FA/RE. En outre, les r sultats obtenus   partir de ces nouveaux  chantillons n'auraient sans doute pas  t  d'une importance cruciale au sein de l'article. En effet, les profils d'expression entre femelles sexu es et asexu es sont quasiment identiques (Figures 2. a, b, c et d de l'article 1), t moignant sans doute d'une accessibilit  globale de la chromatine corr l e avec celle des femelles asexu es au niveau des chromosomes X et des autosomes, n'apportant donc pas d'information contradictoire vis- -vis de l'accessibilit  de la chromatine des individus m le.



## 2. Challenge bioinformatique : des donn es de qualit  moyenne

Les  chantillons FA"RE m les et femelles asexu s s quenc s comportent donc des ratio FA"RE reproductibles et en accord avec les recommandations trouv es dans la litt rature (en moyenne inf rieur   3%) [Simon *et al.*, 2012]. Apr s les  tapes permettant la construction et le s quen age des banques FA"RE-seq r alis es   partir de ces  chantillons, il s'est av r  que l'ensemble de ces banques contenaient une proportion anormale de s quences r p t es (Tableau "").

**Tableau "".** R sultats d'analyse FastQC concernant les s quences surrepr sent es au sein des 8 banques FA"RE et Contr le issues de femelles parth nog n tiques et de m les entiers.

Library	GC Content	Sequence Duplication	Sequence Length	Sequence Quality	Sequence Complexity	Sequence Identity
Gr It t 1	~42%	High	~150	~30	~0.9	~0.9
Gr It t 2	~42%	High	~150	~30	~0.9	~0.9
Gr It t 3	~42%	High	~150	~30	~0.9	~0.9
Gr It t 4	~42%	High	~150	~30	~0.9	~0.9
a 1	~42%	High	~150	~30	~0.9	~0.9
a 2	~42%	High	~150	~30	~0.9	~0.9
a 3	~42%	High	~150	~30	~0.9	~0.9
a 4	~42%	High	~150	~30	~0.9	~0.9

L'analyse FASTQC a r v l  que ces s quences correspondent   des adaptateurs de s quen age issus des kits TruSeq  llumina. Ces derniers correspondent   des s quences courtes d'ADN ligu s de part et d'autre de l'ensemble des fragments d'ADN FA RE ou Contr le lors de la construction des banques, permettant ensuite leur s quen age sur des appareils  llumina de type Hi-seq. Les adaptateurs doivent repr senter en temps normal une proportion inf rieure   1% de la totalit  des s quences ADN constituant une banque de s quen age. Or ici, les adaptateurs de s quen age repr sentent au minimum 6.11% et au maximum 27.88% de la totalit  d'une banque (M le FA RE 3 et Femelle FA RE 1, respectivement, Tableau " "). Ces valeurs correspondent   une perte  quivalente de profondeur de s quen age pour chacune des banques. Le design exp rimental ayant  t  r alis  afin d'obtenir en moyenne 45 millions de reads par  chantillon s quenc , et en perdant en moyenne 22.4% de la profondeur de s quen age au sein de chaque banque, il en r sulte une profondeur de s quen age r duite pour certaines analyses comme l'appel de pics. Ceci est de plus accentu  par la perte de lectures au cours de leur alignement sur le g nome du puceron du pois, puisque ce sont en moyenne moins de 10 millions de lectures qui ont  t  retenues dans chacune des banques,   l'exception des banques m le Contr le et m le FA RE 3 (17.2 et 13.9 millions respectivement, Figure 26). L'ensemble des analyses portant sur le signal FA RE au sein de cet article reposent sur 20.5 millions et 23.6 millions de lectures FA RE chez les femelles et les m les, respectivement.

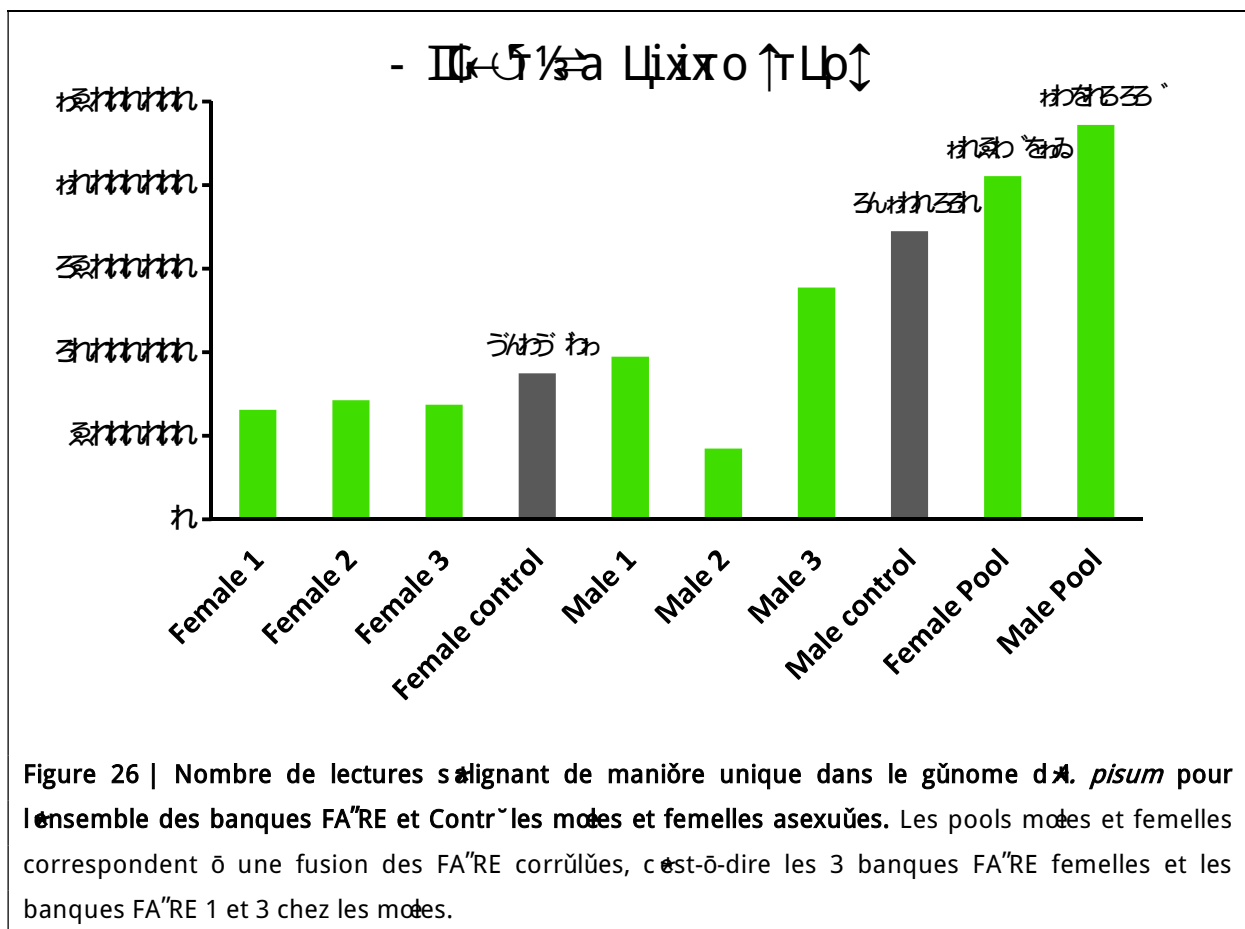


Figure 26 | Nombre de lectures s'alignant de mani re unique dans le g nome d'*A. pisum* pour l'ensemble des banques FA RE et Contr les m les et femelles asexu es. Les pools m les et femelles correspondent   une fusion des FA RE corr l es, c'est- -dire les 3 banques FA RE femelles et les banques FA RE 1 et 3 chez les m les.

La proportion anormale d'adaptateurs est sans doute r v latrice d'une quantit  importante de dim res d'adaptateurs, correspondant   une ligation de deux adaptateurs sans fragments d'ADN FA"RE ou Contr le entre ces derniers. Ce type d'anomalie est souvent expliqu  par une mauvaise s lection de la taille des fragments   s quencer, soit en prenant des fragments d'ADN trop courts fortement enrichis en dim res d'adaptateurs, soit en utilisant des protocoles de s lection de tailles de fragments d'ADN utilisant des billes et ne comportant pas d' tapes suppl mentaires pour l' limination de dim res d'adaptateurs [Bronner *et al.*, 2009 ; Head *et al.*, 2014].

Ces dim res faisant g n ralement plus de 100pb, ils peuvent en effet  tre confondus avec des fragments d'ADN FA"RE ou Contr le de taille similaire sur des profils quantifiant la concentration de fragments d'ADN en fonction de leur taille. La s lection de taille ayant  t  ici effectu e   l'aide de billes, un protocole de s lection de taille des fragments d'ADN par gel en ne gardant que des fragments de plus de 200 nucl otides aurait sans doute permis d' viter cette concentration  lev e en dim res d'adaptateurs, responsable de la faible couverture de certaines banques. Cependant, ceci n'explique pas la tr s faible couverture de la banque FA"RE mode 2 (9.6 millions de lectures), qui a d'ailleurs  t  d termin e comme non reproductible vis- -vis des deux autres banques FA"RE par "DR (cf "V.B.Fig. 3).

Malgr  la faible profondeur de s quen age, nous avons pu  tudier l'accessibilit  de la chromatine entre modes et femelles au niveau des autosomes et du chromosome X. Avec des banques n'ayant pas autant de dim res d'adaptateurs, une meilleure profondeur de s quen age aurait  t  obtenue et aurait permis d'identifier avec plus de pr cision des r gions diff rentiellement accessibles entre modes et femelles. En revanche, cela n'aurait pas alt r  les profils globaux d'ouverture et ainsi la robustesse de nos conclusions.

### 3. Identification de longs ARN non-codants modes-biais s

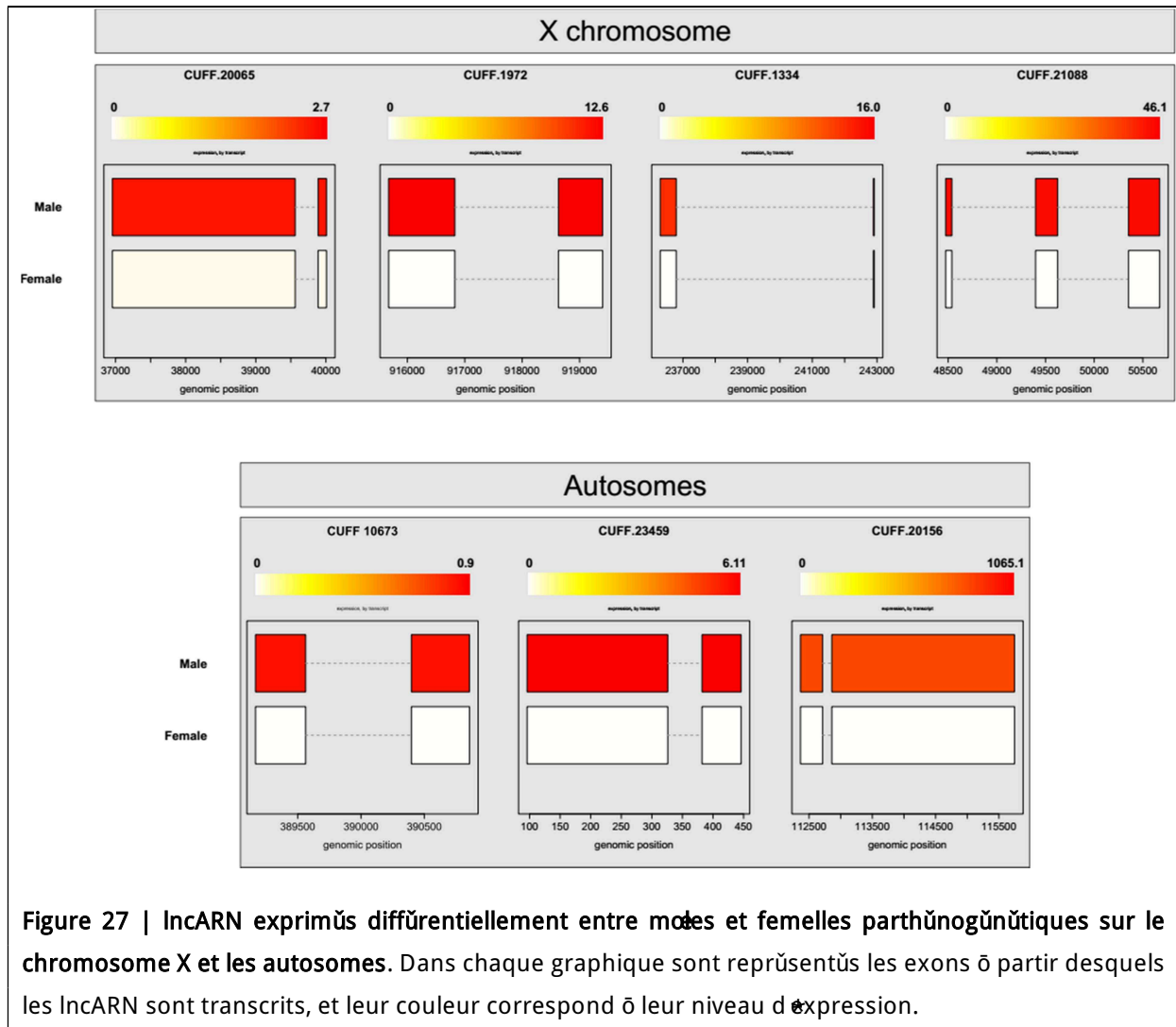
Les lncRNA sont connus pour  tre impliqu s dans des m canismes de compensation de dose, comme le Xist chez les mammif res qui englobe l'ensemble de la chromatine d'un chromosome X afin de le d sactiver transcriptionnellement [Penny *et al.*, 1996], et les lncARN *roX1* et *roX2* chez *D. melanogaster* qui permettent l'assemblage du complexe MSL sur le chromosome X des modes afin d'augmenter la transcription des g nes port s par ce chromosome [Conrad et Akhtar, 2012].

Les lncARN peuvent  tre identifi s   partir de donn es RNA-seq classiques gr ce   des algorithmes tels que FEELnc [Wucher *et al.*, 2017]. Ainsi des analyses compl mentaires sur les donn es d'expression modes et femelles asexu es ont  t  men es afin d'identifier de potentiels lncARN diff rentiels entre ces deux sexes. 518 lncARN ont  t  identifi s   l'aide de FEELnc au sein des banques d'ARNm modes et femelles asexu es fusionn es. Parmi ces lncARN, 7 sont diff rentiellement exprim s entre morphes et ces 7 lncARN sont fortement exprim s chez les modes comparativement aux femelles (Figure 27). Trois d'entre eux sont exprim s au niveau de loci associ s aux autosomes et les quatre lncARN diff rentiels restant sont exprim s au niveau du chromosome X des modes.

Le profil d'expression mâle-biaisé de ces lncARN soulève l'interrogation de leur implication potentielle dans la compensation de dose de *A. pisum*. En effet, on sait que chez *D. melanogaster* l'expression des lncARN roX1 et roX2 est mâle-spécifique [Meller, 2003]. Les 4 lncARN ici identifiés sur le chromosome X des mâles semblent présenter un profil d'expression similaire à roX1 et roX2. Ainsi, il y a deux hypothèses possible quant aux profils d'expression de ces 4 lncARN mâle-spécifiques : soit ils sont impliqués dans la compensation de dose des mâles, soit ils participent à d'autres fonctions et régulations transcriptionnelles.

Concernant la première hypothèse, certains de ces 4 lncARN pourraient par exemple permettre l'assemblage d'un complexe protéique impliqué dans le dépôt de marques histones activant globalement la transcription des gènes associés au X par une augmentation de l'accessibilité de la chromatine, au même titre que les lncARN roX1 et roX2 et le complexe MSL chez la Drosophile. En effet, l'ensemble des protéines constituant ce complexe présentent des homologues chez le puceron du pois, à différents degrés de conservation, comme sugguré par les données supplémentaires 6 et 7 de l'article présenté dans ce chapitre. Ceci constitue donc une hypothèse à tester concernant l'implication de ces lncARN dans des régulations épigénétiques en relation avec la compensation de dose chez les pucerons mâles.

Concernant la seconde hypothèse, il a été identifié d'une part une transcription du chromosome X globalement plus élevée chez les mâles que chez des femelles [Richard *et al.*, 2017a], et d'autre part un enrichissement de gènes mâle-biaisés au cours de l'évolution sur le chromosome X de *A. pisum* [Jaqiury *et al.*, 2013]. Il est donc possible que l'identification de ces lncARN mâle-spécifiques associés au chromosome X ne soit que le reflet du profil d'expression si particulier de ce chromosome chez le puceron du pois, et qu'ils ne soient donc pas directement impliqués dans la compensation de dose des mâles et puissent avoir d'autres fonctions de régulations qui ne sont pas associées à des mécanismes épigénétiques.



# V". Chapitre "": R gulations chromatiennes impliqu es dans le polyph nisme de reproduction du puceron du pois

## A. "ntroduction

### 1. Contexte

Dans le cadre de ce chapitre, nous nous proposons d'aborder les m canismes mol culaires qui r gulent les programmes g n tiques impliqu s dans le changement de mode de reproduction chez le puceron du pois. Des analyses transcriptomiques pass es ont montr  que les premi res  tapes de perception et transduction du signal photop riodique au niveau de la t te des femelles parth nog n tiques modifiaient les profils des transcrits des voies de signalisation neuro-endocrine (insulines, dopamine), des hormones (notamment les hormones juv niles), et des modifications de cuticules, probablement li es au r le de la dopamine dans la scl rotisation de la cuticule [Le Trionnaire *et al.*, 2009]. Des  tudes plus r centes ont initi  la comparaison des profils transcriptomiques des embryons en cours de d veloppement sexu  ou asexu , dans la fen tre de la p riode critique de d veloppement dans laquelle les embryons de stade 17 sont capables de r agir au signal photop riodique et d'orienter leur ph notype vers l'un des modes de reproduction. Les programmes g n tiques identifi s concernent les processus de ovogen se pr coces d crits chez la drosophile, ainsi que la stabilisation ou d gradation d'ARNm, et des modifications de la chromatine (comme *uhrf1*, *suv(4)-E20*) [Gallot *et al.*, 2012]

Dans ce contexte, ce chapitre a pour objectif d' tudier les modifications  ventuelles de la chromatine chez les embryons au cours du polyph nisme de reproduction, afin de tester l'hypoth se que ce processus est r gul  (en partie) par des m canismes  pig n tiques.

### 2. Rappel des objectifs et strat gie

Afin d' tudier les r organisations de la chromatine entre embryons sexu s et asexu s et de comprendre leur implication dans la mise en place du polyph nisme de reproduction des pucerons, nous avons extrait puis s quenc  la chromatine ouverte par FA"RE-seq, ainsi que les ARNm par RNA-seq des quatre derniers stades embryonnaires   devenir sexu  ou asexu . Ces exp riences de g nomique ont  t  men es via l'utilisation d'un protocole permettant de produire de mani re synchrone des embryons sexu s (produisant des gam tes haplo des) et des embryons asexu s (produisant des ovocytes diplo des) sous une condition de photop riode courte, par application ectopique ou non de kinopr ne, un analogue de l'hormone juv nile [Corbitt et Hardie, 1985 ; Gallot *et al.*, 2012]. Des analyses bio-informatiques ont permis de caract riser les r gions chromatiniennes

diff rentiellement ouvertes entre embryons sexu s et asexu s. De plus, nous avons associ s ces r gions ouvertes aux g nes qui peuvent potentiellement r guler afin de corr ler ouverture diff rentielle de l'ADN et expression diff rentielle des ARNm. Ces r sultats sont pr sent s sous forme d'un article en pr paration, et de donn es compl mentaires.

## B. Second article

### Title

Genome-wide analysis of open chromatin and transcriptomes in aphids of sexual and asexual embryos reveals asexual reproduction as the default differentiation program and early activation of lipid metabolism in sexual embryos

### Authors

Gautier Richard, Sylvie Tanguy, Nathalie Leterme-Prunier, Sylvie Hudaverdian, Fabrice Legeai, Anthony Bretaudeau, Denis Tagu, Ga l Le Trionnaire

### Abstract

Living organisms adapt their phenotype to environmental conditions with the help of continuous phenotypic plasticity. Some organisms respond in discrete manners to specific biotic or abiotic stimuli, which correspond to polyphenisms. For example, insects such as aphids display a reproductive polyphenism based on the production of asexual morphs under long photoperiod and sexual morphs under short photoperiod. This polyphenism is associated with a concentration of juvenile hormone in parthenogenetic females, which is high to trigger asexual reproduction in parthenogenetic females embryos, and low to trigger sexual reproduction in these embryos. The initiation of the reproductive polyphenism in aphids has been studied notably by comparing the transcriptomes of sexual and asexual embryos by the analysis of cDNA chips using specific environmental conditions. By rearing parthenogenetic females under short photoperiod (which naturally produce sexual embryos), and treating them, or not, with an analog of juvenile hormone, it is possible to synchronize the production of asexual and sexual embryos under short photoperiod conditions. Using this protocol, a differential expression of genes involved in chromatin reorganizations has been identified between sexual and asexual embryos, suggesting the involvement of chromatin accessibility modifications to set up the sexual and/or asexual reproductions. Here, we studied the chromatin and transcriptomic regulations involved in the setup of the pea aphid reproductive polyphenism by performing FA-RE-seq (open chromatin extraction) and RNA-seq in the last four stages of sexual and asexual embryos, including the last flexible embryo stage that is able to respond to the juvenile hormone titers of their mother, and thus to the hormonal treatment that we applied or not under short photoperiod. We identified differentially open chromatin peaks between sexual and asexual embryos, and correlated them with

the modification of expression of the genes in which these FA"RE peaks are located. By GO enrichment analysis and BLAST, we identified that the first stages determined as sexual embryos display chromatinian and transcriptional regulations of their lipid metabolism, suggesting an important role of lipids in setting up the sexual reproduction. Moreover, we identified that the chromatin organization is more conserved between flexible embryos and the first asexual embryo stage, than between flexible and sexual embryos. These results are surprising considering that under short photoperiod; sexual embryos could be primed, while on the contrary to asexual embryos are produced in response to an artificial hormonal treatment. These, and other results, suggests that the asexual reproduction mode might be the default program for flexible embryos, even undershort photoperiod.

## **Background**

During their life cycle, living organisms constantly need to cope with the fluctuations of their environment by producing adapted phenotypes. Phenotypic plasticity is one way to achieve this goal and this mechanism generates reaction norms, *i.e.* the production by a given genotype of a continuous range of phenotypes in response to the modulation of a given environmental factor [Griffiths et al., 2000]. For example, in insects and more particularly in *Drosophila*, the wings size changes accordingly to the temperature at which the larvae develop [Stanley, 1935]. Thus, the higher the temperature is, the larger the size of the adult wings is, because of an increased cell size, and to a lesser extent, because of an increased number of cells constituting the wings [De Moed and Al., 1997].

The majority of the phenotypic plasticity examples encountered in the nature are continuous. "n other rare and more spectacular cases, phenotypic plasticity can be discrete. "n such cases, called polyphenisms [Mayr, 1963], a genotype produces a defined number of discrete phenotypes in response to changes of an environmental factor. For example, the bee *Apis mellifera* [Evans et Wheeler, 1999] and the ant *Camponotus floridanus* [Bonasio, 2014] display castes polyphenism where individuals show discrete phenotypes associated with specific functions inside the colonies. Such castes can correspond for example to workers, soldiers, explorers and queens. The latter are the only individuals able of breeding within the hive or the anthill, thus ensuring the survival of the colony. "n most social insects, larval nutrition is the major environmental factor determining the caste of individuals. For example in bees, the type of jelly larvae are fed with, either royal or common jelly, induces different genetic programs that determine the fate of larvae to become either a queen or a worker, respectively.

Other insects can display other types of polyphenisms. "n that regard, aphids are another remarkable example since they can display dispersal, caste, or reproductive polyphenisms. The latter corresponds to the capacity of aphids to alternate their mode of reproduction during their annual life cycle according to the seasonal photoperiod [Tagu et al., 2005]. "n spring and summer, aphids reproduce asexually by parthenogenesis, which corresponds to a phenotype adapted to mild and high temperatures, and provoking a rapid colonization of their host plants and thus, agro-ecosystems. "n autumn, males and oviparous sexual females are produced by parthenogenetic females in response

to the decreasing photoperiod: after mating, females lay cold-resistant eggs. This phenotype allows a survival of the progeny in stressful conditions characterized by cold temperatures, as well as a mean to maintain genetic diversity within populations. Aphids are thus able to produce during their life cycle two discrete reproductive phenotypes adapted to seasonal constraints that explain their success as worldwide crop pests. The photoperiod signal perceived by the head of parthenogenetic females is transduced by the neuroendocrine system to the target tissues: the embryos and more particularly the germaria which contain the future oocytes. This signal modifies the embryos developmental trajectory as well as the fate of their germ cells. "ndeed, embryos that will develop as asexual females display diploid oocytes that will differentiate into embryos, while embryos that will develop in sexual morphs display haploid oocytes or spermatocytes, that will differentiate into meiotic gametes once fully developed as adult sexual females or males.

Previous analyses were focused on the effect of photoperiod detection and sensing on the genetic programs expressed in the heads of parthenogenetic females reared under long or short photoperiod [Le Trionnaire *et al.*, 2007b]. Based on differentially expressed transcripts [Le Trionnaire *et al.*, 2009], working hypotheses involving neuroendocrine signaling have been made, including putative roles of insulin related peptides [Huybrechts *et al.*, 2010b] and dopamine pathways in relation with cuticle sclerotization [Gallot *et al.*, 2010 : 201]. More recently, transcriptomes of embryos developing as sexual or asexual females were also compared [Gallot *et al.*, 2012 ; Duncan *et al.*, 2013]. 33 transcripts were thus identified as differentially expressed and among those 7 are potentially involved in oogenesis including homologues of *orb* and *nanos*, two genes involved in germ cell development found in *Drosophila*, 5 in post-translational regulation, 3 in the regulation of the cell cycle including *cyclin J*, an homologue involved in mitosis at the level of embryos in *Drosophila*, and finally 4 transcripts involved in epigenetic regulations, such as post-translational modifications of histones (*suv4-20*, *uhrf1*) or the deposition of histone variants H2B. 3 [Gallot *et al.*, 2012]. These results suggest that genetic programs underlying reproductive polyphenism in aphids are likely to involve epigenetic regulations of chromatin organization between asexual and sexual embryos.

Epigenetic mechanisms have been recently shown to be involved in insects polyphenisms regulation. Epigenetics is defined as a change in the phenotype of an individual without modification of its genotype [Waddington, 1942]. "t involves various mechanisms such as DNA methylation, post-translational histone modifications, or the involvement of some non-coding RNAs. Each of these mechanisms have a direct or indirect impact  for example by the recruitment of protein factors  on the chromatin accessibility by increasing or reducing it, thus increasing or reducing gene transcription, respectively. "n *A. mellifera*, the DNA methylation profiling of 4-day-old larvae revealed that the future workers larvae (fed with worker jelly) had higher levels of methylation in 80% of the regions identified as differentially methylated compared to the larvae destined to become a queen (fed with royal jelly) [Foret *et al.*, 2012]. This suggests that *de novo* DNA methylation could play a major role in bees caste polyphenism. "n ants, an overall differential level of acetylation of histone H3 lysine 27 (H3K27ac) was identified between the major and minor worker castes in *C. floridanus* [Simola *et al.*, 2013b]. "n addition, the gene coding for a CBP homologue in *C. floridanus*, *i.e.*, the protein responsible for the

deposition of H3K27ac in *Drosophila melanogaster* [Spannhoff *et al.*, 2011], was identified as differentially expressed between these two castes, suggesting an important role of this post-translational histone mark in the regulation of ant caste polyphenism. In the case of aphids, the annotation of the genome showed that the machineries required for DNA methylation of chromatin modifications were present [Walsh *et al.*, 2010], but no epigenetic mechanism has yet been identified as related to or implicated in the regulation reproductive polyphenism. In this study we aim at characterizing the chromatin accessibility landscape by Formaldehyde assisted Isolation of Regulatory Elements (FAIRE) that allows the isolation of open chromatin [Simon *et al.*, 2012] and its putative consequences on genes expression (with RNA-seq) during the establishment of embryonic reproductive polyphenism of aphids. To achieve this goal, we specifically studied genome plasticity of embryos destined to reproduce sexually or asexually using the protocol described in Gallot *et al.* (2012) on the aphidian model species *Acyrtosiphon pisum* (pea aphid). We thus performed FAIRE-seq and RNA-seq analyzes on the last four stages of developing embryos (stages 17 to 20) destined to be either sexual or asexual females. We first identified and characterized differentially accessible chromatin regions between asexual and sexual embryos, corresponding to regions putatively regulated by epigenetic mechanisms. We also identified patterns of differential chromatin accessibility correlated with modified gene expression, defining some of the genetic programs at stake in the embryos differentiation towards sexual or asexual reproduction. Finally, our data suggest that the asexual genetic program works as a default program for embryos differentiation before routing towards the sexual developmental program.

## **Material and methods**

### **Aphids rearing, embryos treatment and collection**

Similar conditions as those described by Gallot *et al.*, 2012 were used for aphid rearing. *Acyrtosiphon pisum* LSR1 clonal female individuals are maintained on *Vicia fabae* as parthenogenetic colonies under long photoperiod (16 hours) at 18, C and at low density (5 individuals per plant) during at least 2 generations to prevent the production of winged morphs. To induce the reproductive switch, third instar larvae are moved under short photoperiod (12 hours) at 18, c until they reach adulthood. Their L1 offspring are then isolated: they are the so-called sexuparae, which are parthenogenetic females that once adult will produce sexual individuals (oviparous females and male). The development of these sexuparae females is then synchronized during a 12 hours window during the molting from L3 to L4 stage, in order to keep L4 individuals of roughly the same age to perform subsequent treatments. At this time, the most developed embryos of each sexuparae are at stage 16 out of the 20 stages described in the pea aphid embryogenesis (Miura *et al.* 2003). 24 hours after the synchronization (L4+24h), the most developed embryos are at stage 17, which is the last stage of the embryogenesis that is able to respond to the hormonal treatment. The latter corresponds to the ectopic application of kinoprene, a juvenile hormone (JH) analog, on the sexuparae abdomen. This

hormonal treatment modifies the embryos germ cells fate by changing their differentiation to diploid oocytes, which then develop to embryos (asexual reproduction). Without this treatment under short photoperiod, it has been characterized that the first 40 embryos of the LSR1 sexuparae differentiate into sexual females, and thus display haploid germ cells and oocytes differentiating into unfertilized eggs (Gallot et al. 2012). This protocol allows the production of synchronized sexual and asexual embryos under one unique photoperiod condition. The hormonal treatment of the sexuparae and their embryos thus starts at L4+24h stage by separating the sexuparae in two groups. The first group of sexuparae females is treated with 50% of acetone, applied ectopically on their abdomen, and embryos collected from this female are sexual. The second group of sexuparae is treated ectopically with 400 ng of kinoprene (Sigma Aldrich) diluted in 50  $\mu$ L of acetone, and embryos collected from these individuals will become asexual (Gallot et al. 2012). To collect these two types of embryos, the 6 most developed ones of acetone treated sexuparae were dissected just after the treatment, constituting the first time point of the embryos developmental kinetic of this study. This time point is dubbed T0 and corresponds to embryos of stage 17, as defined by Miura et al. (2003). 24 hours later, the 6 most developed embryos were collected in a similar way for acetone and kinoprene treated sexuparae on a daily basis for 3 days, thus constituting the 2<sup>nd</sup> (T1Sex and T1Asx), 3<sup>rd</sup> (T2Sex and T2Asx) and 4<sup>th</sup> (T3Sex and T3Asx) time point of the kinetic. These time points correspond to the developmental stages 18 (T1), 19 (T2) and 20 (T3). The collected embryos were frozen in liquid nitrogen and stored at -80, C for mRNA or FA<sup>RE</sup> DNA extraction. 60 embryos were collected for mRNA extraction while several hundreds of embryos were collected for each time point for FA<sup>RE</sup> DNA extraction (indicated on Figure 1). This procedure has been repeated 3 times, to generate as much biological replicates of the whole kinetic.

As a control, some acetone and kinoprene treated sexuparae were not dissected in order to let them lay some first instar larvae, corresponding to fully developed embryos. Once they reached adulthood, these individuals were dissected in order to assess the efficiency of the acetone/kinoprene treatment by dissecting their reproductive apparatus: sexual ovaries that are only made of haploid eggs (acetone treated aphids, negative control) or asexual ovaries that are only constituted of embryos at different stages of development (kinoprene treated individuals). The results of the treatment are summarized in Supplementary Data 1

### **mRNA isolation, sequencing and analysis**

Messenger RNAs were isolated from each sample by using the mirVana<sup>®</sup> isolation Kit (ThermoFisher Scientific) according to manufacturer's instructions in order to separate small RNAs from mRNAs. RNA quality was checked on Bioanalyzer (Agilent) and quantified on Nanodrop (Thermo scientific). Messenger RNAs were extracted from the same batches of embryos initially collected for FA<sup>RE</sup> and Control DNA isolations. Prior to sequencing, an equimolar pool of the extracted mRNAs was performed, ending up with 7 mRNAs samples.

Messenger RNAs libraries preparation and 2x100nt paired-end sequencing were performed by the CGEx NGS platform of the Institut Curie using TruSeq Stranded mRNA Library Prep Kit (Illumina) and a

HiSeq 2500 in normal mode, respectively. One replicate for each condition was sequenced this way. We also used RNA-seq data of three biological replicates prepared in the same biological conditions as described above, but with a different extraction and sequencing protocol, generated before the present study (W cher et al., unpublished data). We thus analyzed the RNA-seq data of a total of 28 samples (7 conditions and 4 biological replicates).

All these paired-end libraries were mapped onto the pea aphid genome assembly 2.1 (Aphidbase) using the NCB" gene annotation v2 using STAR (Dobin et al., 2013) with the following parameters: --runMode alignReads --runThreadN 8 --outFilterMultimapNmax 5 --outFilterMismatchNmax 3 --align"ntronMin 10 --align"ntronMax 50000 --alignMatesGapMax 50000 --alignEndsType EndToEnd. The number of mapped reads for each library is reported in Supplementary Data 5. Fragment counts per genes were estimated by Subread featureCounts (Liao et al., 2014) using default parameters. Using R (R Core Team, 2014) with the limma (Phipson et al., 2016; Ritchie et al., 2015) and edgeR (Robinson et al., 2010) packages based on Law et al., 2016, these raw fragment counts were converted to counts per million (CPM). To perform this conversion, expressed genes were filtered by keeping genes with CPM>1 in at least 14 libraries among the 28 analyzed libraries. CPMs were normalized by using edgeR TMM method for Normalization Factor calculation. Graphics supporting the validation of the aforescribed pipeline can be found as Supplementary Data 6.

### **Open chromatin isolation by Formaldehyde Assisted "solation of Regulatory Elements, sequencing and analysis**

Open chromatin of the harvested embryos in the 7 different conditions and three replicates were extracted using the FA"RE protocol of Simon et al., 2012 and similar conditions as the ones described in Richard et al., 2017. Embryos were ground using a Biospec Bio-pulverizer in PBS1X buffer, and then fixed by the addition of 3% of Thermo Scientific Pierce formaldehyde during 8 min. The fixation was then stopped by the addition of glycine at 125 mM. After rinsing and re-suspending the pellet, the tissues were ground with a Tissue Lyser, using Qiagen metallic beads, during five cycles of 2 minutes and 30 seconds, each being interrupted during 2 minutes on ice. The sonication steps were performed using a Bioruptor Plus set in HighLpreset during 12 cycles of 30 seconds, each being interrupted during 30 seconds. The sonicated chromatin was then separated in two equivalent volumes. The first volume was used to extract FA"RE DNA by phenol chloroform extraction using Sigma phenol chloroform. Subsequently to its extraction, the FA"RE DNA was purified using ZYMO Ch"P DNA Clean & Concentrator. For Control DNA extraction, the second half of sonicated chromatin has been unfixed prior to phenol-chloroform extraction following Simon et al. 2012 recommendations. Finally, FA"RE and Control DNA were quantified using a Quantus with the Quantifluor dsDNA kit. The reproducibility of the replicates was assessed before sequencing by calculating the FA"RE/Control ratio described in Simon et al., 2012 (Supplementary Data 2).

FA"RE-seq libraries were prepared using TruSeq Ch"P Library Preparation Kit (Illumina) and sequenced using a 2x100nt protocol in a Hi-seq 2500 device (Illumina) in normal mode by "CGex NGS platform of the "nstitut Curie (Paris, France). Three biological replicates of FA"RE DNA extraction were sequenced

for each one of the 7 conditions, while the Control DNA samples were sequenced as equimolar pools of the three Control DNA extracted concomitantly with the FA"RE samples, for each condition. Subsequently to sequencing, FA"RE and Control reads were mapped using bowtie2 with default parameters [Langmead *et al.*, 2009 ; Langmead et Salzberg, 2012] on the pea aphid genome assembly v2.1 (AphidBase: [http://bipaa.genouest.org/is/aphidbase/acyrthosiphon\\_pisum/](http://bipaa.genouest.org/is/aphidbase/acyrthosiphon_pisum/)). Mapped read numbers are available in Supplementary Data 3. Only uniquely mapped reads with a mapping quality over or equal to 30 were kept using SAMtools (RR"D:SCR\_002105) [Li *et al.*, 2009], following the "DR recommendations [Li *et al.*, 2011b ; Kundaje, 2012]. MACS2 (RR"D:SCR\_013291) [Zhang *et al.*, 2008b] was used to perform the peak calling with the following parameters using control samples: --gsize 541675471 --nomodel --extsize 500 -p 0.05 --keep-dup all -f BEDPE, followed by "DR analyses using a threshold of 0.01 for original replicates, of 0.02 for self-consistency replicates and of 0.0025 for pooled pseudoreplicates. Replicates consistency was then assessed and the two most correlated FA"RE replicates out of the three in each condition were pooled; only the peaks with "DR above the threshold defined by the "DR analysis were kept. Retained FA"RE peaks were then pooled between the 7 conditions using the HOMER mergePeaks module.

The merged FA"RE peaks called by MACS2 and "DR were large, which was needed to better assess the reproducibility between replicates by overlapping the coordinates of peaks. We chose to perform peak calling refinement steps in order to reduce the FA"RE peaks size and to have an overall more precise peak calling. Using parameters described by Schick *et al.*, 2015, peaks were called using MACS2 on the FA"RE pooled replicates selected by the previous "DR analysis with the following parameters, without control samples: --tsize 101 --nomodel --keep-dup all --shift -100 --extsize 200 -p 0.0005 -t \$1 -f BEDPE. The number of nucleotides covered across the genome given by HOMER makeTagDirectory for each FA"RE-seq samples were used as --gsize. FA"RE and Control samples reads were counted for each peak called using HOMER, and only peaks with a FA"RE/Control fold-change over 1.5 were kept. The density of the FA"RE peaks size has then been compared before and after peak-calling refinement (Supplementary Data 4). FA"RE peaks genomic classes were attributed with the following priority: transcription start sites (*TSS*), *Promoter*, genes body (*Gene*) and *Intergenic*. *TSS* were defined as 450bp upstream and 450bp downstream of the predicted first nucleotide of each gene, according to the profile of Figure 4 (5'UTR or first exon, depending on each gene annotation), and promoters were defined as 2000bp upstream of the predicted TSS using the NCB" v2 gene annotation of *A. pisum*. The gene bodies were defined as the sequence between the predicted start and stop codons, including introns and exons.

Input-normalized FA"RE-seq signals were calculated using deepTools across the whole genome for each condition by calculating the mean Pooled FA"RE / Control normalized read counts in windows of 10bp. Subsequently to this calculation, maximum input-normalized FA"RE-seq signal value (abbreviated M" NFS) for each peak was retrieved using deepTools. This M" NFS value allows a simple, efficient and unbiased calculation of the height of a given peak. In order to keep biologically relevant FA"RE peaks when comparing the conditions T0   T1Asx   T1Sex (T0 T1 comparison), T2 Asx   T2 Sex



## Gene Ontology analysis

Since the pea aphid Gene Ontology is not available in many Gene Ontology tools, we chose to perform a BLASTp analysis with a maximum 1E-20 e-value to retrieve all the pea aphid\* homologs of the *Drosophila melanogaster* genome. Genes displaying biologically relevant FA"RE peaks and a drosophila BLASTp hit were analyzed with gcocoa from the gProfiler suite (<http://biit.cs.ut.ee/gprofiler/gcocoa.cgi>) [Reimand *et al.*, 2016]. Heatmaps of GO Terms enriched only in one FA"RE peak class were then calculated with R using log10(p-value) given by gcocoa.

## Results

### Global RNA-seq data analysis

RNA-seq experiments have been conducted on 7 types of embryos, depending on their developmental stage and reproductive status: stage 17 (T0), stage 18 (T1Sex and T1Asx), stage 19 (T2Sex and T2Asx), and stage 20 (T3Sex and T3Asx). The mapping of the RNA-seq libraries on the pea aphid genome assembly v2.1 (Aphidbase) with the NCB" genome annotation 2.0 yielded in average 15.1 ± 6.7 (s.d.) million reads (detailed in Table " ), with sensibly more mapped reads for the fourth replicates, which have been generated specifically for the present study. Transcriptional data collected on the seven types of embryos were analyzed with the help of the pipeline published by Law *et al.* 2016. Quality control graphics are available as Supplementary Data 6, and during such analyses, the correlation of the different biological replicates have been assessed (Figure 2). Using multidimensional scaling on the basis of the normalized Counts Per Million (CPM), accordingly to Law *et al.* 2016, we can observe that sexual and asexual embryos are closely related to each other inside a given developmental stage, *i.e.* biological replicates are grouped by embryos developmental stages rather than by their reproduction type (sexual or asexual).

When comparing biological replicates inside each one of the seven conditions, it appears that all four biological replicates are correlated. The first 3 biological replicates coming from the same sequencing experiment (W cher *et al.*, unpublished) are more closely related together than with the fourth replicate. This slight difference in correlation between replicates might reflect the fact respectively 1 and 3 replicates were sequenced separately, or it might also reflect the difference of sequencing depth of the fourth replicates compared to the other three (Supplementary Data 5). Since these fourth RNA-seq replicates were sequenced at the same time than the FA"RE-seq experiments, and since all RNA-seq data are correlated, it is then possible to compare the FA"RE-seq data to all four RNA-seq biological replicates for further analyses.

### Global FA"RE-seq data analysis and FA"RE peaks assignation to genomic features

In order to assess how the chromatin accessibility is modified during the sexual or asexual differentiation of embryos oocytes, Formaldehyde Assisted Isolation of Regulatory Elements followed by high-throughput sequencing has been performed on three biological replicates for each of the seven types of embryos (*T0*, *T1Sex*, *T1Asx*, *T2 Sex*, *T2 Asx*, *T3 Sex*, *T3 Asx*). The mapping of these libraries on the *A. pisum* genome (Aphidbase assembly v2.1) yielded a mean of  $39.4 \pm 16.0$  (s.d.) million mapped reads for the 21 FA"RE samples, and  $61.0 \pm 6.0$  (s.d.) million mapped reads for the 7 Control samples (Supplementary Data 3). Peaks have been called in each one of the seven conditions using FA"RE and Control DNA with MACS2 followed by Irreproducible Discovery Rate (IDR), in order to correlate biological replicates and keep the two most correlated replicates in each condition to pool them (Figure 3), according to the recommendations of the ENCODE Consortium for ChIP-seq data [Landt *et al.*, 2012]. Two IDR scores have been calculated (see Methods), and only replicates showing a Log2 of these scores inferior to two have been kept, according to IDR recommendations.

The correlation between open chromatin and genes expression were also assessed in each of the seven types of embryos, according to Simon *et al.*, 2012 recommendations (Figure 4). In every condition, the more a gene is expressed, the more its putative Transcription Start Site is accessible. Using this metagenic profiling of chromatin accessibility around every gene's putative TSS, we defined new genomic features for the purpose of attributing FA"RE-seq peaks to distinct genomic locations. The *TSS* class has been defined as 450bp upstream and downstream of every gene putative TSS, while *Promoters* have been defined as 2000bp upstream every *TSS*. The *TSS* class definition is relatively large in terms of length, in order to avoid false-positives concerning the FA"RE peaks attribution to *Promoters* or *Gene*.

After the MACS2 and IDR peak-calling to assess the reproducibility of FA"RE replicates, we performed a peak-calling refinement step in order to shrink the FA"RE peaks size for a more precise merging step afterwards (Supplementary Data 4). Indeed, all peaks called after the refinement were merged into one FA"RE peak set comprising 197 993 peaks, each one being attributed to one or more condition(s) among the seven types of embryos. As a quality control, the overlap of the FA"RE peaks with the four genomic classes (*Promoter*, *TSS*, *Gene* and *Intergenic*) revealed a predominant attribution of FA"RE peaks to the *TSS* class (Figure 5a), as expected for a FA"RE experiment [Simon *et al.*, 2012]. Moreover, the proportion of attribution of these 197 993 merged FA"RE peaks to the genomic classes were assessed (Figure 5b). It appears that FA"RE peaks are more likely to be attributed to *Intergenic* and then to *Gene* classes due to the size of these categories being superior to the size of the *TSS* and *Promoter* classes. This thus strengthens the enrichment in FA"RE peaks we have seen in the TSS Class.

### Analysis of FA"RE peaks assignation to each type of embryo

The merged FA"RE peaks set have been assigned to the relevant samples (among the 7 types of embryo) after a thorough filtration of peaks showing aberrant profiles (see Methods). The 197,993 FA"RE peaks have been assigned in three independent rounds of assignation to perform as much different comparisons (Figure 6). We first compared the transition between T0 and the two T1 conditions (T1Sex and T1Asex embryos) to study the chromatin reorganizations involved in the transition between the flexible embryos of stage 17, and sexual or asexual embryos of stage 18. This comparison also allows us to compare the differential chromatin accessibility of T1Sex and T1Asex embryos. Then, we compared T2Sex with T2Asex embryos and T3Sex with T3Asex embryos in order to follow the differences of embryo development as future sexual or asexual morphs. For each of these three comparisons, the number of peaks assigned to each biological class and the proportion of their attribution to genomic classes were assessed (Figure 6). For each comparison, most of the FA"RE peaks are assigned to the central classes of the Venn Diagrams, with 91.4% assigned to  $T0 \rightarrow T1Sex \rightarrow T1Asex$ , 88.4% assigned to  $T2Sex \rightarrow T2Asex$ , and 90.5% assigned to  $T3Sex \rightarrow T3Asex$  (Figure 6 a, c, e). This indicates that most of the identified FA"RE peaks are likely to be associated with constitutive euchromatin, and only few peaks are condition-specific and are thus putatively involved in shaping the genetic programs involved in the differentiation of embryos and their germ cells. These three peak classes share similar distributions across the genomic features (Figure 6 b, d, f), reflecting their similarity and the distribution of euchromatin throughout the genomic features defined in this study.

Concerning the external part of the Venn diagrams, equivalent FA"RE peaks numbers are assigned to  $T0$ ,  $T1A$  and  $T1K$  classes. Moreover for T2 and T3 time points, more peaks are found specific of asexual embryos than of sexual embryos, especially in stage 20 embryos where  $T3Asex$  FA"RE peaks are over 7 times more represented than  $T3Sex$  FA"RE peaks (Figure 6 a, c, e). FA"RE peaks specific of sexual embryos show a higher proportion of *intergenic* peaks, which could be associated with their lower number since the *intergenic* class is the more represented in the genome. However,  $T0$  peaks display the highest proportion of *intergenic* peaks despite a number of peaks similar to many classes that do not show this trend, namely  $T1Asex$ ,  $T1Sex \rightarrow T1Asex$ , and  $T0 \rightarrow T1Asex$ . The enrichment of intergenic peaks in certain conditions could thus reflect an enhancement in distal regulatory elements rather than just reflecting the higher chance of a high number of FA"RE peaks of a given set to be assigned to the *intergenic* genomic feature considering its size compared to others.

When comparing  $T0 \rightarrow T1Sex$  and  $T0 \rightarrow T1Asex$  FA"RE peaks biological classes, it is clear that less FA"RE peaks are conserved between T0 and T1Sex than between T0 and T1Asex, with 260 versus 743 FA"RE peaks, respectively. Moreover, this observable difference is slightly underestimated since total T1Sex FA"RE peaks are slightly more represented than total T1Asex FA"RE peaks (68 377 versus 67 776). There is thus a tendency showing that T0 is closer to T1Asex in terms of accessible chromatin.

### Global enrichment of functions putatively regulated by FA"RE peaks specific of each type of embryos

GO Terms enrichment have been performed on genes displaying FA"RE peaks belonging to condition specific classes, namely *T0*, *T1Asx*, *T1Sex*, *T2Sex*, *T2Asx*, *T3Sex*, *T3Asx*, or time point specific classes, namely *T1Sex*  $\rightarrow$  *T1Asx*, *T2Sex*  $\rightarrow$  *T2Asx*, *T3Sex*  $\rightarrow$  *T3Asx*. As described in methods, only genes having a *Drosophila* homologue have been analyzed using the GO Terms annotation of this organism, as such approach already gave insightful results in the pea aphid [Richard *et al.*, 2017b]. "n Figure 7, only GO Terms specifically enriched in one condition are displayed for clarity purposes. Considering the fragmentation of the pea aphid genome in over 20 000 scaffolds and the chimerism of these scaffolds [Jaqui ry *et al.*, 2017], it is difficult to assign genes to distal intergenic peaks, which thus have been discarded from further analyses.

For the T0 T1 comparison (Figure 7 a), T0 FA"RE peaks display only one GO Term enriched throughout the complete dataset, which is "egionalization" corresponding to pattern specification processes resulting in the subdivision of an axis or axes in space to define an area in which specific patterns of cell differentiation will take place. The lack of enrichment come from a lack of *Drosophila* homologues sharing similar GO Terms, resulting potentially from a high level of genes encoding uncharacterized proteins. *T1Sex* FA"RE peaks are found in genes having fruit fly homologs involved in "ipid metabolic process", "ellular lipid metabolic process" and "lycerolipid metabolic process", maybe corresponding to the beginning of yolk accumulation in sexual haploid oocytes differentiating into unfertilized eggs. "n *T1Asx* FA"RE peaks, some functions involved in female reproductive apparatus are found enriched such as "epithelial cell development", "eproductive system development", "ovarian follicle cell development" and involved in embryonic development such as "embryonic pattern specification" and "lastoderm segmentation", corresponding to very early embryonic development, and thus maybe to diploid germ cells starting to differentiate into early embryos at stage 18. For *T1Sex*  $\rightarrow$  *T1Asx* peaks, functions involved in general cell operation such as "actin filament-based process", "actin cytoskeleton organization" and "ellular homeostasis" are particularly enriched, which could correspond to peaks specific of the stage 18 embryos maintenance and development.

For T2 and T3 comparisons, the very large peak classes *T2Sex*  $\rightarrow$  *T2Asx* and *T3Sex*  $\rightarrow$  *T3Asx* are taken into account in order to assess which functions are putatively specific of the development of diploid oocytes in the asexual embryos ovaries. Because of these large peak classes, less condition specific GO Terms are found enriched compared to the T0 T1 comparison where the 4 classes taken into account in the heatmaps display around 1000 assigned FA"RE peaks.

For the T2 comparison (Figure 7 b), no specific functions are found enriched in genes displaying *T2Sex*  $\rightarrow$  *T2Asx* FA"RE peaks. Genes showing *T2Sex* FA"RE peaks have GO Terms "ellular macromolecule catabolic process" and "egulation of autophagy", which could be associated with metabolism pathways specific of sexual embryos. Genes with *T2Asx* FA"RE peaks display enriched GO Terms like "epithelial cell development", "euron differentiation" and "euroblast proliferation", which could correspond to the establishment of neurons in the developing early embryos inside stage 19 embryos.

For the T3 comparison (Figure 7 c), genes with *T3Sex* or *T3Sex*  $\rightarrow$  *T3Asx* FA"RE peaks do not display any condition specific GO Terms, while genes with *T3Asx* classes display specific GO Terms such as  $\text{actin filament-based process}$  and  $\text{myofibril assembly}$ , that could be associated to muscle development of the stage 20 embryos, since these functions are enriched comparatively to *T3Sex*  $\rightarrow$  *T3Asx* peaks.

### **"dentification of genes whose expression is putatively regulated by condition specific FA"RE-seq peaks**

Genes displaying a condition specific FA"RE peak in their *body*, *TSS* or *promoter* have been analyzed regarding their expression levels in the conditions of interest. These levels are based on normalized Counts Per Million (CPM), by comparing the expression of each gene in the conditions corresponding to the FA"RE peak class. As an example, for genes displaying *T2Sex*-specific FA"RE peaks,  $\log_2(T2Sex/T2Asx)$  has been represented for each gene, while genes showing *T0*-specific FA"RE peaks,  $\log_2(T0/T1Sex)$  and  $\log_2(T0/T1Asx)$  have been calculated and displayed (Figure 8). Hereafter, down-regulated, up-regulated and differentially expressed genes are defined as such by the help of the 0.56 and -0.56  $\log_2$  fold changes thresholds (dotted lines in Figure 8), corresponding to a minimum fold change of 1.5 between two conditions, since no differential gene expression tests have been performed. This choice has been made considering the low number of differential genes found between sexual and asexual embryos by W cher and colleagues, and by taking into account that differential expression is here supported by the presence of potential open regulatory elements identified by condition-specific FA"RE peaks inferred from FA"RE-seq data displaying high coverage.

Globally, it appears that most of the regulated genes display FA"RE peaks in their *body*, rather than their *TSS* or *promoter*, which corresponds to the usual profile of differential open chromatin sites [Waki *et al.*, 2011].

For *T0* FA"RE peaks, genes are distributed diagonally between  $\log_2(T0/T1Sex)$  and  $\log_2(T0/T1Asx)$ . This indicates a strong correlation of expression between T1Sex and T1Asx embryos, as previously suggested by Figure 2. Thus, many genes are differentially expressed between T0 and T1Sex and T1Asx, with identical expression levels between the T1 conditions and may thus correspond to genes with functions associated with the development of stage 17 embryos contrastingly with the development of stage 18 embryos. It is notably the case of the fruit fly homologues *zye* and *neo*, which encodes proteins involved in the regulation of embryonic cell shape in *Drosophila* [Fernandes *et al.*, 2010].

Genes displaying *T1Sex* and *T1Asx* FA"RE peaks are in majority differentially expressed with the T0 condition with only one gene that is differentially expressed between T1Sex and T1Asx embryos with a fold-change of -0.68, LOC100164454, an homologue of the fruit fly *slimfast*, which displays a *T1Sex*-specific FA"RE peak in its body and which is down-regulated in the T1Sex embryos comparatively to T1Asx embryos. *Slif* encodes a protein involved in the lipid metabolism in *D. melanogaster*. The genome view of this particular gene is displayed Figure 9.

"n genes showing *T2Sex* FA"RE peaks, we found that LOC100164454 *slimfast* is also down regulated with a -1.15 fold-change in *T2Sex* embryos compared to *T2Asx* embryos. LOC100161259 also encodes a cytosolic acetyl-CoA acetyltransferase involved in the fatty acid metabolism [Lynen et Ochoa, 1953], and is found down regulated in *T2Sex* embryos with a -0.57 fold change. LOC100161559 is a fruit fly ~~✶~~ *lysosomal enzyme receptor protein (lerp)* homolog and is found more expressed in *T2Sex* embryos, and less expressed in *T2Asx* embryos (FC of 0.98 or -0.98), thus displaying both a *T2Sex*-specific and a *T2Asx*-specific FA"RE peaks. *Lerp* encodes a protein notably involved in the lysosomal transport [Dennes et al., 2005].

Concerning *T2Asx*-specific FA"RE peaks, one is found in the body of LOC100167113, a homolog of *bond* in *D. melanogaster*, which is down regulated in *T2Asx* embryos. "n the fruit fly, *bond* encodes a protein involved in fatty-acid elongation and meiosis cytokinesis. LOC100569137, an E3-ubiquitin ligase RNF8-like, is also downregulated in *T2Asx* embryos, and could mediate extensive chromatin decondensation [Luijsterburg et al., 2012], thus reducing these global changes in the genome of asexual embryos compared to sexual embryos. Two fruit fly ~~✶~~ genes *lost* and *BicC* are involved in embryos patterning, and two homologs of these genes in the pea aphid genome, respectively LOC100160598 and LOC100168076, are found up regulated in *T2Asx* embryos compared to *T2Sex* embryos while displaying a *T2Asx* FA"RE peak in their body. These genes have already been found as differentially expressed between asexual and sexual embryos by cDNA-chip on the same type of biological material [Gallot et al., 2012].

"n *T3Sex* embryos, only 6 genes are found as differentially expressed comparatively to *T3Asx* embryos. Most of them encode uncharacterized proteins or do not have a homolog in *D. melanogaster*~~✶~~ genome. Still, LOC100570810, a Forkhead domain 59A (fd59a) homolog, is upregulated in *T3Sex* embryos and encompass a *T3Sex*-sepific FA"RE peak in its body. Forkhead domain 59A is a transcription factor that is expressed largely in postmitotic neurons in the central nervous system. "ts roles include nervous system development and egg-laying behavior according to flybase.

"n *T3Asx* embryos, fd59a is found down regulated while displaying a *T3Asx* FA"RE peak in its body, thus encompassing both a *T3Asx* and a *T3Sex* FA"RE peak. LOC100162729, the second most down regulated gene displaying *T3Asx* FA"RE peak in *T3Asx* embryos compared to *T3Sex* embryos, encodes a fruit fly ~~✶~~ *brummer* homolog, which is involved in triglycerides and lipids homeostasis and catabolic process [Gronke et al., 2005 ; Gr nke et al., 2007]. Concerning up regulated genes with *T3Asx* FA"RE peaks in *T3Asx* embryos, LOC100159378 encodes a homolog of *lodestar* involved in embryonic mitosis (flybase), LOC100168823 encodes a homolog of *nudel* involved in embryonic dorsoventral patterning (flybase), LOC100159201 encodes a homolog of *lachesin* involved in morphogenesis of embryonic epithelium [Pilot et al., 2006], LOC100168201 encodes a homolog of *neuroglian*, which is involved in axon and nervous system development, and LOC100164181 encodes a homolog of *watterproof* which is involved in filling the embryonic tracheal system in *D. melanogaster*. All these upregulated genes might putatively participate to the development of very early embryos inside stage 20 asexual

embryos, and *nudel* and *lodestar* have already been found as differentially expressed between asexual and sexual embryos by Gallot et al., 2012.

## **Discussion**

Using FA"RE-seq and RNA-seq, we studied the chromatin reorganization and its potential impact on gene expression during the setup of the reproductive polyphenism in the last four stages of the pea aphid embryogenesis, respectively developing as future sexual or asexual females.

### **Asexual reproduction seems to be the default reproduction mode, even in short photoperiod at the chromatin level**

Here we use a protocol described by [Gallot *et al.*, 2012] which allows the production of synchronized sexual and asexual embryos under a short photoperiod condition, notably by applying ectopically a juvenile hormone (JH) analog on the abdomen of sexuparae individuals to promote the switch from sexual to asexual embryogenesis. We then compared by FA"RE-seq the open chromatin profile of sexual and asexual embryos at different stages of development. By analyzing FA"RE peaks specifically identified in the open chromatin of stage 17 (T0), stage 18 sexual (T1Sex) and stage 18 (T1Asex) asexual embryos, it appears that nearly 3 times less FA"RE peaks are conserved between stage 17 and stage 18 sexual embryos (T1Sex), than between stage 17 and stage 18 asexual embryos (T1Asex). These results suggest that there are more global and profound chromatin reorganizations in the case of the transition from stage 17 embryos to the stage 18 sexual embryos than to asexual embryos. This clear and simple observation might suggest that asexual embryogenesis could be the default genetic program since embryos chromatin organization seems to be less affected than in the case of sexual embryogenesis. Moreover, a FA"RE peak specific of stage 19 asexual embryos might regulate the transcription of LOC100569137 by reducing its expression in stage 19 asexual embryos, resulting in a higher expression in stage 19 sexual embryos. LOC100569137 encodes an E3-ubiquitin ligase RNF8-like, which could mediate extensive chromatin decompaction like its mammalian counterpart [Luijsterburg *et al.*, 2012]. Such chromatin modification might occur in stage 19 sexual embryos where the gene is not down regulated, thus inducing profound chromatin reorganizations that are absent from stage 19 asexual embryos, further differentiating the sexual embryos chromatin profile compared to asexual embryos in relation with flexible embryos.

Altogether, we can thus make the other hypothesis that stage 17 embryos fate is likely to be asexual, even if its mother is placed under short photoperiod, known to trigger the switch towards sexual reproduction. It is well known that under natural conditions, parthenogenetic (sexuparae) females display low levels of JH under short days, which is correlated with the production of sexual morphs. On the contrary under long days, JH titers are higher and are correlated with asexual reproduction [Shikawa *et al.*, 2012]. It is thus possible to propose a new interpretation of the protocol we used in this study. Under short day condition, there is a gradual diminution of JH titers within the sexuparae mothers. At stage 17 (T0), embryos are still flexible which would mean that JH titers are not low enough to trigger the switch. At stage 18 (T1Sex), the critical threshold of JH titers - quantitatively

linked to the gradual perception of a sufficient number of consecutive short days - has probably been reached and would allow the effective switch of embryos from asexual to sexual reproduction. This would be then well correlated with the observation of important chromatin modifications of germline organization. On the contrary, when applying JH at stage 17, JH titers are artificially restored (and Long Days conditions mimicked) and these levels are then probably not low enough to initiate the switch to sexual reproduction when reaching stage 18. In that context, stage 18 embryos would remain asexual (T1Asex) which would fit with the rather conserved chromatin organization between T0 and T1Asex compared with the transition T0-T1Sex. In this hypothetical model stage 17 embryos would thus have an asexual fate and not a sexual fate. The initial interpretation of the protocol [Corbitt et Hardie, 1985] is that JH application is likely to promote asexual embryogenesis and that sexual embryogenesis is the default program under short days. Going into the level of chromatin organization thus allowed refining our understanding of the switch from asexual to sexual embryogenesis. In nature, photoperiod is the necessary and invariable signal that allows the autumnal switch under temperate regions where photoperiod and seasons changes are marked all throughout the years. On the contrary, under tropical regions, photoperiod changes are nearly absent and only asexual lineages can be found in the ecosystems. This simple observation thus fits with the idea that asexual reproduction is a default reproductive program, and that until the stage 17 of embryogenesis, embryos are primed to be asexual. Then only a sufficient decrease in photoperiod, correlated with at least a decrease in JH titers until a critical threshold is a necessary condition to initiate a switch towards sexual embryogenesis, as revealed in this study by the more important chromatin re-organizations between T0 and T1Sex in the absence of JH treatment.

### **Lipid metabolism is strongly regulated in embryos with germ cells differentiating into haploid unfertilized eggs**

By analyzing the GO Terms enrichment of fruit fly homologs of genes underlying *T1Sex* FA<sup>RE</sup> peaks, we found an enrichment of the functions lipid metabolic process, cellular lipid metabolic process and glycerolipid metabolic process. The genes showing such functions might be involved in the yolk accumulation into the haploid oocytes of sexual embryos differentiating to unfertilized eggs, suggesting that stage 18 embryos already begin to accumulate yolk in their haploid oocytes. Such functions might be regulated by putative regulatory elements binding the DNA sequence encompassed in the identified FA<sup>RE</sup> peaks. However no transcription factor binding site databases are available for the pea aphid genome, making such analyzes hazardous.

Moreover, we found a *T1Sex* FA<sup>RE</sup> peak in the body of LOC100164454, a gene coding for a homolog of the fruit fly *slimfast* protein. *Slif* mutants in *Drosophila* display a remobilization of nutrient stores in the fat body, resulting in aggregation of storage vesicles. The down expression of LOC100164454 in stage 18 sexual embryos could thus be involved in the remobilization of lipids stored in embryos fat body [Costechareyre et al., 2012] for the formation of yolk involved in the development of the haploid oocytes into unfertilized eggs.

Moreover, LOC100162729, the homolog of *brummer* in the fruit fly, is down regulated in asexual embryos of stage 20, potentially by transcription factor bindings its *T3Asx*-specific FA"RE peak, and is thus highly expressed in sexual embryos of stage 20. The *brummer* (*bmm*) gene encodes the lipid storage droplet-associated TAG lipase Brummer, a homolog of human adipocyte triglyceride lipase (ATGL). "n *Drosophila*, food deprivation or chronic *bmm* overexpression depletes organismal fat stores *in vivo*, and is thus involved in the release of lipids from the fat body, producing accumulation of lipid droplets in oenocytes [Gutierrez *et al.*, 2007]. The protein encoded by LOC100162729 could thus be involved in the remobilization of lipids from the embryos fat body for yolk accumulation in developing unfertilized eggs.

These results suggest a putative cross-talk between embryo fat body, oenocytes [Tsuchida *et al.*, 2014] and haploid oocytes, regulated by transcription factor binding in open chromatin sites in the case of the pea aphid reproductive polyphenism.

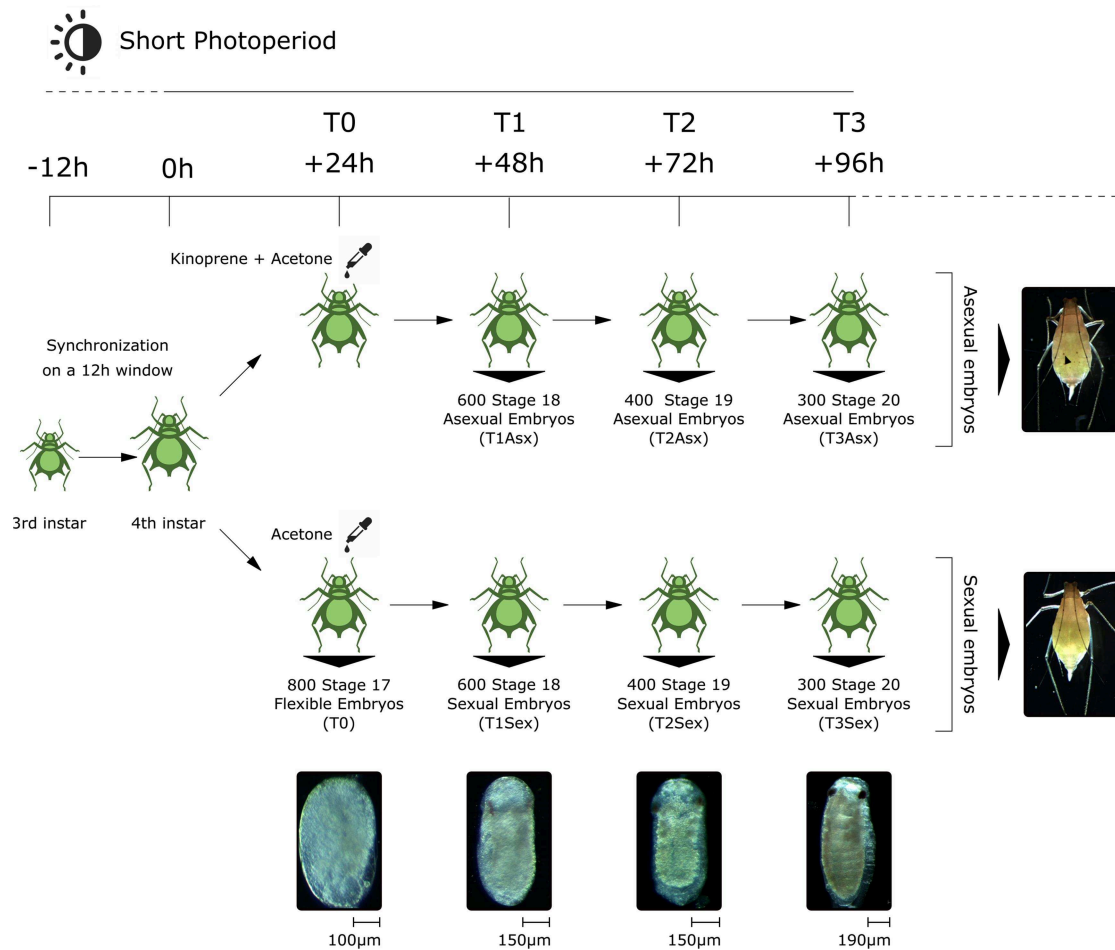
## Conclusion

"n this article we studied the chromatin accessibility regulations and their putative impact on embryonic genetic programs expression during the initiation of the reproductive polyphenism in aphids. We identified the putative involvement of two genes, LOC100164454 and LOC100162729, in the lipid remobilization of the sexual embryos fat body. Such function could be associated with the yolk accumulation involved in the development of the unfertilized eggs inside these embryos. This suggests that yolk accumulation or at least its synthesis might be initiated as soon as an embryo is destined to reproduce sexually, at the stage 18. Moreover, we discovered that the transition from flexible embryos to asexual embryos involve less chromatin reorganizations that in the case of sexual embryos. This suggests that asexual reproduction might be the default genetic program of flexible embryos. Analysis of DNA motifs enrichment and their binding by putative transcription factors in distal intergenic FA"RE peaks with the help of chromosome conformation capture methods is the next step in the understanding of the epigenetic regulation of the genetic programs involved in the reproductive polyphenism in aphids.

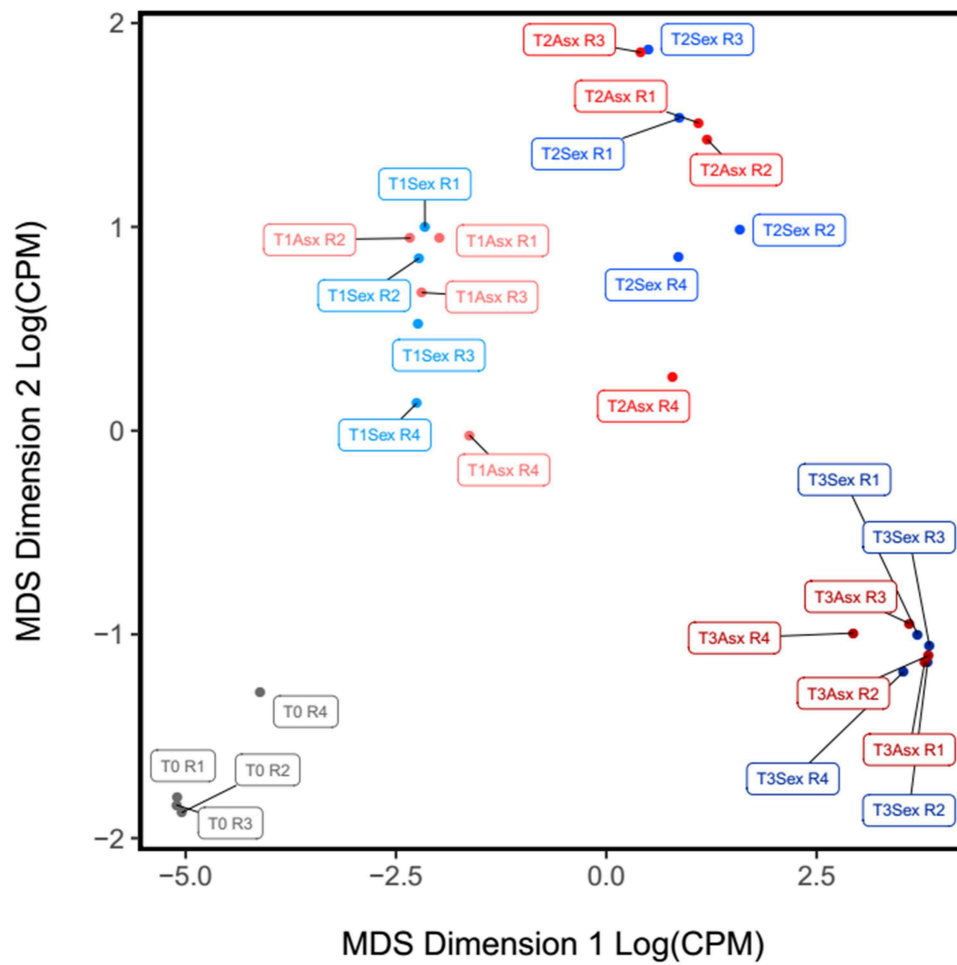
## Acknowledgments

High-throughput sequencing has been performed by the "CGex NGS platform of the "nstitut Curie supported by the grants ANR-10-EQPX-03 (Equipex) and ANR-10-"NBS-09-08 (France G nomique Consortium) from the Agence Nationale de la Recherche ("nvestissements d'Avenir" program), by the Cancropole "le-de-France and by the SiR" C-Curie program - SiR" C Grant "NCA-DGOS- 4654 \*

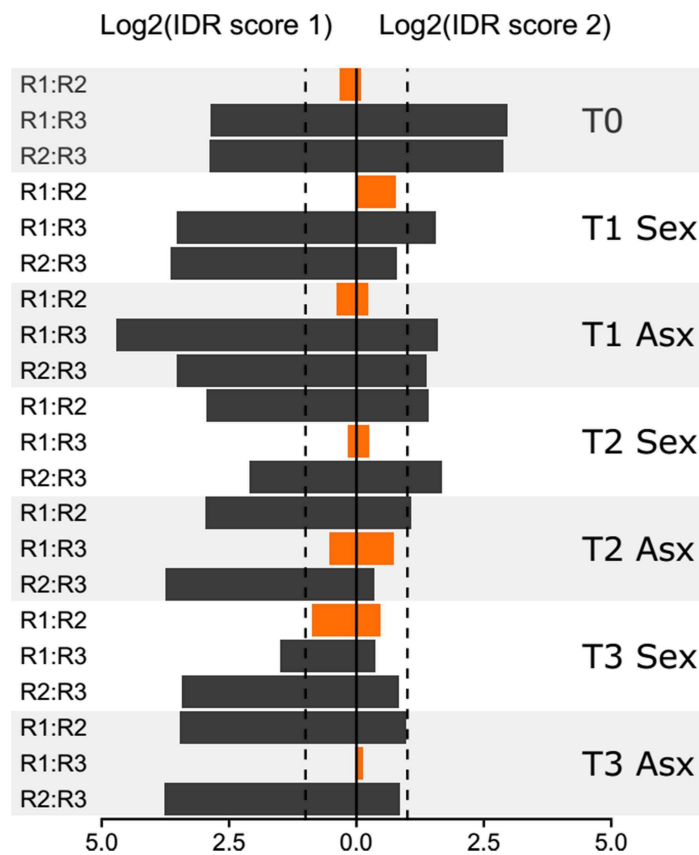
**Figures and Supplementary data**



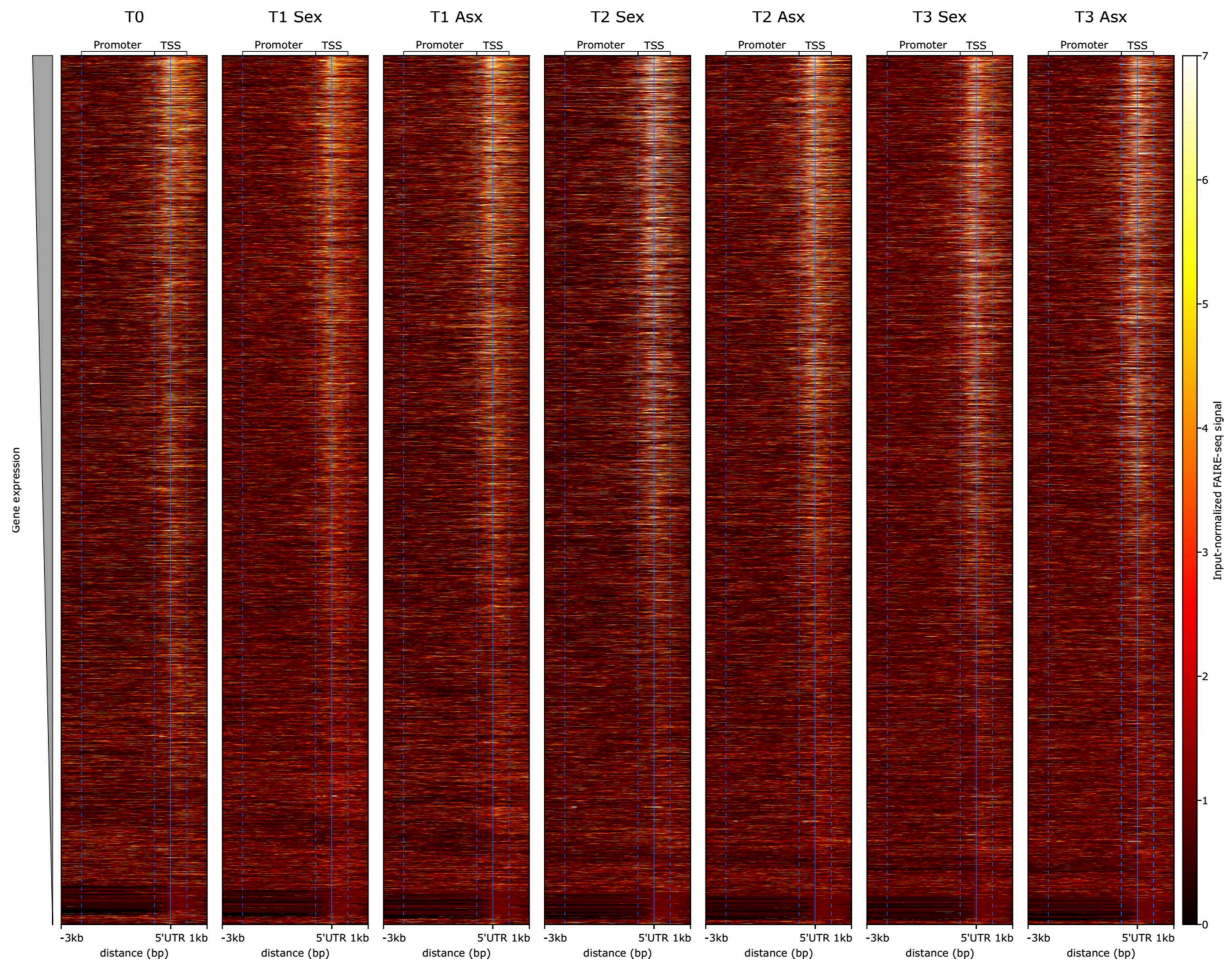
**Figure 1 | Protocol allowing the synchronized production of asexual and sexual embryos under short photoperiod.** As described in methods, the development of sexuparae females and their embryos is synchronized during a 12 hours window. 24 hours after synchronization, sexuparae are treated with acetone or acetone and kinoprene, a juvenile hormone (JH) analog that changes the fate of embryos germ cells to diploid oocytes (asexual reproduction). Embryos of kinoprene treated sexuparae correspond to asexual females displaying diploid oocytes (embryos) once adult, while embryos of acetone treated individuals correspond to sexual females displaying haploid oocytes (eggs) once adult. A portion of acetone treated individuals are dissected for the collection of their 6 most developed embryos, for a total of 800 embryos at stage 17. These embryos constitute the first time point (T0) of both asexual and sexual embryos developmental kinetics. 24 hours after the treatment, kinoprene and acetone treated individuals are dissected for the collection of their 6 most developed embryos during every 24 hours during 3 days, allowing the collection of 600 stage 18 embryos, 400 stage 19 embryos and 300 stage 20 embryos for sexual and asexual kinetics. FA<sup>2</sup>RE-seq has been performed on these embryos, following the number collected indicated for each condition. RNA-seq has been performed on 60 embryos in each condition.



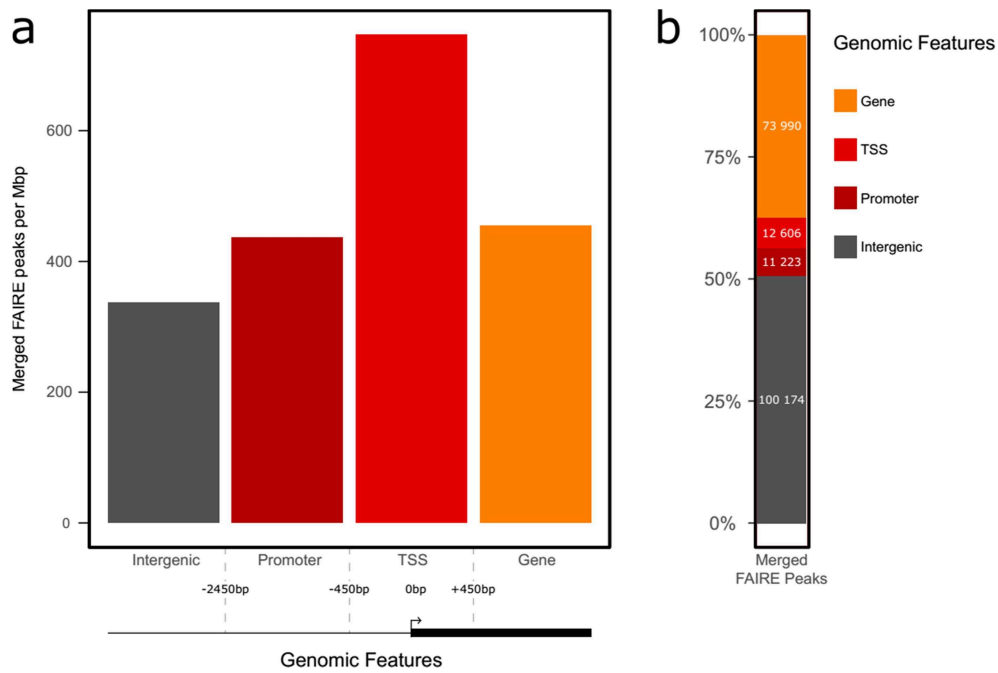
**Figure 2 | RNA-seq biological replicates correlation.** After the calculation of CPM for each RNA-seq libraries (see Methods), replicates have been compared to each other by plotting the Log<sub>2</sub>(CPM) of expressed genes using Multidimensional Scaling (MDS).



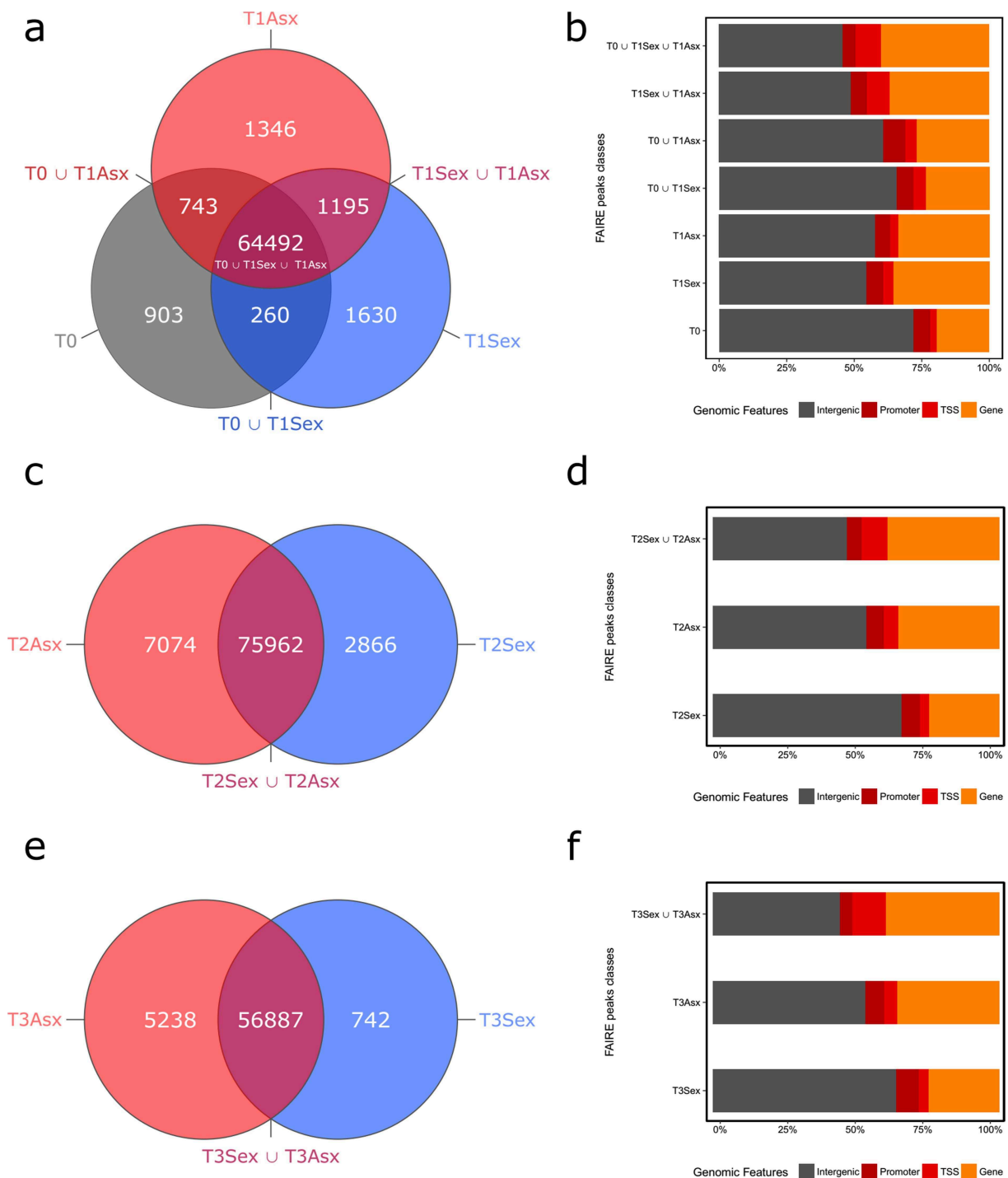
**Figure 3 | Assessment of the reproducibility of FA-RE-seq samples.** Using "reproducible Discovery Rate ("DR), each one of the three biological replicates of a given condition (R1, R2, R3) were compared in a one-versus-one manner on the basis of two different "DR scores: the comparison between the number of FA-RE peaks retained at a certain "DR between original replicates ("DR score 1), and the maximum of the number of FA-RE peaks kept at a given "DR between pooled-pseudoreplicates and original pseudo-replicates ("DR score 2). Samples with high consistency, which have then been pooled for further analyses, are represented in orange, when under the threshold value of 2 suggested by "DR.



**Figure 4 | Definition of TSS and promoter genomic classes by the analysis of open chromatin and gene expression correlation.** The mean input-normalized FAIRE-seq signal have been represented along each gene of *A. pisum* genome using 10bp windows with deepTools [Ram rez *et al.*, 2014, 2016]. Genes have been ranked from the most expressed (top) to the least expressed (bottom) depending on their expression in each condition. The TSS and Promoter genomic classes limit are represented with dashed blue lines, while the putative TSS of each gene is represented by the full blue line in each heatmap. In order to encompass the strong open chromatin signal around the putative TSSs in a genomic class, the *TSS* genomic class was defined as 450bp upstream and downstream the putative TSSs. The *Promoter* genomic class was, however, defined as 2000bp upstream the *TSS* class.



**Figure 5 | FAIRE peaks distribution across genomic features after merging throughout all seven embryos type.** All the peaks retained after the peak-calling were merged into a single peak set comprising 197 993 FAIRE peaks, defined as accessible in some conditions. **a** Distribution of this peak set across the different genomic classes expressed in FAIRE peaks number for one Megabase pair (Mbp), accompanied visually by the definition of each one of these classes: *Intergenic*, *Promoter*, *TSS* and *Gene* (corresponding to all genes \*body, and overlapping with 450bp of the *TSS* class). **b** Proportion of the 197 993 merged peaks across the genomic classes.



**Figure 6 | Number of identified FAIRE peaks by biological and genomic classes.** After merging and assignation to genomic classes, each peak of the unique peak set has been assigned to biologically relevant classes after thorough filtering steps (see Methods). **a, c, e** Number of merged peaks assigned to each biological class for T0 T1 (**a**), T2 (**c**) and T3 (**e**) comparisons between sexual and asexual embryos. **b, d, f** Proportion of assignment to genomic classes (*intergenic*, *promoter*, *TSS* and *gene*) of merged FAIRE peaks for each biological class for the T0 T1 (**b**), T2 (**d**), T3 (**f**) comparisons between sexual and asexual embryos.

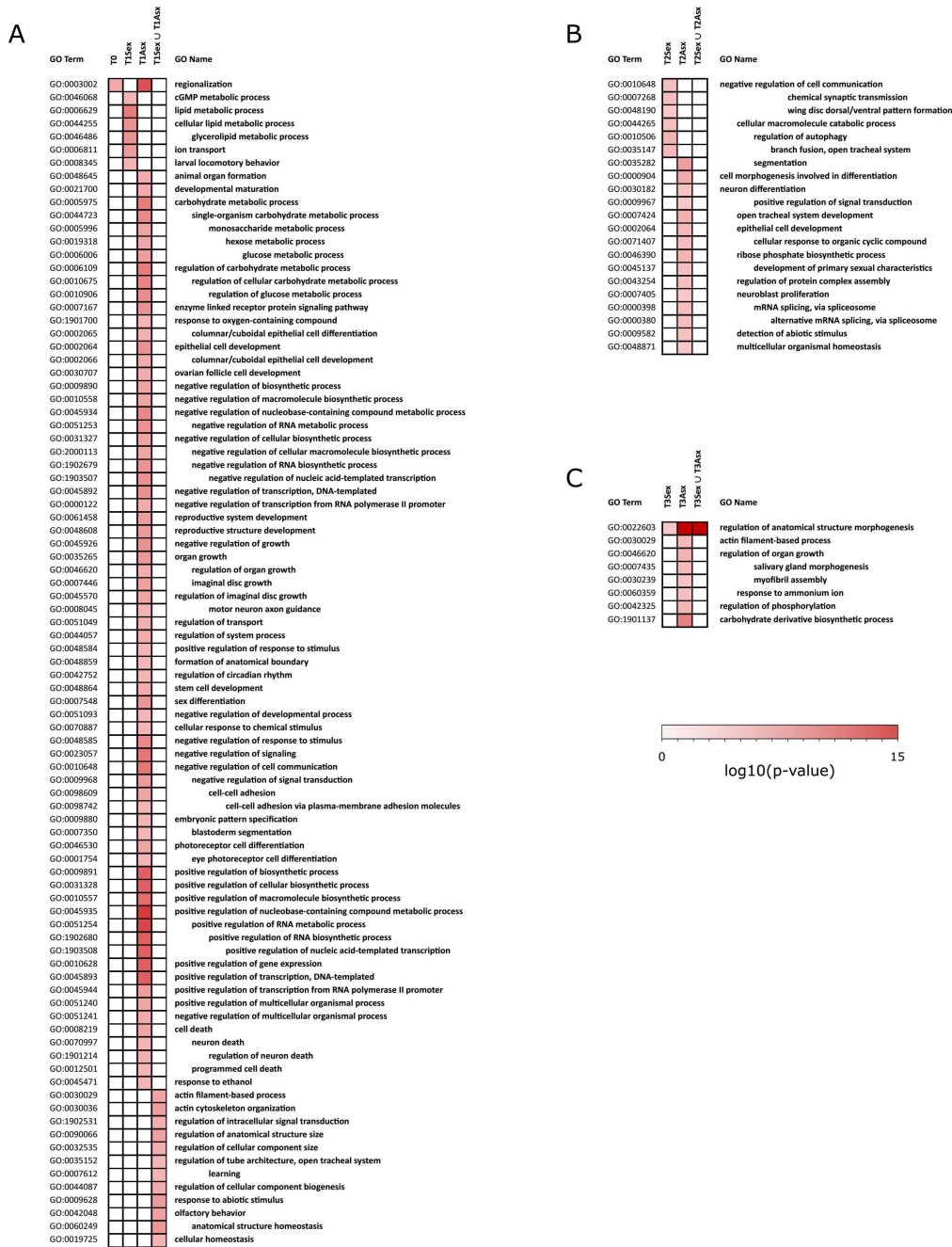
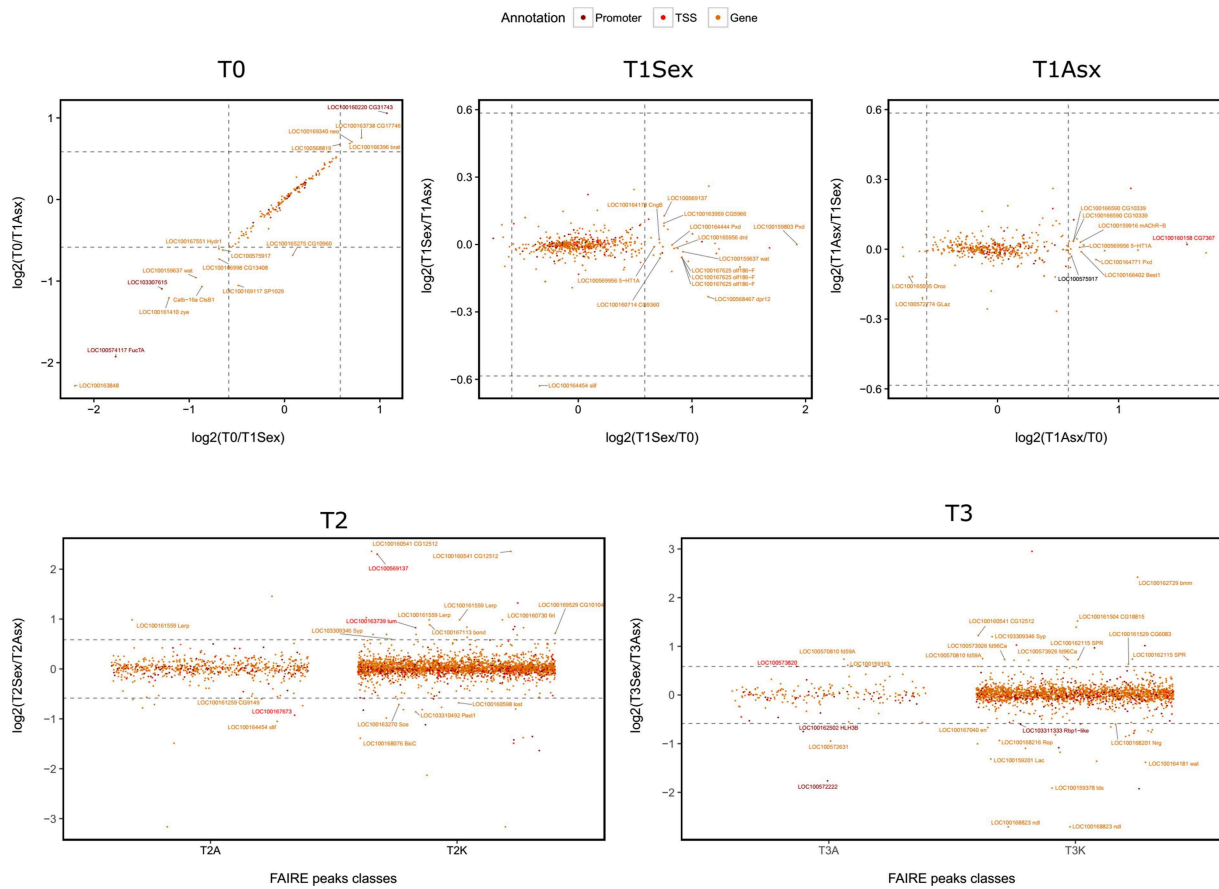
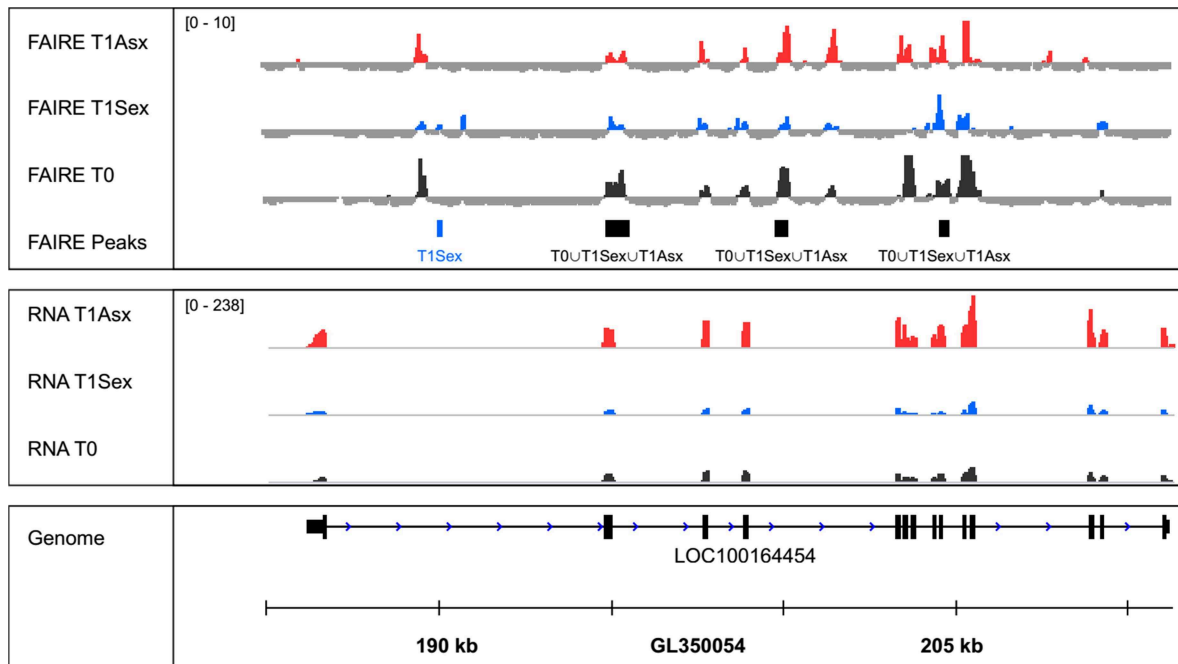


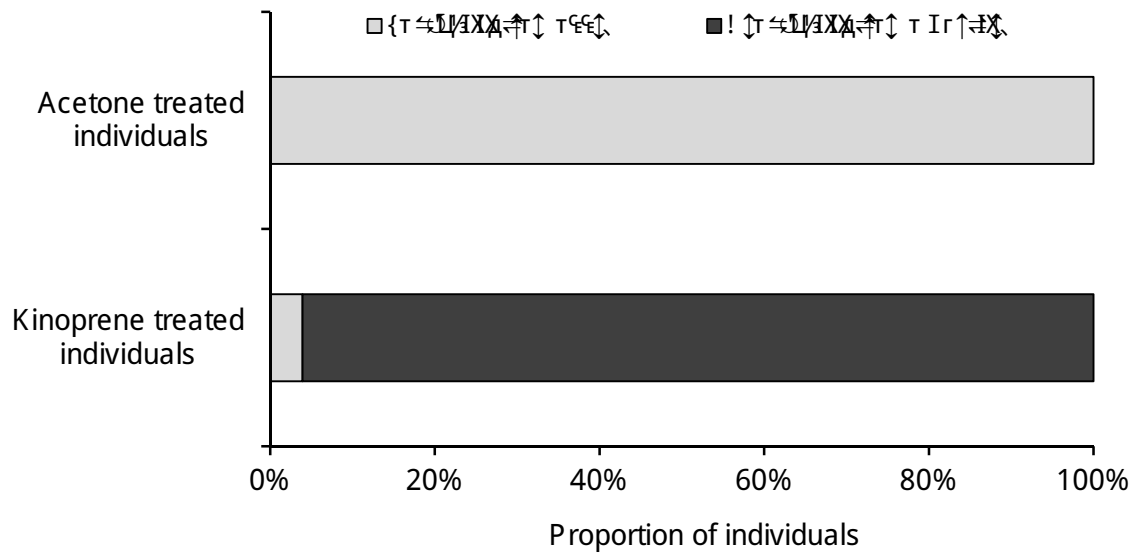
Figure 7 | Gene Ontology analyses of genes displaying condition specific FA<sup>RE</sup> peaks of the T0 T1 (a), T2 (b) and T3 (c) comparisons between sexual and asexual embryos. Heatmaps results are expressed in log<sub>10</sub>(p-value) calculated by gcooa for each GO Term. Only GO Terms displaying a condition specific enrichment are represented.



**Figure 8 | Expression analysis of genes displaying condition specific FAIRE peaks in their promoter (dark red), TSS (red) or body (orange).** Each point represents a FAIRE peak specific of the condition described above the graph, and for which an expressed gene is associated with. Gene expression is represented in  $\log_2(\text{ratio})$  between relevant conditions for each biological class of FAIRE peaks. Gene names and their homologue in *D. melanogaster* is represented for certain differentially expressed genes. Dashed-lines correspond to 0.56 or -0.56 values, which are equivalent to  $\log_2$  values of 1.5 and 0.67 fold-changes, respectively.



**Figure 9 | Genomic view of the LOC100164454 locus, a homologue of the *slimfast* gene in *Drosophila*, by "GV. The differential FA"RE peak putatively associated to the down expression of LOC100164454 is displayed in bleu. FA"RE-seq signal (top) is centered on a value of 3, with all values above 3 displayed with colors depending on the condition like the RNA-seq tracks (T1Asx in red, T1Sex in blue, T0 in black), while values below 3 are represented upside-down in grey.**



Supplementary Data 1 | Reproductive phenotype of the offspring of acetone or kinoprene treated sexuparae reared under short photoperiod.

/IXDQXII	. QX&eQULVwT iXQULVt	a TLI/IXTIX5b!	a TLIQ lW 5b!	Q lW 9
{FET 3u 0k w3	!	3u	w3	3u
{FET 3T 0 3T w3	!	3u	w3	w3
{FET 3T 0 3T w3	!	3u	w3	w3
{FET 3 T 0 3T w3	!	w3	3u	w3
{FET 3 ! 0 3T w3	!	3u	w3	w3
{FET 3u T 0 3T w3	!	u	w3	w3
{FET 3u ! 0 3T w3	!	3u	w3	w3

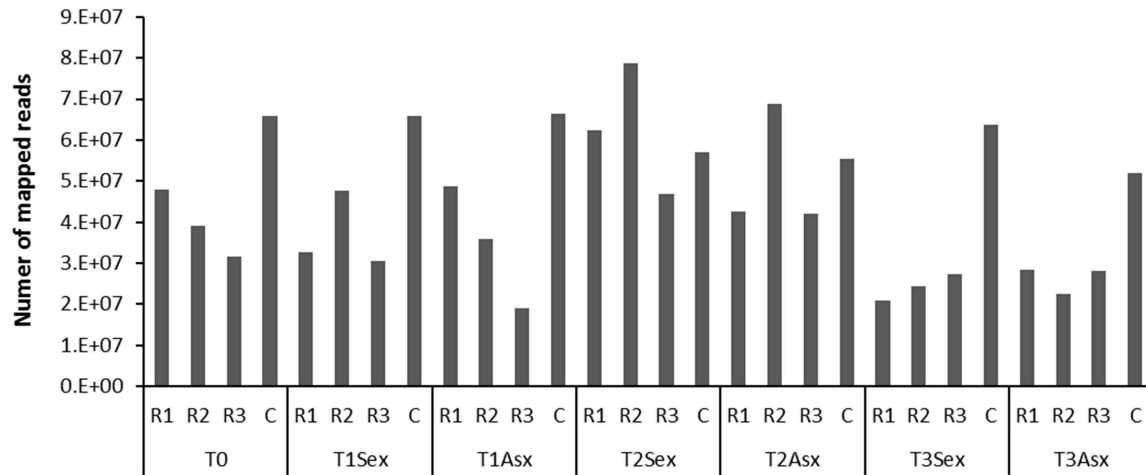
  

/IXDQXII	. QX&eQULVwT iXQULVt	a TLI/IXTIX5b!	a TLIQ lW 5b!	Q lW 9
{FET 3u 0k w3	.	3u	w3	w3
{FET 3T 0 3T w3	.	3u	w3	w3
{FET 3T 0 3T w3	.	3u	w3	w3
{FET 3 T 0 3T w3	/	3u	w3	w3
{FET 3 ! 0 3T w3	/	3u	w3	w3
{FET 3u T 0 3T w3	/	3u	w3	w3
{FET 3u ! 0 3T w3	/	3u	w3	w3

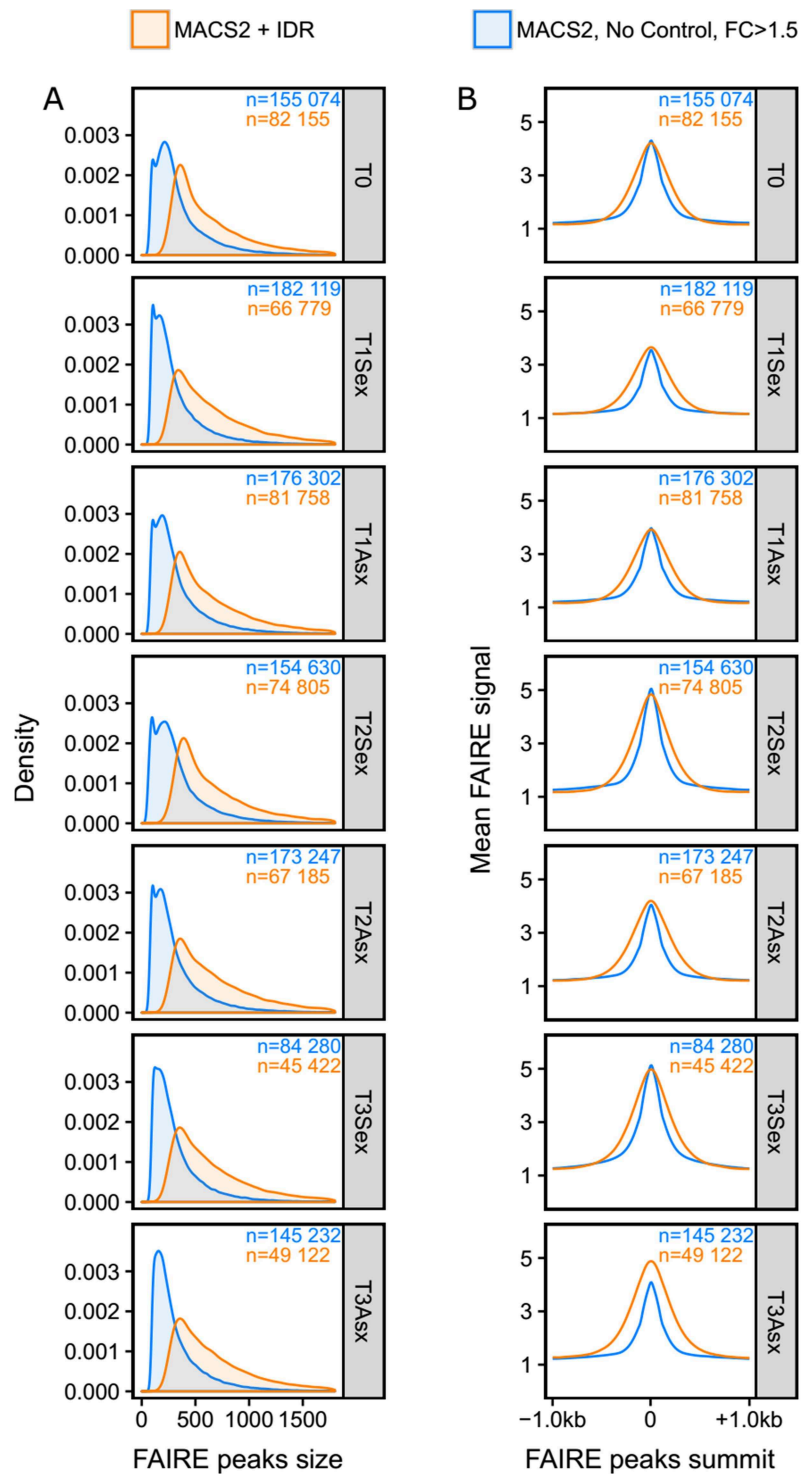
  

/IXDQXII	. QX&eQULVwT iXQULVt	a TLI/IXTIX5b!	a TLIQ lW 5b!	Q lW 9
{FET 3u 0k w3	5	3u	3u	3u
{FET 3T 0 3T w3	5	3u	3u	3u
{FET 3T 0 3T w3	5	3u	3u	3u
{FET 3 T 0 3T w3	9	w3	3u	3u
{FET 3 ! 0 3T w3	9	3u	w3	3u
{FET 3u T 0 3T w3	9	w3	3u	w3
{FET 3u ! 0 3T w3	9	w3	3u	w3

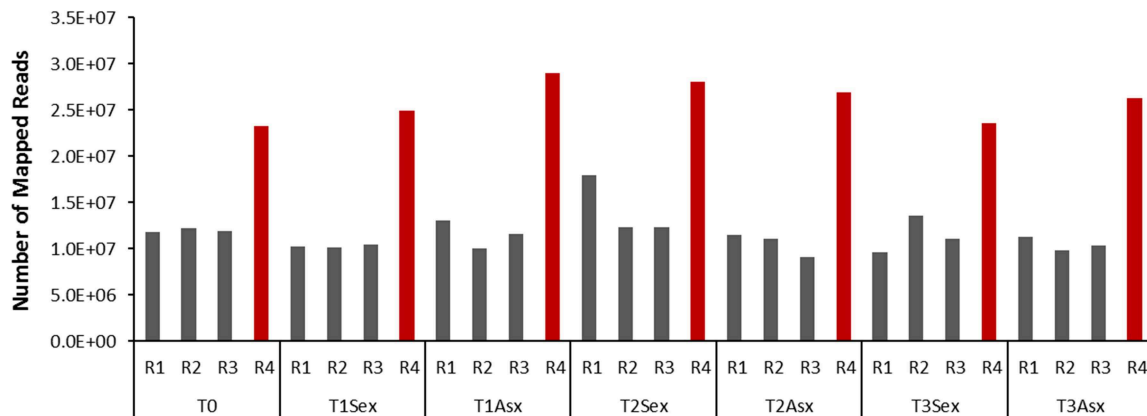
Supplementary Data 2 | FA"RE and Control DNA extraction data sent for sequencing.



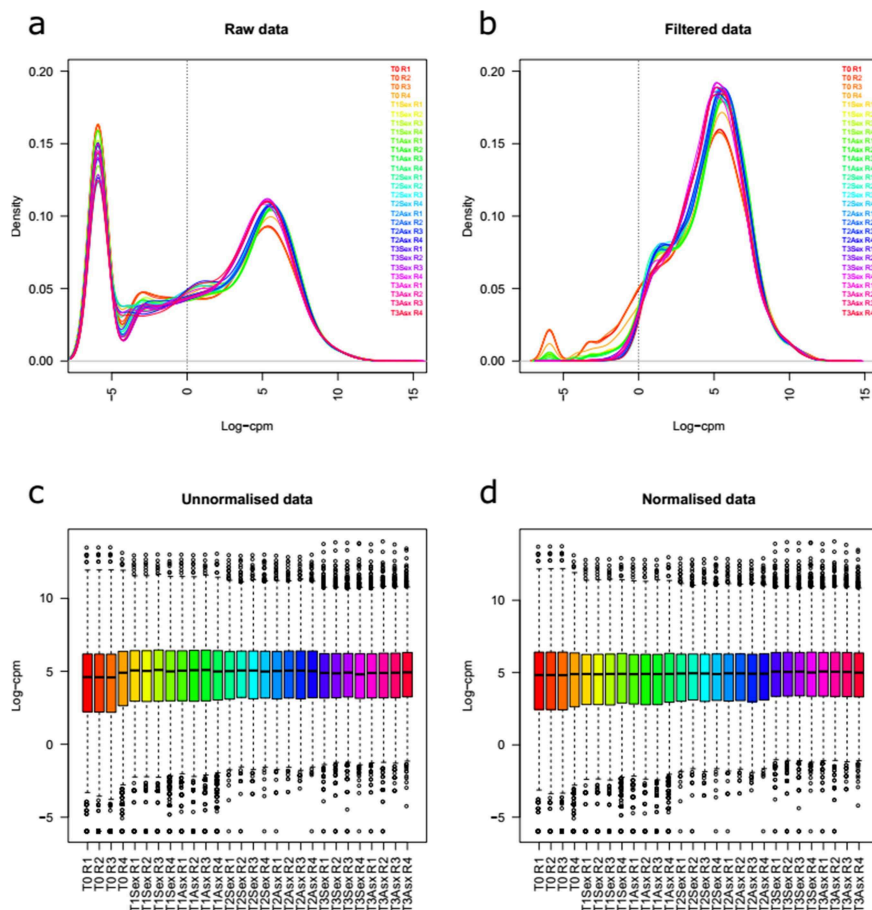
**Supplementary Data 3 | Number of FA"RE-seq mapped reads on the pea aphid genome assembly v2.1 (Aphidbase).** R1, R2 and R3 correspond to FA"RE biological replicates, while C corresponds to an equimolar pool of Control samples extracted concomitantly with the FA"RE samples.



**Supplementary Data 4 | Refinement of FAIRE peaks called with MACS2 and IDR (orange) or MACS2 without control samples followed by a filtering step keeping only peaks showing a ratio of minimum 1.5 between normalized FAIRE counts and Control DNA counts (blue). a** Density plot representing the proportion of FAIRE peaks size. **b** Mean control-normalized FAIRE signal calculated around the summit of FAIRE peaks.



Supplementary Data 5 | Number of RNA-seq mapped reads on the pea aphid genome assembly v2.1 (Aphidbase), with data from W cher et al., unpublished in dark grey, and data sequenced along the FA RE-seq of the present article in red.



Supplementary Data 6 | Quality controls of RNA-seq data analysis. a and b Effect of filtration of unexpressed genes on the density of genes   expression in Log2(CPM). c and d Effect of normalization using EdgeR TMM Method on the distribution of genes   expression in Log2(CPM).

## References

- Bonasio, R. (2014). The role of chromatin and epigenetics in the polyphenisms of ant castes. *Brief. Funct. Genomics* 13, 235–245.
- Corbitt, T.S., and Hardie, J. (1985). Juvenile hormone effects on polymorphism in the pea aphid, *Acyrtosiphon pisum*. *Entomol. Exp. Appl.* 38, 131–135.
- Costechareyre, D., Balmant, S., Condemine, G., and Rahb , Y. (2012). *Dickeya dadantii*, a Plant Pathogenic Bacterium Producing Cyt-Like Entomotoxins, Causes Septicemia in the Pea Aphid *Acyrtosiphon pisum*. *PLoS ONE* 7.
- Dennes, A., Cromme, C., Suresh, K., Kumar, N.S., Eble, J.A., Hahnenkamp, A., and Pohlmann, R. (2005). The Novel *Drosophila* Lysosomal Enzyme Receptor Protein Mediates Lysosomal Sorting in Mammalian Cells and Binds Mammalian and *Drosophila* GGA Adaptors. *J. Biol. Chem.* 280, 12849–12857.
- Duncan, E.J., Leask, M.P., and Dearden, P.K. (2013). The pea aphid (*Acyrtosiphon pisum*) genome encodes two divergent early developmental programs. *Dev. Biol.* 377, 262–274.
- Evans, J.D., and Wheeler, D.E. (1999). Differential gene expression between developing queens and workers in the honey bee, *Apis mellifera*. *Proc. Natl. Acad. Sci.* 96, 5575–5580.
- Fernandes, M., Chanut-Delalande, H., Ferrer, P., Latapie, Y., Waltzer, L., Affolter, M., Payre, F., and Plaza, S. (2010). Zona Pellucida Domain Proteins Remodel the Apical Compartment for Localized Cell Shape Changes. *Dev. Cell* 18, 64–76.
- Foret, S., Kucharski, R., Pellegrini, M., Feng, S., Jacobsen, S.E., Robinson, G.E., and Maleszka, R. (2012). DNA methylation dynamics, metabolic fluxes, gene splicing, and alternative phenotypes in honey bees. *Proc. Natl. Acad. Sci.* 109, 4968–4973.
- Gallot, A., Rispe, C., Leterme, N., Gauthier, J.-P., Jaubert-Possamai, S., and Tagu, D. (2010). Cuticular proteins and seasonal photoperiodism in aphids. *Insect Biochem. Mol. Biol.* 40, 235–240.
- Gallot, A., Shigenobu, S., Hashiyama, T., Jaubert-Possamai, S., and Tagu, D. (2012). Sexual and asexual oogenesis require the expression of unique and shared sets of genes in the insect *Acyrtosiphon pisum*. *BMC Genomics* 13, 76.
- Gronke, S., Mildner, A., Fellert, S., Tennagels, N., Petry, S., Muller, G., Jackle, H., and Kuhnlein, R.P. (2005). Brummer lipase is an evolutionary conserved fat storage regulator in *Drosophila*. *Cell Metab.* 1, 323–330.
- Gronke, S., M ller, G., Hirsch, J., Fellert, S., Andreou, A., Haase, T., Jackle, H., and Kuhnlein, R.P. (2007). Dual lipolytic control of body fat storage and mobilization in *Drosophila*. *PLoS Biol.* 5, e137.
- Gutierrez, E., Wiggins, D., Fielding, B., and Gould, A.P. (2007). Specialized hepatocyte-like cells regulate *Drosophila* lipid metabolism. *Nature* 445, 275–280.
- Huybrechts, J., Bonhomme, J., Minoli, S., Prunier-Leterme, N., Dombrovsky, A., Abdel-Latif, M., Robichon, A., Veenstra, J.A., and Tagu, D. (2010). Neuropeptide and neurohormone precursors in the pea aphid, *Acyrtosiphon pisum*. *Insect Mol. Biol.* 19, 87–95.
- Shikawa, A., Ogawa, K., Gotoh, H., Walsh, T.K., Tagu, D., Brisson, J.A., Rispe, C., Jaubert-Possamai, S., Kanbe, T., Tsubota, T., et al. (2012). Juvenile hormone titre and related gene expression during the change of reproductive modes in the pea aphid. *Insect Mol. Biol.* 21, 49–60.
- Jaqui ry, J., Peccoud, J., Ouisse, T., Legeai, F., Prunier-Leterme, N., Gouin, A., Nouhaud, P., Brisson, J., Bickel, R., Purandare, S., et al. (2017). Disentangling The Causes For Faster-X Evolution in Aphids. *bioRxiv* 125310.
- Kundaje, A. (2012). (2012) ENCODE: TF ChIP-seq peak calling using the irreproducibility Discovery Rate (IRDR) framework - Anshul Kundaje.

Landt, S.G., Marinov, G.K., Kundaje, A., Kheradpour, P., Pauli, F., Batzoglou, S., Bernstein, B.E., Bickel, P., Brown, J.B., Cayting, P., et al. (2012). ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome Res.* 22, 1813–1831.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 10, R25.

Le Trionnaire, G., Jaubert, S., Sabater-Mu oz, B., Benedetto, A., Bonhomme, J., Prunier-Leterme, N., Martinez-Torres, D., Simon, J.-C., and Tagu, D. (2007). Seasonal photoperiodism regulates the expression of cuticular and signalling protein genes in the pea aphid. *Insect Biochem. Mol. Biol.* 37, 1094–1102.

Le Trionnaire, G., Francis, F., Jaubert-Possamai, S., Bonhomme, J., De Pauw, E., Gauthier, J.-P., Haubruge, E., Legeai, F., Prunier-Leterme, N., Simon, J.-C., et al. (2009). Transcriptomic and proteomic analyses of seasonal photoperiodism in the pea aphid. *BMC Genomics* 10, 456.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078–2079.

Li, Q., Brown, J.B., Huang, H., and Bickel, P.J. (2011). Measuring reproducibility of high-throughput experiments. *Ann. Appl. Stat.* 5, 1752–1779.

Luijsterburg, M.S., Acs, K., Ackermann, L., Wiegant, W.W., Bekker-Jensen, S., Larsen, D.H., Khanna, K.K., van Attikum, H., Mailand, N., and Dantuma, N.P. (2012). A new non-catalytic role for ubiquitin ligase RNF8 in unfolding higher-order chromatin structure. *EMBO J.* 31, 2511–2527.

Lynen, F., and Ochoa, S. (1953). Enzymes of fatty acid metabolism. *Biochim. Biophys. Acta* 12, 299–314.

Mayr, E. (1963). *Animal species and evolution*. (Harvard University Press; London: Oxford University Press).

Pilot, F., Philippe, J.M., Lemmers, C., Chauvin, J.P., and Lecuit, T. (2006). Developmental control of nuclear morphogenesis and anchoring by charleston, identified in a functional genomic screen of *Drosophila* cellularisation. *Development* 133, 711–723.

Ram rez, F., D ndar, F., Diehl, S., Gr ning, B.A., and Manke, T. (2014). deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res.* 42, W187–W191.

Ram rez, F., Ryan, D.P., Gr ning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S., D ndar, F., and Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* 44, W160–W165.

Reimand, J., Arak, T., Adler, P., Kolberg, L., Reisberg, S., Peterson, H., and Vilo, J. (2016). g:Profiler – a web server for functional interpretation of gene lists (2016 update). *Nucleic Acids Res.* 44, W83–W89.

Richard, G., Legeai, F., Prunier-Leterme, N., Bretaudeau, A., Tagu, D., Jaqui ry, J., and Le Trionnaire, G. (2017). Dosage compensation and sex-specific epigenetic landscape of the X chromosome in the pea aphid. *Epigenetics Chromatin* 10, 30.

Robinson, J.T., Thorvaldsd ttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative Genomics Viewer. *Nat. Biotechnol.* 29, 24–26.

Schick, S., Fournier, D., Thakurela, S., Sahu, S.K., Garding, A., and Tiwari, V.K. (2015). Dynamics of chromatin accessibility and epigenetic state in response to UV damage. *J Cell Sci* jcs.173633.

Simola, D.F., Ye, C., Mutti, N.S., Dolezal, K., Bonasio, R., Liebig, J., Reinberg, D., and Berger, S.L. (2013). A chromatin link to caste identity in the carpenter ant *Camponotus floridanus*. *Genome Res.* 23, 486–496.

Simon, J.M., Giresi, P.G., Davis, J., and Lieb, J.D. (2012). Using FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) to isolate active regulatory DNA. *Nat. Protoc.* 7, 256–267.

Spannhoff, A., Kim, Y.K., Raynal, N.J.-M., Gharibyan, V., Su, M.-B., Zhou, Y.-Y., Li, J., Castellano, S., Sbardella, G., "ssa, J.-P.J., et al. (2011). Histone deacetylase inhibitor activity in royal jelly might facilitate caste switching in bees. *EMBO Rep.* 12, 238  243.

Tagu, D., Sabater-Mu oz, B., and Simon, J.-C. (2005). Deciphering reproductive polyphenism in aphids. *Invertebr. Reprod. Dev.* 48, 71  80.

Tsuchida, T., Koga, R., Fujiwara, A., and Fukatsu, T. (2014). Phenotypic Effect of  andidatus *Rickettsiella viridis*, La Facultative Symbiont of the Pea Aphid (*Acyrtosiphon pisum*), and "ts "nteraction with a Coexisting Symbiont. *Appl. Environ. Microbiol.* 80, 525  533.

Waddington, C.H. (1942). Canalization of Development and the "nheritance of Acquired Characters. *Nature* 150, 563  565.

Waki, H., Nakamura, M., Yamauchi, T., Wakabayashi, K., Yu, J., Hirose-Yotsuya, L., Take, K., Sun, W., "wabu, M., Okada-"wabu, M., et al. (2011). Global Mapping of Cell Type  pecific Open Chromatin by FA"RE-seq Reveals the Regulatory Role of the NF" Family in Adipocyte Differentiation. *PLOS Genet.* 7, e1002311.

Walsh, T.K., Brisson, J.A., Robertson, H.M., Gordon, K., Jaubert-Possamai, S., Tagu, D., and Edwards, O.R. (2010). A functional DNA methylation system in the pea aphid, *Acyrtosiphon pisum*. *Insect Mol. Biol.* 19, 215  228.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoutte, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., et al. (2008). Model-based analysis of Ch"IP-Seq (MACS). *Genome Biol.* 9, R137.

## C. R sultats compl mentaires

### 1. RNA-seq : un potentiel manque de profondeur de s quen age par rapport aux donn es FA"RE-seq

Dans cette  tude, les donn es RNA-seq et FA"RE-seq ont  t  compar es, mais chaque exp rience implique une grande diff rence de profondeur de s quen age. En effet, les donn es RNA-seq comportent en moyenne  $15.1 \pm 6.7$  millions de lectures mapp es sur le g nome du puceron du pois tandis que les exp riences FA"RE en pr sentent en moyenne  $39.4 \pm 16.0$  millions et les  chantillons Contr le  $61.0 \pm 6.0$  millions (donn es suppl mentaires 3 et 5 de l'article, intervalle de confiance exprim  en  cart type).

Lors de la comparaison des donn es FA"RE-seq et RNA-seq, nous avons trouv  de nombreux pics FA"RE sp cifiques des conditions  tudi es au sein de g nes non-diff rentiellement exprim s. Ainsi, peu de corr lations entre expression et ouverture diff rentielles ont  t  identifi s, particuli rement lors des comparaisons entre embryons de stade 18 asexu s et sexu s, o  seulement une seule corr lation est caract ris e (LOC100164454). Cela pourrait s'expliquer par un manque de profondeur de s quen age des ARNm par rapport aux ADN FA"RE et Contr le, permettant une identification plus pr cise des r gions chromatiniennes putativement r gulatrices comparativement   l'identification de g nes diff rentiellement exprim s. Nous avons donc choisi de ne pas effectuer d'analyse d'expression diff rentielle sur ces donn es RNA-seq afin de ne pas limiter l'analyse des donn es FA"RE-seq par manque potentiel de couverture des donn es d'expression. La production de nouvelles donn es RNA-seq avec un s quen age plus profond pourrait corriger ce probl me. De mani re alternative, il serait possible d'analyser le transcriptome non pas des embryons sexu s ou asexu s, mais directement celui des cellules germinales et ovocytes (haplo des ou diplo des). Cependant, un tel projet n cessiterait de mettre au point un protocole de microdissection laser pour extraire les ARNm de mani re sp cifique de ces tissus. Plus simplement, il serait possible d'analyser le transcriptome de cellules germinales de femelles sexu es et asexu es adultes, bien que ce type d'approche soit plus biais  qu'une microdissection des tissus embryonnaires correspondant. Ces nouvelles donn es permettraient sans doute de mieux corr ler ouverture et expression diff rentielle, dans le but de caract riser de fa on plus exhaustive les r gulations transcriptionnelles mises en jeu entre embryons sexu s et asexu s.

### 2. Identification de sites de fixation de facteurs de transcription

La m thode FA"RE-seq permet l'identification de potentiels  l ments r gulateurs de la transcription, comme des motifs enhancers, silencer ou insulator, par la recherche de l'enrichissement de motifs ADN au niveau des pics FA"RE identifi s. Des analyses pr liminaires d'enrichissement de motifs d'ADN au niveau des pics diff rentiels d'ouverture de la chromatine entre embryons sexu s et asexu s ont donc  t  initi es. Les motifs identifi s ont  t  compar s   une base de donn es de facteurs de transcription (FT) de la Drosophile, comprenant des homologues prot iques du puceron du pois exprim s au niveau des embryons sexu s et asexu s. Les enrichissements de motifs et leur

comparaison avec les motifs reconnus par des FT de la *Drosophila* ont  t  men s   l'aide de MEME-ChIP (param tres par d faut) [Bailey *et al.*, 2009 ; Machanick et Bailey, 2011], tandis que la constitution de la banque de FT conserv s entre *A. pisum* et *D. melanogaster* a  t  r alis e par BLASTp+ [States et Gish, 1994 ; Camacho *et al.*, 2009] en ne conservant que des homologues avec une E-value maximale 10 et une p-value maximale de 1e-20 (Figure 28).

Cette analyse nous permet, par l'interm diaire de l'annotation du g nome de *D. melanogaster*, de caract riser des  v nements de fixation de facteurs de transcription putatifs au sein de r gions de la chromatine ouverte de mani re sp cifique au sein des embryons  tudi s.

Certains facteurs de transcription sont identifi s comme se fixant au niveau de pics FA'RE dans plusieurs types d'embryons, comme c'est le cas de l'homologue de datilografo (DAT), malgr  la sp cificit  des pics FA'RE analys s. Les fonctions de ces facteurs de transcription sont sans doute g n ralistes, comme par exemple odd paired (OPA, homologue de LOC100168291), qui est impliqu  dans l'embryog nese chez la *Drosophila*. Certains facteurs de transcription pr sentent des profils complexes de fixation putative au niveau des motifs d'ADN ouvert enrichis, comme par exemple KLU, DAR1 et SR, identifi s au niveau des embryons T0, T1Asx et T2Sex. Ainsi, la fa on la plus logique d'analyser ces donn es est de s'int resser aux facteurs de transcription reconnaissant des motifs enrichis au niveau de la chromatine ouverte d'un seul type d'embryon.

Seul MED est capable de reconnaitre sp cifiquement les motifs d'ADN enrichis au niveau des pics FA'RE sp cifiques des embryons de stade 17 (T0). MED ayant de nombreux r les dans le d veloppement impliquant le patterning dorso-ventral, le patterning et la prolif ration des disques alaires et la r gulation de l'expression de g nes dans le corps p doncul  du cerveau des larves, il est difficile de d terminer un r le putatif de son homologue LOC100158948, chez *A. pisum* dans le cadre de la mise en place du polyph nisme de reproduction.

<i>D. melanogaster</i> TF	<i>A. pisum</i> homolog	BLASTp p-value	TF DNA Binding Motif	Complete name	TF Motif Enrichment							
					T0	T1Sex	T1Asx	T2Sex	T2Asx	T3Sex	T3Asx	
DATI	LOC100164834	9.10E-167	TTTTTTTTTATTTT	datilografo, isoform D								
OPA	LOC100168291	6.76E-100	GGGGCCATGGGCCCCCGCCCCCCCC	odd paired								
KLU	LOC100163683	3.75E-85	GGGGCCATGGGCCCCCGCCCCCCCC	klumpfuss, isoform A								
DAR1	LOC100159231	2.10E-53	GGGGCCATGGGCCCCCGCCCCCCCC	dendritic arbor reduction 1								
SR	LOC100159519	5.34E-111	GGGGGGGGGGGATG	stripe, isoform C								
SP1	LOC100167763	4.55E-111	GCCCCCCC	Sp1, isoform B								
MED	LOC100158948	3.57E-122	CGCCCGC	medea, isoform B								
BAB1	LOC100163259	1.01E-95	TATTATT	bric a brac 1, isoform B								
BLIMP-1	LOC100159482	3.26E-82	TATCTGTCCGATGGGAAAGTAATCTAGT	Blimp-1, isoform A								
CI	LOC100575006	3.70E-104	CACACAC	cubitus interruptus, isoform B								
TWI	LOC100160123	7.50E-30	CATAGGCG	twist, isoform A								
AB	LOC100165314	1.91E-70	CAGTGCGTATTAGGGGGGGGGGGGG	abrupt, isoform D								
PDM3	LOC100569279	1.82E-36	CAGTGGCGTATTAGGGGGGGGGGGGG	pou domain motif 3, isoform A								
PAD	LOC100165878	1.07E-32	CCCCTCCC	poils au dos								
CWO	LOC100570459	1.28E-29	GGTGACCA	clockwork orange, isoform A								
BTB VII	LOC100167015	3.72E-50	TATATATA	BTB-protein-VII, isoform E								
KAH	LOC100159857	1.23E-76	CCACC	kahuli								
GL	LOC100161008	8.53E-33	CACTGTGA	glass, isoform A								
SCRT	LOC100571267	1.35E-97	AAAAGGTG	scratch, isoform A								
CG12605	LOC100574120	4.68E-96	AAAAGGTG	uncharacterized protein Dmel_CG12605, isoform C								
ODD	LOC100568506	5.25E-45	ACAGTAG	odd skipped								
EIP78C	LOC100161053	2.49E-94	GGGGGTCA	Ecdysone-induced protein 78C, isoform B								
EIP75B	LOC100166941	1.25E-51	GGGGGTCA	Ecdysone-induced protein 75B, isoform C								
HNF4	LOC100168732	4.38E-171	GGGGGTCA	hepatocyte nuclear factor 4, isoform E								
DA	LOC100162653	6.55E-79	ACACAC	daughterless, isoform A								

**Figure 28 | Enrichissement de sites de fixation de facteurs de transcription au sein des pics FA"RE sp cifiques de chacun de types d'embryon  tudi s.** Sont d crits dans cette figure, les facteurs de transcription (TF dans la figure) du g nome de la Drosophile ayant des homologues chez le puceron du pois, qui sont exprim s au niveau des embryons. Ces facteurs de transcription pourraient  ventuellement reconna tre des motifs d'ADN enrichis au niveau des pics FA"RE sp cifiques des embryons de stade 17 (T0), 18 (T1), 19 (T2) et 20 (T3)   d veloppement sexu  (Sex) ou asexu  (Asx). Pour chaque FT ainsi identi f , sont report s le nom des homologues d'*A. pisum*, la p-value de l'analyse BLAST prot ique (niveau de conservation), le motif reconnu par chaque FT dans le g nome de la Drosophile, leur nom complet, ainsi que la pr sence (vert) ou absence (blanc) d'un enrichissement du motif ADN sur lequel ces FT peuvent se fixer au sein des divers types d'embryons.

Concernant les facteurs de transcription reconnaissant des r gions ouvertes au sein de la chromatine des embryons asexu s, sont identi f s de fa on notoire des homologues de KAH et PDM3, deux facteurs de transcription en relation avec le syst me nerveux ; ainsi que DA, une prot ine bHLH de classe " [Gaudet *et al.*, 2010] r gulant notamment de mani re critique la prolif ration des cellules somatiques [Smith *et al.*, 2002]. L'homologue de DA chez le puceron du pois pourrait ainsi jouer un r le dans le d veloppement des ovocytes diplo des en embryons au sein des embryons de stade 20, chez qui des embryons, majoritairement constitu s de cellules somatiques, peuvent d j   atteindre le 7 me stade embryonnaire [Miura *et al.*, 2003]. En revanche, il est surprenant de ne voir ressortir aucun facteur de transcription r pondant   l'hormone juv nile chez les embryons asexu s. En effet, d'apr s les  tudes ayant conduit   la mise en place du protocole ici utilis , l'ajout d'hormone juv nile est d crite comme l' l ment d clenchant le basculement vers cette reproduction asexu e [Corbitt et Hardie, 1985]. Or, aucun facteur de transcription ici d tect  ne semble supporter cette hypoth se.

Concernant la reconnaissance de pics FA"RE sp cifiques des embryons sexu s par des facteurs de transcription, nous avons identifi  des sites putatifs de fixation des facteurs de transcription BL"MP-1, E"P78C et E"P75B, et HNF4. Ce dernier correspond chez la Drosophile   un facteur de transcription r gulant la mobilisation des lipides et la b ta-oxydation des acides gras pendant les stades larvaires, corr lant donc avec la remobilisation des lipides d crite chez les embryons sexu s au sein de l'article par des analyses corr latives d'ouverture et d'expression diff rentielles de g nes. BL"MP-1, E"P78C et E"P75B sont quant   eux des facteurs de transcription inductibles par l' cdysone chez *D. melanogaster*.

BL"MP-1 est un facteur de transcription ayant des r les importants dans la m tamorphose chez la Drosophile, notamment en  tant impliqu  dans le timing de la pupation par la r gulation de la synth se de 20-hydroxyecdysone. En effet, BL"MP-1 inhibe la transcription de *XFtz-f1* au cours du stade pr -pupal. Cette inhibition est lev e lors du passage au stade pupal, permettant alors l'augmentation de la transcription de *shade* qui est impliqu  dans la conversion de l' cdysone en 20-hydroxyecdysone r gulant la transition d veloppementale par l'induction de l'expression de certains g nes [Akagi *et al.*, 2016]. Ainsi BL"MP-1 est impliqu  dans le timing d'une transition d veloppementale par la r gulation de la voie de synth se du 20-hydroxyecdysone ; or des sites de fixations putatifs pour ce FT sont retrouv s de mani re sp cifique dans la chromatine ouverte des embryons sexu s de stade 18, sugg rant ainsi l'implication potentielle de LOC100159482, son homologue chez *A. pisum*, dans la transition entre un embryon flexible de stade 17 et un embryon sexu  de stade 18 via l' cdysone. E"P78C est un r cepteur hormonal nucl aire   l' cdysone associ    E"P75C, notamment impliqu  dans l'ovog n se. Chez *D. melanogaster*, E"P78C est exprim  dans toute la lign e germinale pendant l'ovog n se, et est important pour la production d' ufs et pour le contr le maternel de l'embryog n se pr coce. Par des mutants de type perte de fonction, il a notamment  t  montr  qu'E"P78C favorise la survie des follicules en d veloppement par son activit  au niveau des cellules germinales [Ables *et al.*, 2015]. E"P75B r pond   l' cdysone et est impliqu  dans le choix du d veloppement ou de la d g n rescence des chambres    ufs par l'induction ou la suppression de l'apoptose au sein des larves de stades 8 et 9, en fonction des conditions de nutrition [Terashima *et Bownes*, 2006]. BL"MP-1 et les prot ines E"P ne reconnaissant pas les m mes motifs, il semble que nous ayons d tect  ici une signature forte de facteurs de transcription r pondant   l' cdysone et  tant associ s   l'ovog n se ou   une transition d veloppementale dans le processus du polyph nisme de reproduction chez le puceron du pois. Ainsi, ces r sultats sugg rent que la transition d'un embryon de stade 17 dont le devenir est flexible, vers un embryon de stade 18 sexu  serait r gul e par une voie impliquant l' cdysone, qui serait per ue par des facteurs de transcription d clenchant la diff renciation des cellules germinales des embryons en ovocytes haplo des ( ufs non-fertilis s).

De ces r sultats, bien qu'issus d'analyses pr liminaires, ressortent des conclusions int ressantes, sugg rant d'une part l'absence de l'implication de facteurs de transcription r pondant   l'hormone juv nile chez les embryons asexu s, et d'autre part l'implication de facteurs de transcription r pondant   l' cdysone chez les embryons sexu s, et dont les fonctions sont li es notamment  

Le vovogŕnŕse, bien qu'aucune implication de laecdysone dans le polyphŕnisme de reproduction des pucerons n'a ŕtŕ montrŕ. La signalisation par l'hormone juvŕnile pourrait cependant ŕtre post-traductionnelle, donc non-dŕtectŕe par l'analyse de rŕgulations transcriptionnelles.

## V". Discussion g n rale

Au cours des deux chapitres pr c dents, nous avons pr sent  les r sultats majeurs de ma th se qui vont   pr sent  tre discut s. Dans un premier temps nous rappellerons les avantages et inconv nients du FA"RE-seq   qui constitue le socle m thodologique de cette th se   au regard d'autres m thodes d'analyse de la chromatine ouverte tout en d taillant et discutant les mises au point bio-informatiques qui ont  t  r alis es. Dans un second temps nous discuterons des limites de l'assemblage et de l'annotation du g nome du puceron du pois ainsi que des parti pris qui nous ont permis de nous affranchir de ces limites. Enfin, nous rappellerons les r sultats majeurs obtenus au cours de ma th se, afin de les discuter d'un point de vue int gratif et vis- -vis des donn es bibliographiques existantes tout en envisageant de nouvelles pistes d' tude qui permettraient d'aller plus loin dans la compr hension de l'implication des m canismes  pig n tiques dans la r gulation du mode de reproduction chez les pucerons.

### A. La m thode FA"RE-seq : avantages, inconv nients et limites

#### 1. Choix de la m thode FA"RE-seq vis- -vis du DNase-seq et de l'ATAC-seq

Plusieurs m thodes ont  t  d velopp es   ce jour pour permettre l'identification des r gions accessibles de la chromatine   l' chelle des g nomes, et le <sup>2</sup> Formaldehyde Assisted "solation of Regulatory Elements P suivi d'un s quen age haut-d bit (FA"RE-seq) [Gaulton *et al.*, 2010] utilis  dans cette th se, constitue l'une de ces m thodes. Le s quen age des sites hypersensibles   la DNase " (DNase-seq) [Boyle *et al.*, 2008a] ainsi que <sup>2</sup> l'Assay for Transposase-Accessible Chromatin with highthroughput sequencing P (ATAC-seq) [Buenrostro *et al.*, 2013a] sont deux m thodes alternatives au FA"RE-seq, tandis que le <sup>2</sup> Micrococcal nuclease assisted isolation of nucleosomes P suivi d'un s quen age haut-d bit (MA"NE-seq) [Schones *et al.*, 2008] est une m thode compl mentaire aux trois pr c dentes, puisqu'elle permet l' tude de la position des nucl osomes (ADN non-accessible donc) via une digestion de la chromatine par des nucl ases micrococcales.

L'extraction d'ADN ouvert par le protocole de DNase-seq est bas e sur la digestion de la chromatine par la DNase ", une endonucl ase clivant l'ADN accessible mais laissant intacte l'ADN ferm  associ e   des prot ines et notamment les histones. L'ADN dig r e par la DNase " correspondant   des fragments plus courts que l'ADN intact, une s lection par taille permet de s parer la chromatine ferm e de la chromatine ouverte, puis de proc der   l'extraction et au s quen age haut-d bit de cette derni re [Song et Crawford, 2010]. L'ATAC-seq est une m thode bas e sur la forte affinit  des transposons pour les r gions ouvertes de la chromatine au sein desquelles ils peuvent s'ins rer plus facilement qu'au niveau des r gions ferm es. Ainsi, l'ATAC-seq utilise une transposase Tn5 mut e et hyperactive. La haute activit  de cette transposase permet un clivage efficace de l'ADN libre,

coordonn e   une ligation simultan e de s quences adaptateurs. Les fragments d'ADN portant ces adaptateurs sont ensuite isol s, amplifi s par PCR et utilis s pour du s quenage haut-d bit [Buenrostro *et al.*, 2013a]. Le FA"RE-seq correspond quant   lui   un protocole permettant l'extraction des r gions appauvries en nucl osomes (NDR) gr ce   une fixation au formald hyde de la chromatine suivie de sa fragmentation par sonication et d'une extraction au ph nol-chloroforme, permettant de dissocier ADN ouvert et ADN associ  aux nucl osomes et ainsi d'extraire puis s quencer les NDRs [Simon *et al.*, 2012]. Ces trois m thodes permettent un enrichissement d'ADN ouvert au niveau des TSS qui est corr l  positivement avec l'expression des g nes   les g nes les plus exprim s ayant un TSS plus accessible   ce qui permet de v rifier le succ s des exp rimentations r alis es.

Parmi ces trois m thodes, nous avons fait le choix d s 2012 de la mise au point du FA"RE-seq pour diverses raisons qu'il convient de pr senter. Tout d'abord, d'un point de vue historique le FA"RE-seq, d velopp  pour la premi re fois en 2010 est le successeur du DNase-seq d velopp  en 2008, et permet d'extraire les r gions ouvertes de la chromatine sans r aliser les  tapes fastidieuses de titration d'enzymes, une des  tapes limitante du DNase-seq [Tsompana et Buck, 2014]. De plus, le FA"RE-seq n cessite un nombre de cellules moindre vis- -vis du DNase-seq,   rendement final  quivalent (Figure 29). En outre, nous  tions int ress s par l' tude de la chromatine ouverte extraite   partir de tissus potentiellement congel s dont l'acquisition est d licate, comme cela a  t  le cas des nombreux embryons collect s par vivisection d'individus sexupares dans le cadre du dispositif exp rimental d crit dans le Chapitre ". Le protocole du FA"RE-seq d crit en 2012 [Simon *et al.*, 2012] a notamment  t  d velopp  pour des tissus congel s, contrairement au DNase-seq pour qui le premier protocole pour tissus congel s a  t  publi    la fin de l'ann e 2013 [Ling et Waxman, 2013]. Ainsi, au d but de l'ann e 2013 lors de la r flexion men e par l' quipe autour de la mise au point d'une m thode d' tude de la chromatine ouverte, le FA"RE-seq apparaissait comme une m thode innovante, relativement simple   mettre en place, et adapt e   des tissus vivants ou congel s. LATAC-seq, dont le protocole a  t  publi  en 2013, n cessite bien moins de mat riel biologique de d part pour un rendement final  quivalent au DNase-seq et FA"RE-seq (Figure 29), tout en permettant une  tude  largie d'une grande diversit  d' l ments r gulateurs au sein du g nome (discut  par la suite) comparativement aux deux autres m thodes. Ce sont pour ces raisons que LATAC-seq remplace progressivement le FA"RE-seq et le DNase-seq, dont les utilisations d croissent sensiblement   partir de l'ann e 2017 au profit de LATAC-seq (Figure 30). LATAC-seq aurait ainsi constitu  une m thode de choix pour r pondre aux questions biologiques abord es dans cette th se. Mais son protocole n'ayant  t  d crit que mi-2013, uniquement pour des cultures cellulaires (  tissus vivants  ), et au vue du manque de recul sur la m thode ATAC au moment de la r flexion de l' quipe pour d terminer le protocole d' tude de la chromatine ouverte   mettre au point, le FA"RE-seq constituait un choix logique.

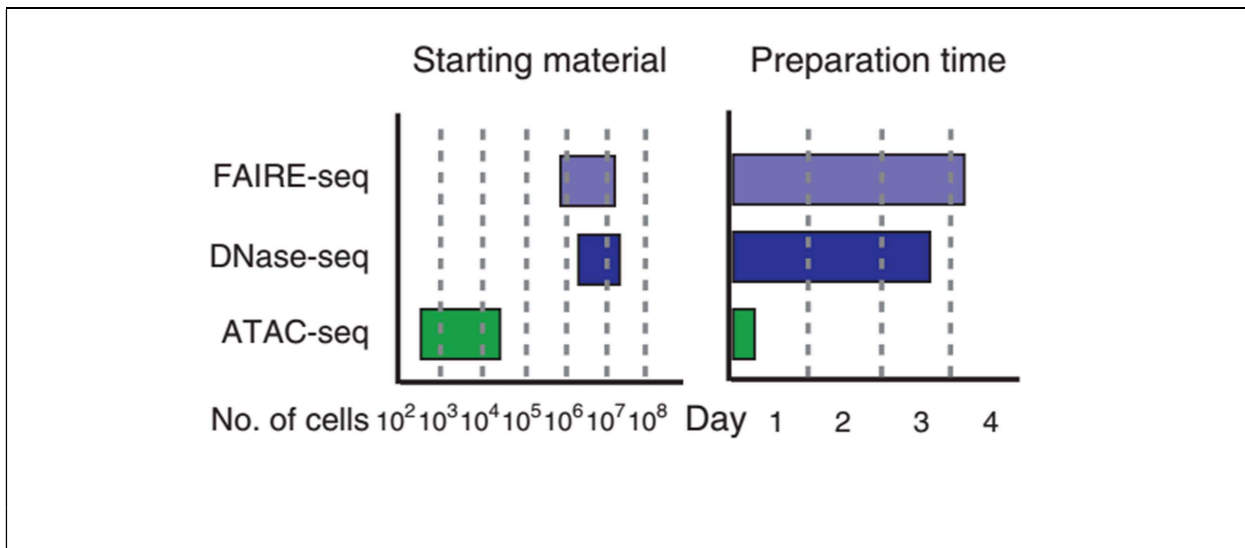


Figure 29 | Comparaison des protocoles FA"RE-seq, DNase-seq et ATAC-seq. Ń rendements Ńgaux, FA"RE-seq et DNase-seq demandent environ la mŃme quantitŃ de matŃriel biologique (exprimŃ ici en nombre de cellules) ainsi qu'environ la mŃme durŃe de prŃparation des Ńchantillons d'ADN. L'ATAC-seq tend Ń remplacer ces deux mŃthodes car elle demande moins de matŃriel biologique de dŃpart, et prŃsente l'avantage d'Ńtre environ 4 fois plus rapide Ń rŃaliser. D'aprŃs [Buenrostro *et al.*, 2013a].

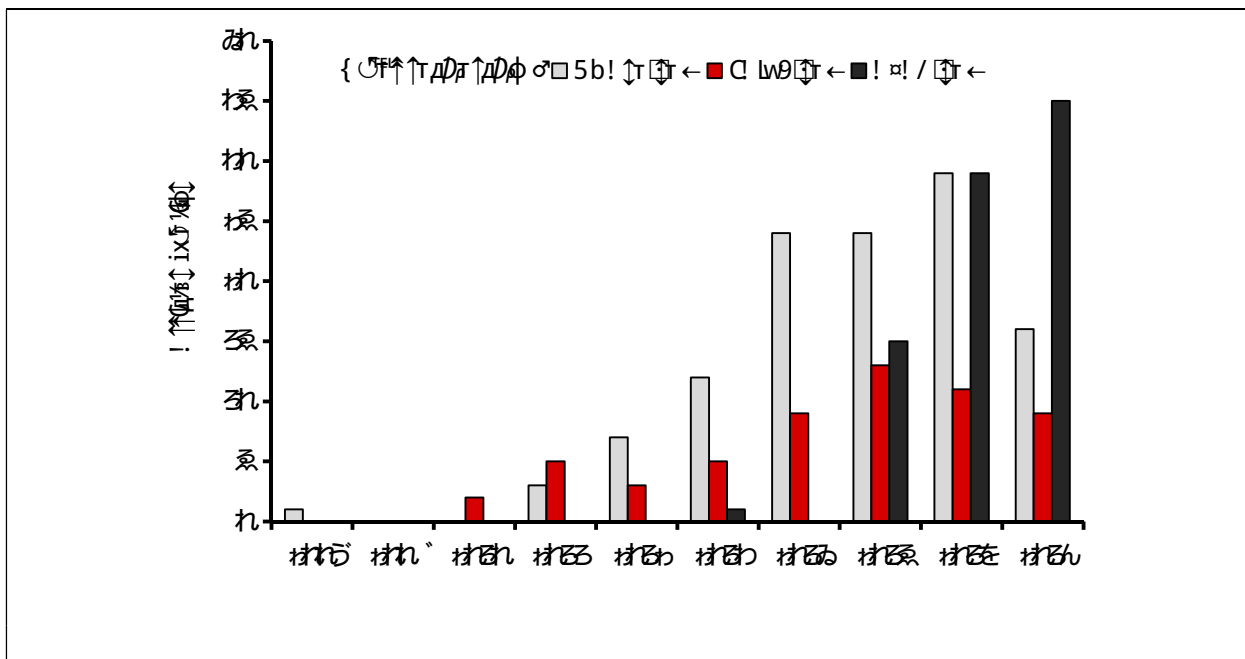


Figure 30 | BibliomŃtrie concernant les sujets DNA-seq, FAIRE-seq et ATAC-seq au sein de Web of Science entre les annŃes 2008 et 2017. Trois recherches indŃpendantes de ces termes ont ŃtŃ effectuŃes au sein de WoS dans la catŃgorie Topic, et le nombre d'articles publiŃs par annŃe comportant le terme recherchŃ a ŃtŃ reportŃ.

Comme nous l'avons constat  au cours des exp rimentations FA"RE d crites aux Chapitres " et "", les rendements FA"RE par rapport aux rendements Contr le (compar  gr ce au ratio FA"RE [Simon *et al.*, 2012] sont variables, notamment   cause des  tapes de fixation de la chromatine au formald hyde dont les r sultats d pendent de nombreux param tres parfois complexes   ma triser tels que la composition des tissus (par exemple en acide gras dans le cadre du Chapitre "), la finesse de broyage des tissus ou la perm abilit  des cellules [Tsompana et Buck, 2014]. Plusieurs r p titions biologiques ont ainsi  t  n cessaires afin d'obtenir des  chantillons reproductibles sur la base du ratio FA"RE/Contr le et des profils de r gions candidates en PCR quantitative (non publi s). Cette m thode peut donc  tre fastidieuse dans le cas de mat riel biologique limitant, comme cela a  t  le cas pour la cin tique de d veloppement embryonnaire qui a n cessit  9 r p titions biologiques compl tes pour 3 r p titions biologiques suffisamment reproductibles d'un point de vue du ratio FA"RE pour  tre s quenc es (Chapitre "").

Concernant l'utilisation de tissus congel s, ils sont notamment utilis s dans le cas d'analyse de la chromatine ouverte chez l'Homme sur des tissus limitant extraits suite   des biopsies, et qui par nature ne peuvent  tre analys s qu'extemporan ment. La cong lation des tissus peut entra ner des modifications de l'agencement de la chromatine au sein des cellules, et donc conduire   des biais d'identification de r gions ouvertes de la chromatine par FA"RE-seq. La comparaison des m thodes FA"RE entre tissus congel s et vivants n'a pas  t  r alis e de mani re sp cifique. A l'inverse, des exp riences de XCh"p-seq  qui permet  galement d'extraire de l'ADN en relation avec son contexte chromatinien (d crit en synth se bibliographique)  ont  t    la fois r alis es sur des h patocytes frais et congel s et ont r v l  une corr lation forte des r sultats, conduisant   des conclusions identiques entre les deux types de protocoles [Savic *et al.*, 2013]. La m thode FA"RE, comme le XCh"p, utilise une  tape de fixation de l'ADN aux prot ines   l'aide de formald hyde. Ce type de fixation limitant les modifications de la structure de la chromatine au cours des exp rimentations FA"RE et XCh"p, celles-ci doivent donc  tre r duites lors du processus de d cong lation des tissus, permettant donc l'obtention de r sultats similaires entre tissus congel s et non-congel s. Ceci justifie le choix de l' quipe au sein de laquelle s'est d roul e ma th se de r aliser les exp rimentations FA"RE-seq sur des tissus congel s.

## 2. L'analyse de donn es FA"RE-seq

### a. FA"RE-seq, DNase-seq, ATAC-seq : signal et r gions identifi es

Concernant les analyses bioinformatiques des trois m thodes d crites pr c demment, FA"RE-seq et ATAC-seq partagent de nombreux points en commun avec les analyses issues du Ch"p-seq, tandis que le DNase-seq pr sente des analyses d di es compte-tenu de la particularit  du signal obtenu. Les r gions identifi es par ces trois m thodes suite aux analyses bioinformatiques sont le plus souvent tr s corr l es, et anti-corr l es avec les r gions identifi es par MA"NE-seq. Cependant, chaque m thode comporte ses propres particularit s quant aux r gions sp cifiques qu'elles permettent d'identifier.

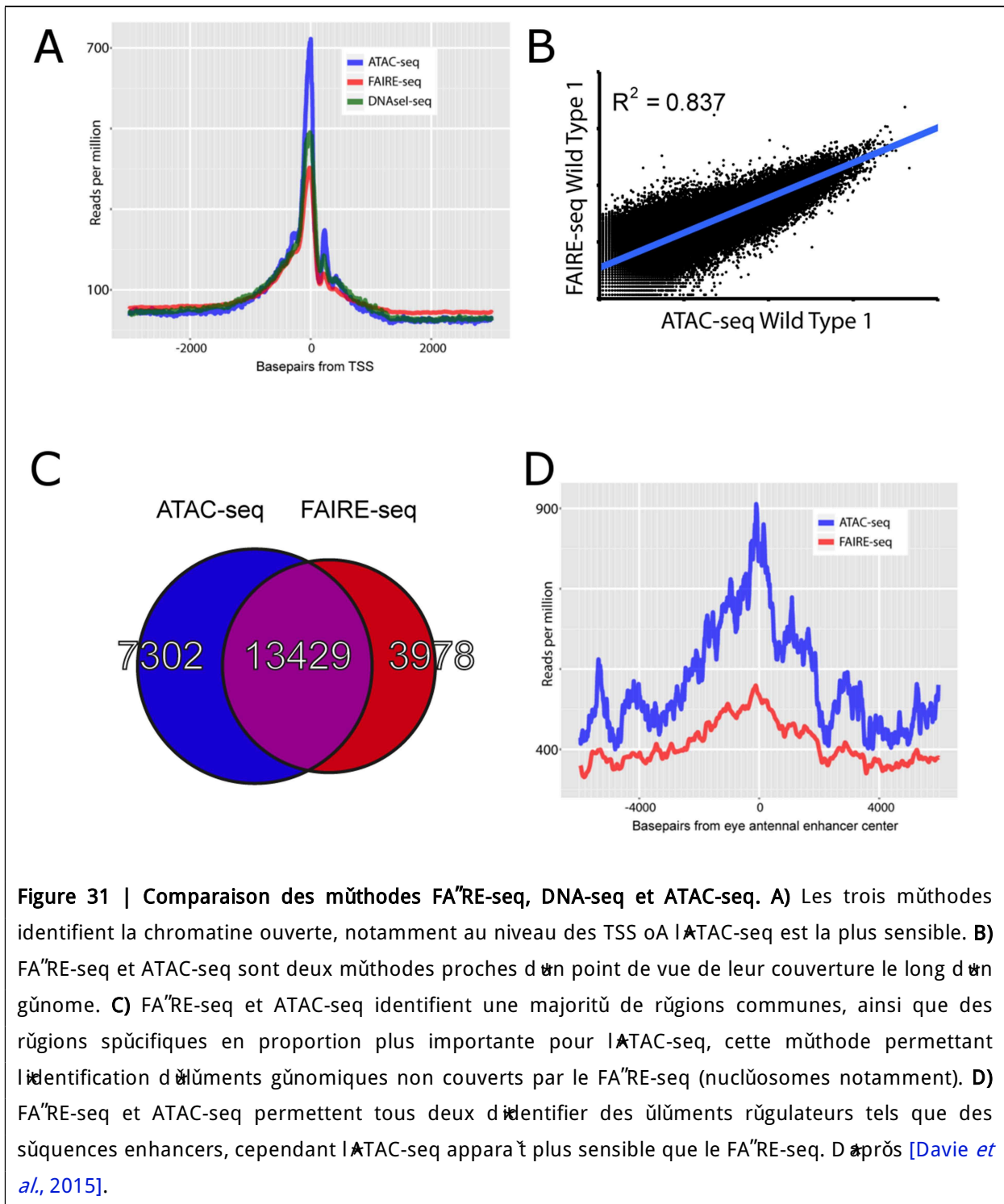
Concernant les avantages et inconv nients du FA"RE d'un point de vue bioinformatique, il permet de se affranchir du probl me des s quences surrepr sent es pouvant  tre enrichies dans le cadre des protocoles faisant intervenir des enzymes tels que le DNase-seq et le MA"NE-seq,  vitant donc des biais dans l'identification des sites de fixation des facteurs de transcription [He *et al.*, 2014]. Cependant le DNase-seq, tout comme l'ATAC-seq, pr sentent un ratio signal/bruit sup rieur   celui du FA"RE-seq. Ceci a pour effet de rendre les analyses bioinformatiques plus directes pour ces deux m thodes vis- vis du FA"RE. Ainsi, les donn es FA"RE-seq, bien que robustes et reproductibles, sont complexes   analyser en raison d'un ratio signal/bruit faible, rendant parfois complexe la dissociation entre r gion ouverte et bruit de fond notamment pour l'identification de pics ouverts de la chromatine.

Les r gions identifi es par FA"RE-seq et DNase-seq sont pour la plupart corr l es. En revanche le FA"RE-seq permet d'identifier des pics sp cifiques non-associ s   des r gions promotrices, tandis que le DNase-seq permet,   l'inverse, d'identifier des r gions sp cifiques des promoteurs [Kumar *et al.*, 2013]. L'ATAC permet quant   lui l'identification de r gions ouvertes de la chromatine corr l es avec le FA"RE-seq et le DNase-seq, ainsi que la position des nucl osomes et des r gions associ es   la fixation de facteurs de transcription proches des r gions accessibles de la chromatine [Tsompana et Buck, 2014]. Cette particularit  peut rendre complexe l'analyse et l'interpr tation des donn es sans autre information que celles apport es par l'ATAC-seq, puisque parmi les diff rents pics identifi s suite aux analyses bioinformatiques, diff rents  l ments g nomiques autres que de la chromatine ouverte peuvent  tre identifi s, y compris des  l ments antagonistes comme les nucl osomes. De plus, cette m thode  tant relativement r cente, l'analyse des donn es s quenc es peut se av rer complexe par manque d'outils sp cifiques   cette m thode,   l'inverse du DNase-seq et du Ch"IP-seq par exemple. Cependant, l'ATAC-seq devenant de plus en plus utilis  pour l' tude de la chromatine ouverte   l' chelle des g nomes, de nouvelles m thodes d'analyse d' di es sont actuellement en cours de d veloppement ([https://github.com/kundajelab/atac dnase pipelines](https://github.com/kundajelab/atac_dnase_pipelines)).

#### a. FA"RE-seq : des analyses complexes et non-uniformis es

Les protocoles d'analyse bioinformatique des donn es FA"RE-seq sont principalement d'riv s des analyses de donn es Ch"IP-seq, en  tant bas  sur un appel de pics suivi d'une assignation des r gions identifi es   diff rents  l ments g nomiques (g nes, promoteurs) mais  galement le calcul d'un signal FA"RE normalis .

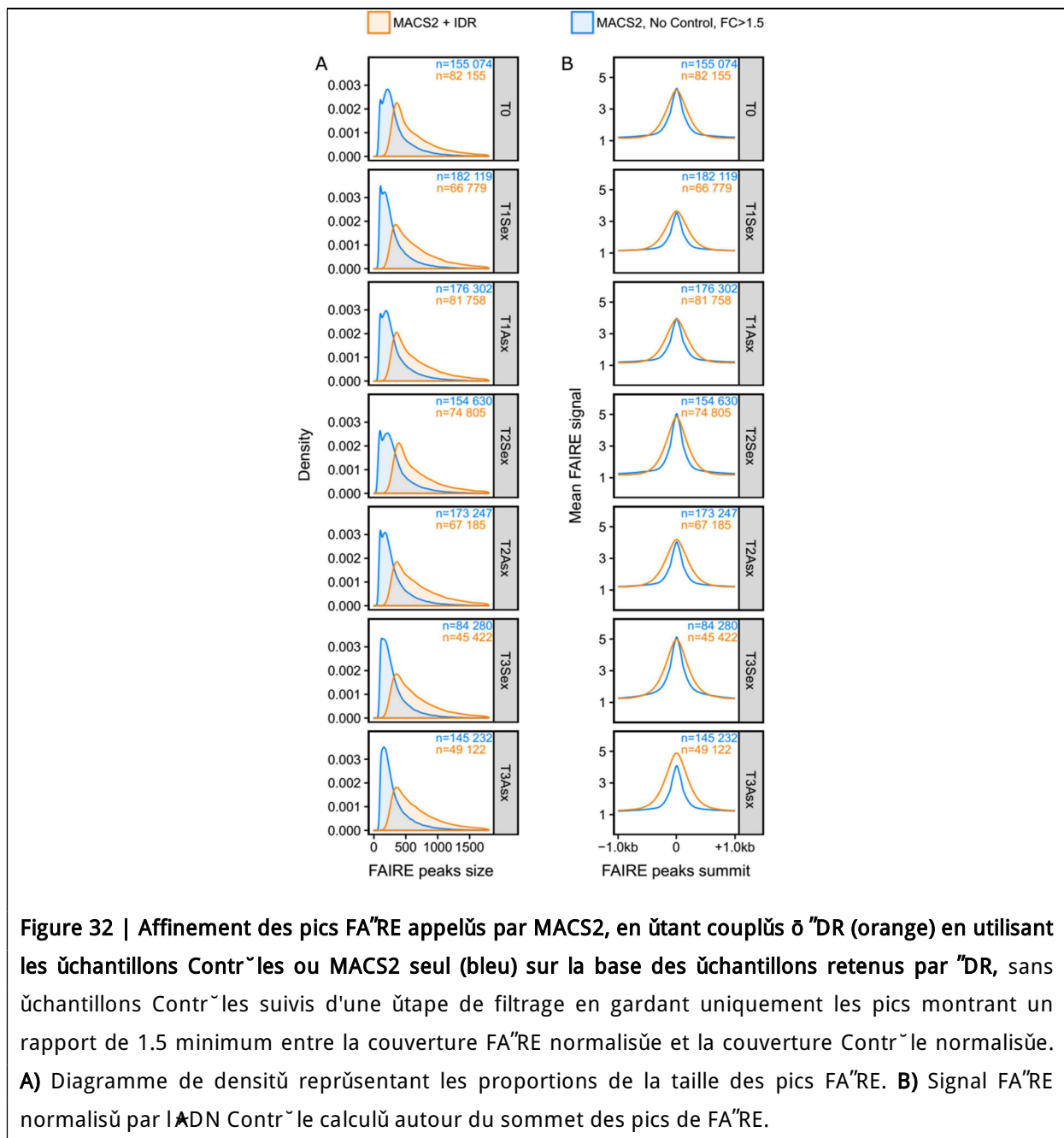
Le calcul du signal FA"RE r alis  par deepTools [Ram rez *et al.*, 2016] permet de quantifier l'accessibilit  de la chromatine sur un certain nombre de r gions d'int r t tels que les TSS des g nes. Cette approche a permis de tirer des conclusions robustes dans le cadre du premier article, car les diff rences d'accessibilit   taient marqu es entre conditions (m le/femelle, chromosome X/autosomes). En revanche, concernant le deuxi me article, d'apr s des analyses pr liminaires et les profils m tag niques pr sent s, il ne semble pas y avoir de diff rence d'accessibilit  globale de la chromatine. Ainsi, c'est l'approche d'appel de pics ouverts de la chromatine qui a  t  utilis e afin de r pondre   la question biologique pos e



La plupart de ces protocoles d'appel de pics impliquent l'utilisation de l'algorithme MACS2 [Zhang *et al.*, 2008b] accompagn  d' DR [Kundaje, 2012] permettant de classer les pics identifi s en fonction de leur reproductibilit  entre r plicas biologiques, tout en comparant la corr lation de ces derniers. Cependant, on retrouve au sein de la litt rature un manque global d'homog n it  des m thodes d'analyses. En effet, la plupart des articles traitant de donn es FA"RE-seq utilisent des m thodes extr mement diff rentes pour traiter des donn es pourtant similaires [Waki *et al.*, 2011 ; Leung *et al.*, 2014 ; Schick *et al.*, 2015], rendant complexe l'optimisation de l'appel de pics sur des donn es FA"RE-seq. En testant de multiples param tres et approches d crites dans divers articles, nous avons mis au point un protocole robuste d'appel de pics, et prenant en compte les limites de l'assemblage du g nome d'*A. pisum* (voir ci-dessous). Ainsi, en se basant sur le protocole d crit par [Kundaje, 2012], utilisant MACS2 et  DR pour des donn es Ch"IP-seq, et sur le protocole d crit par [Schick *et al.*, 2015], nous avons coupl  les analyses de reproductibilit  des r plicas biologiques (MACS2 +  DR)   des analyses de d tection de pics adapt es   la d tection de pics FA"RE fins tels que Z"NBAs seraient capables de les identifier (voir ci-apr s). En effet, les analyses couplant MACS2    DR sans traitement en aval ont tendance   g n rer des pics finaux trop larges pour  tre compar s entre les sept conditions diff rentes dans le Chapitre ". Ainsi nous avons utilis  les r plicas biologiques les mieux corr l es retenus par  DR, puis affin  la d tection des pics en se basant sur le protocole d crit par [Schick *et al.*, 2015]. Cette m thode a ainsi permis de conserver des pics reproductibles, tout en les affinant et en conservant un nombre suffisant par la d tection de pics sp cifiques en aval (Figure 32). Ces analyses peuvent g n rer des faux-positifs qui peuvent par exemple correspondre   un pic FA"RE appel  dans une seule condition, pouvant donc  tre d fini comme sp cifique de cette condition, alors qu'un signal FA"RE beaucoup plus  lev  (d finissant la pr sence d'un pic, ici non appel ) est d tect  au m me locus dans d'autres conditions. Afin d' viter la d tection de faux-positifs, une  tape de filtration des pics en fonction de leur hauteur (signal FA"RE maximal) entre diff rentes conditions a  t  r alis e et est d taill e au sein du deuxi me article, Chapitre ". De par le manque d'homog n it  des analyses FA"RE au sein de la litt rature, plusieurs mois ont  t  n cessaires   la mise au point d'un tel protocole associant un appel de pics fins, robustes, reproductibles, et dont la sp cificit  au sein des conditions  tudi es est v rifi e, tout en prenant en compte les limites du g nome d'*A. pisum* et le faible ratio signal/bruit inh rent   la m thode FA"RE-seq.

C'est dans le but d'unifier les analyses de donn es FA"RE-seq que l'algorithme d'appel des pics Z"NBAs [Rashid *et al.*, 2011], a  t  d velopp . Il propose diff rentes  tapes adapt es au faible ratio signal/bruit des donn es FA"RE-seq et se positionne ainsi comme l'un des principaux algorithmes de ce type  tre d di  au FA"RE, avec comme principale alternative F-seq [Boyle *et al.*, 2008b]. Cependant F-seq n'est disponible que sur un faible nombre de g nomes d'organismes mod les, et n'est donc pas  t  appliqu  au g nome du puceron du pois. Z"NBAs quant   lui n'est pas  t  d velopp  pour supporter des g nomes assembl s en de tr s nombreux scaffolds, comme l'est celui du puceron du pois. Z"NBAs est en effet bas  sur la comparaison deux   deux des chromosomes constituant un g nome, en g n rant un fichier par comparaison. Dans le cas du g nome du puceron du pois et ses 23 925 scaffolds, ce sont plus de 570 millions de fichiers (contre 16 fichiers pour le g nome de la *Drosophile*)

qui auraient d   tre g n r s, rendant impossible leur analyse par Z"NBa a posteriori, compte tenu de son optimisation pour un faible nombre de fichiers. MACS2  tant facilement applicable sur n'importe quel g nome et ayant  t  test  de mani re rigoureuse avec "DR, il n'est pas  tonnant de le retrouver utilis  pour l'analyse de donn es FA"RE-seq, bien qu'il ne soit pas aussi adapt  que Z"NBa au faible ratio signal/bruit du FA"RE, ayant  t  d velopp  avant tout pour l'analyse de donn es Ch"IP-seq. En dehors du probl me du choix des algorithmes et de leurs param tres, le g nome du puceron du pois repr sente en lui-m me un challenge bioinformatique non-n gligeable.





aux femelles qui en comportent deux, et chaque sexe comportant une paire de chaque autosome, une couverture divis e par deux chez les m les par rapport aux femelles permet d'identifier avec pr cision quels scaffolds appartiennent au chromosome X. Cependant, plut ˆt que d'observer des diff rentiels le long de scaffolds entiers, des alternances de r gions appartenant au chromosome X ou aux autosomes ont  t  identifi es   l'int rieur m ˆme de scaffolds. C'est ainsi qu'une nouvelle assignation entre le chromosome X et les autosomes a  t  cr e et a permis de r aliser les analyses comparatives de l'expression des g nes et de l'accessibilit  de la chromatine entre chromosome X et autosomes chez des individus m les et femelles dans le cadre du Chapitre ".

Le g nome du puceron du pois pr sente  galement des particularit s dans l'annotation de ses g nes. En effet, dans la plupart des analyses de g nomique comparative, il appara ˆt que le g nome d'*A. pisum* comporte des expansions importantes dans certaines familles de g nes, ainsi qu'un taux de duplication de g nes  lev  [International Aphid Genomics Consortium, 2010]. Ceci est le reflet d'une r alit  biologique, avec plus de 2000 familles de g nes dupliqu es. Cependant, il n'est pas exclu que certaines expansions de g nes comprennent des formes all liques dues   l'h t rozygotie du clone dont le g nome a  t  s quenc , conduisant   une sur-pr diction de g nes d'  l'all lisme d'un tel g nome [International Aphid Genomics Consortium, 2010]. De plus, le g nome du puceron pr sente actuellement deux annotations de g nes : AphidBase qui comporte 36 990 g nes et l'annotation du NCB" qui comporte 18 568 g nes. Ainsi, l'annotation AphidBase comporte environ deux fois plus de g nes annot s par rapport   l'annotation NCB". Ceci est d'  une part   une sur-pr diction dans le cas d'AphidBase, et d'  une sous-pr diction dans le cas de l'annotation NCB", cette derni re  tant cur e de mani re stringente pour ne conserver que les g nes support s de mani re robuste par des donn es d'expression.

Le g nome du puceron du pois n cessite donc d' ˆtre mieux caract ris . Actuellement, des nouvelles s quences longues ont  t  produites par la technologie PacBio et un nouvel assemblage est en cours afin d' liminer le chim risme (Fabrice Legeai, non publi ). Suite au nouvel assemblage, une annotation automatique sera g n r e incluant tous les RNA-Seq disponibles et les g nes pr dits d'esp ces proches (pucerons, h mipt res) et mod les (drosophile). Ainsi, les futures  tudes de g nomique structurale,  volutive et fonctionnelle pourront ˆtre plus ais es   r aliser et plus exhaustives.

## 2. L'incidence des limites du g nome d'*A. pisum* sur l'analyse de donn es de r gulation de la transcription

### a. Compensation de dose : signal FA"RE et territoire chromatinien

Concernant le premier chapitre portant sur la compensation de dose entre individus m les et femelles, l'ensemble des analyses bioinformatiques se sont articul es autour du calcul du signal FA"RE au niveau de l'ensemble des g nes. Pour rappel, le signal FA"RE correspond au ratio de la couverture FA"RE par la couverture Contr ˆle, chaque couverture  tant normalis e par la profondeur de s quen age. Cette approche pr sente l'avantage d' ˆtre ind pendante de la qualit  d'assemblage du g nome, et ind pendante de l'annotation des g nes, puisque ces derniers sont regroup s par profil

d'expression, et non par fonction. L'annotation AphidBase - pr sentant 13 711 g nes assign s au chromosome X, et 19 232 g nes assign s aux autosomes - a  t  utilis e pour les analyses bioinformatiques pr sent es dans l'article. Cependant, l'annotation NCB" comporte aujourd'hui un nombre bien plus restreint de g nes, avec seulement 5585 g nes assign s au chromosome X et 12 640 g nes assign s aux autosomes (avec 348 g nes dans des r gions non assign es). Afin de v rifier l'ind pendance des r sultats pr sent s dans l'article <sup>2</sup> Dosage compensation and sex-specific epigenetic landscape of the X chromosome in the pea aphid *P. vis- -vis* de l'annotation des g nes du g nome d *A. pisum*, nous avons recalcul  le signal FA"RE moyen le long des g nes associ s au chromosome X et aux autosomes chez les m les et femelles pour chaque annotation (Figure 33). Il s'av re que, bien que les signaux FA"RE soit plus  lev s chez les deux sexes dans le cas de l'annotation NCB", les tendances observ es entre m les et femelles sont ind pendantes de l'annotation utilis e. Ceci d montre la robustesse et l'ind pendance vis- -vis de la qualit  d'assemblage du g nome ou de l'annotation fonctionnelle des g nes du signal FA"RE. Une telle approche quantifiant l'accessibilit  de la chromatine peut donc  tre appliqu e au g nome de nombreux organismes non-mod les.

Le chim risme des scaffolds aurait pu constituer un obstacle   l'analyse des donn es FA"RE-seq et RNA-seq, puisqu'il aurait  t  impossible de pr dire avec pr cision quelles r gions g nomiques et quels g nes sont assign s au chromosome X ou aux autosomes. L'utilisation des donn es d'assignation [Jaqui ry *et al.*, 2017] nous a permis de contourner ce probl me et de mener   bien les analyses comparatives de l'accessibilit  de la chromatine entre chromosome X et autosomes malgr  la fragmentation et le chim risme de l'assemblage du g nome d *A. pisum*. Par contre, il nous est impossible pour le moment de rep rer le chim risme (s'il y en a) entre autosomes.

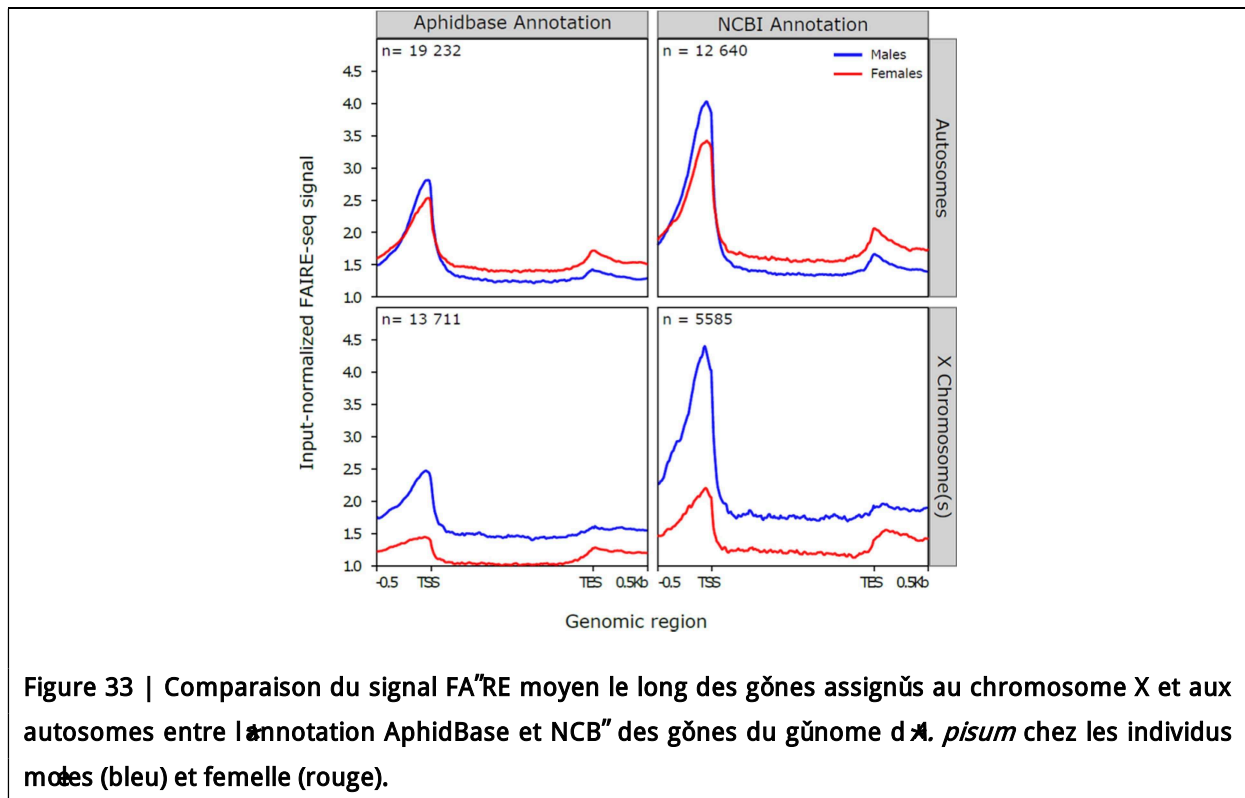
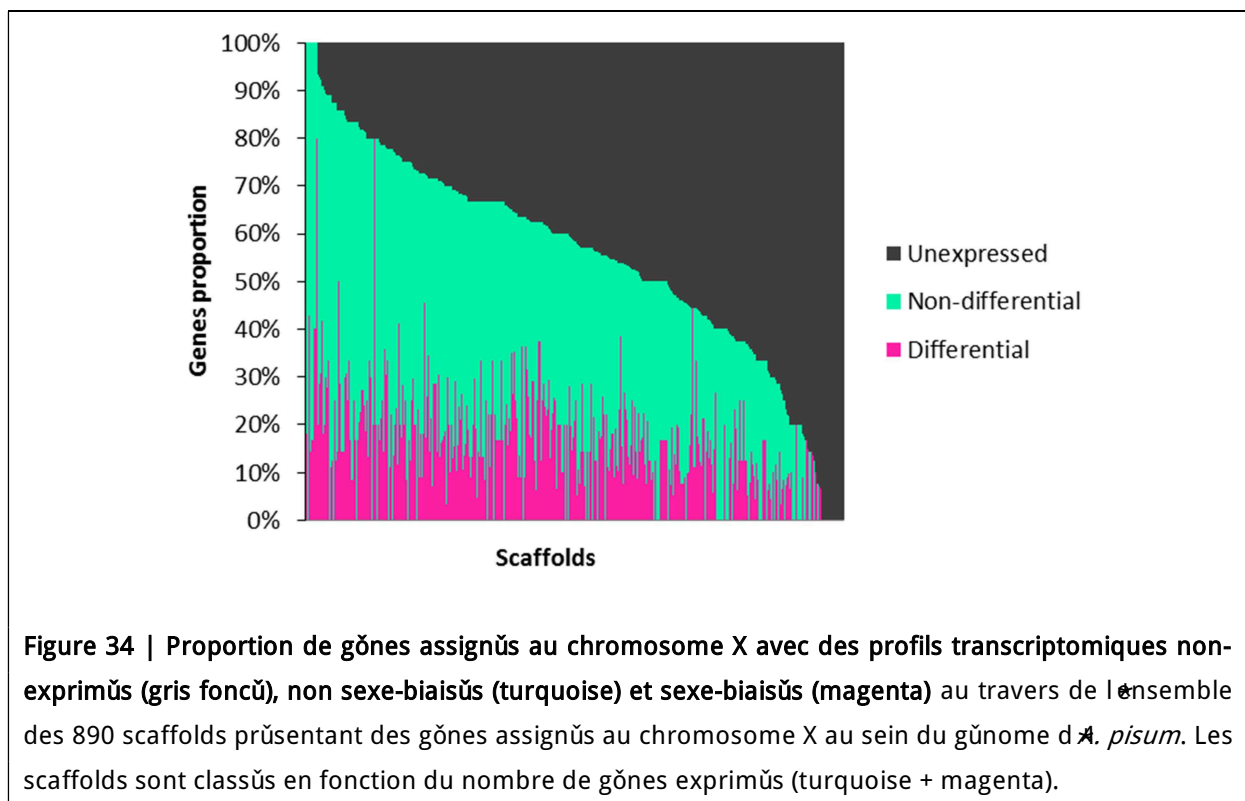


Figure 33 | Comparaison du signal FA"RE moyen le long des g nes assign s au chromosome X et aux autosomes entre l'annotation AphidBase et NCB" des g nes du g nome d *A. pisum* chez les individus m les (bleu) et femelle (rouge).

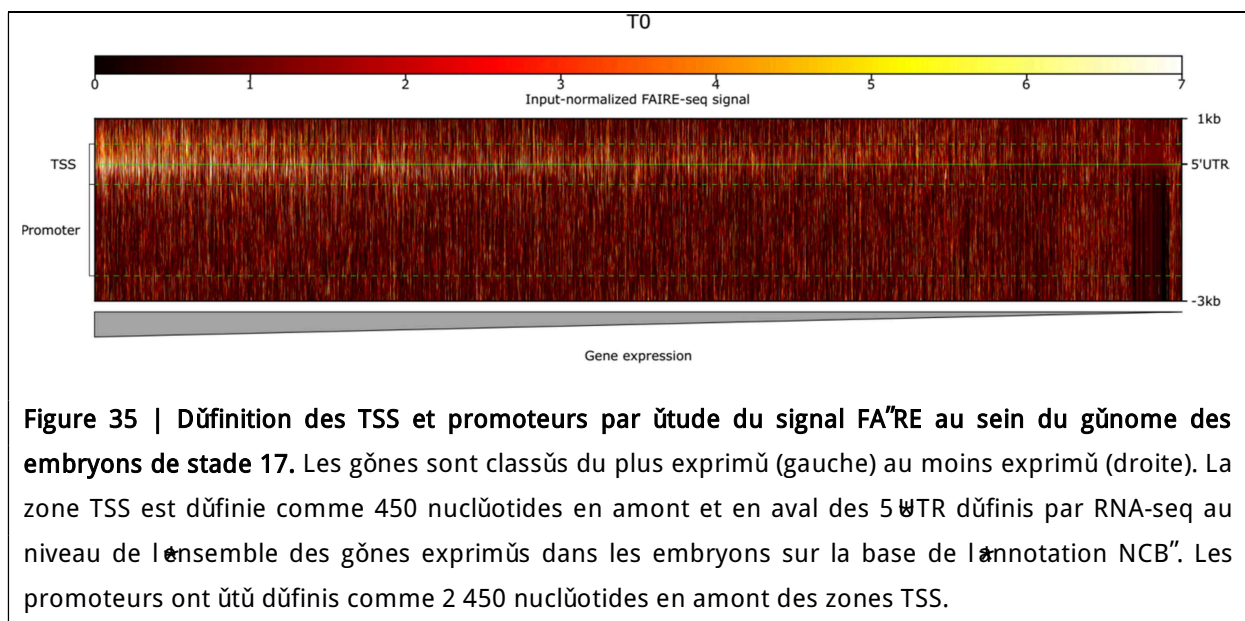
En revanche, il nous a řtř impossible de vřrifier la proximitř physique des gōnes potentiellement compensřs chez les mōles sur le chromosome X. En effet, ces gōnes exprimřs au mōme niveau entre mōles et femelles (alors que les mōles ne possřdent qu'un X) pourraient řventuellement appartenir ř un mōme territoire chromatinien, et donc řtre regroupřs par clusters ensuite rapprochřs les uns des autres ř l'aide des repliements tridimensionnels de la chromatine dans le noyau. Nous avons cependant tentř de le vřrifier par le calcul, pour chaque scaffold du gūnome d *A. pisum*, de la proportion des gōnes non-exprimřs, sexe-biaisřs et non-biaisřs (potentiellement compensřs) assignřs au chromosome X (Figure 34). Les gōnes non-diffřrentiels (non-biaisřs) sont prřsents de maniōre unique sur des scaffolds plus frřquemment que ne le sont les gōnes diffřrentiels (biaisřs), třmoignant d'un potentiel regroupement des gōnes non-biaisřs. Avec l'assemblage actuel du gūnome du puceron du pois, il est difficile d'aller plus loin dans ces analyses.

Des techniques comme le Hi-C permettraient d'avancer dans l'interprřtation de ces donnřes en gūnřrant des matrices de contact et donc la dřfinition des TADs du chromosome X chez des individus mōles et femelles. Il serait alors possible de vřrifier si d'une part les gōnes potentiellement compensřs (non-biaisřs) appartiennent ř un mōme domaine chromosomique tridimensionnel permettant la co-rřgulation de leur expression par le mřcanisme de compensation de dose. D'autre part, la comparaison de la dřfinition des TADs du chromosome X entre mōles et femelles permettrait d'identifier d'řventuelles rřorganisations de la structure de la chromatine entre sexe, qui pourrait supporter la compensation de dose.



b. Polyphūnisme de reproduction : appel de pics et rēgulations longue distance

Un des points essentiels permettant l'identification d'ūlments rēgulateurs ō partir de donnūes FA"RE-seq est d'avoir une annotation prūcise de la position des diffūrents ūlments gūnomiques : sites d'initiation et de terminaison de la transcription, regions promotrices putatives, sites connus de fixation de facteurs de rēgulation de la transcription, afin d'interprūter correctement le rōle putatif de ces ūlments rēgulateurs. Cependant, aucune donnūe permettant la dūfinition des ūlments gūnomiques sus-citūs ne sont disponibles chez le puceron pois. Des banques RNA-seq pleine longueur ou de Global Run-On sequencing (GRO-seq, une mūthode de suivi des transcrits naissant) permettraient la dūfinition prūcise des TSS, desquels sont dūrivūes les prūdictions de rēgions promotrices. Afin d'analyser les donnūes FA"RE-seq nous avons fait le choix de dūfinir des zones potentielles dans lesquels les TSS peuvent ūtre localisūs ō partir du signal FA"RE. En identifiant le locus comportant le signal FA"RE maximum en 5'UTR du dūbut de la sūquence codante des gōnes, puis en comparant les coordonnūes de ces locus entre l'ensemble des gōnes analysūs, nous avons dūfini que les TSS des gōnes d *A. pisum* peuvent se trouver dans une zone de 900 nuclōtides centrūe autour du premier nuclōtode des rēgions 5'UTR putatifs dūfinis par RNA-seq au sein de l'annotation NCB" (Figure 35).

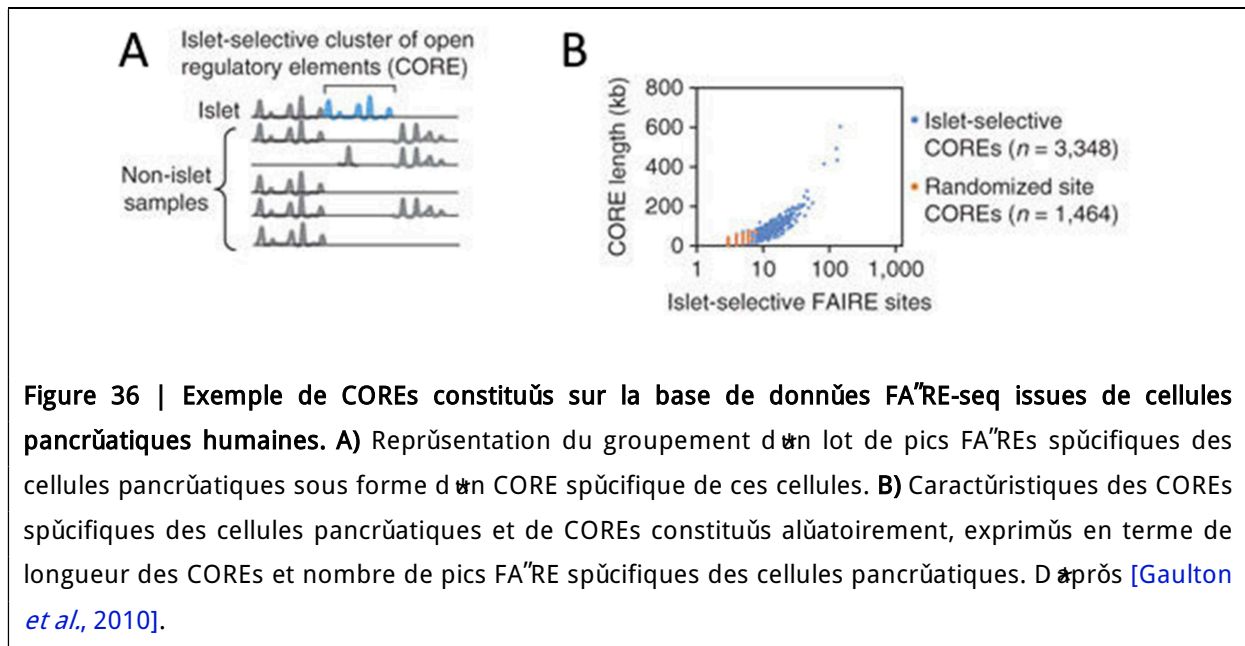


De plus, l'absence d'annotation de sites de fixation de facteurs de transcription (TFBS) dans le gūnome d *A. pisum* constitue un frein ō la paramūtrisation fine des algorithmes d'appel de pics sur des donnūes FA"RE-seq, et donc ō la recherche de TFBS sous ces pics. En effet, l'identification par Ch"IP-seq de sites de fixation d'un ou plusieurs facteurs de transcription connus permettrait de constituer un lot de loci d'intūrūt constituant un contrōle qualitū robuste des paramōtres utilisūs pour l'identification des pics FA"RE. Le but est de paramūtrer l'appel de pics afin de rēcupūrer si possible l'ensemble de ces sites de fixations de ces facteurs de transcription au sein des pics FA"RE. Afin de palier ō ce manque,

nous avons cherch    d finir des pics les plus fins possible   l'aide de MACS2, comme d crit pr c demment, afin d'enrichir avec plus de pr cision les motifs d' l ments r gulateurs putatifs.

La fragmentation et le chim risme de l'assemblage du g nome d'*A. pisum* posent le probl me de la difficult  d'analyser les r gulations de la transcription   longue distance par des  l ments r gulateurs putatifs en *cis* ou *trans* des g nes qu'ils r gulent, situ s au niveau de pics FA"RE distaux interg niques. Ainsi, il est impossible d'attribuer les pics FA"RE purement interg niques   un g ne distant de plus de 10 000 nucl otides, sans crainte de faux positifs. En effet, puisqu'il y a du chim risme entre chromosome X et autosomes au niveau de l'ensemble des scaffolds du g nome, il est fort probable que du chim risme entre les trois autosomes ait aussi eu lieu lors de l'assemblage du g nome. Ainsi, l'assignation d'un pic FA"RE donn    un g ne distant de 20 000 nucl otides ou plus peut conduire   l'assignation d'un pic situ  sur des autosomes diff rents. Malheureusement, sans un meilleur assemblage du g nome ou des donn es suppl mentaires, il est actuellement impossible de contourner ce probl me, sauf en s'int ressant uniquement aux assignations courtes distance. Cependant le FA"RE  tant particuli rement adapt    l'identification de r gions ouvertes distales des promoteurs [Kumar *et al.*, 2013], il est probable qu'une partie de l'information soit perdue.

Sans un g nome aussi fragment , il aurait  t  possible d' tudier ces pics FA"RE interg niques  loign s des promoteurs en constituant des Clusters of Regulatory Elements (COREs) [Gaulton *et al.*, 2010]. Ces  l ments g nomiques correspondent   de larges portions du g nome enrichies en pics FA"RE d tect s de fa on sp cifique au sein d'une condition analys e. Afin de constituer ces COREs, des distances sont calcul es entre chaque pic FA"RE sp cifique d'une condition au niveau de chaque chromosome. Lorsque la distance entre deux pics FA"RE sp cifiques successifs est trop importante, un CORE est cr   avec l'ensemble des pics FA"RE successifs pr c dent. Par comparaison des COREs d finis pour chaque condition, des COREs sp cifiques aux conditions  tudi es sont g n r s. Ceux-ci peuvent atteindre quelques centaines de milliers de nucl otides, et englobent donc potentiellement des centaines de g nes (Figure 36). "Il est ensuite possible de calculer des enrichissements d'ontologie des g nes (GO Terms) par CORE sp cifique d'une condition, et ainsi analyser de mani re int gr e les fonctions des g nes potentiellement r gul s en cluster par des  l ments r gulateurs distaux se situant au niveau des pics FA"RE interg niques." "ci,   cause des limitations de l'assemblage du g nome du puceron du pois, il n'est pas possible de constituer des COREs sans grouper des pics FA"RE potentiellement situ s sur diff rents chromosomes.



**Figure 36 | Exemple de COREs constitués sur la base de données FAIRE-seq issues de cellules pancréatiques humaines. A)** Représentation du groupement d'un lot de pics FAIREs spécifiques des cellules pancréatiques sous forme d'un CORE spécifique de ces cellules. **B)** Caractéristiques des COREs spécifiques des cellules pancréatiques et de COREs constitués aléatoirement, exprimés en terme de longueur des COREs et nombre de pics FAIRE spécifiques des cellules pancréatiques. D'après [Gaulton *et al.*, 2010].

En conservant l'assemblage fragmenté du génome du puceron, il serait tout de même possible d'assigner les éléments régulateurs se trouvant sous les pics FAIRE à l'ensemble des gènes qui les régulent grâce à des expériences de Hi-C. Cette méthode nous permettrait de vérifier quels gènes sont à proximité des pics FAIRE spécifiques d'une condition, ou encore de vérifier que des gènes régulés et des pics FAIRE spécifiques se trouvent au sein de mêmes domaines topologiques. Il serait ainsi plus aisé d'accorder de l'importance à un motif régulateur putatif localisé sous un pic FAIRE, et aux facteurs de transcription pouvant s'y fixer. En outre, la technique Hi-C permettrait d'améliorer l'assemblage du génome du puceron [Flot *et al.*, 2015] en étant couplé avec d'autres méthodes comme le séquençage d'ADN avec les lectures de plusieurs milliers de nucléotide par la technique PAC-BIO [Lapp *et al.*, 2017]. De plus, compte tenu de l'utilité de cette méthode pour l'étude des TADs en lien avec les territoires chromatinien du chromosome X des mouches, le Hi-C constitue une perspective de choix pour les études génomiques chez le puceron du pois.

### c. Polyphénisme de reproduction : une annotation incomplète du génome

Afin de permettre des annotations fonctionnelles du génome du puceron du pois, il est nécessaire de se référer à des génomes modèles bien annotés comme celui de *D. melanogaster*. Plusieurs approches comme la recherche de domaines conservés et les ontologies sont utilisées pour comparer *in silico* les génomes d'*A. pisum* avec celui d'autres organismes. Il est néanmoins difficile d'interpréter certains résultats expérimentaux (listes de gènes différentiellement exprimés par exemple) sur la base de la fonction des gènes si celle-ci n'a pas été caractérisée expérimentalement. Comme c'est le cas chez de nombreux organismes non-modèle et le puceron ne fait pas exception, la transgénération stable n'est pas disponible. Un protocole de mutagenèse aléatoire à l'EMS a été décrit [Tagu *et al.*, 2014b] mais il reste inadapté au criblage de phénotypes aussi complexes que le mode de reproduction. Plusieurs études rapportent également l'utilisation du RNAi afin de diminuer le niveau

d'accumulation d'un transcrit donn  [Mutti *et al.*, 2006 ; Jaubert-Possamai *et al.*, 2007 ; Sapountzis *et al.*, 2014], mais les r sultats sont tr s d pendants du g ne  tudi , des techniques utilis es ou encore du tissu cibl , et les ph notypes observ s sont souvent tr s g n raux (mortalit  accrue, perte de f condit ).

Cependant, *D. melanogaster* et *A. pisum* sont deux insectes qui comportent un certain nombre de g nes tr s bien conserv s, environ 4500 d'apr s nos analyses par BLASTp (non publi es). Afin de r aliser des analyses GO gr ce aux nombreux outils disponibles en ligne tels que GOzilla [Eden *et al.*, 2009] ou gprofiler [Reimand *et al.*, 2007, 2016], nous avons pris le parti d'utiliser les g nes homologues de *D. melanogaster* chez *A. pisum*, en ne conservant que les homologues avec une E-value inf rieure   10 et une p-value inf rieure   1e-20. De cette fa on, nous avons pu r aliser des analyses GO donnant des r sultats satisfaisants   la fois dans le cadre de la compensation de dose et dans le cadre du polyph nisme de reproduction   l'aide de GOzilla et de gprofiler, ce dernier pr sentant l'avantage de pouvoir comparer plusieurs listes de g nes de grande taille et de ne pas extraire que les GO Terms diff rentiellement enrichis. A titre d'exemple, dans le cadre du premier article, concernant les g nes male-biais s localis s sur le chromosome X, nous avons retrouv  un enrichissement de GO Terms en relation avec la motilit  et la production de spermatozo des, tandis que dans le cadre du polyph nisme de reproduction, nous avons notamment identifi  un enrichissement de fonctions associ es au m tabolisme des lipides dans des embryons sexu s pr sentant probablement d j  des  ufs non-fertilis s en cours de formation, et pour lesquels une accumulation de vitellog nine   compos e en grande partie de lipides  est indispensable, confirmant la robustesse de ces approches.

Concernant le manque d'annotation des sites de facteur de transcription (FT) dans le g nome d'*A. pisum*, et afin d'analyser la fixation des FT au niveau des motifs d'ADN enrichis sous les pics FA"RE identifi s, nous avons utilis  les informations obtenues sur les facteurs de transcriptions de *D. melanogaster*. L  encore, des BLAST ont  t  effectu s afin de ne conserver que les facteurs de transcription conserv s entre le g nome du puceron du pois de *D. melanogaster*. Ceci est possible gr ce   la conservation au travers de l' volution des motifs d'ADN reconnus par les facteurs de transcription [Nitta *et al.*, 2015]. Gr ce   l'enrichissement de motifs ADN au sein des pics FA"RE sp cifiques de chaque type d'embryons dans le cadre de la compensation de dose,   l'aide de l'algorithme MEME [Machanick et Bailey, 2011], nous avons identifi  des sites de fixations putatifs de FT de la drosophile, et dont les domaines catalytique et de reconnaissance   l'ADN sont conserv s. Des analyses compl mentaires visant   faire le lien entre ces TFBS et les g nes qu'ils r gulent doivent  tre r alis es.

## C. Cycle de vie des pucerons : polyphŃnisme et compensation de dose

Maintenant que nous avons discutŃ des problŃmes techniques rencontrŃs au cours de ma thŃse et des solutions mises en place ayant permis de les rŃsoudre, nous allons discuter les rŃsultats obtenus ainsi que leurs implications vis-Ń-vis de la comprŃhension des mŃcanismes Ń la base du cycle de vie des pucerons et de leur reproduction.

### 1. Un mŃcanisme de compensation de dose permettrait d'Ńquilibrer la transcription du chromosome X vis-Ń-vis des autosomes chez les pucerons males

Au niveau transcriptomique, nous avons identifiŃ chez les males une expression Ńquivalente entre X et autosomes aprŃs filtration des gŃnes les moins exprimŃs. Ceci est dŃ Ń la forte concentration en gŃnes non-exprimŃs Ń l'Ńchelle de l'individu entier au niveau du chromosome X. Chez les femelles, nous avons relevŃ des niveaux d'expression globaux trŃs faibles sur le chromosome X, correspondant sans doute Ń la masculinisation de ce dernier chez *A. pisum* [JaquiŃry *et al.*, 2013]. En effet, parmi les nombreux gŃnes male-biaisŃs identifiŃs, la majoritŃ d'entre eux sont male-spŃcifiques (environ 700 sur 931). Ainsi, seule une trŃs faible proportion de gŃnes sont exprimŃs sur le chromosome X des femelles, dŃ un ratio X/A infŃrieur Ń 0.5 chez ces derniŃres.

Au niveau de la chromatine, nous avons identifiŃ une accessibilitŃ accrue de celle-ci au niveau du TSS des gŃnes portŃs par le seul X des males, comparativement aux deux X des femelles. Cependant nous avons identifiŃ une ouverture similaire au niveau des autosomes entre males et femelles, posant la question d'une accessibilitŃ accrue chez les males ou diminuŃe chez les femelles de la chromatine du chromosome X. D'aprŃs le profil d'accessibilitŃ des gŃnes males non-biaisŃs, le seul chromosome X des males prŃsente une accessibilitŃ accrue de la chromatine par rapport aux autosomes, ainsi qu'en comparaison avec les autosomes et le chromosome X des femelles. Ceci suggŃre que seule une partie de la chromatine du chromosome X, probablement un territoire chromatinien spŃcifique, comporterait une dŃcompaction trŃs importante chez les males au niveau de gŃnes compensŃs. Ces derniers pourraient Ńtre regroupŃs sous forme de clusters, comme suggŃrŃ par la figure 34, ou dans des TADs, pour lesquels il faudrait rŃaliser des expŃriences de capture de la conformation tridimensionnelle de la chromatine. Cependant, ces seules donnŃes de chromatine ouverte ne nous renseignent pas sur le mŃcanisme ŃpigŃnŃtique Ń l'origine de cette dŃcompaction de la chromatine du chromosome X des males dans le cadre de la compensation de dose chez le puceron du pois.

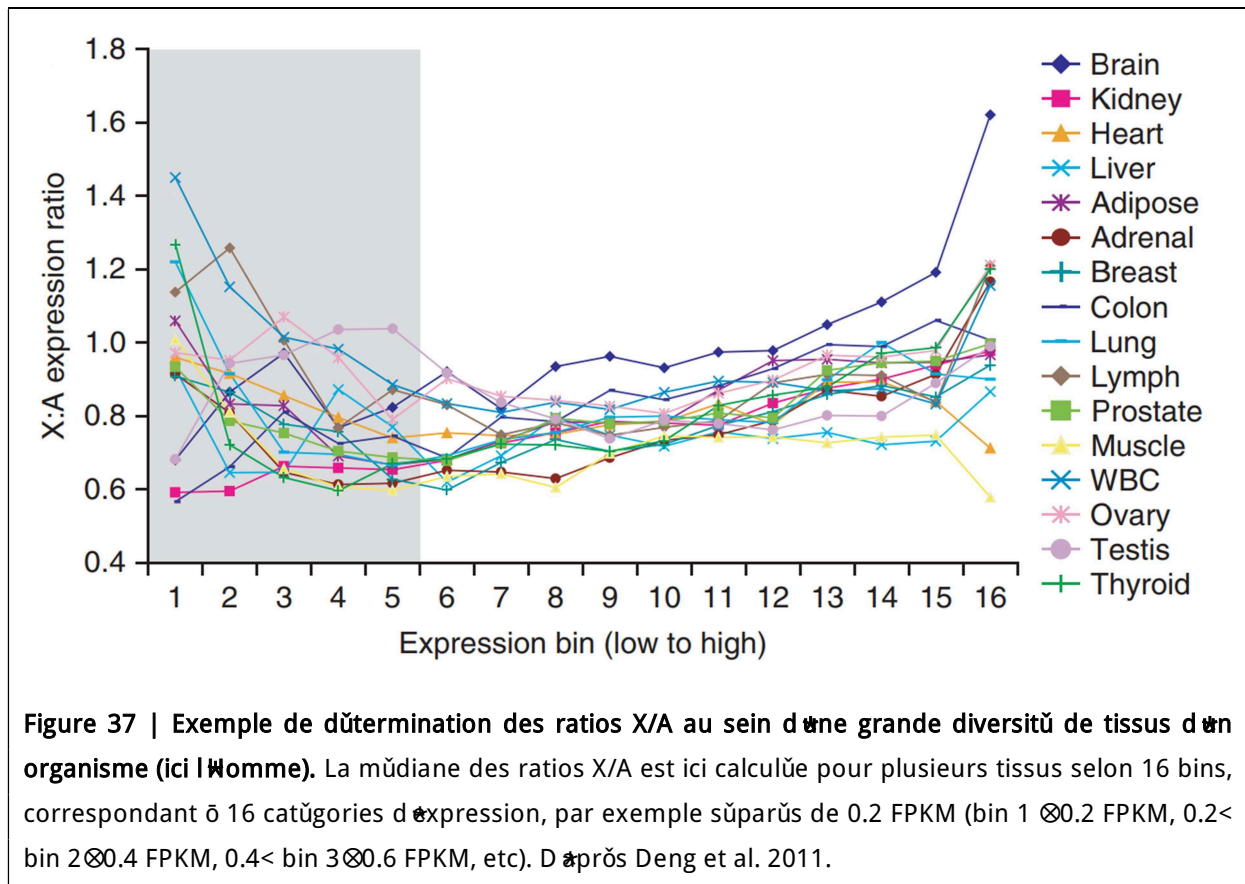
Cette accessibilitŃ accrue pourrait sans doute Ńtre associŃe Ń l'enrichissement d'une ou de plusieurs marques histones actives sur l'unique chromosome X des males, l'hypothŃse des marques rŃpressives sur les X des femelles Ńtant Ń priori rejetŃe grace aux rŃsultats de FA"RE-seq dŃcrits ci-avant. Ainsi, ces modifications post-traductionnelles des histones pourraient Ńtre globalement plus enrichies chez le male que chez la femelle, ou bien enrichies de maniŃre spŃcifique sur le X des males, comme c'est le

cas de la marque H4K16ac chez la Drosophile [Gelbart *et al.*, 2009]. Des exp riences de Ch<sup>3</sup>P-seq permettraient la caract risation de ces marques sur le X des m les et des femelles.

A l'instar de la Drosophile, il est effet possible qu'un complexe prot ique assembl  de mani re sp cifique sur le chromosome X des pucerons m les permette un enrichissement de cette ou de ces marques histones actives   l'origine de l'augmentation de l'accessibilit  de la chromatine du seul X des m les, dans le but d'augmenter sa transcription et de l' quilibrer avec celle des autosomes. Il est possible que le complexe MSL de *D. melanogaster* soit conserv  chez *A. pisum*. Une mutag nese cibl e par CRISPR-CAS9 d'un des g nes codant les prot ines le constituant (MSL1, MSL2, MSL3, MOF et MLE), puis l'analyse du sex-ratio de la descendance de femelles sexupares mut es permettrait de caract riser l'importance d'une de ces prot ines dans un m canisme indispensable   la survie des m les, mais pas   celle des femelles. En outre, l'assemblage sp cifique du complexe MSL putatif des pucerons pourrait  tre permis par 1 ou plusieurs des 4 lncARN exprim s sp cifiquement sur le X des m les, de la m me mani re que les lncARN *roX1* et *roX2* permettent l'assemblage du complexe MSL chez la Drosophile.

D'un point de vue  volutif, il serait int ressant de v rifier que d'autres esp ces de pucerons pr sentent un profil transcriptomique et chromatinien similaire   celui de *A. pisum* afin de g n raliser les r sultats obtenus chez cet organisme, avec par exemple la r alisation de RNA-seq et de FA<sup>2</sup>RE-seq chez *Myzus persicae* et *Diuraphis noxia*, deux pucerons pr sentant une h t rogam tie m le et un g nome assembl  et annot . Une comparaison avec *Daktulosphaira vitifoliae*, un insecte dont le g nome est s quenc  et assembl , pourrait  galement  tre int ressant, puisque celui-ci est beaucoup plus basal et permettrait donc de v rifier la conservation d'un tel m canisme de compensation de dose sur des temps  volutifs plus importants.

Enfin, il faudrait tester de mani re fine et pr cise la compensation de dose chez *A. pisum*. En effet, les donn es de RNA-seq publi es dans le premier article pr sent  dans cette th se sont issues d'individus entiers m les et femelles. Ceci peut constituer un biais dans l'analyse de la compensation de dose puisque les ARNm sont extraits   partir d'un m lange de tissus somatiques et germinaux, alors que la compensation de dose n'est parfois pas pr sente au niveau des cellules germinales, comme c'est le cas chez les mammif res [Deng *et al.*, 2011]. Pour ce faire, une analyse compl te de donn es RNA-seq issues de tissus somatiques (t tes et pattes des trois morphes sexuels par exemple), et de tissus principalement germinaux tels que les testicules et le p nis, les germaria et ovocytes pr coces (haplo des ou diplo des selon le sexe) de femelles asexu es et sexu es pourraient  tre men es. Ces analyses se baseraient tout d'abord sur une identification des g nes tissus-sp cifiques par analyse diff rentielle d'expression. Apr s  limination de l'ensemble de ces g nes tissus-sp cifiques dans les donn es d'expression de chaque tissu, il serait alors possible de tester la compensation de dose, dans chaque tissu, par classement des g nes selon leur expression puis par s paration de ces g nes en groupes de nombres  gaux (m thode du <sup>2</sup> binning B), puis par calcul des ratio X/A dans chacun de ces groupes, pour chaque tissu, tel que d crit par Deng et al. 2011 (Figure 37).



## 2. Le polyphŃnisme de reproduction des pucerons est potentiellement associŃ Ń diffŃrentes voies de signalisation neuro-endocrine

Au cours du polyphŃnisme de reproduction des pucerons, en rŃponse Ń la diminution de la photopŃriode, sont gŃnŃrŃes des femelles sexupares capables de produire des femelles sexuŃes, de males et de femelles asexuŃes. La production de ces morphes par un mŃme individu Ń partir de cellules germinales initialement identiques suggŃre une reprogrammation de la transcription de nombreux gŃnes au niveau des cellules des germaria, suite Ń la transduction du signal photopŃriodique par des voies neuro-endocrine. Parmi les gŃnes diffŃrentiellement exprimŃs entre embryons Ń reproduction sexuŃe et embryons Ń reproduction asexuŃe (de faŃon antŃrieure Ń cette thŃse) nous pouvons citer Ubiquitin Like With PHD And Ring Finger Domains 1 (*uhrf1*) et Suppressor Of Variegation 4-20 Homolog 1 (*suv(4)-20H1*) [Gallot *et al.*, 2012], deux gŃnes codant des protŃines impliquŃes dans des modifications de la structure de la chromatine. Les programmes gŃnŃtiques Ń l'origine de l'initiation du polyphŃnisme de reproduction impliquent donc des rŃorganisations chromatiniennes que nous avons cherchŃ Ń caractŃriser entre embryons sexuŃs et asexuŃs.

Afin d'Ńtudier ces processus, nous avons utilisŃ un protocole permettant la production d'embryons Ń dŃveloppement sexuŃ ou asexuŃ sous les mŃmes conditions de photopŃriode Ń l'aide d'un traitement

hormonal par un analogue de l'hormone juvŕnile sur des femelles sexupares placŕes en photopŕiode courte. Ce protocole vise ŕ rŕtablir un niveau d'hormone juvŕnile ŕquivalent ŕ celui mesurŕ chez des femelles parthŕnogŕnŕtiques placŕes en jours longs qui produisent des embryons ŕ reproduction asexuŕe [Shikawa *et al.*, 2012], suggŕrant donc une voie de diffŕrenciation des embryons vers une reproduction asexuŕe maintenue par des niveaux ŕlevŕs d'hormone juvŕnile. Nous avons cherchŕ ŕ caractŕriser les modifications de l'accessibilitŕ de la chromatine au sein de ces embryons par FA"RE-seq, et l'impact de ces modifications de l'ouverture de la chromatine sur la transcription de gŕnes, en caractŕrisant leurs fonctions, et en tentant d'analyser les ŕlŕments rŕgulateurs putatifs pouvant se fixer au niveau de rŕgions chromatiniennes diffŕrentiellement accessibles entre embryons sexuŕs et asexuŕs.

En analysant globalement par la GO la fonction de gŕnes au sein desquels se trouvent des rŕgions chromatiniennes spŕcifiquement accessibles en fonction du type d'embryon, nous avons identifiŕ chez les embryons sexuŕs de stade 18 un enrichissement de gŕnes associŕs au mŕtabolisme des lipides. L'expression de ces gŕnes ŕtant potentiellement rŕgulŕe par des ŕlŕments situŕs au niveau des pics FA"RE identifiŕs, nous avons corrŕlŕ l'ouverture spŕcifique de ces rŕgions gŕnomiques avec le niveau d'expression de gŕnes en fonction du mode de reproduction et du stade de dŕveloppement des embryons. Au stade 18, nous n'avons identifiŕ qu'un seul gŕne diffŕrentiel entre embryons sexuŕs et asexuŕs : un homologue de la protŕine codŕe par le gŕne *slimfast* de la *Drosophile* qui est sous-exprimŕ chez les embryons sexuŕs, tandis qu'au stade 20, le second gŕne le plus sous-exprimŕ chez les embryons asexuŕs par rapport aux embryons sexuŕs est un homologue de *brummer*. Ces deux gŕnes sont impliquŕs dans le mŕtabolisme des lipides et chez la *Drosophile*, une sous-expression de *slimfast* conduit ŕ l'agrŕgation de vŕsicules de stockage de lipides, tandis qu'une surexpression de *brummer* conduit ŕ une remobilisation des lipides stockŕs dans le corps adipeux. Si la fonction de ces gŕnes est conservŕe chez *A. pisum*, il est probable que ces gŕnes soient impliquŕs dans la synthŕse et ŕventuellement le stockage de lipides au sein des ŕufs non fertilisŕs en cours de formation au sein des embryons sexuŕs. En outre, l'ŕtude des facteurs de transcription pouvant potentiellement se fixer au niveau des pics FA"RE spŕcifiques des embryons sexuŕs a rŕvŕlŕ un enrichissement de motifs reconnus putativement par un homologue protŕique de HNF4, un facteur de transcription impliquŕ dans la rŕgulation de gŕnes impliquŕs dans la mobilisation des lipides et la bŕta-oxydation des acides gras pendant le dŕveloppement larvaire. Il serait cependant intŕressant de vŕrifier dans les promoteurs de quels gŕnes ce facteur de transcription pourrait se fixer afin d'en rŕguler l'expression. Nos analyses suggŕrent donc l'implication potentielle de rŕorganisations de la structure de la chromatine dans la rŕgulation de la transcription de gŕnes, et dans la fixation de facteurs de transcription qui auraient pour effet de modifier le mŕtabolisme des lipides chez les embryons sexuŕs, comparativement aux embryons asexuŕs. Ceci suggŕre une possible modification du stockage ou tout du moins de la remobilisation des lipides au niveau des lignŕes germinales, dans le cadre de la mise en place du phŕnotype sexuŕ.

En outre, nous avons dŕveloppŕ plusieurs hypothŕses conduisant ŕ la conclusion que la reproduction asexuŕe semble ŕtre la trajectoire dŕveloppementale par dŕfaut pour les embryons de femelles

sexupares. Il s'avŃre que la transition entre les embryons dont la destinŃe reproductive n'est pas encore dŃfinie (stade 17/T0) et les embryons Ń destinŃe reproductive asexuŃe (stade 18/T1Asx) implique moins de reconfigurations de l'accessibilitŃ de la chromatine que la transition vers des embryons Ń destinŃe reproductive sexuŃe (stade 18/T1Sex). Ces rŃsultats ont ŃtŃ obtenus alors que les femelles sexupares sont placŃes en jours courts, et produisent donc naturellement des embryons Ń destinŃe reproductive sexuŃe, tandis que les embryons Ń destinŃe asexuŃe sont obtenus par perturbation de l'Ńquilibre hormonal des femelles par dŃpŃt de quantitŃ importante d'un analogue d'hormone juvŃnile sur leur abdomen. Ainsi, malgrŃ un potentiel <sup>2</sup> priming p des embryons vers une reproduction sexuŃe en rŃponse Ń la diminution graduelle de la photopŃriode, la transition vers cette reproduction nŃcessite plus de modifications de la structure de la chromatine, ou tout du moins une moins grande conservation de la structure de la chromatine ouverte, que pour la transition vers une reproduction asexuŃe. Ceci suggŃre que la transition vers l'asexualitŃ serait le programme par dŃfaut pour un embryon de stade 17. Cette hypothŃse est renforcŃe par l'absence de dŃtection de motifs reconnus par des facteurs de transcriptions inductibles par l'hormone juvŃnile au niveau de la chromatine des embryons asexuŃs. Ces rŃsultats suggŃrent l'absence d'une voie spŃcifique de diffŃrenciation vers des embryons asexuŃs dŃclenchŃe de maniŃre directe par l'hormone juvŃnile.

Concernant la transition vers un mode de reproduction sexuŃ, les embryons encore flexibles au stade 17 doivent opŃrer des changements importants dans l'organisation de leur chromatine et ceci est supportŃ par la surexpression chez les embryons sexuŃs de stade 19 d'un homologue de l'Ń ubiquitine ligase RNF8, impliquŃe chez les mammifŃres dans des dŃcompactions de la chromatine. De maniŃre surprenante, les motifs d'ADN enrichis au niveau des pics de chromatine ouverte identifiŃs chez les embryons sexuŃs pourraient correspondre Ń des sites de fixations de facteurs de transcription inductibles par l'ecdysone. Parmi ces FT, nous avons identifiŃ des homologues de BLMP-1, E'P75B et E'P78C. *Blimp* est impliquŃ chez la Drosophile dans le timing de la pupation en rŃgulant la synthŃse de 20-hydroxyecdysone (20-E) Ń partir d'ecdysone au sein du corps adipeux via la rŃgulation transcriptionnelle de *Xftz-f1*, rŃgulant lui-mŃme l'expression de *shade* [Akagi *et al.*, 2016]. E'P78C et E'P75B sont quant Ń eux impliquŃs dans l'ovogŃnŃse, avec E'P78C qui favorise la survie des follicules en dŃveloppement par son activitŃ au niveau des cellules germinales, et E'P75B qui joue un rŃle dans le choix du dŃveloppement ou de la dŃgŃnŃrescence des chambres Ń Ńufs par l'induction ou la suppression de l'apoptose au sein des larves de stades 8 et 9, chez la Drosophile [Terashima *et Bownes*, 2006 ; Ales *et al.*, 2015]. Ainsi ces 3 facteurs de transcription permettraient de rŃguler la synthŃse de l'ecdysone, et de permettre la mise en place de l'ovogŃnŃse au sein des embryons sexuŃs. Une voie de rŃponse Ń l'ecdysone serait donc responsable de l'initiation de la reproduction sexuŃe chez des embryons de femelles placŃes en jours courts. Si aucune Ńtude n'atteste du rŃle du l'ecdysone dans la rŃgulation du mode de reproduction, l'ecdysone a ŃtŃ montrŃe comme impliquŃe dans la dŃtermination du polyphŃnisme de dispersion des pucerons, avec une diminution de la proportion d'individus ailŃs dans la descendance de femelles traitŃs avec de l'ecdysone ou un analogue, et Ń l'inverse une production accrue d'individus ailŃs chez des femelles traitŃes avec des inhibiteurs de la voie de l'ecdysone [Vellichirammal *et al.*, 2017]. Hormones juvŃniles et ecdysones

sont connues pour intervenir ensemble sur de nombreux processus physiologiques  $\ni$  comme la mue. Notre Ńtude permet d'ajouter l'ecdysone comme Ńventuel rŃgulateur du polyphŃnisme du mode de reproduction, en plus des hormones juvŃniles.

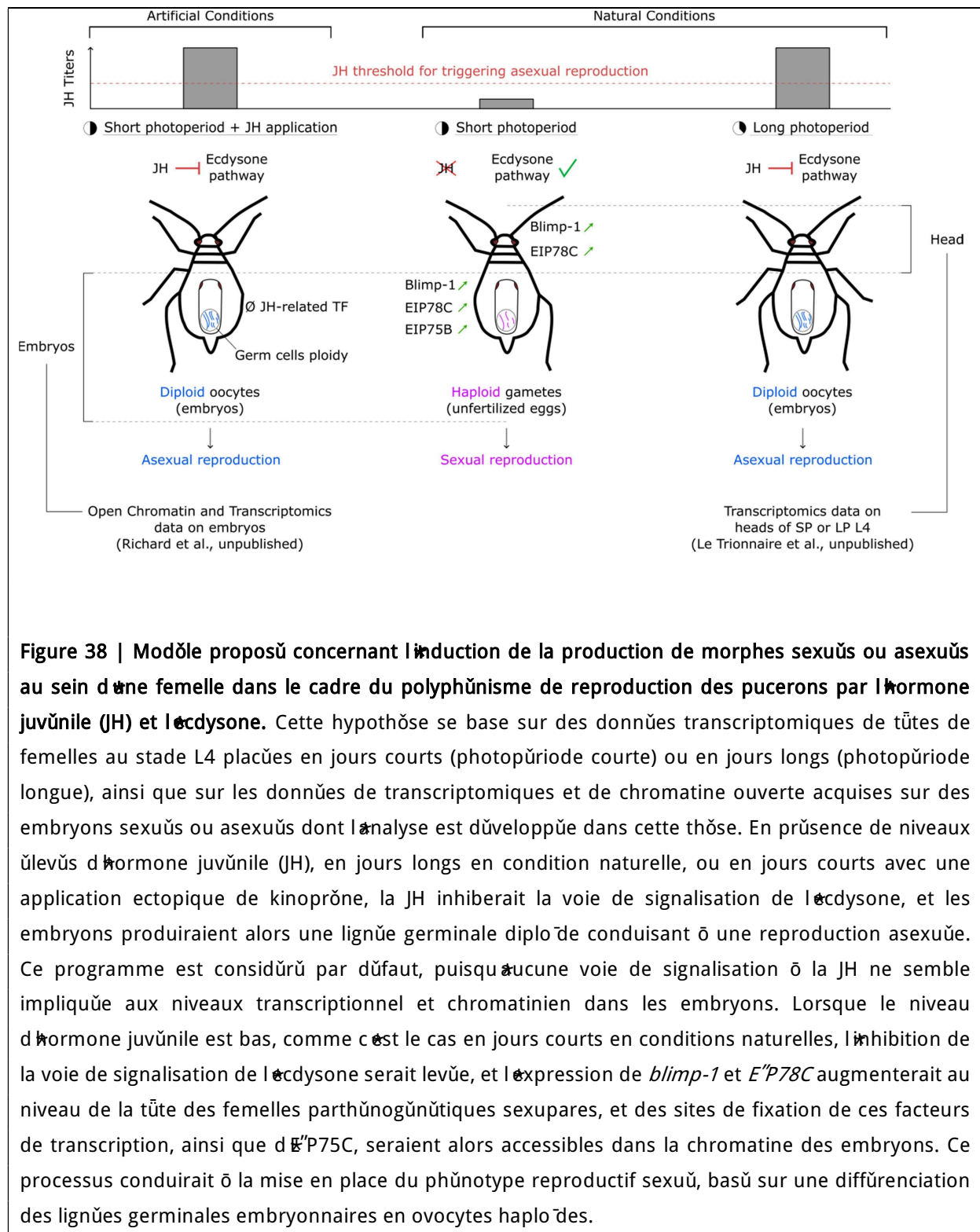
Chez les insectes il est connu que les voies de signalisation de l'hormone juvŃnile et de l'ecdysone sont antagonistes pour de nombreux processus et c'est souvent la rŃsultante de cet antagonisme qui permet le bon dŃveloppement ou le bon fonctionnement de ce processus. Par exemple, au niveau des intestins d'*Aedes aegypti*, d'*Heliothis virescens* et de *Tribolium castaneum*, l'expression des gŃnes *EcR*, *usp*, *E74*, *E75*, *HR3*, *broad*, *E93*, *dronc* et *drice* est induite par le 20-hydroxyecdysone en l'absence d'hormone juvŃnile, et permet le dŃveloppement des intestins par une prolifŃration cellulaire. Ce processus est bloquŃ en prŃsence d'hormone juvŃnile, montrant ainsi la capacitŃ de l'hormone juvŃnile Ń inhiber les voies de signalisation de l'ecdysone [Wu *et al.*, 2006 ; Parthasarathy *et al.*, 2008 ; Parthasarathy et Palli, 2008 ; Jindra *et al.*, 2013]. "Il est donc possible de proposer l'hypothŃse selon laquelle, chez le puceron du pois et concernant le polyphŃnisme de reproduction, l'hormone juvŃnile n'agisse en jours longs que comme un inhibiteur de la voie de signalisation de l'ecdysone conduisant Ń une reproduction sexuŃe des embryons en l'absence d'hormone juvŃnile en jours courts (Figure 38).

En plus d'Ńtre supportŃe au niveau des embryons, cette hypothŃse est Ńgalement validŃe par l'analyse prŃliminaire de transcriptomes de tŃtes de femelles parthŃnogŃnŃtiques au 4<sup>Ńme</sup> stade larvaire placŃes en jours courts (sexupares) ou en jours longs, oŃ les homologues des facteurs de transcription en relation avec l'ecdysone, *BLMP-1* et *E'P75B*, sont surexprimŃs au niveau des tŃtes de femelles placŃes en jours courts (Tableau V), qui sont donc destinŃes Ń produire une descendance Ń reproduction sexuŃe.

Les rŃsultats que nous avons gŃnŃrŃs suggŃrent donc qu'en photopŃriode longue, l'hormone juvŃnile est synthŃtisŃe en quantitŃ suffisante pour inhiber la voie de signalisation de l'ecdysone, conduisant ainsi Ń la production d'embryons asexuŃs prŃsentant des ovocytes diploŃdes, sans intervention d'une voie de signalisation particuliŃre en direction des embryons, faisant de la reproduction asexuŃe une trajectoire dŃveloppementale par dŃfaut. En photopŃriode courte, la synthŃse d'hormone juvŃnile diminue, levant l'inhibition de la voie de signalisation de l'ecdysone, permettant une augmentation de la transcription de *blimp-1* et de *E'P75B* au niveau de la tŃte des femelles et des embryons et/ou de leur lignŃe germinale, conduisant Ń la production d'embryons sexuŃs prŃsentant des ovocytes haploŃdes. Afin de vŃrifier cette hypothŃse, il faudrait rŃaliser une titration de l'ecdysone au niveau de femelles placŃes en jours courts et en jours longs, ainsi que leur traitement avec de l'ecdysone (ou un analogue) ou avec des inhibiteurs de l'ecdysone, puis Ńtudier l'impact de ces traitement hormonaux sur le phŃnotype de la descendance.

**Tableau V | LOC100161053 et LOC100159482, des homologues respectifs de E<sup>7</sup>P78C et de Blimp-1 de *D. melanogaster*, sont diffŃrentiellement exprimŃs entre des tŃtes de femelles au 4<sup>Ńme</sup> stade larvaire placŃes en jours long (LD) ou en jours courts (SD). Un protocole d'analyse d'expression diffŃrentielle similaire Ń celui dŃcrit par [Law *et al.*, 2016] a ŃtŃ rŃalisŃ.**

Aphid gene	Drosophila	L4-LD	L4-SD	FC	p-value
LOC100161053	Eip78C	48,70	129,64	1,42	5,04E-04
LOC100159482	Blimp-1	30,14	183,39	2,60	4,97E-04



**Figure 38 | ModŃle proposŃ concernant l'induction de la production de morphes sexuŃs ou asexuŃs au sein d'une femelle dans le cadre du polyphŃnisme de reproduction des pucerons par l'hormone juvŃnile (JH) et l'ecdysone.** Cette hypothŃse se base sur des donnŃes transcriptomiques de tŃtes de femelles au stade L4 placŃes en jours courts (photopŃriode courte) ou en jours longs (photopŃriode longue), ainsi que sur les donnŃes de transcriptomiques et de chromatine ouverte acquises sur des embryons sexuŃs ou asexuŃs dont l'analyse est dŃveloppŃe dans cette thŃse. En prŃsence de niveaux ŃlevŃs d'hormone juvŃnile (JH), en jours longs en condition naturelle, ou en jours courts avec une application ectopique de kinoprŃne, la JH inhiberait la voie de signalisation de l'ecdysone, et les embryons produiraient alors une lignŃe germinale diploŃde conduisant Ń une reproduction asexuŃe. Ce programme est considŃrŃ par dŃfaut, puisque aucune voie de signalisation Ń la JH ne semble impliquŃe aux niveaux transcriptionnel et chromatinien dans les embryons. Lorsque le niveau d'hormone juvŃnile est bas, comme c'est le cas en jours courts en conditions naturelles, l'inhibition de la voie de signalisation de l'ecdysone serait levŃe, et l'expression de *blimp-1* et *EIP78C* augmenterait au niveau de la tŃte des femelles parthŃnogŃnŃtiques sexupares, et des sites de fixation de ces facteurs de transcription, ainsi que de *EIP75C*, seraient alors accessibles dans la chromatine des embryons. Ce processus conduirait Ń la mise en place du phŃnotype reproductif sexuŃ, basŃ sur une diffŃrenciation des lignŃes germinales embryonnaires en ovocytes haploŃdes.

## V'''. Perspectives et conclusion

Ma thèse visait à comprendre l'implication des réorganisations de la chromatine et leur impact sur la transcription du génome du puceron du pois dans le cadre de son cycle de vie, et plus particulièrement de son polyphénisme de reproduction et de sa compensation de dose. Ainsi j'ai identifié un mécanisme putatif de compensation de dose spécifique aux individus mâles et basé sur une augmentation de l'accessibilité de la chromatine de leur unique chromosome X. De plus, j'ai été amené à développer des pistes concernant d'une part l'implication putative de l'ecdysone dans la mise en place du phénotype reproductif sexuel dans le cadre du polyphénisme de reproduction et l'aide de facteurs de transcription inductibles par l'ecdysone dont l'expression est différentielle au niveau des têtes de femelles, et dont des sites d'ouverture de la chromatine sont spécifiques des embryons sexuels. D'autre part il s'avère que ces embryons sexuels présentent dès leur détermination sexuelle une modification des programmes génétiques associés aux lipides qui serait régulée via des réorganisations fines de la chromatine. Enfin, de par l'absence d'une voie de signalisation de l'hormone juvénile chez les embryons asexuels, j'ai suggéré un rôle d'inhibition de la voie de l'ecdysone par l'hormone juvénile plus que l'induction directe d'un programme génétique par la JH permettant la reproduction asexuelle des embryons. La plupart des analyses menées durant ma thèse se sont heurtées aux limites du génome du puceron du pois qui seront sans doute levées prochainement par l'utilisation de méthodes telles que le PacB<sup>2</sup>O et le Hi-C qui permettront un nouvel assemblage du génome. Le Hi-C permettra aussi d'étudier la structure tridimensionnelle de la chromatine du chromosome X des mâles et des femelles dans le cadre de la compensation de dose, afin d'identifier d'éventuels TADs différentiels, ou bien un regroupement des gènes compensés du chromosome X qui seraient alors régulés en cluster par un mécanisme épigénétique augmentant l'accessibilité de ces gènes et leur transcription. La définition précise du mécanisme permettant cette compaction de la chromatine du X des mâles pourra être réalisée par l'étude de marques histones actives et répressives au niveau de ce chromosome par ChIP-seq. Concernant le polyphénisme de reproduction, le Hi-C permettrait d'analyser les pics intergénomiques distaux des gènes, et donc d'obtenir une vision plus exhaustive des régulations transcriptionnelles ayant lieu au niveau de la chromatine des embryons sexuels et asexuels. Enfin, le développement de méthodes de mutagenèse ciblées telles que le CRISPR (récemment développé dans l'équipe), permettra de réaliser des études de génétique fonctionnelle chez les pucerons, amorçant l'annotation fonctionnelle du génome de ces hémiptères, et permettant la mutation ciblée de gènes d'intérêt dans le cadre de la compensation de dose tels que les homologues de *MSL*, et dans le polyphénisme de reproduction comme par exemple les homologues de *blimp-1* ou *E<sup>78C</sup>*. Les données de chromatine ouverte et d'expression génomiques au cours de ma thèse continueront donc de prendre leur sens au cours des prochaines années grâce au développement de nouvelles méthodes, permettant une vision plus intégrative des régulations transcriptionnelles impliquées aux différents niveaux du cycle de vie des pucerons.

## "X. R f rences

ABLES E.T., BO'S K.E., GARC'A C.A., et DRUMMOND-BARBOSA D., 2015,   Ecdysone response gene E78 controls ovarian germline stem cell niche formation and follicle survival in *Drosophila*.  , *Developmental Biology*, vol. 400, n, 1 : 33  42.

AKAG' K., SARHAN M., SULTAN A.R., N'SH'DA H., KO'E A., NAKAYAMA T., et UEDA H., 2016,   A biological timer in the fat body comprising Blimp-1,  ftz-f1 and Shade regulates pupation timing in *Drosophila melanogaster*.  , *Development*, vol. 143, n, 13 : 2410  2416.

ALEKSEYENKO Artyom A., PENG Shouyong, LARSCHAN Erica, GORCHAKOV Andrey A., LEE Ok-Kyung, KHARCHENKO Peter, MCGRATH Sean D., WANG Charlotte ., MARD'S Elaine R., PARK Peter J., et KURODA Mitzi ., 2008,   A sequence motif within chromatin entry sites directs MSL establishment on the *Drosophila* X chromosome  , *Cell*, vol. 134, n, 4 : 599  609.

ALLAN J., HARTMAN P. G., CRANE-ROB'NSON C., et AV'LES F. X., 1980,   The structure of histone H1 and its location in chromatin  , *Nature*, vol. 288, n, 5792 : 675  679.

ALL'S C. David et JENUWE'N Thomas, 2016,   The molecular hallmarks of epigenetic control  , *Nature Reviews Genetics*, vol. 17, n, 8 : 487  500.

AMMAR E. D. et NAULT L. R., 1985,   Assembly and accumulation sites of maize mosaic virus in its planthopper vector  , *Intervirology*, vol. 24, n, 1 : 33  41.

AOK' Shigeyuki, 1977,   *Colophina clematis* (Homoptera, Pemphigidae), an aphid Species with   Soldiers  ,  , vol. 45, n, 2 : 276  282.

AOK' Shigeyuki, 1980,   Occurrence of a simple labor in a gall aphid, *Pemphigus dorocola* (Homoptera, Pemphigidae)  ,  , vol. 48, n, 1 : 71  73.

AOK' Shigeyuki, 1987,   Evolution of sterile soldiers in aphids  , *Animal societies: theories and facts* : 53  55.

ARATAN' Satoko, KAGEYAMA Yuji, NAKAMURA Akira, FUJ'TA Hidetoshi, FUJ' Ryoji, N'SH'OKA Kusuki, et NAKAJ'MA Toshihiro, 2008,   MLE activates transcription via the minimal transactivation domain in *Drosophila*  , *International Journal of Molecular Medicine*, vol. 21, n, 4 : 469  476.

AXEL Richard, CEDAR Howard, et FELSENFELD Gary, 1973,   Synthesis of Globin Ribonucleic Acid from Duck-Reticulocyte Chromatin "n Vitro  , *Proceedings of the National Academy of Sciences of the United States of America*, vol. 70, n, 7 : 2029  2032.

BA'LEY Timothy L., BODEN Mikael, BUSKE Fabian A., FR'TH Martin, GRANT Charles E., CLEMENT' Luca, REN Jingyuan, L' Wilfred W., et NOBLE William S., 2009,   MEME Suite: tools for motif discovery and searching  , *Nucleic Acids Research*, vol. 37, n, suppl\_2 : W202  W208.

BAKER Dean A., NOLAN Tony, F'SCHER Bettina, P'NDER Alex, CR'SANT' Andrea, et RUSSELL Steven, 2011,   A comprehensive gene expression atlas of sex- and tissue-specificity in the malaria vector, *Anopheles gambiae*  , *BMC Genomics*, vol. 12 : 296.

BALLAS Nurit et MANDEL Gail, 2005,   The many faces of REST oversee epigenetic programming of neuronal genes  , *Current Opinion in Neurobiology*, vol. 15, n, 5 : 500  506.

BANERJĀ J., RUSCONĀ S., et SCHAFFNER W., 1981, <sup>2</sup> Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences p, *Cell*, vol. 27, n, 2 Pt 1 : 299 308.

BARBERĀ M., MENGUAL B., COLLANTES-ALEGRE J. M., CORTĀS T., GONZĀLEZ A., et MART NEZ-TORRES D., 2013, <sup>2</sup> Identification, characterization and analysis of expression of genes encoding arylalkylamine N-acetyltransferases in the pea aphid *Acyrtosiphon pisum* p, *Insect Molecular Biology*, vol. 22, n, 6 : 623 634.

BARBERĀ Miquel, COLLANTES-ALEGRE Jorge Mariano, et MART NEZ-TORRES David, 2017, <sup>2</sup> Characterisation, analysis of expression and localisation of circadian clock genes from the perspective of photoperiodism in the aphid *Acyrtosiphon pisum* p, *Insect Biochemistry and Molecular Biology*, vol. 83 : 54 67.

BARDEN J. A., BENNETTS B. H., DOS REMEDĀS C. G., HAMBLY B. D., MĀKĀ M., et PHĀLLĀPS L., 1988, <sup>2</sup> Structure and function of contractile proteins in muscle fibres p, *Australian Paediatric Journal*, vol. 24 Suppl 1 : 31 33.

BARTH Teresa K. et MHOF Axel, 2010, <sup>2</sup> Fast signals and slow marks: the dynamics of histone modifications p, *Trends in Biochemical Sciences*, vol. 35, n, 11 : 618 626.

BASHAW Greg J. et BAKER Bruce S., 1997, <sup>2</sup> The Regulation of the *Drosophila* msl-2 Gene Reveals a Function for Sex-lethal in Translational Control p, *Cell*, vol. 89, n, 5 : 789 798.

BAULCOMBE David, 2004, <sup>2</sup> RNA silencing in plants p, *Nature*, vol. 431, n, 7006 : 356 363.

BELL Adam C., WEST Adam G., et FELSENFELD Gary, 1999, <sup>2</sup> The protein CTCF is required for the enhancer blocking activity of vertebrate insulators p, *Cell*, vol. 98, n, 3 : 387 396.

BELMONT A. S. et BRUCE K., 1994, <sup>2</sup> Visualization of G1 chromosomes: a folded, twisted, supercoiled chromonema model of interphase chromatid structure p, *The Journal of Cell Biology*, vol. 127, n, 2 : 287 302.

BELTON Jon-Matthew, MCCORD Rachel Patton, GĀBCUS Johan Harmen, NAUMOVA Natalia, ZHAN Ye, et DEKKER Job, 2012, <sup>2</sup> Hi-C: a comprehensive technique to capture the conformation of genomes p, *Methods (San Diego, Calif.)*, vol. 58, n, 3 : 268 276.

BENTON TG et FOSTER WA, 1992, <sup>2</sup> Altruistic housekeeping in a social aphid p, *Proceedings of the Royal Society of London B: Biological Sciences*, vol. 247, n, 1320 : 199 202.

BERNSTĀN E., CAUDY A. A., HAMMOND S. M., et HANNON G. J., 2001, <sup>2</sup> Role for a bidentate ribonuclease in the initiation step of RNA interference p, *Nature*, vol. 409, n, 6818 : 363 366.

BESTOR T. H., 1990, <sup>2</sup> DNA methylation: evolution of a bacterial immune function into a regulator of gene expression and genome structure in higher eukaryotes p, *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, vol. 326, n, 1235 : 179 187.

BESTOR T. H., 2000, <sup>2</sup> The DNA methyltransferases of mammals p, *Human Molecular Genetics*, vol. 9, n, 16 : 2395 2402.

BEWĀCK Adam J., VOGEL Kevin J., MOORE Allen J., et SCHMĀTZ Robert J., 2017, <sup>2</sup> Evolution of DNA Methylation across Insects p, *Molecular Biology and Evolution*, vol. 34, n, 3 : 654 665.

BĀCKHART Derek M., ROSEN Benjamin D., KOREN Sergey, SAYRE Brian L., HASTĀE Alex R., CHAN Saki, LEE Joyce, LAM Ernest T., LĀCHKO Āvan, SULLĀVAN Shawn T., BURTON Joshua N., HUSON Heather J., NYSTROM John C., KELLEY Christy M., HUTCHĀSON Jana L., ZHOU Yang, SUN Jiajie, CRĀSĀ Alessandra, PONCE DE LEĀN F. Abel,

SCHWARTZ John C., HAMMOND John A., WALDBESER Geoffrey C., SCHROEDER Steven G., LU George E., DUNHAM Maitreya J., SHENDURE Jay, SONSTEGARD Tad S., PHILLIPPY Adam M., VAN TASSELL Curtis P., et SMITH Timothy P. L., 2017, Single-molecule sequencing and chromatin conformation capture enable de novo reference assembly of the domestic goat genome, *Nature Genetics*, vol. 49, n. 4 : 643-650.

BIRD A., 1992, The essentials of DNA methylation, *Cell*, vol. 70, n. 1 : 5-8.

BIRD A. P. et WOLFFE A. P., 1999, Methylation-induced repression--belts, braces, and chromatin, *Cell*, vol. 99, n. 5 : 451-454.

BIRNEY Ewan, STAMATOYANNOPOULOS John A., DUTTA Anindya, GUGH Roderic, GNGERAS Thomas R., MARGULIES Elliott H., WENG Zhiping, SNYDER Michael, DERMITZAKIS Emmanouil T., THURMAN Robert E., KUEHN Michael S., TAYLOR Christopher M., NEPH Shane, KOCH Christoph M., ASTHANA Saurabh, MALHOTRA Ankit, ADZHUBEI Ivan, GREENBAUM Jason A., ANDREWS Robert M., FLICEK Paul, BOYLE Patrick J., CAO Hua, CARTER Nigel P., CLELLAND Gayle K., DAVS Sean, DAY Nathan, DHAM Pawandeep, DILLON Shane C., DORSCHNER Michael O., FEGLER Heike, GERES Paul G., GOLDY Jeff, HAWRYLYCZ Michael, HAYDOCK Andrew, HUMBERT Richard, JAMES Keith D., JOHNSON Brett E., JOHNSON Ericka M., FRUM Tristan T., ROSENZWEIG Elizabeth R., KARNAN Neerja, LEE Kirsten, LEFEBVRE Gregory C., NAVAS Patrick A., NER FIDENCIO, PARKER Stephen C. J., SABO Peter J., SANDSTROM Richard, SHAFER Anthony, VETRE David, WEAVER Molly, WILCOX Sarah, YU1 Man, COLLINS Francis S., DEKKER Job, LEB Jason D., TULLIUS Thomas D., CRAWFORD Gregory E., SUNYAEV Shamil, NOBLE William S., DUNHAM Ian, DENOEUDE France, REYMOND Alexandre, KAPRANOV Philipp, ROZOWSKY Joel, ZHENG Deyou, CASTELO Robert, FRANKSH Adam, HARROW Jennifer, GHOSH Srinika, SANDELIN Albin, HOFACKER Ivo L., BAERTSCH Robert, KEEFE Damian, DIKE Sujit, CHENG Jill, HIRSCH Heather A., SEKNGER Edward A., LAGARDE Julien, ABRIL Josep F., SHAHAB Atif, FLAMM Christoph, FRED Claudia, HACKERMILLER Jorg, HERTEL Jana, LINDEMEYER Manja, MISSAL Kristin, TANZER Andrea, WASHETL Stefan, KORBEL Jan, EMANUELSSON Olof, PEDERSEN Jakob S., HOLROYD Nancy, TAYLOR Ruth, SWARBRECK David, MATTHEWS Nicholas, DICKSON Mark C., THOMAS Daryl J., WEIRAUCH Matthew T., GILBERT James, DRENKOW Jorg, BELL Ian, ZHAO Xiaodong, SRINIVASAN K. G., SUNG Wing-Kin, OOHONG Sain, CHU Kuo Ping, FOSSAC Sylvain, ALOTO Tyler, BRENT Michael, PACTHER Lior, TRESS Michael L., VALENCIA Alfonso, CHOO Siew Woh, CHOO Chiou Yu, UCLA Catherine, MANZANO Caroline, WYSS Carine, CHEUNG Evelyn, CLARK Taane G., BROWN James B., GANESH Madhavan, PATEL Sandeep, TAMMANA Hari, CHRAST Jacqueline, HENRICHSEN Charlotte N., KAI Chikatoshi, KAWA Jun, NAGALAKSHMI Ugrappa, WU Jiaqian, LIAN Zheng, LIAN Jin, NEWBURGER Peter, ZHANG Xueqing, BICKEL Peter, MATTICK John S., CARNINCI Piero, HAYASHIZAKI Yoshihide, WEISSMAN Sherman, HUBBARD Tim, MYERS Richard M., ROGERS Jane, STADLER Peter F., LOWE Todd M., WEI Chia-Lin, RUAN Yijun, STRUHL Kevin, GERSTEIN Mark, ANTONARAKIS Stylianos E., FU Yutao, GREEN Eric D., KARAIJZ Ulax, SEPEL Adam, TAYLOR James, LEFER Laura A., WETTERSTRAND Kris A., GOOD Peter J., FEINGOLD Elise A., GUYER Mark S., COOPER Gregory M., ASHMENOS George, DEWEY Colin N., HOU Minmei, NIKOLAEV Sergey, MONTOYA-BURGOS Juan J., LIJTYNOJA Ari, WHELAN Simon, PARDI Fabio, MASSINGHAM Tim, HUANG Haiyan, ZHANG Nancy R., HOLMES Ian, MULLIKIN James C., URETA-VEDAL Abel, PATEN Benedict, SERINGHAUS Michael, CHURCH Deanna, ROSENBLOOM Kate, KENT W. James, STONE Eric A., BATZOGLOU Serafim, GOLDMAN Nick, HARDISON Ross C., HAUSSLER David, MILLER Webb, SADOW Arend, TRINKLEIN Nathan D., ZHANG Zhengdong D., BARRERA Leah, STUART Rhona, KING David C., AMEUR Adam, ENROTH Stefan, BEDA Mark C., KIM Jonghwan, BHINGE Akshay A., JIANG Nan, LU Jun, YAO Fei, VEGA Vinsensius B., LEE Charlie W. H., NG Patrick, YANG Annie, MOQTADER Zarmik, ZHU Zhou, XU Xiaoqin, SQUAZZO Sharon, OBERLEY Matthew J., NIMAN David, SINGER Michael A., RICHMOND Todd A., MUNN Kyle J., RADA-GLESAS Alvaro, WALLERMAN Ola, KOMOROWSKI Jan, FOWLER Joanna C., COUTTET Phillippe, BRUCE Alexander W., DOVEY Oliver M., ELLIS Peter D., LANGFORD Cordelia F., NIX David A., EUSKRICHEN Ghia, HARTMAN Stephen, URBAN Alexander E., KRAUS Peter, CALCAR Sara Van, HENTZMAN Nate, KIM Tae Hoon, WANG Kun, QU Chunxu, HON Gary, LUNA Rosa, GLASS Christopher K., ROSENFELD M. Geoff, ALDRED Shelley Force, COOPER Sara J., HALEES Anason, LIAN Jane M., SHULHA Hennady P., ZHANG Xiaoling, XU Mousheng, HADAR Jaafar N. S., YU Yong, BIRNEY Ewan, VYER Vishwanath R., GREEN Roland D., WADELINUS Claes, FARNHAM Peggy J., REN Bing, HARTE Rachel A., HENRICHS Angie S., TRUMBOWER Heather, CLAWSON Hiram, HILLMAN-JACKSON Jennifer, ZWEIG Ann S., SMITH Kayla, THAKKAPALLAYAL Archana, BARBER Galt, KUHN Robert M., KAROLCHAK Donna, ARMENGOL Lluís, BIRD Christine P., BAKKER Paul W. de, KERN Andrew D., LOPEZ-

BIGAS Nuria, MARTIN Joel D., STRANGER Barbara E., WOODROFFE Abigail, DAVYDOV Eugene, D'AMAS Antigone, EYRAS Eduardo, HALLGRIMSDOTTIR Ingileif B., HUPPERT Julian, ZODY Michael C., ABECAS Gonçalo R., ESTVILL Xavier, BOUFFARD Gerard G., GUAN Xiaobin, HANSEN Nancy F., DOL Jacquelyn R., MADURO Valerie V. B., MASKER Baishali, MCDOWELL Jennifer C., PARK Morgan, THOMAS Pamela J., YOUNG Alice C., BLAKESLEY Robert W., MUZNY Donna M., SODERGREN Erica, WHEELER David A., WORLEY Kim C., JIANG Huaiyang, WEINSTOCK George M., GIBBS Richard A., GRAVES Tina, FULTON Robert, MARDIS Elaine R., WILSON Richard K., CLAMP Michele, CUFF James, GNERRE Sante, JAFFE David B., CHANG Jean L., LINDBLAD-TOH Kerstin, LANDER Eric S., KORABNE Maxim, NEFEDOV Mikhail, OSOEGAWA Kazutoyo, YOSHINAGA Yuko, ZHU Baoli, et JONG Pieter J. de, 2007, <sup>2</sup> Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project  *Nature*, vol. 447, n. 7146 : 799-816.

BLACKMAN R. L., 1978, <sup>2</sup> Early development of the parthenogenetic egg in three species of aphids (homoptera: Aphididae)  *International Journal of Insect Morphology and Embryology*, vol. 7, n. 1 : 33-44.

BLACKMAN R. L., 1980, <sup>2</sup> Chromosome numbers in the Aphididae and their taxonomic significance  *Systematic Entomology*, vol. 5, n. 1 : 7-25.

BLACKMAN R. L. et HALES D. F., 1986, <sup>2</sup> Behaviour of the X chromosomes during growth and maturation of parthenogenetic eggs of *Amphorophora tuberculata* (Homoptera, Aphididae), in relation to sex determination  *Chromosoma*, vol. 94, n. 1 : 59-64.

BLACKMAN ROGER L, 1987, <sup>2</sup> Reproduction, cytogenetics and development  *Aphids: their biology, natural enemies and control*, vol. 2 : 163-195.

BOERJAN Bart, SAS Filip, ERNST Ulrich R., TOBACK Julie, LEMERE Filip, VANDEGEHUCHTE Michiel, JANSSEN Colin, BADSCO Liesbeth, MARCHAL Elisabeth, VERLINDEN Heleen, SCHOofs Liliane, et DE LOOF Arnold, 2011, <sup>2</sup> Locust phase polyphenism: does epigenetic precede endocrine regulation?  *GENERAL AND COMPARATIVE ENDOCRINOLOGY*, vol. 173, n. 1 : 120-128.

BONASO Roberto, 2014, <sup>2</sup> The role of chromatin and epigenetics in the polyphenisms of ant castes  *Briefings in Functional Genomics*, vol. 13, n. 3 : 235-245.

BONASO Roberto, LI Qiye, LIAN Jinmin, MUTT Navdeep S., JIN Lijun, ZHAO Hongmei, ZHANG Pei, WEN Ping, XIANG Hui, DING Yun, JIN Zonghui, SHEN Steven S., WANG Zongji, WANG Wen, WANG Jun, BERGER Shelley L., LEBERG JArgen, ZHANG Guojie, et REINBERG Danny, 2012, <sup>2</sup> Genome-wide and caste-specific DNA methylomes of the ants *Camponotus floridanus* and *Harpegnathos saltator*  *Current biology*, vol. 22, n. 19 : 1755-1764.

BONASO Roberto, ZHANG Guojie, YE Chaoyang, MUTT Navdeep S., FANG Xiaodong, QIN Nan, DONAHUE Greg, YANG Pengcheng, LI Qiye, LI Cai, ZHANG Pei, HUANG Zhiyong, BERGER Shelley L., REINBERG Danny, WANG Jun, et LEBERG JArgen, 2010, <sup>2</sup> Genomic comparison of the ants *Camponotus floridanus* and *Harpegnathos saltator*  *Science (New York, N.Y.)*, vol. 329, n. 5995 : 1068-1071.

BOSTICK Magnolia, KIM Jong Kyong, ESTÈVE Pierre-Olivier, CLARK Amander, PRADHAN Sriharsa, et JACOBSEN Steven E., 2007, <sup>2</sup> UHRF1 plays a role in maintaining DNA methylation in mammalian cells  *Science (New York, N.Y.)*, vol. 317, n. 5845 : 1760-1764.

BOYLE Alan P., DAVIS Sean, SHULHA Hennady P., MELTZER Paul, MARGULIES Elliott H., WENG Zhiping, FUREY Terrence S., et CRAWFORD Gregory E., 2008a, <sup>2</sup> High-Resolution Mapping and Characterization of Open Chromatin across the Genome  *Cell*, vol. 132, n. 2 : 311-322.

BOYLE Alan P., GU"NEY Justin, CRAWFORD Gregory E., et FUREY Terrence S., 2008b, <sup>2</sup> F-Seq: a feature density estimator for high-throughput sequence tags *Þ*, *Bioinformatics (Oxford, England)*, vol. 24, n, 21 : 2537 <sup>2</sup>538.

BRAULT Vůronique, BLANC Stůphane, et JACQUOT Emmanuel, 2007, <sup>2</sup> How aphids transmit virus diseases to plants *Þ*, *Biofutur*, vol. 247 : 40 <sup>2</sup>4.

BRAULT Vůronique, UZEST Maryline, MONS"ON Baptiste, JACQUOT Emmanuel, et BLANC Stůphane, 2010, <sup>2</sup> Aphids as transport devices for plant viruses *Þ*, *Comptes Rendus Biologies*, vol. 333, n, 6 : 524 <sup>2</sup>38.

BRENNECKE Julius, MALONE Colin D., ARAV"N Alexei A., SACH"DANANDAM Ravi, STARK Alexander, et HANNON Gregory J., 2008, <sup>2</sup> An epigenetic role for maternally inherited piRNAs in transposon silencing *Þ*, *Science (New York, N.Y.)*, vol. 322, n, 5906 : 1387 <sup>2</sup>392.

BROCKDORFF N., ASHWORTH A., KAY G. F., MCCABE V. M., NORR"S D. P., COOPER P. J., SW"FT S., et RASTAN S., 1992, <sup>2</sup> The product of the mouse Xist gene is a 15 kb inactive X-specific transcript containing no conserved ORF and located in the nucleus *Þ*, *Cell*, vol. 71, n, 3 : 515 <sup>2</sup>26.

BROCKDORFF Neil et TURNER Bryan M., 2015, <sup>2</sup> Dosage Compensation in Mammals *Þ*, *Cold Spring Harbor Perspectives in Biology*, vol. 7, n, 3 : a019406.

BRONNER "raad F., QUAL" Michael A., TURNER Daniel J., et SWERDLOW Harold, 2009, <sup>2</sup> "mproved Protocols for "llumina Sequencing *Þ*, *Current protocols in human genetics / editorial board, Jonathan L. Haines ...[et al.]*, vol. 0 18.

BROWER-TOLAND Brent, F"NDLEY Seth D., J"ANG Ling, L"U Li, Y"N Hang, DUS Monica, ZHOU Pei, ELG"N Sarah C.R., et L"N Haifan, 2007, <sup>2</sup> Drosophila P"W" associates with chromatin and interacts directly with HP1a *Þ*, *Genes & Development*, vol. 21, n, 18 : 2300 <sup>2</sup>311.

BROWN C. J., BALLAB"O A., RUPERT J. L., LAFREN"ERE R. G., GROMPE M., TONLORENZ" R., et W"LLARD H. F., 1991, <sup>2</sup> A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome *Þ*, *Nature*, vol. 349, n, 6304 : 38 <sup>2</sup>4.

BROWN C. J., HENDR"CH B. D., RUPERT J. L., LAFREN"ERE R. G., X"NG Y., LAWRENCE J., et W"LLARD H. F., 1992, <sup>2</sup> The human X"ST gene: analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus *Þ*, *Cell*, vol. 71, n, 3 : 527 <sup>2</sup>42.

BUENROSTRO Jason, G"RES" Paul, ZABA Lisa, CHANG Howard, et GREENLEAF William, 2013a, <sup>2</sup> Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position *Þ*, *Nature Methods*, vol. 10, n, 12 : 1213 <sup>2</sup>218.

BUENROSTRO Jason, 2013b, <sup>2</sup> Transposition of native chromatin for multimodal regulatory analysis and personal epigenomics *Þ*, *Nature methods*, vol. 10, n, 12 : 1213 <sup>2</sup>218.

BUENROSTRO Jason, WU Beijing, CHANG Howard, et GREENLEAF William, 2015, <sup>2</sup> ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide *Þ*, *Current protocols in molecular biology / edited by Frederick M. Ausubel ...[et al.]*, vol. 109 : 21.29.1-21.29.9.

BIHLER Marc, 2009, <sup>2</sup> RNA turnover and chromatin-dependent gene silencing *Þ*, *Chromosoma*.

BIHLER Marc, VERDEL Andrů, et MOAZED Danesh, 2006, <sup>2</sup> Tethering R"TS to a nascent transcript initiates RNAi- and heterochromatin-dependent gene silencing *Þ*, *Cell*, vol. 125, n, 5 : 873 <sup>2</sup>86.

BULL James J et CHARNOV Eric L, 1977, <sup>2</sup> Heredity - Abstract of article: Changes in the heterogametic mechanism of sex determination p, *Heredity*, vol. 39, n, 1 : 1  14.

CA" Yong, J"n Jingji, SWANSON Selene K., COLE Michael D., CHO" Seung Hyuk, FLORENS Laurence, WASHBURN Michael P., CONAWAY Joan W., et CONAWAY Ronald C., 2010, <sup>2</sup> Subunit composition and substrate specificity of a MOF-containing histone acetyltransferase distinct from the male-specific lethal (MSL) complex p, *The Journal of Biological Chemistry*, vol. 285, n, 7 : 4268  272.

CAMACHO Christiam, COULOUR"s George, AVAGYAN Vahram, MA Ning, PAPADOPOULOS Jason, BEALER Kevin, et MADDEN Thomas L., 2009, <sup>2</sup> BLAST+: architecture and applications p, *BMC bioinformatics*, vol. 10 : 421.

CERBO Vincenzo Di, MOHN Fabio, RYAN Daniel P., MONTELL"ER Emilie, KACEM Salim, TROPBERGER Philipp, KALL"s Eleni, HOLZNER Monika, HOERNER Leslie, FELDMANN Angelika, R"CHTER Florian Martin, BANN"STER Andrew J., M"TTLER Gerhard, M"CHAELS Jens, KHOCHB"n Saadi, FE"l Robert, SCHUEBELER Dirk, OWEN-HUGHES Tom, DAUJAT Sylvain, et SCHNE"DER Robert, 2014, <sup>2</sup> Acetylation of histone H3 at lysine 64 regulates nucleosome dynamics and facilitates transcription p, *eLife*, vol. 3 : e01632.

CHANG Chun-Che, LEE Wen-Chih, COOK Charles E., L"n Gee-Way, et CHANG Tschining, 2004, <sup>2</sup> Germ-plasm specification and germline development in the parthenogenetic pea aphid *Acyrtosiphon pisum*: Vasa and Nanos as markers p, *International Journal of Developmental Biology*, vol. 50, n, 4 : 413  21.

CHAPMAN Reginald Frederick, 1998, *The insects: structure and function*, Cambridge university press.

CHARLESWORTH Brian, 1996, <sup>2</sup> The evolution of chromosomal sex determination and dosage compensation p, *Current Biology*, vol. 6, n, 2 : 149  62.

CHEN De-Qiao, CAMPBELL Bruce C., et PURCELL Alexander H., 1996, <sup>2</sup> A New Rickettsia from a Herbivorous Insect, the Pea Aphid *Acyrtosiphon pisum* (Harris) p, *Current Microbiology*, vol. 33, n, 2 : 123  28.

CHONG J. A., TAP"A-RAM REZ J., K"m S., TOLEDO-ARAL J. J., ZHENG Y., BOUTROS M. C., ALTSHULLER Y. M., FROHMAN M. A., KRANER S. D., et MANDEL G., 1995, <sup>2</sup> REST: a mammalian silencer protein that restricts sodium channel gene expression to neurons p, *Cell*, vol. 80, n, 6 : 949  57.

CHU Chia-Ying et RANA Tariq M., 2007, <sup>2</sup> Small RNAs: regulators and guardians of the genome p, *Journal of Cellular Physiology*, vol. 213, n, 2 : 412  19.

CHUREAU Corinne, CHANTALAT Sophie, ROM"TO Antonio, GALVAN" Ang lique, DURET Laurent, AVNER Philip, et ROUGEULLE Claire, 2011, <sup>2</sup> Ftx is a non-coding RNA which affects Xist expression and chromatin structure within the X-inactivation center region p, *Human Molecular Genetics*, vol. 20, n, 4 : 705  18.

CLARK Michael B., JOHNSTON Rebecca L., "NOSTROZA-PONTA Mario, FOX Archa H., FORT"n Ellen, MOSCATO Pablo, D"NGER Marcel E., et MATT"CK John S., 2012, <sup>2</sup> Genome-wide analysis of long noncoding RNA stability p, *Genome Research*, vol. 22, n, 5 : 885  98.

CLEMENTS Adrienne, POUX Arienne N., LO Wan-Sheng, P"LLUS Lorraine, BERGER Shelley L., et MARMORSTE"n Ronen, 2003, <sup>2</sup> Structural Basis for Histone and Phosphohistone Binding by the GCN5 Histone Acetyltransferase p, *Molecular Cell*, vol. 12, n, 2 : 461  73.

COCKL"n Ross R. et WANG Mu, 2003, <sup>2</sup> Identification of Methylation and Acetylation Sites on Mouse Histone H3 Using Matrix-Assisted Laser Desorption/Ionization Time-of-Flight and Nano-electrospray Ionization Tandem Mass Spectrometry p, *Journal of Protein Chemistry*, vol. 22, n, 4 : 327  34.

COLL"NGS Clayton K., WADDELL Peter J., et ANDERSON John N., 2013, <sup>2</sup> Effects of DNA methylation on nucleosome stability p, *Nucleic Acids Research*, vol. 41, n, 5 : 2918  931.

CONRAD Thomas et AKHTAR Asifa, 2012, <sup>2</sup> Dosage compensation in *Drosophila melanogaster*: epigenetic fine-tuning of chromosome-wide transcription *p*, *Nature Reviews Genetics*, vol. 13, n, 2 : 123 [1134](#).

CONSORTIUM The ENCODE Project, 2012, <sup>2</sup> An integrated encyclopedia of DNA elements in the human genome *p*, *Nature*, vol. 489, n, 7414 : 57 [174](#).

COOPER Sara J, TRINKLEIN Nathan D, ANTON Elizabeth D, NGUYEN Loan, et MYERS Richard M, 2006, <sup>2</sup> Comprehensive analysis of transcriptional promoter structure and function in 1% of the human genome *p*, *Genome research*, vol. 16, n, 1 : 1 [10](#).

CORBETT Terry S. et HARDY Jim, 1985, <sup>2</sup> Juvenile hormone effects on polymorphism in the pea aphid, *Acyrtosiphon pisum* *p*, *Entomologia Experimentalis et Applicata*, vol. 38, n, 2 : 131 [135](#).

COSGROVE Michael S., BOEKE Jef D., et WOLBERGER Cynthia, 2004, <sup>2</sup> Regulated nucleosome mobility and the histone code *p*, *Nature Structural & Molecular Biology*, vol. 11, n, 11 : 1037 [1043](#).

COSTANZI Carl et PEHRSON John R., 1998, <sup>2</sup> Histone macroH2A1 is concentrated in the inactive X chromosome of female mammals *p*, *Nature*, vol. 393, n, 6685 : 599 [601](#).

COSTECHAREYRE Denis, BALMAND S verine, CONDEMI NE Guy, et RAHBO Yvan, 2012, <sup>2</sup> *Dickeya dadantii*, a Plant Pathogenic Bacterium Producing Cyt-Like Entomotoxins, Causes Septicemia in the Pea Aphid *Acyrtosiphon pisum* *p*, *PLoS ONE*, vol. 7, n, 1.

CRIDGE Andrew G., LEASK Megan P., DUNCAN Elizabeth J., et DEARDEN Peter K., 2015, <sup>2</sup> What Do Studies of Insect Polyphenisms Tell Us about Nutritionally-Triggered Epigenomic Changes and Their Consequences? *p*, *Nutrients*, vol. 7, n, 3 : 1787 [1797](#).

CSANKOVSKY G., PANNING B., BATES B., PEHRSON J. R., et JANSCH R., 1999, <sup>2</sup> Conditional deletion of Xist disrupts histone macroH2A localization but not maintenance of X inactivation *p*, *Nature Genetics*, vol. 22, n, 4 : 323 [324](#).

CUI Kairong et ZHAO Keji, 2012, <sup>2</sup> Genome-wide approaches to determining nucleosome occupancy in metazoans using MNase-Seq *p*, *Methods in Molecular Biology (Clifton, N.J.)*, vol. 833 : 413 [419](#).

CZECH Benjamin et HANNON Gregory J., 2016, <sup>2</sup> One Loop to Rule Them All: The Ping-Pong Cycle and piRNA-Guided Silencing *p*, *Trends in Biochemical Sciences*, vol. 41, n, 4 : 324 [337](#).

DABAN Joan-Ramon, 2011, <sup>2</sup> Electron microscopy and atomic force microscopy studies of chromatin and metaphase chromosome structure *p*, *Micron (Oxford, England: 1993)*, vol. 42, n, 8 : 733 [750](#).

D'ALESSANDRO Rosalba, KLAJN Andrijana, et MELDOLESI Jacopo, 2009, <sup>2</sup> Expression of dense-core vesicles and of their exocytosis are governed by the repressive transcription factor NRSF/REST *p*, *Annals of the New York Academy of Sciences*, vol. 1152 : 194 [200](#).

DALMAY T., HAMILTON A., RUDD S., ANGELL S., et BAULCOMBE D. C., 2000, <sup>2</sup> An RNA-dependent RNA polymerase gene in *Arabidopsis* is required for posttranscriptional gene silencing mediated by a transgene but not by a virus *p*, *Cell*, vol. 101, n, 5 : 543 [553](#).

DAVIE Kristofer, JACOBS Jelle, ATKINS Mardelle, POTTER Delphine, CHRISTAENS Valerie, HALDER Georg, et AERTS Stein, 2015, <sup>2</sup> Discovery of Transcription Factors and Regulatory Regions Driving In Vivo Tumor Development by ATAC-seq and FAIRE-seq Open Chromatin Profiling *p*, *PLoS Genetics*, vol. 11, n, 2.

DE MOED G. H., DE JONG G., et SCHARLOO W., 1997, <sup>2</sup> The phenotypic plasticity of wing size in *Drosophila melanogaster*: the cellular basis of its genetic variation *p*, *Heredity*, vol. 79 ( Pt 3) : 260 [267](#).

DEDRYVER C-A, LE GALL C J-F, MAH O F, S MON J-C, et DEDRYVER F, 2013,   The genetics of obligate parthenogenesis in an aphid species and its consequences for the maintenance of alternative reproductive modes  , *Heredity*, vol. 110, n, 1 : 39  15.

DEDRYVER Charles-Antoine, 2007,   Pucerons   des d g ts et des hommes  , *Biofutur*, n, 279 : 22  25.

DEDRYVER Charles-Antoine, HULLE Maurice, GALL C Jean-Fran ois Le, CA LLAUD Marina C., et S MON Jean-Christophe, 2001,   Coexistence in space and time of sexual and asexual populations of the cereal aphid *Sitobion avenae*  , *Oecologia*, vol. 128, n, 3 : 379  388.

DEKKER Job, 2006,   The three  Ls of chromosome conformation capture: controls, controls, controls  , *Nature Methods*, vol. 3, n, 1 : 17  21.

DEKKER Job, R PPE Karsten, DEKKER Martijn, et KLECKNER Nancy, 2002,   Capturing chromosome conformation  , *Science (New York, N.Y.)*, vol. 295, n, 5558 : 1306  1311.

DENG Xinxian, H ATT Joseph B., NGUYEN Di Kim, ERCAN Sevinc, STURG LL David, H LL ER LaDeana W., SCHLES NGER Felix, DAV S Carrie A., RE NKE Valerie J., G NGERAS Thomas R., SHENDURE Jay, WATERSTON Robert H., OL VER Brian, L EB Jason D., et D STECHE Christine M., 2011,   Evidence for compensatory upregulation of expressed X-linked genes in mammals, *Caenorhabditis elegans* and *Drosophila melanogaster*  , *Nature Genetics*, vol. 43, n, 12 : 1179  1185.

DENNES A., CROMME C., SURESH K., KUMAR N.S., EBLE J.A., HAHNENKAMP A., et POHLMANN R., 2005,   The Novel *Drosophila* Lysosomal Enzyme Receptor Protein Mediates Lysosomal Sorting in Mammalian Cells and Binds Mammalian and *Drosophila* GGA Adaptors.  , *Journal of Biological Chemistry*, vol. 280, n, 13 : 12849  12857.

DERR EN Thomas, JOHNSON Rory, BUSSOTT  Giovanni, TANZER Andrea, DJEBAL  Sarah, T LGNER Hagen, GUERNEC Gregory, MART N David, MERKEL Angelika, KNOWLES David G., LAGARDE Julien, VEERAVALL  Lavanya, RUAN Xiaolan, RUAN Yijun, LASSMANN Timo, CARN NC  Piero, BROWN James B., L POV CH Leonard, GONZALEZ Jose M., THOMAS Mark, DAV S Carrie A., SH EKHATTAR Ramin, G NGERAS Thomas R., HUBBARD Tim J., NOTREDAME Cedric, HARROW Jennifer, et GU G  Roderic, 2012,   The GENCODE v7 catalog of human long noncoding RNAs: Analysis of their gene structure, evolution, and expression  , *Genome Research*, vol. 22, n, 9 : 1775  1789.

DEVA AH Ballachanda N., CASE-BORDEN Chanelle, GEGONNE Anne, HSU Chih Hao, CHEN Qingrong, MEERZAMAN Daoud, DEY Anup, OZATO Keiko, et S NGER Dinah S., 2016,   BRD4 is a Histone Acetyltransferase that Evicts Nucleosomes from Chromatin  , *Nature structural & molecular biology*, vol. 23, n, 6 : 540  548.

DEVL N Robert H. et NAGAHAMA Yoshitaka, 2002,   Sex determination and sex differentiation in fish: an overview of genetic, physiological, and environmental influences  , *Aquaculture*, vol. 208, n, 3 : 191  196.

D CKMAN Mark J., KUCHARSK  Robert, MALESZKA Ryszard, et HURD Paul J., 2013,   Extensive histone post-translational modification in honey bees  , *Insect Biochemistry and Molecular Biology*, vol. 43, n, 2 : 125  137.

D EDER CHS Sven, 2014,   The four dimensions of noncoding RNA conservation  , *Trends in genetics: T G*, vol. 30, n, 4 : 121  123.

D XON A. F. G. et HOWARD M. T., 1986,   Dispersal in Aphids, A Problem in Resource Allocation  , in *Insect Flight*, Springer, Berlin, Heidelberg : 145  151.

DIXON Jesse R., SELVARAJ Siddarth, YUE Feng, KIM Audrey, LI Yan, SHEN Yin, HU Ming, LIU Jun S., et REN Bing, 2012, <sup>2</sup> Topological domains in mammalian genomes identified by analysis of chromatin interactions *p*, *Nature*, vol. 485, n. 7398 : 376-380.

DOUGLAS A. E., 1998, <sup>2</sup> Nutritional interactions in insect-Microbial Symbioses: Aphids and Their Symbiotic Bacteria Buchnera *p*, *Annual Review of Entomology*, vol. 43, n. 1 : 17-37.

DU Z. Q., LI L., LIU L., WANG X. F., et ZHOU G., 2007, <sup>2</sup> EVALUATION OF APHID TRANSMISSION ABILITIES AND VECTOR TRANSMISSION PHENOTYPES OF BARLEY YELLOW DWARF VIRUSES IN CHINA *p*, *Journal of Plant Pathology*, vol. 89, n. 2 : 251-259.

DUDCHENKO Olga, BATRA Sanjit S., OMER Arina D., NYQUIST Sarah K., HOEGER Marie, DURAND Neva C., SHAMIM Muhammad S., MACHOL Ido, LANDER Eric S., ADEN Aviva Presser, et ADEN Erez Lieberman, 2017, <sup>2</sup> De novo assembly of the Aedes aegypti genome using Hi-C yields chromosome-length scaffolds *p*, *Science (New York, N.Y.)*, vol. 356, n. 6333 : 92-95.

DUNCAN Elizabeth J., LEASK Megan P., et DEARDEN Peter K., 2013, <sup>2</sup> The pea aphid (*Acyrtosiphon pisum*) genome encodes two divergent early developmental programs *p*, *Developmental Biology*, vol. 377, n. 1 : 262-274.

EDEN Eran, NAVON Roy, STEINFELD Israel, LEVISON Doron, et YAKHINI Zohar, 2009, <sup>2</sup> GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists *p*, *BMC Bioinformatics*, vol. 10 : 48.

EHRICH Melanie, GAMA-SOSA Miguel A., CARREIRA Laura H., LJUNGDAHL Lars G., KUO Kenneth C., et GEHRKE Charles W., 1985, <sup>2</sup> DNA methylation in thermophilic bacteria: N<sup>4</sup>-methylcytosine, 5-methylcytosine, and N<sup>5</sup> methyladenine *p*, *Nucleic Acids Research*, vol. 13, n. 4 : 1399-1412.

ESSENBERG Joel C. et REUTER Gunter, 2009, <sup>2</sup> Cellular mechanism for targeting heterochromatin formation in Drosophila *p*, *International Review of Cell and Molecular Biology*, vol. 273 : 1-17.

ELBASHIR Sayda M., LENDECKEL Winfried, et TUSCHL Thomas, 2001, <sup>2</sup> RNA interference is mediated by 21- and 22-nucleotide RNAs *p*, *Genes & Development*, vol. 15, n. 2 : 188-200.

ELLEGREN Hans, 2011, <sup>2</sup> Sex-chromosome evolution: recent progress and the influence of male and female heterogamety *p*, *Nature Reviews Genetics*, vol. 12, n. 3 : 157-166.

ELLIS Peggy E., 1959, <sup>2</sup> Some factors influencing phase characters in the nymphs of the locust, *Locusta migratoria migratorioides* (R. and F.) *p*, *Insectes Sociaux*, vol. 6, n. 1 : 21-39.

EL-SHAMM Mahmoud, PONTER Dominique, LAHMY Sylvie, BRAUN Laurence, PICART Claire, VEGA Danielle, HAKIM Mohamed-Ali, JACOBSEN Steven E., COOKE Richard, et LAGRANGE Thierry, 2007, <sup>2</sup> Reiterated WG/GW motifs form functionally and evolutionarily conserved ARGONAUTE-binding platforms in RNAi-related components *p*, *Genes & Development*, vol. 21, n. 20 : 2539-2544.

ERNST Jason et KELLIS Manolis, 2010, <sup>2</sup> Discovery and characterization of chromatin states for systematic annotation of the human genome *p*, *Nature biotechnology*, vol. 28, n. 8 : 817-825.

ERNST Ulrich R., HEIL Matthias B. Van, DEPUYDT Geert, BOERJAN Bart, LOOF Arnold De, et SCHOOF Liliene, 2015, <sup>2</sup> Epigenetics and locust life phase transitions *p*, *Journal of Experimental Biology*, vol. 218, n. 1 : 88-99.

EVANS Jay D. et WHEELER Diana E., 1999, <sup>2</sup> Differential gene expression between developing queens and workers in the honey bee, *Apis mellifera* ♀, *Proceedings of the National Academy of Sciences*, vol. 96, n, 10 : 5575 ~~5~~580.

FAT"CA Alessandro et BOZZON" "rene, 2014, <sup>2</sup> Long non-coding RNAs: new players in cell differentiation and development ♀, *Nature Reviews Genetics*, vol. 15, n, 1 : 7 ~~2~~1.

FELLE Max, JOPP"EN Saskia, N"EMETH Attila, D"ERME"ER Sarah, THALHAMMER Verena, DOBNER Thomas, KREMMER Elisabeth, KAPPLER Roland, et L"INGST Gernot, 2011, <sup>2</sup> The USP7/Dnmt1 complex stimulates the DNA methylation activity of Dnmt1 and regulates the stability of UHRF1 ♀, *Nucleic Acids Research*, vol. 39, n, 19 : 8355 ~~8~~365.

FELLER Christian, PRESTEL Matthias, HARTMANN Holger, STRAUB Tobias, SIJD"NG Johannes, et BECKER Peter B., 2012, <sup>2</sup> The MOF-containing NSL complex associates globally with housekeeping genes, but activates only a defined subset ♀, *Nucleic Acids Research*, vol. 40, n, 4 : 1509 ~~1~~522.

FENG Suhua, COKUS Shawn J., ZHANG Xiaoyu, CHEN Pao-Yang, BOST"CK Magnolia, GOLL Mary G., HETZEL Jonathan, JA"AN Jayati, STRAUSS Steven H., HALPERN Marnie E., UKOMADU Chinweike, SADLER Kirsten C., PRADHAN Sriharsa, PELLEGR"AN" Matteo, et JACOBSEN Steven E., 2010, <sup>2</sup> Conservation and divergence of methylation patterning in plants and animals ♀, *Proceedings of the National Academy of Sciences of the United States of America*, vol. 107, n, 19 : 8689 ~~8~~694.

FERNANDES "sabelle, CHANUT-DELANDE Hůl"one, FERRER Pierre, LATAP"E Yvan, WALTZER Lucas, AFFOLTER Markus, PAYRE Fran"ois, et PLAZA Serge, 2010, <sup>2</sup> Zona Pellucida Domain Proteins Remodel the Apical Compartment for Localized Cell Shape Changes ♀, *Developmental Cell*, vol. 18, n, 1 : 64 ~~1~~6.

F"L"POW"CZ Witold, BHATTACHARYYA Suvendra N., et SONENBERG Nahum, 2008, <sup>2</sup> Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? ♀, *Nature Reviews. Genetics*, vol. 9, n, 2 : 102 ~~1~~14.

F"L"POVA Galina N, FAGERL"E Sara, KLENOVA Elena M, MYERS Cena, DEHNER Yvonne, GOODW"AN Graham, NE"MAN Paul E, COLL"NS Steve J, et LOBANENKOV Victor V, 1996, <sup>2</sup> An exceptionally conserved transcriptional repressor, CTCF, employs different combinations of zinc fingers to bind diverged promoter sequences of avian and mammalian c-myc oncogenes. ♀, *Molecular and cellular biology*, vol. 16, n, 6 : 2802 ~~2~~813.

FLEMM"NG Walther, 1882, *Zellsubstanz, kern und zelltheilung..*, Leipzig, F. C. W. Vogel.

FLOT Jean-Fran"ois, MAR"E-NELLY Hervů, et KOSZUL Romain, 2015, <sup>2</sup> Contact genomics: scaffolding and phasing (meta)genomes using chromosome 3D physical signatures ♀, *FEBS letters*, vol. 589, n, 20 Pt A : 2966 ~~2~~974.

FORET Sylvain, KUCHARSK" Robert, PELLEGR"AN" Matteo, FENG Suhua, JACOBSEN Steven E., ROB"NSON Gene E., et MALESZKA Ryszard, 2012, <sup>2</sup> DNA methylation dynamics, metabolic fluxes, gene splicing, and alternative phenotypes in honey bees ♀, *Proceedings of the National Academy of Sciences*, vol. 109, n, 13 : 4968 ~~4~~973.

FOURN"ER Alexandra, SASA" Nobuhiro, NAKAO Mitsuyoshi, et DEFOSSEZ Pierre-Antoine, 2012, <sup>2</sup> The role of methyl-binding proteins in chromatin organization and epigenome maintenance ♀, *Briefings in Functional Genomics*, vol. 11, n, 3 : 251 ~~2~~64.

FRENCH V., FEAST M., et PARTR"DGE L., 1998, <sup>2</sup> Body size and cell size in *Drosophila*: the developmental response to temperature ♀, *Journal of Insect Physiology*, vol. 44, n, 11 : 1081 ~~1~~089.

FUKATSU Takema, SARJYA A, et SHBAO Harunobu, 2005, <sup>2</sup> Soldier caste with morphological and reproductive division in the aphid tribe Nipponaphidini *», Insectes sociaux*, vol. 52, n. 2 : 132-138.

FUKATSU Takema, TSUCHIDA Tsutomu, NAKOH Naruo, et KOGA Ryuichi, 2001, <sup>2</sup> Spiroplasma Symbiont of the Pea Aphid, *Acyrtosiphon pisum* (Insecta: Homoptera) *», Applied and Environmental Microbiology*, vol. 67, n. 3 : 1284-1291.

GALLOT Aurore, RÈPE Claude, LETERME Nathalie, GAUTHIER Jean-Pierre, JAUBERT-POSSAMA Stéphanie, et TAGU Denis, 2010, <sup>2</sup> Cuticular proteins and seasonal photoperiodism in aphids *», Insect Biochemistry and Molecular Biology*, vol. 40, n. 3 : 235-240.

GALLOT Aurore, SHIGENOBU Shuji, HASHIYAMA Tomomi, JAUBERT-POSSAMA Stéphanie, et TAGU Denis, 2012, <sup>2</sup> Sexual and asexual oogenesis require the expression of unique and shared sets of genes in the insect *Acyrtosiphon pisum* *», BMC Genomics*, vol. 13 : 76.

GARDNER Alessandro, 2017, <sup>2</sup> Global Run-On Sequencing (GRO-Seq) *», Methods in Molecular Biology (Clifton, N.J.)*, vol. 1468 : 111-120.

GASPAR-MATA Alexandre, ALAJEM Adi, POLESSO Fanny, SRIDHARAN Rupa, MASON Michael, HEDERSBACH Amy, RAMALHO-SANTOS João, MCMANUS Michael T., PLATH Kathrin, MESHORER Eran, et RAMALHO-SANTOS Miguel, 2009, <sup>2</sup> Chd1 regulates open chromatin and pluripotency of embryonic stem cells *», Nature*, vol. 460, n. 7257 : 863-868.

GASZNER Miklos et FELSENFELD Gary, 2006, <sup>2</sup> Insulators: exploiting transcriptional and epigenetic mechanisms *», Nature reviews. Genetics*, vol. 7, n. 9 : 703.

GAUDET P., LYSTONE M., et THOMAS P., 2010, <sup>2</sup> Gene Ontology annotation inferences using phylogenetic trees. *GO Reference Genome Project.* *»*

GAULTON Kyle J., NAMMO Takao, PASQUALI Lorenzo, SIMON Jeremy M., GRES Paul G., FOGARTY Marie P., PANHUIS Tami M., MIECZKOWSKI Piotr, SECCHI Antonio, BOSCO Domenico, BERNEY Thierry, MONTANYA Eduard, MOHLKE Karen L., LEB Jason D., et FERRER Jorge, 2010, <sup>2</sup> A map of open chromatin in human pancreatic islets *», Nature Genetics*, vol. 42, n. 3 : 255-259.

GAUTHIER Jean-Pierre, OUTREMAN Yannick, MEUZET Lucie, et SIMON Jean-Christophe, 2015, <sup>2</sup> Bacterial Communities Associated with Host-Adapted Populations of Pea Aphids Revealed by Deep Sequencing of 16S Ribosomal DNA *», PLOS ONE*, vol. 10, n. 3 : e0120664.

GAVIN David P. et SHARMA Rajiv P., 2010, <sup>2</sup> Histone modifications, DNA methylation, and schizophrenia *», Neuroscience and Biobehavioral Reviews*, vol. 34, n. 6 : 882-888.

GEKAKIS Nicholas, SAEZ Lino, DELAHAYE-BROWN Anne-Marie, MYERS Michael P., SEHGAL Amita, YOUNG Michael W., et WEITZ Charles J., 1995, <sup>2</sup> Isolation of timeless by PER Protein Interaction: Defective Interaction Between timeless Protein and Long-Period Mutant PER<sup>L</sup> *», Science*, vol. 270 : 811-815.

GELBART Marnie E., LARSCHAN Erica, PENG Shouyong, PARK Peter J., et KURODA Mitzi M., 2009, <sup>2</sup> Drosophila MSL complex globally acetylates H4K16 on the male X chromosome for dosage compensation *», Nature Structural & Molecular Biology*, vol. 16, n. 8 : 825-832.

GERACE Erica L., HALIC Mario, et MOAZED Danesh, 2010, <sup>2</sup> The Methyltransferase Activity of Clr4Suv39h triggers RNAi independently of Histone H3K9 Methylation *», Molecular cell*, vol. 39, n. 3 : 360-372.

GERSHENZON Naum M et OSHKHESELYA P, 2004, <sup>2</sup> Synergy of human Pol III core promoter elements revealed by statistical sequence analysis *», Bioinformatics*, vol. 21, n. 8 : 1295-1300.

G LLESP E Peter J., GAMBUS Agnieszka, et BLOW J. Julian, 2012, 2 Preparation and use of Xenopus egg extracts to study DNA replication and chromatin associated proteins  , *Methods (San Diego, Calif.)*, vol. 57, n, 2 : 203  13.

G RES  Paul G., K M Jonghwan, MCDAN ELL Ryan M.,  YER Vishwanath R., et L EB Jason D., 2007, 2 FA RE (Formaldehyde-Assisted  solation of Regulatory Elements) isolates active regulatory elements from human chromatin  , *Genome Research*, vol. 17, n, 6 : 877  85.

G RES  Paul G. et L EB Jason D., 2009, 2  solation of active regulatory elements from eukaryotic chromatin using FA RE (Formaldehyde Assisted  solation of Regulatory Elements)  , *Methods (San Diego, Calif.)*, vol. 48, n, 3 : 233  39.

GOLL Mary Grace et BESTOR Timothy H., 2005, 2 Eukaryotic cytosine methyltransferases  , *Annual Review of Biochemistry*, vol. 74 : 481  14.

GRATAN  Loretta, 2014, *Plant Phenotypic Plasticity in Response to Environmental Factors*, <https://www.hindawi.com/archive/2014/208747/> (consult  le 29 juillet 2017).

GR FF THS Anthony JF, M LLER Jeffrey H., SUZUK  David T., LEWONT N Richard C., et GELBART William M., 2000, 2 *Norm of reaction and phenotypic distribution*  .

GRONKE S., M LDNER A., FELLERT S., TENNAGELS N., PETRY S., MULLER G., JACKLE H., et KUHNLE N R.P., 2005, 2 Brummer lipase is an evolutionary conserved fat storage regulator in Drosophila.  , *Cell Metabolism*, vol. 1, n, 5 : 323  30.

GRIJNKE S., M LLER G., H RSCH J., FELLERT S., ANDREOU A., HAASE T., J CKLE H., et K HNLE N R.P., 2007, 2 Dual lipolytic control of body fat storage and mobilization in Drosophila.  , *PLoS Biology*, vol. 5, n, 6 : e137.

GROSSN KLAUS Ueli et PARO Renato, 2014, 2 Transcriptional Silencing by Polycomb-Group Proteins  , *Cold Spring Harbor Perspectives in Biology*, vol. 6, n, 11 : a019331.

GUAN D., FACTOR D., L U Yu, WANG Z., et KAO H.-Y., 2013, 2 The epigenetic regulator UHRF1 promotes ubiquitination-mediated degradation of the tumor-suppressor protein promyelocytic leukemia protein  , *Oncogene*, vol. 32, n, 33 : 3819  828.

GUAY Jean-Fr d ric, BOUDREAU Simon, M CHAUD Dominique, et CLOUT ER Conrad, 2009, 2  mpact of environmental stress on aphid clonal resistance to parasitoids: Role of Hamiltonella defensa bacterial symbiosis in association with a new facultative symbiont of the pea aphid  , *Journal of  nsect Physiology*, vol. 55, n, 10 : 919  26.

GUO Junjie U., SU Yijing, ZHONG Chun, M NG Guo-li, et SONG Hongjun, 2011, 2 Hydroxylation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain  , *Cell*, vol. 145, n, 3 : 423  34.

GUT ERREZ Eugenio, W GG NS David, F ELD NG Barbara, et GOULD Alex P., 2007, 2 Specialized hepatocyte-like cells regulate Drosophila lipid metabolism  , *Nature*, vol. 445, n, 7125 : 275  80.

GUTTMAN Mitchell, AM T  do, GARBER Manuel, FRENCH Courtney, L N Michael F., FELDSEER David, HUARTE Maite, ZUK Or, CAREY Bryce W., CASSADY John P., CAB L  Moran N., JAEN SCH Rudolf, M KKELSEN Tarjei S., JACKS Tyler, HACOHEN Nir, BERNSTE N Bradley E., KELL S Manolis, REGEV Aviv, R NN John L., et LANDER Eric S., 2009, 2 Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals  , *Nature*, vol. 458, n, 7235 : 223  27.

GUTTMAN Mitchell et RENN John L., 2012, <sup>2</sup> Modular regulatory principles of large non-coding RNAs *Nature*, vol. 482, n. 7385 : 339 [146](#).

HAHN Matthew W. et LANZARO Gregory C., 2005, <sup>2</sup> Female-biased gene expression in the malaria mosquito *Anopheles gambiae* *Current Biology*, vol. 15, n. 6 : R192 [193](#).

HALL Michael A., SHUNDROVSKY Alla, BA Lu, FULBRIGHT Robert M., L's John T., et WANG Michelle D., 2009, <sup>2</sup> High-resolution dynamic mapping of histone-DNA interactions in a nucleosome *Nature Structural & Molecular Biology*, vol. 16, n. 2 : 124 [129](#).

HALLACR Erinc et AKHTAR Asifa, 2009, <sup>2</sup> X chromosomal regulation in flies: when less is more *Chromosome Research*, vol. 17, n. 5 : 603 [619](#).

HAMADA Fumika N., PARK Peter J., GORDADZE Polina R., et KURODA Mitzi ., 2005, <sup>2</sup> Global regulation of X chromosomal genes by the MSL complex in *Drosophila melanogaster* *Genes & Development*, vol. 19, n. 19 : 2289 [2294](#).

HAMMOND S. M., BOETTCHER S., CAUDY A. A., KOBAYASHI R., et HANNON G. J., 2001, <sup>2</sup> Argonaute2, a link between genetic and biochemical analyses of RNAi *Science (New York, N.Y.)*, vol. 293, n. 5532 : 1146 [1150](#).

HANSEN Jeffrey C., 2002, <sup>2</sup> Conformational dynamics of the chromatin fiber in solution: determinants, mechanisms, and functions *Annual Review of Biophysics and Biomolecular Structure*, vol. 31 : 361 [392](#).

HARDIE Jim, 1987, <sup>2</sup> Juvenile hormone stimulation of oocyte development and embryogenesis in the parthenogenetic ovaries of an aphid, *Aphis fabae* *International journal of invertebrate reproduction and development*, vol. 11, n. 2 : 189 [202](#).

HARGREAVES Diana C et CRABTREE Gerald R, 2011, <sup>2</sup> ATP-dependent chromatin remodeling: genetics, genomics and mechanisms *Cell Research*, vol. 21, n. 3 : 396 [420](#).

HE Housheng Hansen, MEYER Clifford A., HU Sheng an Shawn, CHEN Mei-Wei, ZANG Chongzhi, L'u Yin, RAO Prakash K., FEI Teng, XU Han, LONG Henry, L'u X. Shirley, et BROWN Myles, 2014, <sup>2</sup> Refined DNase-seq protocol and data analysis reveals intrinsic bias in transcription factor footprint identification *Nature Methods*, vol. 11, n. 1 : 73 [78](#).

HE Yu-Fei, L' Bin-Zhong, L' Zheng, L'u Peng, WANG Yang, TANG Qingyu, D'NG Jianping, J'A Yingying, CHEN Zhangcheng, L' Lin, SUN Yan, L' Xiuxue, DA' Qing, SONG Chun-Xiao, ZHANG Kangling, HE Chuan, et XU Guo-Liang, 2011, <sup>2</sup> Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA *Science (New York, N.Y.)*, vol. 333, n. 6047 : 1303 [1307](#).

HEAD Steven R., KOMORI H. Kiyomi, LAMERE Sarah A., WH'ENANT Thomas, VAN N'EUWERBURGH Filip, SALOMON Daniel R., et ORDOUKHAN'AN Phillip, 2014, <sup>2</sup> Library construction for next-generation sequencing: Overviews and challenges *BioTechniques*, vol. 56, n. 2 : 61 [passim](#).

HENRY Lee M., PECCOUD Jean, S'MON Jean-Christophe, HADFIELD Jarrod D., MA'DEN Martin J. C., FERRARI Julia, et GODFRAY H. Charles J., 2013, <sup>2</sup> Horizontally Transmitted Symbionts and Host Colonization of Ecological Niches *Current Biology*, vol. 23, n. 17 : 1713 [1717](#).

H'LFKER A., H'LFKER-KLE'NER D., PANNUT' A., et LUCCHESI J. C., 1997, <sup>2</sup> mof, a putative acetyl transferase gene related to the Tip60 and MOZ human genes and to the SAS genes of yeast, is required for dosage compensation in *Drosophila* *The EMBO journal*, vol. 16, n. 8 : 2054 [2060](#).

H"R"ART Edwige et VERDEL Andrů, 2013, <sup>2</sup> Long noncoding RNA-based chromatin control of germ cell differentiation: a yeast perspective p, *Chromosome Research*, vol. 21, n, 6 1 : 653 663.

HOLOCH Daniel et MOAZED Danesh, 2015, <sup>2</sup> RNA-mediated epigenetic regulation of gene expression p, *Nature Reviews Genetics*, vol. 16, n, 2 : 71 84.

HONG Eun-Jin Erica, V"LL"ON Judit, GERACE Erica L., GYG" Steven P., et MOAZED Danesh, 2005, <sup>2</sup> A cullin E3 ubiquitin ligase complex associates with Rik1 and the Clr4 histone H3-K9 methyltransferase and is required for RNAi-mediated heterochromatin formation p, *RNA biology*, vol. 2, n, 3 : 106 111.

HOPPELER Hans, BAUM Oliver, LURMAN Glenn, et MUELLER Matthias, 2011, <sup>2</sup> Molecular mechanisms of muscle plasticity with exercise p, *Comprehensive Physiology*, vol. 1, n, 3 : 1383 1412.

HOPPELER Hans et FL"CK Martin, 2002, <sup>2</sup> Normal mammalian skeletal muscle and its phenotypic plasticity p, *The Journal of Experimental Biology*, vol. 205, n, Pt 15 : 2143 2152.

HORN Peter J., BAST"E Jean-Noůl, et PETERSON Craig L., 2005, <sup>2</sup> A Rik1-associated, cullin-dependent E3 ubiquitin ligase is essential for heterochromatin formation p, *Genes & Development*, vol. 19, n, 14 : 1705 1714.

HOU Chunhui, L" Li, Q"N Zhaohui S., et CORCES Victor G., 2012, <sup>2</sup> Gene density, transcription, and insulators contribute to the partition of the Drosophila genome into physical domains p, *Molecular Cell*, vol. 48, n, 3 : 471 484.

HUANG Ting-Hua, FAN Bin, ROTHSCH"LD Max F., HU Zhi-Liang, L" Kui, et ZHAO Shu-Hong, 2007, <sup>2</sup> MiRFinder: an improved approach and software implementation for genome-wide fast microRNA precursor scans p, *BMC Bioinformatics*, vol. 8 : 341.

HUANG Xiao A., Y"N Hang, SWEENEY Sarah, RAHA Debasish, SNYDER Michael, et L"N Haifan, 2013, <sup>2</sup> A major epigenetic programming mechanism guided by piRNAs p, *Developmental Cell*, vol. 24, n, 5 : 502 516.

HUTV"GNER Gy"rgy et ZAMORE Phillip D., 2002, <sup>2</sup> A microRNA in a multiple-turnover RNAi enzyme complex p, *Science (New York, N.Y.)*, vol. 297, n, 5589 : 2056 2060.

HUYBRECHTS J., BONHOMME J., M"NOL" S., PRUN"ER-LETERME N., DOMBROVSKY A., ABDEL-LAT"EF M., ROB"CHON A., VEENSTRA J. A., et TAGU D., 2010a, <sup>2</sup> Neuropeptide and neurohormone precursors in the pea aphid, *Acyrtosiphon pisum* p, *Insect Molecular Biology*, vol. 19 : 87 95.

HUYBRECHTS J., 2010b, <sup>2</sup> Neuropeptide and neurohormone precursors in the pea aphid, *Acyrtosiphon pisum* p, *Insect Molecular Biology*, vol. 19 : 87 95.

"INTERNAT"ONAL APH"D GENOM"CS CONSORT"UM, 2010, <sup>2</sup> Genome sequence of the pea aphid *Acyrtosiphon pisum* p, *PLoS biology*, vol. 8, n, 2 : e1000313.

"PSARO Jonathan J., HAASE Astrid D., KNOTT Simon R., JOSHUA-TOR Leemor, et HANNON Gregory J., 2012, <sup>2</sup> The structural biochemistry of Zucchini implicates it as a nuclease in piRNA biogenesis p, *Nature*, vol. 491, n, 7423 : 279 283.

"SH"KAWA A., OGAWA K., GOTOH H., WALSH T. K., TAGU D., BR"SSON J. A., R"PE C., JAUBERT-POSSAMA" S., KANBE T., TSUBOTA T., SH"OTSUK" T., et M"URA T., 2012, <sup>2</sup> Juvenile hormone titre and related gene expression during the change of reproductive modes in the pea aphid p, *Insect Molecular Biology*, vol. 21, n, 1 : 49 50.

"TI Yosiaki, 1989, <sup>2</sup> The evolutionary biology of sterile soldiers in aphids p, *Trends in ecology & evolution*, vol. 4, n, 3 : 69 73.

JAEN'SCH Rudolf et B'RD Adrian, 2003, <sup>2</sup> Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals *Þ*, *Nature Genetics*, vol. 33 : 245 254.

JAQU'ERY Julie, PECCOUD Jean, OU'SSE Tiphaine, LEGEA" Fabrice, PRUN'ER-LETTERME Nathalie, GOU"AN Anais, NOUHAUD Pierre, BR'SSON Jennifer, B'CKEL Ryan, PURANDARE Swapna, POULA"AN Julie, BATTAL Christophe, LEMATRE Claire, M'EUZET Lucie, TR'ONNA"RE Gael Le, S'MON Jean-Christophe, et R'SPE Claude, 2017, <sup>2</sup> Disentangling The Causes For Faster-X Evolution "n Aphids *Þ*, *bioRxiv*: 125310.

JAQU'ERY Julie, R'SPE Claude, ROZE Denis, LEGEA" Fabrice, TR'ONNA"RE GaúL Le, STOECKEL Solenn, M'EUZET Lucie, S"LVA Corinne Da, POULA"AN Julie, PRUN'ER-LETTERME Nathalie, SEGURENS Búatrice, TAGU Denis, et S'MON Jean-Christophe, 2013, <sup>2</sup> Masculinization of the X Chromosome in the Pea Aphid *Þ*, *PLOS Genetics*, vol. 9, n, 8 : e1003690.

JAQU'ERY Julie, STOECKEL Solenn, LAROSE Chloé, NOUHAUD Pierre, R'SPE Claude, M'EUZET Lucie, BONHOMME JoúL, MAHEO Frúdürrique, LEGEA" Fabrice, GAUTH'ER Jean-Pierre, PRUN'ER-LETTERME Nathalie, TAGU Denis, et S'MON Jean-Christophe, 2014, <sup>2</sup> Genetic Control of Contagious Asexuality in the Pea Aphid *Þ*, *PLOS Genetics*, vol. 10, n, 12 : e1004838.

JAUBERT-POSSAMA" Stúphanie, LE TR'ONNA"RE GaúL, BONHOMME JoúL, CHR'STOPH'DES Georges K., R'SPE Claude, et TAGU Denis, 2007, <sup>2</sup> Gene knockdown by RNAi in the pea aphid *Acyrtosiphon pisum* *Þ*, *BMC biotechnology*, vol. 7 : 63.

JEON Junhyun, CHO" Jaeyoung, LEE Gir-Won, PARK Sook-Young, HUH Aram, DEAN Ralph A., et LEE Yong-Hwan, 2015, <sup>2</sup> Genome-wide profiling of DNA methylation provides insights into epigenetic regulation of fungal development in a plant pathogenic fungus, *Magnaporthe oryzae* *Þ*, *Scientific Reports*, vol. 5 : srep08567.

J"ANG Xiaofang, B"EDLER James K., Q" Yumin, HALL Andrew Brantley, et Tu Zhijian, 2015, <sup>2</sup> Complete Dosage Compensation in *Anopheles stephensi* and the Evolution of Sex-Biased Genes in Mosquitoes *Þ*, *Genome Biology and Evolution*, vol. 7, n, 7 : 1914 1924.

J"ANG Yong-Hui, SAHOO Trilochan, M"CHAE"LS Ron C., BERCOV"CH Dani, BRESSLER Jan, KASHORK Catherine D., L"U Qian, SHAFFER Lisa G., SCHROER Richard J., STOCKTON David W., SP"ELMAN Richard S., STEVENSON Roger E., et BEAUDET Arthur L., 2004, <sup>2</sup> A mixed epigenetic/genetic model for oligogenic inheritance of autism with a limited role for UBE3A *Þ*, *American Journal of Medical Genetics. Part A*, vol. 131, n, 1 : 1 10.

J"NDRA Marek, PALL" Subba R., et R'DD"ORD Lynn M., 2013, <sup>2</sup> The juvenile hormone signaling pathway in insect development *Þ*, *Annual Review of Entomology*, vol. 58 : 181 204.

JOHN Sam, SABO Peter J., CANF"ELD Theresa K., LEE Kristen, VONG Shinny, WEAVER Molly, WANG Hao, V"ERSTRA Jeff, REYNOLDS Alex P., THURMAN Robert E., et STAMATOYANNOPOULOS John A., 2013, <sup>2</sup> Genome-scale mapping of DNase " hypersensitivity *Þ*, *Current Protocols in Molecular Biology*, vol. Chapter 27 : Unit 21.27.

JONES Dennis, 1969, <sup>2</sup> The reaction of formaldehyde with unsaturated fatty acids during histological fixation *Þ*, *The Histochemical Journal*, vol. 1, n, 5 : 459 191.

JONES P. A. et TAKA" D., 2001, <sup>2</sup> The role of DNA methylation in mammalian epigenetics *Þ*, *Science (New York, N.Y.)*, vol. 293, n, 5532 : 1068 1070.

KADLEC Jan, HALLACL" Erinc, L"PP Michael, HOLZ Herbert, SANCHEZ-WEATHERBY Juan, CUSACK Stephen, et AKHTAR Asifa, 2011, <sup>2</sup> Structural basis for MOF and MSL3 recruitment into the dosage compensation complex by MSL1 *Þ*, *Nature Structural & Molecular Biology*, vol. 18, n, 2 : 142 149.

KALANTRY Sundeep, 2011, 2 Recent Advances in X-Chromosome Inactivation P, *Journal of cellular physiology*, vol. 226, n, 7 : 1714-1718.

KARAGANN Panagiota, AMAT Larbi, QIN Jun, et WONG Jiemin, 2008, 2 CBP90, a novel methyl K9 H3 binding protein linking protein ubiquitination with heterochromatin formation P, *Molecular and Cellular Biology*, vol. 28, n, 2 : 705-717.

KAUSHIK Kriti, LEONARD Vincent Elvin, KV Shamsudheen, LALWAN Mukesh Kumar, JALAL Saakshi, PATOWARY Ashok, JOSH Adita, SCARVINOD, et SVASUBBU Sridhar, 2013, 2 Dynamic Expression of Long Non-Coding RNAs (lncRNAs) in Adult Zebrafish P, *PLOS ONE*, vol. 8, n, 12 : e83616.

KEBEDE Adam F., SCHNEIDER Robert, et DAUJAT Sylvain, 2015, 2 Novel types and sites of histone modifications emerge as players in the transcriptional regulation contest P, *The FEBS journal*, vol. 282, n, 9 : 1658-1674.

KELLEY Richard L, SOLOVYEVA Irina, LYMAN Laura M, RICHMAN Ron, SOLOVYEV Victor, et KURODA Mitzi, 1995, 2 Expression of Msl-2 causes assembly of dosage compensation regulators on the X chromosomes and female lethality in *Drosophila* P, *Cell*, vol. 81, n, 6 : 867-877.

KELLEY Richard L., WANG Jiwu, BELL Leslie, et KURODA Mitzi, 1997, 2 Sex lethal controls dosage compensation in *Drosophila* by a non-splicing mechanism P, *Nature*, vol. 387, n, 6629 : 195-199.

KHANDELWAL Akanksha, BACCOLLA Albino, VASQUEZ Karen M., et JAN Aklank, 2015, 2 Long non-coding RNA: A new paradigm for lung cancer P, *Molecular Carcinogenesis*, vol. 54, n, 11 : 1235-1251.

KHARCHENKO Peter V., ALEKSEYENKO Artyom A., SCHWARTZ Yuri B., MINODA Aki, RIDDLE Nicole C., ERNST Jason, SABO Peter J., LARSCHAN Erica, GORCHAKOV Andrey A., GU Tingting, LINDER-BASSO Daniela, PLACHETKA Annette, SHANOWER Gregory, TOLSTORUKOV Michael Y., LUQUETTE Lovelace J., XU Ruibin, JUNG Youngsook L., PARK Richard W., BISHOP Eric P., CANFIELD Theresa K., SANDSTROM Richard, THURMAN Robert E., MACALPINE David M., STAMATOYANNOPOULOS John A., KELLIS Manolis, ELGIN Sarah C. R., KURODA Mitzi, PIRROTTA Vincenzo, KARPEN Gary H., et PARK Peter J., 2011, 2 Comprehensive analysis of the chromatin landscape in *Drosophila melanogaster* P, *Nature*, vol. 471, n, 7339 : 480-485.

KOOKE Rik, JOHANNES Frank, WARDENAAR Ren, BECKER Frank, ETCHEVERRY Mathilde, COLOT Vincent, VREUGDENHIL Dick, et KEURENTJES Joost J. B., 2015, 2 Epigenetic basis of morphological variation and phenotypic plasticity in *Arabidopsis thaliana* P, *The Plant Cell*, vol. 27, n, 2 : 337-348.

KOUZARDES Tony, 2007, 2 Chromatin Modifications and Their Function P, *Cell*, vol. 128, n, 4 : 693-705.

KRAUSS Veiko et REUTER Gunter, 2011, 2 DNA methylation in *Drosophila*--a critical evaluation P, *Progress in Molecular Biology and Translational Science*, vol. 101 : 177-191.

KRISTENSEN M., LANGERKAMP A., KNUDSEN G. P., WEBER B. L., BKRRESEN-DALE A. L., et ORSTAVIK K. H., 2002, 2 High frequency of skewed X inactivation in young breast cancer patients P, *Journal of Medical Genetics*, vol. 39, n, 1 : 30-33.

KRVEGA Ivan et DEAN Ann, 2012, 2 Enhancer and promoter interactions--long distance calls P, *Current Opinion in Genetics & Development*, vol. 22, n, 2 : 79-85.

KUCHARSKA R., MALESZKA J., FORET S., et MALESZKA R., 2008, 2 Nutritional Control of Reproductive Status in Honeybees via DNA Methylation P, *Science*, vol. 319, n, 5871 : 1827-1830.

KUMAR Vibhor, MURATAN Masafumi, RAYAN Nirmala Arul, KRAUS Petra, LUFK Thomas, NG Huck Hui, et PRABHAKAR Shyam, 2013, <sup>2</sup> Uniform, optimal signal processing of mapped deep-sequencing data * , Nature Biotechnology*, vol. 31, n. 7 : 615  22.

KUNDAJE A, 2012, (2012) ENCODE: TF ChIP-seq peak calling using the "reproducibility Discovery Rate (DR) framework - Anshul Kundaje, <https://sites.google.com/site/anshulkundaje/projects/idr> (consult  le 24 juillet 2017).

KUZNETSOVA V. G. et SHAPOSHNIKOV G. Kh, 1973, <sup>2</sup> chromosome numbers of the aphids (Homoptera, Aphidinea) of the world fauna * , Entomological review*.

LAGOS-QUINTANA M., RAUHUT R., LENDECKEL W., et TUSCHL T., 2001, <sup>2</sup> Identification of novel genes coding for small expressed RNAs * , Science (New York, N.Y.)*, vol. 294, n. 5543 : 853  58.

LAGRANGE Thierry, KAPANOS Achillefs N, TANG Hong, REINBERG Danny, et EBRGHT Richard H, 1998, <sup>2</sup> New core promoter element in RNA polymerase III-dependent transcription: sequence-specific DNA binding by transcription factor III B * , Genes & development*, vol. 12, n. 1 : 34  4.

LA C Y, BAUMANN L, et BAUMANN P, 1994, <sup>2</sup> Amplification of trpEG: adaptation of Buchnera aphidicola to an endosymbiotic association with aphids. * , Proceedings of the National Academy of Sciences of the United States of America*, vol. 91, n. 9 : 3819  823.

LA Eric C., TOMANCAK Pavel, WILLIAMS Robert W., et RUBIN Gerald M., 2003, <sup>2</sup> Computational identification of Drosophila microRNA genes * , Genome Biology*, vol. 4, n. 7 : R42.

LAM Kin Chung, MHLPFORDT Friederike, VAQUERZAS Juan M., RAJA Sunil Jayaramaiah, HOLZ Herbert, LUSCOMBE Nicholas M., MANKE Thomas, et AKHTAR Asifa, 2012, <sup>2</sup> The NSL complex regulates housekeeping genes in Drosophila * , PLoS genetics*, vol. 8, n. 6 : e1002736.

LAMB K. P., 1959, <sup>2</sup> Composition of the honeydew of the aphid Brevicoryne brassicae (L.) feeding on swedes (Brassica napobrassica DC.) * , Journal of Insect Physiology*, vol. 3, n. 1 : 1  3.

LANDT Stephen G., MARNOV Georgi K., KUNDAJE Anshul, KHERADPOUR Pouya, PAUL Florencia, BATZOGLOU Serafim, BERNSTEIN Bradley E., BICKEL Peter, BROWN James B., CAYTNG Philip, CHEN Yiwen, DESALVO Gilberto, EPSTEIN Charles, FISHER-AYLOR Katherine M., EUSKUCHEN Ghia, GERSTEIN Mark, GERTZ Jason, HARTEMINK Alexander J., HOFFMAN Michael M., IYER Vishwanath R., JUNG Youngsook L., KARMAKAR Subhradip, KELLIS Manolis, KHARCHENKO Peter V., LI Qunhua, LIU Tao, LIU X. Shirley, MA Lijia, MILOSAVLJEVIC Aleksandar, MYERS Richard M., PARK Peter J., PAZIN Michael J., PERRY Marc D., RAHA Debasish, REDDY Timothy E., ROZOWSKY Joel, SHORESH Noam, SODOW Arend, SLATTERY Matthew, STAMATOYANNOPOULOS John A., TOLSTORUKOV Michael Y., WHITE Kevin P., XIA Simon, FARNHAM Peggy J., LEB Jason D., WOLD Barbara J., et SNYDER Michael, 2012, <sup>2</sup> ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia * , Genome Research*, vol. 22, n. 9 : 1813  831.

LANGMEAD Ben et SALZBERG Steven L, 2012, <sup>2</sup> Fast gapped-read alignment with Bowtie 2 * , Nature methods*, vol. 9, n. 4 : 357  59.

LANGMEAD Ben, TRAPNELL Cole, POP Mihai, et SALZBERG Steven L, 2009, <sup>2</sup> Ultrafast and memory-efficient alignment of short DNA sequences to the human genome * , Genome Biology*, vol. 10, n. 3 : R25.

LAPP S. A., GERALDO J. A., CHEN J.-T., AY F., PAKALA S. B., BATUGEDARA G., HUMPHREY J., MAHPIC CONSORTIUM, DEBARRY J. D., LE ROCH K. G., GALINSKI M. R., et KISSINGER J. C., 2017, <sup>2</sup> PacBio assembly of a Plasmodium knowlesi genome sequence with Hi-C correction and manual annotation of the SCAvar gene family * , Parasitology*: 1  4.

- LARSHAN Erica, ALEKSEYENKO Artyom A., GORTCHAKOV Andrey A., PENG Shouyong, L' Bing, YANG Pok, WORKMAN Jerry L., PARK Peter J., et KURODA Mitzi ", 2007, <sup>2</sup> MSL complex is attracted to genes marked by H3K36 trimethylation using a sequence-independent mechanism *p*, *Molecular Cell*, vol. 28, n, 1 : 121 [ű33](#).
- LAU Alyssa C., NABESH"MA Kentaro, et CSANKOVSK" Gy" rgyi, 2014, <sup>2</sup> The C. elegans dosage compensation complex mediates interphase X chromosome compaction *p*, *Epigenetics & Chromatin*, vol. 7 : 31.
- LAU N. C., L" M L. P., WE"NSTE" N E. G., et BARTEL D. P., 2001, <sup>2</sup> An abundant class of tiny RNAs with probable regulatory roles in Caenorhabditis elegans *p*, *Science (New York, N.Y.)*, vol. 294, n, 5543 : 858 [ű62](#).
- LAW Charity W., ALHAMDOOSH Monther, SU Shian, SMYTH Gordon K., et R" TCH" E Matthew E., 2016, <sup>2</sup> RNA-seq analysis is easy as 1-2-3 with limma, Glimma and edgeR *p*, *F1000Research*, vol. 5 : 1408.
- LAW Julie A. et JACOBSEN Steven E., 2010, <sup>2</sup> Establishing, maintaining and modifying DNA methylation patterns in plants and animals *p*, *Nature Reviews. Genetics*, vol. 11, n, 3 : 204 [ű20](#).
- LAWRENCE Moyra, DAUJAT Sylvain, et SCHNE" DER Robert, 2016, <sup>2</sup> Lateral Thinking: How Histone Modifications Regulate Gene Expression *p*, *Trends in genetics: T" G*, vol. 32, n, 1 : 42 [ű6](#).
- LE TR" ONNA" RE G., FRANC" S F., JAUBERT-POSSAMA" S., BONHOMME J., DE PAUW E., GAUTH" ER J-P, HAUBRUGE E., LEGEA" F., PRUN" ER-LETERME N., S" MON J-C, TANGUY S., et TAGU D., 2009, <sup>2</sup> Transcriptomic and proteomic analyses of seasonal photoperiodism in the pea aphid *p*, *BMC Genomics*, vol. 10 : 456.
- LE TR" ONNA" RE G., JAUBERT S., SABATER-MUNOZ B., BENEDETTO A., BONHOMME J., PRUN" ER-LETERME N., MART" NEZ-TORRES D., S" MON J. C., et TAGU D., 2007a, <sup>2</sup> Seasonal photoperiodism regulates the expression of cuticular and signalling protein genes in the pea aphid *p*, *Insect biochemistry and molecular biology*.
- LE TR" ONNA" RE G., JAUBERT S., SABATER-MU" E OZ B., BENEDETTO A., BONHOMME J., PRUN" ER-LETERME N., MART" NEZ-TORRES D., S" MON J. -C., et TAGU D., 2007b, <sup>2</sup> Seasonal photoperiodism regulates the expression of cuticular and signalling protein genes in the pea aphid *p*, *Insect Biochemistry and Molecular Biology*, vol. 37, n, 10 : 1094 [ű102](#).
- LE TR" ONNA" RE Gael, WUCHER Valentin, et TAGU Denis, 2013, <sup>2</sup> Genome expression control during the photoperiodic response of aphids *p*, *Physiological Entomology*, vol. 38, n, 2 : 117 [ű25](#).
- LEE Chee-Gun, RE" CHMAN Trevor W., BA" K Tina, et MATHEWS Michael B., 2004, <sup>2</sup> MLE Functions as a Transcriptional Regulator of the roX2 Gene *p*, *Journal of Biological Chemistry*, vol. 279, n, 46 : 47740 [ű7745](#).
- LEE J. T., DAV" DOW L. S., et WARSHAWSKY D., 1999, <sup>2</sup> Tsix, a gene antisense to Xist at the X-inactivation centre *p*, *Nature Genetics*, vol. 21, n, 4 : 400 [ű04](#).
- LEE R. C. et AMBROS V., 2001, <sup>2</sup> An extensive class of small RNAs in Caenorhabditis elegans *p*, *Science (New York, N.Y.)*, vol. 294, n, 5543 : 862 [ű64](#).
- LEE Rosalind C., FE" NBAUM Rhonda L., et AMBROS Victor, 1993, <sup>2</sup> The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14 *p*, *Cell*, vol. 75, n, 5 : 843 [ű54](#).
- LEES A. D., 1959, <sup>2</sup> The role of photoperiod and temperature in the determination of parthenogenetic and sexual forms in the aphid Megoura viciae Buckton *p*, *Journal of Insect Physiology*, vol. 3, n, 2 : 92 [ű17](#).

LEES A. D., 1960, <sup>2</sup> The role of photoperiod and temperature in the determination of parthenogenetic and sexual forms in the aphid *Megoura viciae* Buckton <sup>1960</sup>. The operation of the interval timer in young clones <sup>1960</sup>, *Journal of Insect Physiology*, vol. 4, n. 2 : 154 <sup>1960</sup>175.

LEES A. D., 1973, <sup>2</sup> Photoperiodic time measurement in the aphid *Megoura viciae* <sup>1973</sup>, *Journal of Insect Physiology*, vol. 19, n. 12 : 2279 <sup>1973</sup>316.

LEHNERTZ Bernhard, UEDA Yoshihide, DERJACK Alwin A. H. A., BRAUNSCHWEG Ulrich, PEREZ-BURGOS Laura, KUBCEK Stefan, CHEN Taiping, L' EN, JENUWEN Thomas, et PETERS Antoine H. F. M., 2003, <sup>2</sup> Suv39h-mediated histone H3 lysine 9 methylation directs DNA methylation to major satellite repeats at pericentric heterochromatin <sup>2003</sup>, *Current biology: CB*, vol. 13, n. 14 : 1192 <sup>2003</sup>200.

LEUNG Amy, PARKS Brian W., DU Juan, TRAC Candi, SETTEN Ryan, CHEN Yin, BROWN Kevin, LUS'S Aldons J., NATARAJAN Rama, et SCHONES Dustin E., 2014, <sup>2</sup> Open Chromatin Profiling in Mice Livers Reveals Unique Chromatin Variations Induced by High Fat Diet <sup>2014</sup>, *Journal of Biological Chemistry*, vol. 289, n. 34 : 23557 <sup>2014</sup>3567.

LEVINE Michael et TJAN Robert, 2003, <sup>2</sup> Transcription regulation and animal diversity <sup>2003</sup>, *Nature*, vol. 424, n. 6945 : 147.

L' Fang, SCHMANN Anja H., et SCOTT and Maxwell J., 2008, <sup>2</sup> Incorporation of the Noncoding roX RNAs Alters the Chromatin-Binding Specificity of the Drosophila MSL1/MSL2 Complex <sup>2008</sup>, *Molecular and Cellular Biology*, vol. 28, n. 4 : 1252.

L' Fei, MARTENSEN Rob, et CANDE W. Zacheus, 2011a, <sup>2</sup> Coordination of DNA replication and histone modification by the Rik1-Dos2 complex <sup>2011</sup>, *Nature*, vol. 475, n. 7355 : 244 <sup>2011</sup>48.

L' Heng, HANDSAKER Bob, WYSOKER Alec, FENNEL Tim, RUAN Jue, HOMER Nils, MARTH Gabor, ABECAS'S Goncalo, et DURBIN Richard, 2009, <sup>2</sup> The Sequence Alignment/Map format and SAMtools <sup>2009</sup>, *Bioinformatics*, vol. 25, n. 16 : 2078 <sup>2009</sup>079.

L' Qunhua, BROWN James B., HUANG Haiyan, et BCKEL Peter J., 2011b, <sup>2</sup> Measuring reproducibility of high-throughput experiments <sup>2011</sup>, *The Annals of Applied Statistics*, vol. 5, n. 3 : 1752 <sup>2011</sup>779.

L' Ruifeng, LU Yifang, L' Tingting, et L' Cheng, 2016, <sup>2</sup> 3Disease Browser: A Web server for integrating 3D genome and disease-associated chromosome rearrangement data <sup>2016</sup>, *Scientific Reports*, vol. 6 : srep34651.

L' Sung-Chou, PAN Chao-Yu, et L'N Wen-chang, 2006, <sup>2</sup> Bioinformatic discovery of microRNA precursors from human ESTs and introns <sup>2006</sup>, *BMC Genomics*, vol. 7 : 164.

L' Yuanyuan et TOLLEFSBOL Trygve O., 2011, <sup>2</sup> DNA methylation detection: Bisulfite genomic sequencing analysis <sup>2011</sup>, *Methods in molecular biology (Clifton, N.J.)*, vol. 791 : 11 <sup>2011</sup>1.

L'EBERMAN-A'DEN Erez, VAN BERKUM Nynke L., WLLAMS Louise, MAKAEV Maxim, RAGOCZY Tobias, TELL'NG Agnes, AM'T 'do, LAJO'E Bryan R., SABO Peter J., DORSCHNER Michael O., SANDSTROM Richard, BERNSTE'N Bradley, BENDER M. A., GROUD'NE Mark, GN'RKE Andreas, STAMATOYANNOPOULOS John, M'RNY Leonid A., LANDER Eric S., et DEKKER Job, 2009, <sup>2</sup> Comprehensive mapping of long-range interactions reveals folding principles of the human genome <sup>2009</sup>, *Science (New York, N.Y.)*, vol. 326, n. 5950 : 289 <sup>2009</sup>93.

L'N Hong, GUPTA Vibhor, VERM'LYEA Matthew D., FALC'AN' Francesco, LEE Jeannie T., O'NE'LL Laura P., et TURNER Bryan M., 2007, <sup>2</sup> Dosage Compensation in the Mouse Balances Up-Regulation and Silencing of X-Linked Genes <sup>2007</sup>, *PLOS Biology*, vol. 5, n. 12 : e326.

L'NG Guoyu et WAXMAN David J., 2013, <sup>2</sup> Isolation of Nuclei for use in Genome-wide DNase Hypersensitivity Assays to Probe Chromatin Structure *Methods in molecular biology (Clifton, N.J.)*, vol. 977.

L'STER Ryan, O'MALLEY Ronan C., TONT'F'L'PP'N' Julian, GREGORY Brian D., BERRY Charles C., M'LLAR A. Harvey, et ECKER Joseph R., 2008, <sup>2</sup> Highly integrated single-base resolution maps of the epigenome in Arabidopsis *Cell*, vol. 133, n. 3 : 523 <sup>36</sup>.

L'U Xue, HAO Lili, L' Dayong, ZHU Lihuang, et HU Songnian, 2015, <sup>2</sup> Long Non-coding RNAs and Their Biological Roles in Plants *Genomics, Proteomics & Bioinformatics*, vol. 13, n. 3 : 137 <sup>47</sup>.

LUCCHES' John C., 1993, <sup>2</sup> How widespread is the phenomenon of dosage compensation? *Seminars in Developmental Biology*, vol. 4, n. 2 : 141 <sup>45</sup>.

LUGER K., M' DER A. W., R'CHMOND R. K., SARGENT D. F., et R'CHMOND T. J., 1997, <sup>2</sup> Crystal structure of the nucleosome core particle at 2.8 Å resolution *Nature*, vol. 389, n. 6648 : 251 <sup>60</sup>.

LUGER Karolin, DECHASSA Mekonnen L., et TREMETH'CK David J., 2012, <sup>2</sup> New insights into nucleosome and chromatin structure: an ordered state or a disordered affair? *Nature reviews. Molecular cell biology*, vol. 13, n. 7 : 436 <sup>47</sup>.

LUGER Karolin et HANSEN Jeffrey C., 2005, <sup>2</sup> Nucleosome and chromatin fiber dynamics *Current Opinion in Structural Biology*, vol. 15, n. 2 : 188 <sup>96</sup>.

LU'JSTERBURG Martijn S., ACS Klara, ACKERMANN Leena, W'EGANT Wouter W., BEKKER-JENSEN Simon, LARSEN Dorthe H., KHANNA Kum Kum, VAN ATT'KUM Haico, MA'LAND Niels, et DANTUMA Nico P., 2012, <sup>2</sup> A new non-catalytic role for ubiquitin ligase RNF8 in unfolding higher-order chromatin structure *The EMBO journal*, vol. 31, n. 11 : 2511 <sup>527</sup>.

LUO Chongyuan, S'DOTE David J., ZHANG Yi, KERSTETTER Randall A., M'CHAEEL Todd P., et LAM Eric, 2013, <sup>2</sup> Integrative analysis of chromatin states in Arabidopsis identified potential regulatory mechanisms for natural antisense transcript production *The Plant Journal*, vol. 73, n. 1 : 77 <sup>90</sup>.

LYKO Frank, FORET Sylvain, KUCHARSK' Robert, WOLF Stephan, FALCKENHAYN Cassandra, et MALESZKA Ryszard, 2010, <sup>2</sup> The Honey Bee Epigenomes: Differential Methylation of Brain DNA in Queens and Workers *PLOS Biology*, vol. 8, n. 11 : e1000506.

LYKO Frank, RAMSAHOYE Bernard H., et JAEN'SCH Rudolf, 2000, <sup>2</sup> Development: DNA methylation in Drosophila melanogaster *Nature*, vol. 408, n. 6812 : 538 <sup>40</sup>.

LYMAN L. M., COPPS K., RASTELL' L., KELLEY R. L., et KURODA M. ., 1997, <sup>2</sup> Drosophila Male-Specific Lethal-2 Protein: Structure/Function Analysis and Dependence on Msl-1 for Chromosome Association *Genetics*, vol. 147, n. 4 : 1743 <sup>753</sup>.

LYNEN F. et OCHOA S., 1953, <sup>2</sup> Enzymes of fatty acid metabolism *Biochimica Et Biophysica Acta*, vol. 12, n. 1 <sup>14</sup> : 299 <sup>14</sup>.

LYON Mary F., 1961, <sup>2</sup> Gene Action in the X-chromosome of the Mouse (*Mus musculus* L.) *Nature*, vol. 190, n. 4773 : 372 <sup>73</sup>.

MA Jin-Biao, YE Keqiong, et PATEL Dinshaw J., 2004, <sup>2</sup> Structural basis for overhang-specific small interfering RNA recognition by the PAZ domain *Nature*, vol. 429, n. 6989 : 318 <sup>22</sup>.

MA Jin-Biao, YUAN Yu-Ren, MEISTER Gunter, PEI Yi, TUSCHL Thomas, et PATEL Dinshaw J., 2005, <sup>2</sup> Structural basis for 5' end-specific recognition of guide RNA by the *A. fulgidus* Piwi protein *p*, *Nature*, vol. 434, n, 7033 : 666 [670](#).

MA Lina, BAJC Vladimir B., et ZHANG Zhang, 2013, <sup>2</sup> On the classification of long non-coding RNAs *p*, *RNA biology*, vol. 10, n, 6 : 925 [933](#).

MACHANCK Philip et BAILEY Timothy L., 2011, <sup>2</sup> MEME-ChIP: motif analysis of large DNA datasets *p*, *Bioinformatics*, vol. 27, n, 12 : 1696 [1697](#).

MAGILL Jane M. et MAGILL Clint W., 1989, <sup>2</sup> DNA methylation in fungi *p*, *Developmental Genetics*, vol. 10, n, 2 : 63 [69](#).

MAHAJAN Shivani et BACHTROG Doris, 2015, <sup>2</sup> Partial Dosage Compensation in Strepsiptera, a Sister Group of Beetles *p*, *Genome Biology and Evolution*, vol. 7, n, 2 : 591 [600](#).

MANDROL M., BZZARO D., GUST M., MANCARD G. C., et BANCH U., 1999, <sup>2</sup> The role of rDNA genes in X chromosome association in the aphid *Acyrtosiphon pisum* *p*, *Genome*, vol. 42, n, 3 : 381 [386](#).

MANDROL Mauro et BORSATT Federica, 2007, <sup>2</sup> Analysis of heterochromatic epigenetic markers in the holocentric chromosomes of the aphid *Acyrtosiphon pisum* *p*, *Chromosome Research*, vol. 15, n, 8 : 1015.

MANOHAR Mridula, MOONEY Alex M., NORTH Justin A., NAKKULA Robin J., PCKNG Jonathan W., EDON Annick, FSHEL Richard, PORTER Michael G., et OTTESEN Jennifer J., 2009, <sup>2</sup> Acetylation of histone H3 at the nucleosome dyad alters DNA-histone binding *p*, *The Journal of Biological Chemistry*, vol. 284, n, 35 : 23312 [23321](#).

MARGUERON Raphael, TROJER Patrick, et REINBERG Danny, 2005, <sup>2</sup> The key to development: interpreting the histone code? *p*, *Current Opinion in Genetics & Development*, vol. 15, n, 2 : 163 [176](#).

MARSHALL GRAVES Jennifer A. et SHETTY Swathi, 2001, <sup>2</sup> Sex from W to Z: Evolution of vertebrate sex chromosomes and sex determining genes *p*, *Journal of Experimental Zoology*, vol. 290, n, 5 : 449 [462](#).

MATHERS Thomas C., CHEN Yazhou, KATHAKOTTL Gemy, LEGEA Fabrice, MUGFORD Sam T., BAA-PUYOULET Patrice, BRETAUDEAU Anthony, CLAVJO Bernardo, COLELLA Stefano, COLLIN Olivier, DALMAY Tamas, DERRIN Thomas, FENG Honglin, GABALDIN Toni, JORDAN Anna, JULCA Irene, KETTLES Graeme J., KOWTANICH Krissana, LAVENER Dominique, LENZ Paolo, LOPEZ-GOMOLLON Sara, LOSKA Damian, MAPLESON Daniel, MAUMUS Florian, MOXON Simon, PRICE Daniel R. G., SUGO Akiko, VAN MUNSTER Manuella, UZEST Marilyne, WATTE Darren, JANDER Georg, TAGU Denis, WILSON Alex C. C., VAN OOSTERHOUT Cock, SWARBRECK David, et HOGENHOUT Saskia A., 2017, <sup>2</sup> Erratum to: Rapid transcriptional plasticity of duplicated gene clusters enables a clonally reproducing aphid to colonise diverse plant species *p*, *Genome Biology*, vol. 18.

MATZKE Marjori A. et MOSHER Rebecca A., 2014, <sup>2</sup> RNA-directed DNA methylation: an epigenetic pathway of increasing complexity *p*, *Nature Reviews. Genetics*, vol. 15, n, 6 : 394 [408](#).

MAYR E., 1963, *Animal species and evolution.*, Harvard University Press; London: Oxford University Press.

MEISTER Gunter, 2013, <sup>2</sup> Argonaute proteins: functional insights and emerging roles *p*, *Nature Reviews Genetics*, vol. 14, n, 7 : 447 [459](#).

ME"STER Gunter, LANDTHALER Markus, PATKAN"OWSKA Agnieszka, DORSETT Yair, TENG Grace, et TUSCHL Thomas, 2004, <sup>2</sup> Human Argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs *þ*, *Molecular Cell*, vol. 15, n, 2 : 185 ¶97.

MELLER V. H., GORDADZE P. R., PARK Y., CHU X., STUCKENHOLZ C., KELLEY R. L., et KURODA M. ", 2000, <sup>2</sup> Ordered assembly of roX RNAs into MSL complexes on the dosage-compensated X chromosome in *Drosophila* *þ*, *Current Biology*, vol. 10, n, 3 : 136 ¶43.

MELLER Victoria H., 2003, <sup>2</sup> "nitiation of dosage compensation in *Drosophila* embryos depends on expression of the roX RNAs *þ*, *Mechanisms of Development*, vol. 120, n, 7 : 759 ¶67.

MENDJAN Sascha, TA"PALE Mikko, K"ND Jop, HOLZ Herbert, GEBHARDT Philipp, SCHELDER Malgorzata, VERMEULEN Michiel, BUSCA"NO Alessia, DUNCAN Kent, MUELLER Juerg, W"LM Matthias, STUNNENBERG Henk G., SAUMWEBER Harald, et AKHTAR Asifa, 2006, <sup>2</sup> Nuclear pore components are involved in the transcriptional regulation of dosage compensation in *Drosophila* *þ*, *Molecular Cell*, vol. 21, n, 6 : 811 ¶23.

MERCER Tim R., D"NGER Marcel E., SUNK"Ñ Susan M., MEHLER Mark F., et MATT"CK John S., 2008, <sup>2</sup> Specific expression of long noncoding RNAs in the mouse brain *þ*, *Proceedings of the National Academy of Sciences of the United States of America*, vol. 105, n, 2 : 716 ¶21.

MEYER Barbara J., 2005, <sup>2</sup> X-Chromosome dosage compensation *þ*, *WormBook*.

M"LES Peter W., 1999, <sup>2</sup> Aphid saliva *þ*, *Biological Reviews*, vol. 74, n, 1 : 41 ¶5.

M"LLER Joanna A. et W"DOM Jonathan, 2003, <sup>2</sup> Collaborative competition mechanism for gene activation in vivo *þ*, *Molecular and Cellular Biology*, vol. 23, n, 5 : 1623 ¶632.

M"LN Beatrice, SUN Yezhou, CHANG Weizhong, CREASY Todd, MAHURKAR Anup, SHETTY Amol, NURM"NSKY Dmitry, et NURM"NSKAYA Maria, 2014, <sup>2</sup> Map of open and closed chromatin domains in *Drosophila* genome *þ*, *BMC Genomics*, vol. 15 : 988.

M"RNY Leonid A., 2010, <sup>2</sup> Nucleosome-mediated cooperativity between transcription factors *þ*, *Proceedings of the National Academy of Sciences*, vol. 107, n, 52 : 22534 ¶2539.

M"URA Toru, BRAENDLE Christian, SH"NGLETON Alexander, S"SK Geoffroy, KAMBHAMPAT" Srinivas, et STERN David L., 2003, <sup>2</sup> A comparison of parthenogenetic and sexual embryogenesis of the pea aphid *Acyrtosiphon pisum* (Hemiptera: Aphidoidea) *þ*, *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*, vol. 295B, n, 1 : 59 ¶1.

MOAZED Danesh, 2011, <sup>2</sup> Mechanisms for the inheritance of chromatin states *þ*, *Cell*, vol. 146, n, 4 : 510 ¶18.

MONTLLOR Clytia B., MAXMEN Amy, et PURCELL Alexander H., 2002, <sup>2</sup> Facultative bacterial endosymbionts benefit pea aphids *Acyrtosiphon pisum* under heat stress *þ*, *Ecological Entomology*, vol. 27, n, 2 : 189 ¶95.

MORALES Violette, STRAUB Tobias, NEUMANN Martin F, MENGUS Gabrielle, AKHTAR Asifa, et BECKER Peter B, 2004, <sup>2</sup> Functional integration of the histone acetyltransferase MOF into the dosage compensation complex *þ*, *The EMBO Journal*, vol. 23, n, 11 : 2258 ¶268.

MORAN Nancy A., RUSSELL Jacob A., KOGA Ryuichi, et FUKATSU Takema, 2005, <sup>2</sup> Evolutionary Relationships of Three New Species of Enterobacteriaceae Living as Symbionts of Aphids and Other "nsects *þ*, *Applied and Environmental Microbiology*, vol. 71, n, 6 : 3302 ¶310.

MOTAMED" Mohammad R., ER"CA HONG Eun-Jin, L" Xue, GERBER Scott, DEN"SON Carilee, GYG" Steven, et MOAZED Danesh, 2008, <sup>2</sup> HP1 Proteins Form Distinct Complexes and Mediate Heterochromatic Gene Silencing by Non-Overlapping Mechanisms *p*, *Molecular cell*, vol. 32, n, 6 : 778 790.

MOTAMED" Mohammad R., VERDEL Andrů, COLMENARES Serafin U., GERBER Scott A., GYG" Steven P., et MOAZED Danesh, 2004, <sup>2</sup> Two RNAi complexes, R"TS and RDRC, physically interact and localize to noncoding centromeric RNAs *p*, *Cell*, vol. 119, n, 6 : 789 802.

MULLER H.J., 1932, <sup>2</sup> Further studies on the nature and causes of gene mutations. *p*, *Proceedings of the Sixth "International Congress of Genetics, "thaca, New York.*, vol. 1 : 213 255.

MUTT" Navdeep S., PARK Yoonseong, REESE John C., et REECK Gerald R., 2006, <sup>2</sup> RNAi knockdown of a salivary transcript leading to lethality in the pea aphid, *Acyrtosiphon pisum* *p*, *Journal of "nsect Science (Online)*, vol. 6 : 1 7.

NAGARAJAN Raman P., HOGART Amber R., GWYE Ynnez, MART" N Michelle R., et LASALLE Janine M., 2006, <sup>2</sup> Reduced MeCP2 expression is frequent in autism frontal cortex and correlates with aberrant MECP2 promoter methylation *p*, *Epigenetics*, vol. 1, n, 4 : e1-11.

NAGY Peter L., CLEARY Michael L., BROWN Patrick O., et L"EB Jason D., 2003, <sup>2</sup> Genomewide demarcation of RNA polymerase "" transcription units revealed by physical fractionation of chromatin *p*, *Proceedings of the National Academy of Sciences of the United States of America*, vol. 100, n, 11 : 6364 6369.

NAM Jin-Wu, SH" N Ki-Roo, HAN Jinju, LEE Yoontae, K" M V. Narry, et ZHANG Byoung-Tak, 2005, <sup>2</sup> Human microRNA prediction through a probabilistic co-learning model of sequence and structure *p*, *Nucleic Acids Research*, vol. 33, n, 11 : 3570 3581.

DE NAPOLES Mariana, MERMOUD Jacqueline E., WAKAO Rika, TANG Y. Amy, ENDOH Mitusuhiro, APPANAH Ruth, NESTEROVA Tatyana B., S" LVA Jose, OTTE Arie P., V" DAL Miguel, KOSEK" Haruhiko, et BROCKDORFF Neil, 2004, <sup>2</sup> Polycomb group proteins Ring1A/B link ubiquitylation of histone H2A to heritable gene silencing and X inactivation *p*, *Developmental Cell*, vol. 7, n, 5 : 663 676.

NAULT L. R., 1997, <sup>2</sup> Arthropod Transmission of Plant Viruses: a New Synthesis *p*, *Annals of the Entomological Society of America*, vol. 90, n, 5 : 521 541.

NAUMOVA N., SM" TH E. M., ZHAN Y., et DEKKER J., 2012, <sup>2</sup> Analysis of long-range chromatin interactions using Chromosome Conformation Capture *p*, *Methods (San Diego, Calif.)*, vol. 58, n, 3.

NEGUEMBOR Maria Victoria, XYNOS Alexandros, ONORAT" Maria Cristina, CACC" A Roberta, BORTOLANZA Sergia, GOD" O Cristina, P" STON" Mariaelena, CORONA Davide F., SCHOTTA Gunnar, et GABELL" N" Davide, 2013, <sup>2</sup> FSHD muscular dystrophy region gene 1 binds Suv4-20h1 histone methyltransferase and impairs myogenesis *p*, *Journal of Molecular Cell Biology*, vol. 5, n, 5 : 294 307.

NG James C. K. et FALK Bryce W., 2006, <sup>2</sup> Virus-Vector "nteractions Mediating Nonpersistent and Semipersistent Transmission of Plant Viruses *p*, *Annual Review of Phytopathology*, vol. 44, n, 1 : 183 212.

NG James C. K. et PERRY Keith L., 2004, <sup>2</sup> Transmission of plant viruses by aphid vectors *p*, *Molecular Plant Pathology*, vol. 5, n, 5 : 505 511.

NGUYEN Di Kim et D" STECHE Christine M., 2006, <sup>2</sup> Dosage compensation of the active X chromosome in mammals *p*, *Nature Genetics*, vol. 38, n, 1 : 47 53.

NICHOLSON Scott J, NICKERSON Michael L, DEAN Michael, SONG Yan, HOYT Peter R, RHEE Hwanseok, KIM Changhoon, et PUTERKA Gary J, 2015, <sup>2</sup> The genome of *Diuraphis noxia*, a global aphid pest of small grains *P*, *BMC Genomics*, vol. 16.

NISHIMASU Hiroshi, NISHIZU Hirotosugu, SAITO Kuniaki, FUKUHARA Satoshi, KAMATAN Miharuru K., BONNEFOND Luc, MATSUMOTO Naoki, NISHIZAWA Tomohiro, NAKANAGA Keita, AOKI Junken, NISHITANI Ryuichiro, SUGIYAMA Haruhiko, SUGIYAMA Mikiko C., et NUREKI Osamu, 2012, <sup>2</sup> Structure and function of Zucchini endoribonuclease in piRNA biogenesis *P*, *Nature*, vol. 491, n<sup>o</sup> 7423 : 284-287.

NITTA Kazuhiro R., JOLMA Arttu, YIN Yimeng, MORGUNOVA Ekaterina, KIVIOJA Teemu, AKHTAR Junaid, HENS Korneel, TOIVONEN Jarkko, DEPLANCKE Bart, FURLONG Eileen E. M., et TAPALE Jussi, 2015, <sup>2</sup> Conservation of transcription factor binding specificities across 600 million years of bilateria evolution *P*, *eLife*, vol. 4 : e04837.

NOLL M., 1974, <sup>2</sup> Subunit structure of chromatin *P*, *Nature*, vol. 251, n<sup>o</sup> 5472 : 249-251.

NORA Elphège P., LAJOIE Bryan R., SCHULZ Edda G., GORGETT Luca, OKAMOTO Takuhiro, SERVANT Nicolas, PLOUT Tristan, VAN BERKUM Nynke L., MEISING Johannes, SEDAT John, GRIBNAU Joost, BARLLOT Emmanuel, BLATHGEN Nils, DEKKER Job, et HEARD Edith, 2012, <sup>2</sup> Spatial partitioning of the regulatory landscape of the X-inactivation center *P*, *Nature*, vol. 485, n<sup>o</sup> 7398 : 381-385.

NOVOTNÉ Jana, HAVELKA Jan, STARH Petr, KOUTECKÝ Petr, et VTKOVÉ Magda, 2011, <sup>2</sup> Karyotype analysis of the Russian wheat aphid, *Diuraphis noxia* (Kurdjumov) (Hemiptera: Aphididae) reveals a large X chromosome with rRNA and histone gene families *P*, *Genetica*, vol. 139, n<sup>o</sup> 3 : 281-289.

OGBOURNE Steven et ANTAL'S Toni M, 1998, <sup>2</sup> Transcriptional control and the role of silencers in transcriptional regulation in eukaryotes *P*, *Biochemical Journal*, vol. 331, n<sup>o</sup> 1 : 1-4.

OLINS A. L. et OLINS D. E., 1974, <sup>2</sup> Spheroid chromatin units (v bodies) *P*, *Science (New York, N.Y.)*, vol. 183, n<sup>o</sup> 4122 : 330-332.

OLIVER Kerry M., RUSSELL Jacob A., MORAN Nancy A., et HUNTER Martha S., 2003, <sup>2</sup> Facultative bacterial symbionts in aphids confer resistance to parasitic wasps *P*, *Proceedings of the National Academy of Sciences*, vol. 100, n<sup>o</sup> 4 : 1803-1807.

ONEILL Laura P., SPOTSWOOD Hugh T., FERNANDO Milan, et TURNER Bryan M., 2008, <sup>2</sup> Differential loss of histone H3 isoforms mono-, di- and tri-methylated at lysine 4 during X-inactivation in female embryonic stem cells *P*, *Biological Chemistry*, vol. 389, n<sup>o</sup> 4 : 365-370.

OO'Lezanne et WOOD 'an C., 2007, <sup>2</sup> Chromatin crosstalk in development and disease: lessons from REST *P*, *Nature Reviews. Genetics*, vol. 8, n<sup>o</sup> 7 : 544-554.

ORLANDO Enzo, 1974, <sup>2</sup> Sex Determination in *Megoura viciae* Buckton (homoptera Aphididae) *P*, *Monitore Zoologico Italiano - Italian Journal of Zoology*, vol. 8, n<sup>o</sup> 1 : 61-70.

PAGE Robert E., SCHEINER Ricarda, ERBER Joachim, et AMDAM Gro V., 2006, <sup>2</sup> The Development and Evolution of Division of Labor and Foraging Specialization in a Social Insect (*Apis mellifera* L.) *P*, *Current topics in developmental biology*, vol. 74 : 253-286.

PAL Arka et V'OSO Beatriz, 2015, <sup>2</sup> The X Chromosome of Hemipteran Insects: Conservation, Dosage Compensation and Sex-Biased Expression *P*, *Genome Biology and Evolution*, vol. 7, n<sup>o</sup> 12 : 3259-3268.

PALMER M. J., RICHMAN R., RICHTER L., et KURODA M., 1994, <sup>2</sup> Sex-specific regulation of the male-specific lethal-1 dosage compensation gene in *Drosophila* *P*, *Genes & Development*, vol. 8, n<sup>o</sup> 6 : 698-706.

- PARRISH S., FLEENOR J., XU S., MELLO C., et FIRE A., 2000, <sup>2</sup> Functional anatomy of a dsRNA trigger: differential requirement for the two trigger strands in RNA interference *Cell*, vol. 102, n. 5 : 1077-1087.
- PARTHASARATHY R. et PALLI Subba R., 2008, <sup>2</sup> Proliferation and differentiation of intestinal stem cells during metamorphosis of the red flour beetle, *Tribolium castaneum* *Developmental Dynamics: An Official Publication of the American Association of Anatomists*, vol. 237, n. 4 : 893-908.
- PARTHASARATHY R., TAN A., BAN H., et PALLI Subba R., 2008, <sup>2</sup> Transcription factor broad suppresses precocious development of adult structures during larval-pupal metamorphosis in the red flour beetle, *Tribolium castaneum* *Mechanisms of Development*, vol. 125, n. 3 : 299-313.
- PATRIDGE Janet F., SCOTT Kristin S. C., BANNISTER Andrew J., KOUZARDES Tony, et ALLSHIRE Robin C., 2002, <sup>2</sup> cis-acting DNA from fission yeast centromeres mediates histone H3 methylation and recruitment of silencing factors and cohesin to an ectopic site *Current biology: CB*, vol. 12, n. 19 : 1652-1660.
- PAWELETZ Neidhard, 2001, <sup>2</sup> Walther Flemming: pioneer of mitosis research *Nature Reviews Molecular Cell Biology*, vol. 2, n. 1 : 72-75.
- PENER Meir Paul et SIMPSON Stephen J., 2009, <sup>2</sup> Locust Phase Polyphenism: An Update *Advances in Insect Physiology*, vol. 36 : 1-72.
- PENNING SARI, ALLAN James, et DAVEY Colin S., 2005, <sup>2</sup> DNA methylation, nucleosome formation and positioning *Briefings in Functional Genomics & Proteomics*, vol. 3, n. 4 : 351-361.
- PENNY G. D., KAY G. F., SHEARDOWN S. A., RASTAN S., et BROCKDORFF N., 1996, <sup>2</sup> Requirement for Xist in X chromosome inactivation *Nature*, vol. 379, n. 6561 : 131-137.
- PESSA Eugénie, MAKINO Takashi, BALLY-BECHET Marc, MCLYSAGHT Aoife, et MARAS Gabriel A. B., 2012, <sup>2</sup> Mammalian X chromosome inactivation evolved as a dosage-compensation mechanism for dosage-sensitive genes on the X chromosome *Proceedings of the National Academy of Sciences*, vol. 109, n. 14 : 5346-5351.
- PRIKE Nathan et FOSTER William A, 2008, <sup>2</sup> The ecology of altruism in a clonal insect *Ecology of social evolution* : 37-46.
- PLOT F., PHILIPPE J.M., LEMMERS C., CHAUVIN J.P., et LECUIT T., 2006, <sup>2</sup> Developmental control of nuclear morphogenesis and anchoring by charleston, identified in a functional genomic screen of *Drosophila* cellularisation. *Development*, vol. 133, n. 4 : 711-723.
- POMBO Ana et DILLON Niall, 2015, <sup>2</sup> Three-dimensional genome architecture: players and mechanisms *Nature Reviews Molecular Cell Biology*, vol. 16, n. 4 : 245-257.
- PONTS Nadia, HARRIS Elena Y., PRUDHOMME Jacques, WICK van, ECKHARDT-LUDKA Colleen, HICKS Glenn R., HARDMAN Gary, LONARDI Stefano, et LE ROCH Karine G., 2010, <sup>2</sup> Nucleosome landscape and control of transcription in the human malaria parasite *Genome Research*, vol. 20, n. 2 : 228-238.
- PRESTEL Matthias, FELLER Christian, STRAUB Tobias, MÜLLHNER Heike, et BECKER Peter B., 2010, <sup>2</sup> The activation potential of MOF is constrained for dosage compensation *Molecular Cell*, vol. 38, n. 6 : 815-826.
- PRINCE Eldon G., KRKLAND Donna, et DEMUTH Jeffery P., 2010, <sup>2</sup> Hyperexpression of the X Chromosome in Both Sexes Results in Extensive Female Bias of X-Linked Genes in the Flour Beetle *Genome Biology and Evolution*, vol. 2 : 336-346.

QUINN Jeffrey J. et CHANG Howard Y., 2016, <sup>2</sup> Unique features of long non-coding RNA biogenesis and function *p*, *Nature Reviews Genetics*, vol. 17, n, 1 : 47 [162](#).

RADMAN-LVAJA Marta et RANDO Oliver J., 2010, <sup>2</sup> Nucleosome positioning: How is it established, and why does it matter? *p*, *Developmental Biology*, vol. 339, n, 2 : 258 [266](#).

RAJA Sunil Jayaramaiah, CHARAPATSA "ryna, CONRAD Thomas, VAQUERZAS Juan M., GEBHARDT Philipp, HOLZ Herbert, KADLEC Jan, FRATERMAN Sven, LUSCOMBE Nicholas M., et AKHTAR Asifa, 2010, <sup>2</sup> The nonspecific lethal complex is a transcriptional regulator in *Drosophila* *p*, *Molecular Cell*, vol. 38, n, 6 : 827 [841](#).

RAJAKUMARA Eerappa, WANG Zhentian, MA Honghui, HU Lulu, CHEN Hao, L"n Yan, GUO Rui, WU Feizhen, L" Haitao, LAN Fei, SH" Yujiang Geno, XU Yanhui, PATEL Dinshaw J., et SH" Yang, 2011, <sup>2</sup> PHD finger recognition of unmodified histone H3R2 links UHRF1 to regulation of euchromatic gene expression *p*, *Molecular Cell*, vol. 43, n, 2 : 275 [284](#).

RAM REZ Fidel, D NDAR Friederike, D"ehl Sarah, GR N"NG Bj rn A., et MANKE Thomas, 2014, <sup>2</sup> deepTools: a flexible platform for exploring deep-sequencing data *p*, *Nucleic Acids Research*, vol. 42, n, W1 : W187 [191](#).

RAM REZ Fidel, RYAN Devon P., GR N"NG Bj rn, BHARDWAJ Vivek, K"LPERT Fabian, R"CHTER Andreas S., HEYNE Steffen, D NDAR Friederike, et MANKE Thomas, 2016, <sup>2</sup> deepTools2: a next generation web server for deep-sequencing data analysis *p*, *Nucleic Acids Research*, vol. 44, n, W1 : W160 [165](#).

RASH" D Naim U., G"RES" Paul G., "BRAH" M Joseph G., SUN Wei, et L"EB Jason D., 2011, <sup>2</sup> Z" NBA integrates local covariates with DNA-seq data to identify broad and narrow regions of enrichment, even within amplified genomic regions *p*, *Genome Biology*, vol. 12 : R67.

RATTNER Barbara P. et MELLER Victoria H., 2004, <sup>2</sup> *Drosophila* male-specific lethal 2 protein controls sex-specific expression of the roX genes *p*, *Genetics*, vol. 166, n, 4 : 1825 [1832](#).

REAM Thomas S., HAAG Jeremy R., W"ERZB"CK" Andrzej T., N"CORA Carrie D., NORBECK Angela D., ZHU Jian-Kang, HAGEN Gretchen, GU"LF OYLE Thomas J., PASA-TOL" Ljiljana, et P"KAARD Craig S., 2009, <sup>2</sup> Subunit compositions of the RNA-silencing enzymes Pol "V and Pol V reveal their origins as specialized forms of RNA polymerase "" *p*, *Molecular Cell*, vol. 33, n, 2 : 192 [203](#).

RE"MAND J ri, ARAK Tabet, ADLER Priit, KOLBERG Liis, RE"SBERG Sulev, PETERSON Hedi, et V"LO Jaak, 2016, <sup>2</sup> g:Profiler   a web server for functional interpretation of gene lists (2016 update) *p*, *Nucleic Acids Research*, vol. 44, n, W1 : W83 [89](#).

RE"MAND J ri, KULL Meelis, PETERSON Hedi, HANSEN Jaanus, et V"LO Jaak, 2007, <sup>2</sup> g:Profiler   a web-based toolset for functional profiling of gene lists from large-scale experiments *p*, *Nucleic Acids Research*, vol. 35, n, Web Server issue : W193 [200](#).

R"CHARD Gautier, LEGEA" Fabrice, PRUN"ER-LETERME Nathalie, BRETAUDEAU Anthony, TAGU Denis, JAQU"RY Julie, et LE TR"ONNA"RE Ga l, 2017a, <sup>2</sup> Dosage compensation and sex-specific epigenetic landscape of the X chromosome in the pea aphid *p*, *Epigenetics & Chromatin*, vol. 10 : 30.

     , 2017b, <sup>2</sup> Dosage compensation and sex-specific epigenetic landscape of the X chromosome in the pea aphid *p*, *Epigenetics & Chromatin*, vol. 10 : 30.

R"DER JR S. D., SR"NVASAN D. G., et H"LGARTH R. S., 2010, <sup>2</sup> Chromatin-remodelling proteins of the pea aphid, *Acyrtosiphon pisum* (Harris) *p*, *Insect Molecular Biology*, vol. 19 : 201 [214](#).

R"ETHOVEN Jean-Jack M., 2010, <sup>2</sup> Regulatory Regions in DNA: Promoters, Enhancers, Silencers, and "nsulators p, in *Computational Biology of Transcription Factor Binding*, Humana Press, Totowa, NJ : 33 12.

R"NN John L. et CHANG Howard Y., 2012, <sup>2</sup> Genome regulation by long noncoding RNAs p, *Annual Review of Biochemistry*, vol. 81 : 145 166.

R"SPE Claude et P"ERRE Jean-Sűbastien, 1998, <sup>2</sup> Coexistence between Cyclical Parthenogens, Obligate Parthenogens, and "ntermediates in a Fluctuating Environment p, *Journal of Theoretical Biology*, vol. 195, n, 1 : 97 110.

R"ST"MI K" Ari, YL"KORKALA Olavi, PESONEN Kristina, PERHEENTUPA Jaako, et V"NKKA Lasse, 1991, <sup>2</sup> Human milk stimulates prostacyclin production by cultured human vascular endothelial cells p, *The Journal of Clinical Endocrinology & Metabolism*, vol. 72, n, 3 : 623 627.

R"ZZO Jason M. et S"NHSA Satrajit, 2014, <sup>2</sup> Analyzing the global chromatin structure of keratinocytes by MNase-seq p, *Methods in Molecular Biology (Clifton, N.J.)*, vol. 1195 : 49 69.

ROBERTSON Keith D., 2005, <sup>2</sup> DNA methylation and human disease p, *Nature Reviews. Genetics*, vol. 6, n, 8 : 597 610.

ROB"NSON James T., THORVALDSDH TT"R Helga, W"CKLER Wendy, GUTTMAN Mitchell, LANDER Eric S., GETZ Gad, et MES"ROV Jill P., 2011a, <sup>2</sup> "ntegrative Genomics Viewer p, *Nature biotechnology*, vol. 29, n, 1 : 24 26.

ROB"NSON Katie L., TOH"D"-ESFAHAN" Donya, LO Nathan, S"MPSON Stephen J., et SWORD Gregory A., 2011b, <sup>2</sup> Evidence for widespread genomic methylation in the migratory locust, *Locusta migratoria* (Orthoptera: Acrididae) p, *PloS One*, vol. 6, n, 12 : e28167.

ROSS Robert J., WE"NER Molly M., et L"HAIFAN, 2014, <sup>2</sup> P"W" proteins and P"W"-interacting RNAs in the soma p, *Nature*, vol. 505, n, 7483 : 353 359.

RUSSELL Jacob A., WELDON Stephanie, SM"TH Andrew H., K"MYUNG Kyungsun L., HU Yi, UKAS"K Piotr, DOLL Steven, ANASTOPOULOS "oannis, NOV"Matthew, et OL"VER Kerry M., 2013, <sup>2</sup> Uncovering symbiont-driven genetic diversity across North American pea aphids p, *Molecular Ecology*, vol. 22, n, 7 : 2045 2059.

SAAYMAN Sheena, ACKLEY Amanda, TURNER Anne-Marie W., FAM"GL"ETT" Marylinda, BOSQUE Alberto, CLEMSON Matthew, PLANELLES Vicente, et MORR"S Kevin V., 2014, <sup>2</sup> An H"V-encoded antisense long noncoding RNA epigenetically regulates viral transcription p, *Molecular Therapy: The Journal of the American Society of Gene Therapy*, vol. 22, n, 6 : 1164 1175.

SAKAMOTO H, "NOUE K, H"GUCH" , ONO Y, et SH"URA Y, 1992, <sup>2</sup> Control of Drosophila Sex-lethal pre-mRNA splicing by its own female-specific product. p, *Nucleic Acids Research*, vol. 20, n, 21 : 5533 5540.

SAPOUNTZ"S Panagiotis, DUPORT Gabrielle, BALMAND Sűverine, GAGET Karen, JAUBERT-POSSAMA" Stűphanie, FEBVAY Gűrard, CHARLES Hubert, RAHB" Yvan, COLELLA Stefano, et CALEVRO Federica, 2014, <sup>2</sup> New insight into the RNA interference response against cathepsin-L gene in the pea aphid, *Acyrtosiphon pisum*: Molting or gut phenotypes specifically induced by injection or feeding treatments p, *Insect Biochemistry and Molecular Biology*, vol. 51 : 20 22.

SAV"C Daniel, GERTZ Jason, JA"N Preti, COOPER Gregory M, et MYERS Richard M, 2013, <sup>2</sup> Mapping genome-wide transcription factor binding sites in frozen tissues p, *Epigenetics & Chromatin*, vol. 6 : 30.

SCARBOROUGH Claire L., FERRAR" Julia, et GODFRAY H. C. J., 2005, <sup>2</sup> Aphid Protected from Pathogen by Endosymbiont p, *Science*, vol. 310, n, 5755 : 1781 1781.

SCHICK Sandra, FOURNIER David, THAKURELA Sudhir, SAHU Sanjeeb Kumar, GARDING Angela, et THAR Vijay K., 2015, <sup>2</sup> Dynamics of chromatin accessibility and epigenetic state in response to UV damage *Cell Sci*: jcs.173633.

SCHOENHERR C. J. et ANDERSON D. J., 1995, <sup>2</sup> Silencing is golden: negative regulation in the control of neuronal gene transcription *Current Opinion in Neurobiology*, vol. 5, n. 5 : 566-571.

SCHONES Dustin E., CUI Kairong, CUDDAPAH Suresh, ROH Tae-Young, BARSK Artem, WANG Zhibin, WEI Gang, et ZHAO Keji, 2008, <sup>2</sup> Dynamic Regulation of Nucleosome Positioning in the Human Genome *Cell*, vol. 132, n. 5 : 887-898.

SCHWARTZENTRUBER Jeremy, KORSHUNOV Andrey, LIU Xiao-Yang, JONES David T. W., PFAFF Elke, JACOB Karine, STURM Dominik, FONTEBASSO Adam M., QUANG Dong-Anh Khuong, TIJNJES Martje, HOVESTADT Volker, ALBRECHT Steffen, KOOL Marcel, NANTEL Andre, KONERMANN Carolin, LINDROTH Anders, JAGER Natalie, RAUSCH Tobias, RYZHOVA Marina, KORBEL Jan O., HELSCHER Thomas, HAUSER Peter, GARAM Miklos, KLEKNER Almos, BOGNAR Laszlo, EBINGER Martin, SCHUHMANN Martin U., SCHEURLEN Wolfram, PEKRUN Arnulf, FRUHWALD Michael C., ROGGENDORF Wolfgang, KRAMM Christoph, DIJKEN Matthias, ATKINSON Jeffrey, LEPAGE Pierre, MONTPETIT Alexandre, ZAKRZEWSKA Magdalena, ZAKRZEWSKI Krzysztof, LEBERSKI Pawel P., DONG Zhifeng, SEGEL Peter, KULOZIK Andreas E., ZAPATKA Marc, GUHA Abhijit, MALKIN David, FELSBERG Jörg, REIFENBERGER Guido, VON DEMLING Andreas, CHIMURA Koichi, COLLINS V. Peter, WITTT Hendrik, MÜLDE Till, WITTT Olaf, ZHANG Cindy, CASTELO-BRANCO Pedro, LICHTER Peter, FAURY Damien, TABOR Uri, PLASS Christoph, MAJEWSKI Jacek, PFISTER Stefan M., et JABADO Nada, 2012, <sup>2</sup> Driver mutations in histone H3.3 and chromatin remodelling genes in paediatric glioblastoma *Nature*, vol. 482, n. 7384 : 226-231.

SELKER E. U., 1997, <sup>2</sup> Epigenetic phenomena in filamentous fungi: useful paradigms or repeat-induced confusion? *Trends in genetics: T<sub>G</sub>*, vol. 13, n. 8 : 296-301.

SEQUEIRA-MENDES Joana, ARAGI EZ Irene, PEREZ Ramon, MENDEZ-GONZALEZ Raul, ZHANG Xiaoyu, JACOBSEN Steven E., BASTOLLA Ugo, et GUTERREZ Crisanto, 2014, <sup>2</sup> The Functional Topography of the Arabidopsis Genome is Organized in a Reduced Number of Linear Motifs of Chromatin States *The Plant Cell*, vol. 26, n. 6 : 2351-2366.

SEXTON Tom, YAFFE Eitan, KENIGSBERG Ephraim, BANTIGNES Frédéric, LEBLANC Benjamin, HOCHMAN Michael, PARRINELLO Hugues, TANAY Amos, et CAVALLI Giacomo, 2012, <sup>2</sup> Three-dimensional folding and functional organization principles of the Drosophila genome *Cell*, vol. 148, n. 3 : 458-472.

SHEN Li, WU Hao, DINH Dinh, YAMAGUCHI Shinpei, DALESSIO Ana C., FUNG Ho-Lim, ZHANG Kun, et ZHANG Yi, 2013, <sup>2</sup> Genome-wide analysis reveals TET- and TDG-dependent 5-methylcytosine oxidation dynamics *Cell*, vol. 153, n. 3 : 692-706.

SHI Yuan Yuan, HUANG Zachary Y., ZENG Zhi Jiang, WANG Zi Long, WU Xiao Bo, et YAN Wei Yu, 2011, <sup>2</sup> Diet and Cell Size Both Affect Queen-Worker Differentiation through DNA Methylation in Honey Bees (*Apis mellifera*, Apidae) *PLOS ONE*, vol. 6, n. 4 : e18808.

SHIBAO Harunobu, KITSUKAKE Mayako, MATSUYAMA Shigeru, FUKATSU Takema, et SHIMADA Masakazu, 2010, <sup>2</sup> Mechanisms regulating caste differentiation in an aphid social system *Communicative & Integrative Biology*, vol. 3, n. 1 : 1-5.

SHINGLETON Alexander W., SISK Geoffroy C., et STERN David L., 2003, <sup>2</sup> Diapause in the pea aphid (*Acyrtosiphon pisum*) is a slowing but not a cessation of development *BMC Developmental Biology*, vol. 3 : 7.

SHOGREN-KNAAK Michael, SHIM Haruhiko, SUN Jian-Min, PAZIN Michael J., DAVIE James R., et PETERSON Craig L., 2006, <sup>2</sup> Histone H4-K16 Acetylation Controls Chromatin Structure and Protein Interactions *Science*, vol. 311, n. 5762 : 844-847.

SJEN T., FLEENOR J., SUMMER F., THIJSEN K. L., PARRISH S., THOMAS L., PLASTERK R. H., et FIRE A., 2001, <sup>2</sup> On the role of RNA amplification in dsRNA-triggered gene silencing *Cell*, vol. 107, n. 4 : 465-476.

SILVA Jose, MAK Winifred, ZVETKOVA Iona, APPANAH Ruth, NESTEROVA Tatyana B., WEBSTER Zoe, PETERS Antoine H. F. M., JENUWEIN Thomas, OTTE Arie P., et BROCKDORFF Neil, 2003, <sup>2</sup> Establishment of histone h3 methylation on the inactive X chromosome requires transient recruitment of Eed-Enx1 polycomb group complexes *Developmental Cell*, vol. 4, n. 4 : 481-495.

SIMOLA Daniel F., WISSLER Lothar, DONAHUE Greg, WATERHOUSE Robert M., HELMKAMPF Martin, ROUX Julien, NYGAARD Sanne, GLASTAD Karl M., HAGEN Darren E., VILJAKINEN Lumi, REESE Justin T., HUNT Brendan G., GRAUR Dan, ELHAKK Eran, KRIVENTSEVA Evgenia V., WEN Jiayu, PARKER Brian J., CASH Elizabeth, PRIVMAN Eyal, CHILDERS Christopher P., MUÑOZ-TORRES Monica C., BOOMSMA Jacobus J., BORNBERG-BAUER Erich, CURRIE Cameron R., ELSK Christine G., SUEN Garret, GOODSMAN Michael A. D., KELLER Laurent, LEBERG Jørgen, RAWLS Alan, REINBERG Danny, SMITH Chris D., SMITH Chris R., TSUTSU Neil, WURM Yannick, ZDOBNOV Evgeny M., BERGER Shelley L., et GADAU Jørgen, 2013a, <sup>2</sup> Social insect genomes exhibit dramatic evolution in gene composition and regulation while preserving regulatory features linked to sociality *Genome Research*, vol. 23, n. 8 : 1235-1247.

SIMOLA Daniel F., YE Chaoyang, MUTT NAVDEEP S., DOLEZAL Kelly, BONASO Roberto, LEBERG Jørgen, REINBERG Danny, et BERGER Shelley L., 2013b, <sup>2</sup> A chromatin link to caste identity in the carpenter ant *Camponotus floridanus* *Genome Research*, vol. 23, n. 3 : 486-496.

SIMON Jean-Christophe, RUSPE Claude, et SUNNUCKS Paul, 2002, <sup>2</sup> Ecology and evolution of sex in aphids *Trends in Ecology & Evolution*, vol. 17, n. 1 : 34-39.

SIMON Jean-Christophe, STOECKEL Solenn, et TAGU Denis, 2010, <sup>2</sup> Evolutionary and functional insights into reproductive strategies of aphids *Comptes Rendus Biologies*, vol. 333, n. 6 : 488-496.

SIMON Jeremy M., GRES Paul G., DAVIS Ian J., et LEB Jason D., 2012, <sup>2</sup> Using FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) to isolate active regulatory DNA *Nature protocols*, vol. 7, n. 2 : 256-267.

SIMON Jeremy M., 2013, <sup>2</sup> A detailed protocol for formaldehyde-assisted isolation of regulatory elements (FAIRE) *Current Protocols in Molecular Biology*, vol. Chapter 21 : Unit21.26.

SIMPSON S. J., DESPLAND E., HIGLE B. F., et DODGSON T., 2001, <sup>2</sup> Gregarious behavior in desert locusts is evoked by touching their back legs *Proceedings of the National Academy of Sciences*, vol. 98, n. 7 : 3895-3897.

SOM Mikiko C., SATO Kaoru, PEZC Dubravka, et ARAVIN Alexei A., 2011, <sup>2</sup> Piwi-interacting small RNAs: the vanguard of genome defence *Nature Reviews. Molecular Cell Biology*, vol. 12, n. 4 : 246-258.

SMALE Stephen T et BALTMORE David, 1989, <sup>2</sup> The initiator as a transcription control element *Cell*, vol. 57, n. 1 : 103-113.

SMITH Chris R., MUTT NAVDEEP S., JASPER W. Cameron, NADU Agni, SMITH Christopher D., et GADAU Jørgen, 2012, <sup>2</sup> Patterns of DNA Methylation in Development, Division of Labor and Hybridization in an Ant with Genetic Caste Determination *PLOS ONE*, vol. 7, n. 8 : e42433.

SMITH Gilbert, CHEN Yun-Ru, BLISSARD Gary W., et BRISCOE Adriana D., 2014, <sup>2</sup> Complete Dosage Compensation and Sex-Biased Gene Expression in the Moth *Manduca sexta* ♀, *Genome Biology and Evolution*, vol. 6, n. 3 : 526–537.

SMITH John E., CUMMINGS Craig A., et CRONMILLER Claire, 2002, <sup>2</sup> Daughterless coordinates somatic cell proliferation, differentiation and germline cyst survival during follicle formation in *Drosophila* ♀, *Development (Cambridge, England)*, vol. 129, n. 13 : 3255–3267.

SONG Ji-Joon, LEE Jidong, TOLLA Niraj H., SCHNEIDERMAN Jonathan, SMITH Stephanie K., MARTENSSEN Robert A., HANNON Gregory J., et JOSHUA-TOR Leemor, 2003, <sup>2</sup> The crystal structure of the Argonaute2 PAZ domain reveals an RNA binding motif in RNAi effector complexes ♀, *Nature Structural Biology*, vol. 10, n. 12 : 1026–1032.

SONG Lingyun et CRAWFORD Gregory E., 2010, <sup>2</sup> DNase-seq: a high-resolution technique for mapping active gene regulatory elements across the genome from mammalian cells ♀, *Cold Spring Harbor protocols*, vol. 2010, n. 2 : pdb.prot5384.

SPANNHOFF Astrid, KIM Yong Kee, RAYNAL Noel J.-M., GHARIBYAN Vazganush, SU Ming-Bo, ZHOU Yue-Yang, LI Jia, CASTELLANO Sabrina, SBARDELLA Gianluca, BOISSA Jean-Pierre J., et BEDFORD Mark T., 2011, <sup>2</sup> Histone deacetylase inhibitor activity in royal jelly might facilitate caste switching in bees ♀, *EMBO reports*, vol. 12, n. 3 : 238–243.

SPRUJT Cornelia G., GNERLICH Felix, SMITS Arne H., PFAFFENEDER Toni, JANSEN Pascal W. T. C., BAUER Christina, MANNZEL Martin, WAGNER Mirko, MULLER Markus, KHAN Fariha, EBERL H. Christian, MENSINGA Anneloes, BRINKMAN Arie B., LEPIKOV Konstantin, MULLER Udo, WALTER Jörn, BOELEN Rolf, VAN NINGEN Hugo, LEONHARDT Heinrich, CARELL Thomas, et VERMEULEN Michiel, 2013, <sup>2</sup> Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives ♀, *Cell*, vol. 152, n. 5 : 1146–1159.

STANLEY Willard F., 1935, <sup>2</sup> The effect of temperature upon wing size in *Drosophila* ♀, *Journal of Experimental Zoology*, vol. 69, n. 3 : 459–495.

STATES D. J. et GISH W., 1994, <sup>2</sup> Combined use of sequence similarity and codon bias for coding region identification ♀, *Journal of Computational Biology: A Journal of Computational Molecular Cell Biology*, vol. 1, n. 1 : 39–50.

STEEL C. G. et LEES A. D., 1977, <sup>2</sup> The role of neurosecretion in the photoperiodic control of polymorphism in the aphid *Megoura viciae* ♀, *Journal of Experimental Biology*, vol. 67, n. 1 : 117–135.

STERN David L et FOSTER William A, 1996, <sup>2</sup> The evolution of soldiers in aphids ♀, *Biological Reviews*, vol. 71, n. 1 : 27–79.

STEVENS Tim J., LANDO David, BASU Srinjan, ATKINSON Liam P., CAO Yang, LEE Steven F., LEEB Martin, WOHLFAHRT Kai J., BOUCHER Wayne, O'SHAUGHNESSY-KIRWAN Aoife, CRAMARD Julie, FAURE Andre J., RALSER Meryem, BLANCO Enrique, MOREY Lluís, SANSÓ Miriam, PALAYRET Matthieu G. S., LEHNER Ben, DI CROCE Luciano, WUTZ Anton, HENDRICH Brian, KLENERMAN Dave, et LAUE Ernest D., 2017, <sup>2</sup> 3D structures of individual mammalian genomes studied by single-cell Hi-C ♀, *Nature*, vol. 544, n. 7648 : 59–64.

STEVENSON R. D., 1985, <sup>2</sup> Body Size and Limits to the Daily Range of Body Temperature in Terrestrial Ectotherms ♀, *The American Naturalist*, vol. 125, n. 1 : 102–117.

STRAHL B. D. et ALLIS C. D., 2000, <sup>2</sup> The language of covalent histone modifications ♀, *Nature*, vol. 403, n. 6765 : 41–45.

STRAUB Tobias, G LLAN Gregor D., MA ER Verena K., et BECKER Peter B., 2005,   The Drosophila MSL complex activates the transcription of target genes  , *Genes & Development*, vol. 19, n, 19 : 2284  2288.

STRAUB Tobias, GR MAUD Charlotte, G LLAN Gregor D., M TTERWEGER Angelika, et BECKER Peter B., 2008,   The Chromosomal High-Affinity Binding Sites for the Drosophila Dosage Compensation Complex  , *PLoS Genetics*, vol. 4, n, 12 : e1000302.

STRUHL K., KADOSH D., KEAVENEY M., KURAS L., et MOQTADER  Z., 1998,   Activation and Repression Mechanisms in Yeast  , *Cold Spring Harbor Symposia on Quantitative Biology*, vol. 63 : 413  422.

SUG YAMA Tomoyasu, CAM Hugh P., SUG YAMA Rie, NOMA Ken-ichi, ZOFALL Martin, KOBAYASH  Ryuji, et GREWAL Shiv  . S., 2007,   SHREC, an effector complex for heterochromatic transcriptional silencing  , *Cell*, vol. 128, n, 3 : 491  504.

SUZUK  Miho M. et B RD Adrian, 2008,   DNA methylation landscapes: provocative insights from epigenomics  , *Nature Reviews. Genetics*, vol. 9, n, 6 : 465  476.

SYED Sajad Hussain, GOUTTE-GATTAT Damien, BECKER Nils, MEYER Sam, SHUKLA Manu Shubhdarshan, HAYES Jeffrey J., EVERAERS Ralf, ANGELOV Dimitar, BEDNAR Jan, et D M TROV Stefan, 2010,   Single-base resolution mapping of H1-nucleosome interactions and 3D organization of the nucleosome  , *Proceedings of the National Academy of Sciences of the United States of America*, vol. 107, n, 21 : 9620  9625.

SYLVESTER Edward S., 1969,   Evidence of transovarial passage of the sowthistle yellow vein virus in the aphid *Hyperomyzus lactucae*  , *Virology*, vol. 38, n, 3 : 440  446.

TAGU Denis, COLBOURNE John K., et N GRE Nicolas, 2014a,   Genomic data integration for ecological and evolutionary traits in non-model organisms  , *BMC Genomics*, vol. 15 : 490.

TAGU Denis, LE TR ONNA RE Ga  l, TANGUY Sylvie, GAUTH ER Jean-Pierre, et HUYNH Jean-Ren , 2014b,   EMS Mutagenesis in the Pea Aphid *Acyrtosiphon pisum*  , *G3: Genes/Genomes/Genetics*, vol. 4, n, 4 : 657  667.

TAGU Denis, SABATER-MU  OZ Beatriz, et S MON Jean-Christophe, 2005,   Deciphering reproductive polyphenism in aphids  , *Invertebrate Reproduction & Development*, vol. 48, n, 1  3 : 71  80.

TA WO Oluwatosin, W LSON Gareth A., MORR S Tiffany, SE SENBERGER Stefanie, RE K Wolf, PEARCE Daniel, BECK Stephan, et BUTCHER Lee M., 2012,   Methylome analysis using MeD P-seq with low DNA concentrations  , *Nature Protocols*, vol. 7, n, 4 : 617  636.

TAKAYAMA Sachiko, DHAHB  Joseph, ROBERTS Adam, MAO Guanxiong, HEO Seok-Jin, PACTHER Lior, MART N David  .K., et BOFFELL  Dario, 2014,   Genome methylation in *D. melanogaster* is found at specific short motifs and is independent of DNMT2 activity  , *Genome Research*, vol. 24, n, 5 : 821  830.

TALBERT Paul B. et HEN KOFF Steven, 2010,   Histone variants--ancient wrap artists of the epigenome  , *Nature Reviews. Molecular Cell Biology*, vol. 11, n, 4 : 264  275.

TERASH MA J. et BOWNES M., 2006,   E75A and E75B have opposite effects on the apoptosis/development choice of the Drosophila egg chamber.  , *Cell Death and Differentiation*, vol. 13, n, 3 : 454  464.

THOMA F., KOLLER T., et KLUG A., 1979,   Involvement of histone H1 in the organization of the nucleosome and of the salt-dependent superstructures of chromatin  , *The Journal of Cell Biology*, vol. 83, n, 2 Pt 1 : 403  427.

TAN Di, SUN Sha, et LEE Jeannie T., 2010, <sup>2</sup> The long noncoding RNA, Jpx, is a molecular switch for X chromosome inactivation *Cell*, vol. 143, n. 3 : 390-403.

TILL Susanne, LEJEUNE Erwan, THERMANN Rolf, BORTFELD Miriam, HOTHORN Michael, ENDERLE Daniel, HEINRICH Constanze, HENTZE Matthias W., et LADURNER Andreas G., 2007, <sup>2</sup> A conserved motif in Argonaute-interacting proteins mediates functional interactions through the Argonaute P<sup>W</sup> domain *Nature Structural & Molecular Biology*, vol. 14, n. 10 : 897-903.

TJALLING W. F. et ESCH Th. Hogen, 1993, <sup>2</sup> Fine structure of aphid stylet routes in plant tissues in correlation with EPG signals *Physiological Entomology*, vol. 18, n. 3 : 317-328.

TROPBERGER Philipp, POTT Sebastian, KELLER Claudia, KAMENARZ-GDULA Kinga, CARON Matthieu, RICHTER Florian, LI Guohong, MÜLLER Gerhard, LIU Edison T., BILHLER Marc, MARGUERON Raphael, et SCHNEIDER Robert, 2013, <sup>2</sup> Regulation of Transcription through Acetylation of H3K122 on the Lateral Surface of the Histone Octamer *Cell*, vol. 152, n. 4 : 859-872.

TROPBERGER Philipp et SCHNEIDER Robert, 2013, <sup>2</sup> Scratching the (lateral) surface of chromatin regulation by histone modifications *Nature Structural & Molecular Biology*, vol. 20, n. 6 : 657-661.

TSAI Chen-Hsun, LIAO Rick, CHOU Brendan, PALUMBO Michael, et CONTRERAS Lydia M., 2015, <sup>2</sup> Genome-Wide Analyses in Bacteria Show Small-RNA Enrichment for Long and Conserved Intergenic Regions *Journal of Bacteriology*, vol. 197, n. 1 : 40-50.

TSOMPANA Maria et BUCK Michael J., 2014, <sup>2</sup> Chromatin accessibility: a window into the genome *Epigenetics & Chromatin*, vol. 7 : 33.

TSUCHIDA Tsutomu, KOGA Ryuichi, FUJWARA Akiko, et FUKATSU Takema, 2014, <sup>2</sup> Phenotypic Effect of *Candidatus Rickettsiella viridis*, a Facultative Symbiont of the Pea Aphid (*Acyrtosiphon pisum*), and its Interaction with a Coexisting Symbiont *Applied and Environmental Microbiology*, vol. 80, n. 2 : 525-533.

TSUCHIDA Tsutomu, KOGA Ryuichi, et FUKATSU Takema, 2004, <sup>2</sup> Host Plant Specialization Governed by Facultative Symbiont *Science*, vol. 303, n. 5666 : 1989-1989.

TSUCHIDA Tsutomu, KOGA Ryuichi, HORIKAWA Mitsuyo, TSUNODA Tetsuto, MAOKA Takashi, MATSUMOTO Shogo, SIMON Jean-Christophe, et FUKATSU Takema, 2010, <sup>2</sup> Symbiotic Bacterium Modifies Aphid Body Color *Science*, vol. 330, n. 6007 : 1102-1104.

TURNER Bryan, 2001, <sup>2</sup> ChIP with Native Chromatin: Advantages and Problems Relative to Methods Using Cross-Linked Material *in Mapping Protein/DNA Interactions by Cross-Linking*, Paris, Institut national de la santé et de la recherche médicale.

ULVELING Damien, FRANCASTEL Claire, et HUBØ Florent, 2011, <sup>2</sup> When one is better than two: RNA with dual functions *Biochimie*, vol. 93, n. 4 : 633-644.

UVAROV B. P, ANTI-LOCUST RESEARCH CENTRE (GREAT BRITAIN), et CENTRE FOR OVERSEAS PEST RESEARCH (GREAT BRITAIN), 1966, *Grasshoppers and locusts: a handbook of general acridology*, Cambridge, Published for the Anti-Locust Research Centre [by Cambridge] University Press.

UZEST Marilyne, GARGAN Daniel, DOMBROVSKY Aviv, CAZEVEILLE Chantal, COT Didier, et BLANC Stéphane, 2010, <sup>2</sup> The microstyle: a newly described anatomical structure in aphid stylets *Arthropod Structure & Development*, vol. 39, n. 4 : 221-229.

VAN BORTLE Kevin, NICHOLS Michael H., LI Li, ONG Chin-Tong, TAKENAKA Naomi, QIN Zhaohui S., et CORCES Victor G., 2014, Insulator function and topological domain border strength scale with architectural protein occupancy *Genome Biology*, vol. 15, n. 6 : R82.

VAZ NUNES M et HARDY E, 2000, The effect of temperature on the photoperiodic clock and counter-clock of a Scottish clone of the vetch aphid, *Megoura viciae* *Journal of Insect Physiology*, vol. 46, n. 5 : 727-733.

VELLICHAMMAL Neetha Nanoth, GUPTA Purba, HALL Tannice A., et BRINSON Jennifer A., 2017, Ecdysone signaling underlies the pea aphid transgenerational wing polyphenism *Proceedings of the National Academy of Sciences*, vol. 114, n. 6 : 1419-1423.

VENKATESH Swaminathan et WORKMAN Jerry L., 2015, Histone exchange, chromatin structure and the regulation of transcription *Nature Reviews Molecular Cell Biology*, vol. 16, n. 3 : 178-189.

VERDEL Andr , JIA Songtao, GERBER Scott, SUGIYAMA Tomoyasu, GYGIS Steven, GREWAL Shiv S., et MOAZED Danesh, 2004, RNAi-mediated targeting of heterochromatin by the RITS complex *Science (New York, N.Y.)*, vol. 303, n. 5658 : 672-676.

VETTESE-DADEY M, GRANT P A, HEBBES T R, CRANE-ROBINSON C, ALLS C D, et WORKMAN J L, 1996, Acetylation of histone H4 plays a primary role in enhancing transcription factor binding to nucleosomal DNA in vitro. *The EMBO Journal*, vol. 15, n. 10 : 2508-2518.

VILUDA-MARTOS M., RUZ-NAVAJAS Y., FERNANDEZ-LHPEZ J., et P REZ- LVAREZ J.a., 2008, Functional Properties of Honey, Propolis, and Royal Jelly *Journal of Food Science*, vol. 73, n. 9 : R117-124.

VOGT Franka, REUTER Michael, KASARUHO Anisa, SCHULZ Eike C., PILLAI Ramesh S., et BARABAS Orsolya, 2012, Crystal structure of the primary piRNA biogenesis factor Zucchini reveals similarity to the bacterial PLD endonuclease Nuc *RNA*, vol. 18, n. 12 : 2128-2134.

VOLNY Veronica P. et GORDON Deborah M., 2002, Genetic basis for queen-worker dimorphism in a social insect *Proceedings of the National Academy of Sciences*, vol. 99, n. 9 : 6108-6111.

WADDINGTON C. H., 1942, Canalization of Development and the Inheritance of Acquired Characters *Nature*, vol. 150 : 563-565.

WADDINGTON C. H., 2012, The epigenotype. 1942 *International Journal of Epidemiology*, vol. 41, n. 1 : 10-13.

WAKAI Hironori, NAKAMURA Masahiro, YAMAUCHI Toshimasa, WAKABAYASHI Ken-ichi, YU Jing, HIRASE-YOTSUYA Lisa, TAKE Kazumi, SUN Wei, IWABU Masato, OKADA IWABU Miki, FUJITA Takanori, AOYAMA Tomohisa, TSUTSUMI Shuichi, UEKI Kohjiro, KODAMA Tatsuhiko, SAKAI Juro, ABURATAN HIROYUKI, et KADOWAKI Takashi, 2011, Global Mapping of Cell Type-Specific Open Chromatin by FAIRE-seq Reveals the Regulatory Role of the NF-  Family in Adipocyte Differentiation *PLOS Genetics*, vol. 7, n. 10 : e1002311.

WALSH T. K., BRINSON J. A., ROBERTSON H. M., GORDON K., JAUBERT-POSSAMA S., TAGU D., et EDWARDS O. R., 2010, A functional DNA methylation system in the pea aphid, *Acyrtosiphon pisum* *Insect Molecular Biology*, vol. 19 : 215-228.

WANG Xianhui et KANG Le, 2014, Molecular Mechanisms of Phase Change in Locusts *Annual Review of Entomology*, vol. 59, n. 1 : 225-244.

WANG Yanli, JURANEK Stefan, L' Haitao, SHENG Gang, WARDLE Greg S., TUSCHL Thomas, et PATEL Dinshaw J., 2009, <sup>2</sup> Nucleation, propagation and cleavage of target RNAs in Ago silencing complexes *Þ*, *Nature*, vol. 461, n, 7265 : 754 1761.

WANG Ying, JORDA Mireia, JONES Peter L., MALESZKA Ryszard, L'NG Xu, ROBERTSON Hugh M., M'ZZEN Craig A., PE'NADO Miguel A., et ROB'NSON Gene E., 2006, <sup>2</sup> Functional CpG Methylation System in a Social 'nsect *Þ*, *Science*, vol. 314, n, 5799 : 645 147.

WANG Yuqiu, FAN Xiuduo, L'N Fang, HE Guangming, TERZAGH' William, ZHU Danmeng, et DENG Xing Wang, 2014, <sup>2</sup> Arabidopsis noncoding RNA mediates control of photomorphogenesis by red light *Þ*, *Proceedings of the National Academy of Sciences of the United States of America*, vol. 111, n, 28 : 10359 10364.

WE'NTRAUB H. et GROUD'NE M., 1976, <sup>2</sup> Chromosomal subunits in active genes have an altered conformation *Þ*, *Science (New York, N.Y.)*, vol. 193, n, 4256 : 848 156.

WELCH Mat et L'STER Ryan, 2014, <sup>2</sup> Epigenomics and the control of fate, form and function in social insects *Þ*, *Current Opinion in 'nsect Science*, vol. 1 : 31 18.

WENGER Jacob A., CASSONE Bryan J., LEGEA' Fabrice, JOHNSTON J. Spencer, BANSAL Raman, YATES Ashley D., COATES Brad S., PAV'NATO Vitor A. C., et M'CHEL Andy, 2017, <sup>2</sup> Whole genome sequence of the soybean aphid, *Aphis glycines* *Þ*, *'nsect Biochemistry and Molecular Biology*.

W'DOM J., 1986, <sup>2</sup> Physicochemical studies of the folding of the 100 A nucleosome filament into the 300 A filament. Cation dependence *Þ*, *Journal of Molecular Biology*, vol. 190, n, 3 : 411 124.

W'GHTMAN B., HA '., et RUVKUN G., 1993, <sup>2</sup> Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in *C. elegans* *Þ*, *Cell*, vol. 75, n, 5 : 855 162.

W'LK'NSON Gerald S., JOHNS Philip M., METHENY Jackie D., et BAKER Richard H., 2013, <sup>2</sup> Sex-Biased Gene Expression during Head Development in a Sexually Dimorphic Stalk-Eyed Fly *Þ*, *PLOS ONE*, vol. 8, n, 3 : e59826.

W'LL Torsten, TJALL'NG' W. Fred, THIJNNESSEN Alexandra, et BEL Aart J. E. van, 2007, <sup>2</sup> Molecular sabotage of plant defense by aphid saliva *Þ*, *Proceedings of the National Academy of Sciences*, vol. 104, n, 25 : 10536 10541.

W'LSON E. O., 1971, *The insect societies.*, Cambridge, Massachusetts, USA, Harvard University Press[Distributed by Oxford University Press.

WOLMAN M. et GRECO J., 1952, <sup>2</sup> The Effect of Formaldehyde on Tissue Lipids and on Histochemical Reactions for Carbonyl Groups *Þ*, *Stain Technology*, vol. 27, n, 6 : 317 124.

WU Hao et ZHANG Yi, 2014, <sup>2</sup> Reversing DNA Methylation: Mechanisms, Genomics, and Biological Functions *Þ*, *Cell*, vol. 156, n, 1 : 45 18.

WU Yu, PARTHASARATHY R., BA' Hua, et PALL' Subba R., 2006, <sup>2</sup> Mechanisms of midgut remodeling: Juvenile hormone analog methoprene blocks midgut metamorphosis by modulating ecdysone action *Þ*, *Mechanisms of Development*, vol. 123, n, 7 : 530 147.

WUCHER Valentin, LEGEA' Fabrice, HODAN Beno ť, R'ZK Guillaume, LAGOUTTE Lă titia, LEEB Tosso, JAGANNATHAN Vidhya, CAD'EU Edouard, DAV'D Audrey, LOH' Hannes, C'RERA Susanna, FREDHOLM Merete, BOTHEREL Nadine, LEEGWATER Peter A. J., LE BøGUEC Cŭline, F'ETEN Hille, JOHNSON Jeremy, ALFIJLD' Jessica, ANDRø Catherine, L'NDBLAD-TOH Kerstin, H'TTE Christophe, et DERR'EN Thomas, 2017, <sup>2</sup> FEELnc: a tool for

long non-coding RNA annotation and its application to the dog transcriptome  , *Nucleic Acids Research*, vol. 45, n. 8 : e57.

WYATT G. R., 1951,   The purine and pyrimidine composition of deoxyribose nucleic acids  , *Biochemical Journal*, vol. 48, n. 5 : 584-590.

YANG Chuhu, BOLOTIN Eugene, JIANG Tao, SLADEK Frances M, et MARTINEZ Ernest, 2007,   Prevalence of the initiator over the TATA box in human and yeast genes and identification of DNA motifs enriched in human TATA-less core promoters  , *Gene*, vol. 389, n. 1 : 52-55.

YOON Je-Hyun, KIM Jiyong, et GOROSPE Myriam, 2015,   Long noncoding RNA turnover  , *Biochimie*, vol. 117 : 15-21.

ZAMORE Phillip D, TUSCHL Thomas, SHARP Phillip A, et BARTEL David P, 2000,   RNAi: Double-Stranded RNA Directs the ATP-Dependent Cleavage of mRNA at 21 to 23 Nucleotide Intervals  , *Cell*, vol. 101, n. 1 : 25-33.

ZARATYNU Mikel, CASTEL Stephane E., RIVINE Danielle V., KLOC Anna, REN Jie, LI Fei, DE CASTRO Elisa, MARIN Laura, CHANG An-Yun, GOTO Derek, CANDE W. Zacheus, ANTEQUERA Francisco, ARCANGOLI Benoit, et MARTENSSEN Robert A., 2011,   RNAi promotes heterochromatic silencing through replication-coupled release of RNA Pol III  , *Nature*, vol. 479, n. 7371 : 135-138.

ZARET Kenneth S. et CARROLL Jason S., 2011,   Pioneer transcription factors: establishing competence for gene expression  , *Genes & Development*, vol. 25, n. 21 : 2227-2241.

ZENG Hongkui, QIAN Zuwei, MYERS Michael P., et ROSBASH Michael, 1996,   A light-entrainment mechanism for the Drosophila circadian clock  , *Nature*, vol. 380, n. 6570 : 129-135.

ZERA A. J. et DENNO R. F., 1997,   Physiology and ecology of dispersal polymorphism in insects  , *Annual Review of Entomology*, vol. 42 : 207-230.

ZHANG Ke, MOSCH Kerstin, FISCHLE Wolfgang, et GREWAL Shiv H. S., 2008a,   Roles of the Clr4 methyltransferase complex in nucleation, spreading and maintenance of heterochromatin  , *Nature Structural & Molecular Biology*, vol. 15, n. 4 : 381-388.

ZHANG Yong, LIU Tao, MEYER Clifford A., ECKHOUTE Jerome, JOHNSON David S., BERNSTEIN Bradley E., NUSBAUM Chad, MYERS Richard M., BROWN Myles, LI Wei, et LIU X. Shirley, 2008b,   Model-based analysis of ChIP-Seq (MACS)  , *Genome Biology*, vol. 9, n. 9 : R137.

ZHANG Yu-Chan, LIAO Jian-You, LI Ze-Yuan, YU Yang, ZHANG Jin-Ping, LI Quan-Feng, QU Liang-Hu, SHU Wen-Sheng, et CHEN Yue-Qin, 2014,   Genome-wide screening and functional analysis identify a large number of long noncoding RNAs involved in the sexual reproduction of rice  , *Genome Biology*, vol. 15 : 512.

ZIEGLER R., 2013,   Lipid synthesis by ovaries and fat body of Aedes aegypti (Diptera: Culicidae)  , *EJE*, vol. 94, n. 3 : 385-391.





Les pucerons sont des hémiptères ravageurs des cultures agronomiques particulièrement adaptés à leur environnement. *Acyrtosiphon pisum* (le puceron du pois) présente un cycle de vie basé sur l'alternance d'une reproduction sexuée ou asexuée en réponse à la photopériode. Ils présentent ainsi un polyphénisme de reproduction aboutissant à la formation de trois phénotypes distincts : femelles asexuées, femelles sexuées, et mâles. Ces derniers étant obtenus par élimination d'un chromosome X, *A. pisum* est une espèce hétérogamétique mâle présentant un système chromosomique XX chez les femelles et XO chez les mâles. Le déséquilibre du nombre de chromosome X entre mâles et femelles engendré par cette hétérogamétié nécessite chez certains organismes d'être corrigé par des mécanismes de compensation de dose. Les polyphénismes et compensation de dose impliquent chez d'autres organismes des régulations transcriptionnelles notamment régulées par l'accessibilité de la chromatine. Ma thèse vise ainsi à étudier le polyphénisme de reproduction et la compensation de dose des pucerons sous l'angle d'analyses bio-informatiques de données d'expression des gènes (RNA-seq) et d'accessibilité de la chromatine (FAIRE-seq) dans le but de caractériser l'impact des mécanismes épigénétiques dans ces deux processus biologiques fondamentaux du cycle de vie des pucerons. Les résultats développés dans ma thèse ont permis de montrer d'une part la présence d'une compensation de dose chez le puceron du pois au niveau transcriptomique, supportée par une accessibilité accrue de la chromatine de l'unique X des mâles dans les cellules somatiques. D'autre part, des sites spécifiques d'ouverture de la chromatine entre embryons sexués ou asexués semblent participer à la définition de leur mode de reproduction en modulant l'expression de certains gènes et en permettant la fixation de facteurs de transcriptions. Leur analyse montre l'implication de l'ecdysone comme nouvelle voie hormonale déclenchant putativement la reproduction sexuée en réponse à une photopériode courte. Ces travaux suggèrent l'importance des régulations chromatinienne dans l'adaptation des pucerons à leur environnement et dans leur cycle de vie.

Mots clés : Puceron du pois, polyphénisme de reproduction, compensation de dose, accessibilité de la chromatine, transcription, génomique.

Aphids are hemipterous crops pests that are particularly adapted to their environment. *Acyrtosiphon pisum* (pea aphid) displays a life cycle based on the alternation of sexual or asexual reproduction in response to photoperiod. They thus exhibit a reproductive polyphenism resulting in the formation of three distinct phenotypes: asexual females, sexual females, and males. The latter being obtained by elimination of an X chromosome, *A. pisum* is a male heterogametic species with a XX chromosomal system in females and XO in males. The X chromosome number between males and females caused by this heterogamy requires in some organisms to be corrected by dosage compensation mechanisms. Polyphenisms and dosage compensation both involve in other organisms transcriptional regulations that are notably regulated by the chromatin accessibility regulations. My thesis aims to study the reproductive polyphenism and dosage compensation in aphids in the context of bioinformatic analyzes of gene expression data (RNA-seq) and chromatin accessibility (FAIRE-seq) in order to characterize the impact of epigenetic mechanisms in these two fundamental biological processes of the aphid life cycle. The results developed in my thesis have shown, on one hand, the presence of dose compensation in pea aphid at the transcriptomic level, which is supported by increased chromatin accessibility of the males' single X in somatic cells. On the other hand, specific sites of chromatin opening between sexual and asexual embryos seem to participate in the definition of their reproduction mode by modulating the expression of certain genes and by allowing the fixation of transcription factors. Their analysis shows the involvement of ecdysone as a new hormonal pathway that may trigger sexual reproduction in response to a short photoperiod. This work suggests the importance of chromatin regulation in aphids' environmental adaptation and in their life cycle.

Key words: Pea aphid, reproductive polyphenism, dosage compensation, chromatin accessibility, transcription, genomics.

## RÉSUMÉ

### Régulations chromatinienne et transcriptionnelles impliquées dans le cycle de vie du puceron du pois

Les pucerons sont des hémiptères ravageurs des cultures agronomiques particulièrement adaptés à leur environnement. *Acyrtosiphon pisum* (le puceron du pois) présente un cycle de vie basé sur l'alternance d'une reproduction sexuée ou asexuée en réponse à la photopériode. Ils présentent ainsi un polyphénisme de reproduction aboutissant à la formation de trois phénotypes distincts : femelles asexuées, femelles sexuées, et mâles. Ces derniers étant obtenus par élimination d'un chromosome X, *A. pisum* est une espèce hétérogamétique mâle présentant un système chromosomique XX chez les femelles et XO chez les mâles. Le déséquilibre du nombre de chromosome X entre mâles et femelles engendré par cette hétérogaméte nécessite chez certains organismes d'être corrigé par des mécanismes de compensation de dose. Les polyphénismes et compensation de dose impliquent chez d'autres organismes des régulations transcriptionnelles notamment régulées par l'accessibilité de la chromatine. Ma thèse vise ainsi à étudier le polyphénisme de reproduction et la compensation de dose des pucerons sous l'angle d'analyses bioinformatiques de données d'expression des gènes (RNA-seq) et d'accessibilité de la chromatine (FAIRE-seq) dans le but de caractériser l'impact des mécanismes épigénétiques dans ces deux processus biologiques fondamentaux du cycle de vie des pucerons. Les résultats développés dans ma thèse ont permis de montrer d'une part la présence d'une compensation de dose chez le puceron du pois au niveau transcriptomique, supportée par une accessibilité accrue de la chromatine de l'unique X des mâles dans les cellules somatiques. D'autre part, des sites spécifiques d'ouverture de la chromatine entre embryons sexués ou asexués semblent participer à la définition de leur mode de reproduction en modulant l'expression de certains gènes et en permettant la fixation de facteurs de transcriptions. Leur analyse montre l'implication de l'ecdysone comme nouvelle voie hormonale déclenchant putativement la reproduction sexuée en réponse à une photopériode courte. Ces travaux suggèrent l'importance des régulations chromatinienne dans l'adaptation des pucerons à leur environnement et dans leur cycle de vie.

**Mots-clés** : puceron du pois, polyphénisme de reproduction, compensation de dose, accessibilité de la chromatine, transcription, génomique.

## ABSTRACT

### Chromatin and transcriptional regulations involved in the pea aphid's life cycle

Aphids are hemipterous crops pests that are particularly adapted to their environment. *Acyrtosiphon pisum* (pea aphid) displays a life cycle based on the alternation of sexual or asexual reproduction in response to photoperiod. They thus exhibit a reproductive polyphenism resulting in the formation of three distinct phenotypes: asexual females, sexual females, and males. The latter being obtained by elimination of an X chromosome, *A. pisum* is a male heterogametic species with a XX chromosomal system in females and XO in males. The X chromosome number between males and females caused by this heterogamy requires in some organisms to be corrected by dosage compensation mechanisms. Polyphenisms and dosage compensation both involve in other organisms transcriptional regulations that are notably regulated by the chromatin accessibility regulations. My thesis aims to study the reproductive polyphenism and dosage compensation in aphids in the context of bioinformatic analyzes of gene expression data (RNA-seq) and chromatin accessibility (FAIRE-seq) in order to characterize the impact of epigenetic mechanisms in these two fundamental biological processes of the aphid life cycle. The results developed in my thesis have shown, on one hand, the presence of dose compensation in pea aphid at the transcriptomic level, which is supported by increased chromatin accessibility of the males' single X in somatic cells. On the other hand, specific sites of chromatin opening between sexual and asexual embryos seem to participate in the definition of their reproduction mode by modulating the expression of certain genes and by allowing the fixation of transcription factors. Their analysis shows the involvement of ecdysone as a new hormonal pathway that may trigger sexual reproduction in response to a short photoperiod. This work suggests the importance of chromatin regulation in aphids' environmental adaptation and in their life cycle.

**Keywords**: Pea aphid, reproductive polyphenism, dosage compensation, chromatin accessibility, transcription, genomics.



AGROCAMPUS OUEST • Institut national supérieur des sciences agronomiques, agroalimentaires, horticoles et du paysage  
65 rue de Saint-Brieuc - CS84215 - F-35042 Rennes Cedex  
Tél. : 02 23 48 50 00  
[www.agrocampus-ouest.fr](http://www.agrocampus-ouest.fr)

