



Université
de Toulouse

THÈSE

En vue de l'obtention du
DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par *l'Université Toulouse III - Paul Sabatier*
Discipline ou spécialité : *Génétique Moléculaire*

Présentée et soutenue par Baanannou Aissette
Le 17/12/2012

Titre : *Etude de la régulation transcriptionnelle
des gènes bab au cours du développement des
appendices chez la drosophile*

JURY

Pascale Belenguer (Pr) Présidente
François Karch (Pr) Rapporteur
James Castelli-Gair Hombría (Dr) Rapporteur
Nathalie Dostatni (Dr) Rapporteur
Jean-Louis Couderc (Dr) Examineur
Henri-Marc Bourbon (Dr)
Muriel Boube (Dr)

Ecole doctorale : *Biologie, Santé, Biotechnologies*
Unité de recherche : *Centre de Biologie du Développement*
Directeur(s) de Thèse : *Henri-Marc Bourbon et Muriel Boube (co-directeurs)*

Remerciements

Tous ceux qui l'ont vécu le savent : la thèse est un moment particulier de la vie, une expérience unique qui est à la fois difficile, enrichissante et passionnante. C'est aussi une fenêtre sur le monde fantastique de la recherche et une opportunité exceptionnelle pour s'apercevoir de l'extrême beauté et la complexité de la vie. Je tiens donc à travers cette espace à remercier tous ceux qui m'ont permis de près ou de loin de vivre cette expérience qui me marquera à jamais.

En tout premier lieu, je tiens à exprimer ma profonde gratitude au docteur Henri-Marc Bourbon qui m'a accueilli dans son laboratoire en 2008 pour réaliser cette thèse. Merci Henri pour ton soutien, ta confiance, ta compréhension. Merci pour avoir partagé avec moi tes connaissances et ton expérience.

Mes remerciements vont également à Muriel Boube qui m'a co-encadré au cours de cette thèse et qui m'a apporté beaucoup. Merci pour toutes les pauses café et toute la joie que tu nous apporte chaque matin.

Je tiens ensuite à remercier sincèrement les membres du jury pour l'honneur qu'ils me font en acceptant de juger ce travail.

Je remercie très sincèrement David Cribbs pour sa disponibilité ces précieux conseils et les discussions très intéressantes que j'ai eues avec lui.

Je remercie tout particulièrement Luis Humberto Mojica Vazquez pour ses encouragements, son aide et surtout pour sa contribution au projet bab.

J'adresse aussi mes remerciements les plus chaleureux à tous les membres de l'équipe «Mediateur» qui m'ont permis d'évoluer dans une ambiance très sympathique. Je pense notamment à Christian, Sandra et Clément. Je souhaite sincèrement bon courage aux thésards en espérant qu'ils vivront une expérience aussi enrichissante que la mienne.

Un grand merci à l'ensemble du personnel du CBD et plus particulièrement aux membres du service de laverie, milieu et informatique ainsi que ceux de la plate-forme d'imagerie.

Parmi toutes les personnes exceptionnelles que j'ai rencontrées au CBD, je tiens à remercier très chaleureusement Daniel, Fred et Julien. Je remercie tout particulièrement mon tuteur Serge qui m'a bien aidé à la construction de cette histoire.

Je pense notamment à tous les copains du Basket Julien Nico, Ahmad,...

Je n'oublierai sûrement pas de remercier Bart Deplancke pour son accueil dans son laboratoire lors des expériences du crible simple-hybride, pour sa disponibilité et sa gentillesse. Merci également aux membres de son équipe : Hens, JD, Carine et...!

Cette thèse n'aurait jamais pu avoir lieu sans le soutien et le financement du CNRS, de la région Midi-Pyrénées de l'université Paul Sabatier et de l'ARC. Un grand merci

Naturellement, j'adresse d'immenses remerciements à ma famille, ma belle-famille pour leur compréhension, leur patience et leur soutien.

Résumé

Afin de mieux comprendre (i) les mécanismes moléculaires régissant le contrôle de l'expression des gènes impliqués dans le développement et (ii) les réseaux génétiques s'établissant en réponse à des processus de signalisation cellulaire, de plus en plus d'études s'intéressent aux éléments régulateurs en *cis* des gènes. C'est dans ce cadre que s'inscrit cette thèse intitulée "Contrôle de l'expression transcriptionnelle des gènes *bric-à-brac* lors de la morphogénèse des appendices chez le modèle drosophile". Le locus *bric-à-brac* (*bab*) chez *Drosophila melanogaster* se compose de deux gènes paralogues, *bab1* et *bab2*, codant pour des protéines nucléaires à domaine BTB, et qui (i) sont exprimés dans les disques imaginaux de pattes et d'antenne selon des patrons très similaires et (ii) sont requis pour la segmentation normale des parties distales des pattes et des antennes, avec un rôle prédominant pour *bab2*.

Au cours de ma thèse, je me suis intéressé aux mécanismes moléculaires contrôlant l'expression de *bab2* dans les disques imaginaux de pattes et d'antenne, selon 4 ou 2 anneaux concentriques, respectivement. A partir (i) de constructions transgéniques exprimant un gène rapporteur GFP et insérées dans un site spécifique et (ii) d'un test de sauvetage phénotypique d'un mutant dépourvu de fonction *bab*, par un fragment d'ADN génomique (BAC) sauvage ou muté par recombineering, j'ai montré que l'expression de *bab2* dans les disques de pattes et d'antenne dépend d'un module *cis*-régulateur (CRM) unique de 0,56 kilobases, que nous avons appelé *LAE*, pour "*Leg and Antennal Enhancer*". Par ailleurs, j'ai pu montrer que le *LAE* de *D. melanogaster* est interchangeable avec la région équivalente de *D. virilis*, révélant ainsi une conservation fonctionnelle chez ces deux espèces de drosophilidae séparées de 60 millions d'années. Moyennant des délétions et des mutations dirigées (en "Linker scanning") du *LAE*, j'ai identifié des motifs d'ADN agissant positivement ou négativement soit dans tout le domaine d'expression de *bab2* ou seulement dans des sous-domaines (proximaux ou distaux), et ceci de manière similaire dans les disques de pattes et d'antenne. Enfin, par des expériences de perte et de gain de fonction *in vivo*, ainsi que des analyses biochimiques *in vitro*, j'ai montré (i) que la protéine à homéodomaine Distal-less (Dll) s'associe à une région conservée de l'enhancer *LAE* pour activer directement *bab2* au niveau de tous les anneaux, (ii) tandis que la protéine à doigt de zinc Rotund lie une région non-strictement conservée chez les Drosophilidae, seulement nécessaire pour une activation maximale dans les anneaux proximaux.

Abstract

To better understand (i) the molecular mechanisms controlling the transcriptional expression of developmental genes and (ii) to establish genetic networks responding to cell signaling pathways, more and more studies are focused on the identification and characterization of *cis*-regulatory DNA elements. In this area, my thesis aimed to study “the transcriptional regulation of *bric-a-brac* genes during appendages morphogenesis in the *Drosophila* model”. The *Drosophila melanogaster bric-a-brac (bab)* locus consists of two paralogous BTB protein encoding genes, *bab1* and *bab2*, (i) expressed similarly in leg and antennal imaginal discs (ii) and required for normal segmentation of distal legs and antenna, with a prominent role for *bab2*.

During my thesis, I studied the molecular mechanisms controlling *bab2* expression in leg and antennal imaginal discs, along four or two concentric rings of epithelial cells, respectively. Using a site-targeted GFP reporter assay and phenotypic rescue of a null *bab* mutant with a wild-type or mutated (by recombineering) BAC genomic DNA fragment, I showed that restricted *bab2* expression in the distal primordia of legs and antenna is strictly dependent on a single 0.56 kb *cis*-regulatory module (CRM), termed *LAE*, for “Leg and Antennal Enhancer”. Significantly, I showed that the *D. virilis LAE* is fully functional in *D. melanogaster*, establishing functional conservation among two Drosophilidae separated by 60 million years. Through *LAE* deletions and linker scanning analyses, I identified DNA motifs acting either positively or negatively in all or a subset (proximal versus distal) of the antennal and leg tissues. Finally, loss and gain of function *in vivo* experiments, as well as *in vitro* electrophoretic mobility shift assays, revealed that the Distal-less homeodomain protein binds a conserved region to activate directly *bab2* in all rings, whereas the Rotund zinc finger protein, which associates to a non-conserved region, is specifically required for direct full activation in the proximal rings.

ABBREVIATIONS

a1-5 : segments antennaires
APO : après la ponte de l'œuf
A/P : antéropostérieur
Abd-A : *Abdominal-A*
Abd-B : *Abdominal-B*
ADN : Acide Désoxyribonucléique
al : *aristales*
Antp : *Antennapedia*
ap : *apterous*
AP-2 : Activator protein-2
ar : arista
ARN : Acide Ribonucléique
ARNm : ARN messagers
ARN pol I-III : ARN polymérase
ARNr : ARN ribosomique
ARNt : ARN de transfert
ato : *atonal*
bab : *bric à brac*
bib : *big brain*
brk : *brinker*
BTB : Bric-a-brac, Tramtrack, Broad-complex
BX-C : Complexe Bithorax
CBP : CREB-binding protein
Cbf1 : centromere-binding factor
ci : *cubitus interruptus*
CRM : *Cis-Regulatory Module*
ct : *cut*
CTCF : CCCTC-binding Factor
CTD : Carboxy-Terminal Domain (domaine carboxy-terminal)
cx : coxa
D/V : dorso-ventral
dac : *dachshund*
dan : *distal antenna*
danr : *distal antenna related*
dl : *delta*
Dll : *Distal-less*
DNMT : ADN méthyltransférases,
DPE : downstream promoter element
dpp : *decapentaplegic*
E (spl) : Enhancer of Split
EGFR : *Epidermal growth factor receptor*
en : *engrailed*
fng : *fringe*
esg : *escagot*
exd : *extradenticle*
Fe : fémur
FT : facteur de transcription spécifique
GR : Glucocorticoid Receptor
HAT : Histones Acétyl-Transférase
HDM : Histones Deméthylases
HDAC : Histone dé-acétylases
hh : *hedgehog*
HKMT : Histone lysine méthyltransferases
HMT : Histone Methyl Transférase

Hox : gène *homéotique*
hth : *homothorax*
Kb : *Kilobases*
L3 : troisième stade larvaire
Mad : *Mothers against dpp*
MAPK : Mitogen-Activated Protein kinase
MBD2/3 :
MED : complexe médiateur
MPT : modification post-traductionnelle
N : *Notch*
nub : nubbin
Odd : *odd-skipped*
P/D : Proximodistal
P-TEFb : Positive Transcription Elongation Factor b
pb : paire de bases
Pc : *Polycomb*
Pc-G : Polycomb Group
PCR : Polymerase Chain Reaction
PIC : Preinitiation complex (Complexe de préinitiation)
PRE : PcG Response Element
pri : polished rice
PRMT : protéine arginine méthyltransférase
ptc : *patched*
rn : *rotund*
RMT : arginine méthyltransferases
roe : *roughened eye*
rRNA : ARN ribosomique
sal : *spalt*
salr : *spalt-related*
Scr : *Sex combs reduced*
ser : *serrate*
sna : *snail*
ss : *spineless-aristapedia*
t1-5 : tarsi d'une patte
T1 : patte prothoracique
T2 : patte mésothoracique
T3 : patte métathoracique
TAD : domaine transactivateur
TAF : TBP-Associated Factor
tal : *tarsal-less /pri*
TBP : TATA-Binding Protein
TF : Transcription Factor
Ti : tibia
Tr : trochanter
TRE : TrxG Response Element
trx : *trithorax*
TrxG : Groupe Trithorax
tsh : *teashirt*
Ubx : *Ultrabithorax*
vn : *vein*
vg : *vestigial*
wg : *wingless*
y : *yellow*

Sommaire

Préambule

Introduction

1. La transcription	3
1.1. Structure de l'ARN polymérase II (PolII).....	3
1.2. Les étapes de la transcription	5
1.3. « Paused » PolII	7
1.4. Le Complexe Médiateur	9
2. Régulation de la transcription	13
2.1. Eléments régulateurs en <i>cis</i>	13
2.1.1. Enhancers.....	13
✓ Motifs de séquence et architecture des enhancers	13
✓ Interaction enhancer-promoteur	15
✓ Régulation modulaire des gènes du développement.....	17
✓ Shadow enhancer	21
✓ Intégration de l'activité de plusieurs enhancers	23
✓ Evolution des séquences <i>cis</i> -régulatrices	23
2.1.2. Les insulateurs	31
✓ Régulation de l'interaction enhancer promoteur.....	31
✓ Mise en place de domaines chromatinien	33
2.2. Eléments régulateurs en <i>trans</i>	35
2.2.1. Activation de la transcription.....	37
✓ Interaction avec la machinerie transcriptionnelle basale	37
✓ Coopérativité entre les FTs.....	39
2.2.2. Répression de la transcription	41
2.2.3. Modulation de la spécificité de séquence reconnue par les FTs par des cofacteurs.....	45
3. La chromatine	45
3.1. Régulation de la structure chromatinienne	47
3.1.1. Méthylation de l'ADN	47
3.1.2. Modifications post-traductionnelles des histones	49
✓ Méthylation	51
✓ Acétylation.....	53
✓ Phosphorylation	55
✓ Biotinylation	55
3.1.3. Variants d'histones	55
3.1.4. Remodelage de la chromatine	57
3.2. Groupes Polycomb et Trithorax	59
3.2.1. Activation par les protéines TrxG.....	61
3.2.2. Répression par les protéines PcG.....	61
4. Développement de la patte et de l'antenne	65
4.1. Structure de la patte et de l'antenne	65
4.2. Disques imaginaux	65
4.2.1. Origine embryonnaire des disques imaginaux de pattes	67
4.2.2. Origine embryonnaire du disque imaginal de l'antenne	71
4.3. Mise en place des axes dans les disques de pattes et d'antenne	71
4.3.1. Mise en place de l'axe antéro-postérieur (A/P)	73
4.3.2. Mise en place de l'axe dorso-ventral (D/V)	73
4.3.3. Mise en place de l'axe proximo-distal (P/D)	75
✓ Subdivision précoce: Hth/Dac/Dll.....	75
✓ Régulation de l'expression de <i>dac</i> et <i>Dll</i> par les voies Wg et Dpp.....	77
✓ Etablissement progressif des régions tarse et prétarse (via l'action des FTs Rn, Bar, Al, C15/Cll et dLIM1..).	79
4.4. Formation des jonctions entre les segments des pattes.....	85
4.4.1. Activation de la voie Notch	85
4.4.2. Effecteurs de la voie Notch.....	87

4.4.3. Rôle de Notch dans la croissance des pattes	89
4.4.4. Etablissement de la frontière au niveau des jonctions	89
✓ dans les 'vraies' jonctions	89
✓ dans les jonctions tarsales	91
5. Identité antenne versus patte	93
5.1. <i>spineless (ss)</i>	95
5.2. <i>distal antenna (dan)</i> et <i>distal antenna related (danr)</i>	99
5.3. <i>cut (ct)</i>	99
5.4. Les gènes <i>spalt (sal)</i> / <i>spalt-related (salr)</i>	101
5.5. <i>atonal (ato)</i>	101
6. Les gènes <i>bab</i>	101
6.1. Expression et fonctions des gènes <i>bab</i>	103
6.2. Régulation de l'expression des gènes <i>bab</i> dans la patte et l'antenne	109
6.2.1. Activateur de l'expression des gènes <i>bab</i>	109
✓ Dll	109
✓ Ss	111
6.2.2. Répresseurs de l'expression des gènes <i>bab</i>	113
✓ Dac	113
✓ Bowl	115
✓ Voie EGFR	119
✓ Sal et Hth répresseurs antenne spécifique des gènes <i>bab</i>	119

Résultats

Partie I: Article

Partiell:

1. Crible simple hybride	181
1.1. Principe.	181
1.2. Construction des souches stables	183
1.3. Test d'auto activation et sélection des souches	185
1.4. Crible simple hybride	187
1.5. Résultats du crible simple hybride	189
2. Caractérisation fonctionnelle d'autres séquences <i>cis</i>-régulatrices spécifiques de l'activation distale	191
2.1. La boîte CR2 du LAE est requise pour l'activation de la région distale.	191
2.2. La région intensité 3' contient des régions activatrices distales	191
3. Evaluation du rôle de Bowl dans la répression des gènes <i>bab</i>	193
4. Cll/C15 un inhibiteur potentiel des gènes <i>bab</i> dans la région distale	195
5. Recherche de facteurs conferant l'intensité du signal	197
5.1. Dll, un déterminant de la force du signal ?	197
5.2. Les gènes <i>bab</i> s'autorégulent ?	199
5.2.1. Expression du LAE-GFP en contexte mutant pour <i>bab</i>	199
5.2.2. Recherche de sites de fixation potentiels des protéines <i>bab</i>	199
5.3. Sp1 régulateur spécifique patte des gènes <i>bab</i> ?	201

Discussion

1. Un CRM commun pour les tissus de pattes et d'antenne	203
2. Le LAE : un modele pour l'etude du fonctionnement d'un CRM	205
3. L'homéoprotéine Dll est un activateur direct de <i>bab2</i> dans les pattes et l'antenne	207
4. Un rôle spécifique de la protéine à doigt de zinc Rotund dans la régulation proximale de <i>bab2</i>	209
5. <i>bab1</i> et <i>bab2</i>, deux gènes sous le contrôle d'un même enhancer ?	211
6. Conservation du LAE chez les Drosophilidae et au-delà	213
7. Organisation tridimensionnelle du locus <i>bab</i>	213
8. Rôle du LAE dans le développement des peignes sexuels	215
9. Rn activateur direct de <i>bab2</i> et répresseur indirect	217
10. Modele	219

Sommaire des figures

Figure 1. Structure de la PolII.	2
Figure 2. Etapes de l'initiation de la transcription PolII.	4
Figure 3. PolII en pause et sa libération du promoteur proximal.	6
Figure 4. Structure modulaire du complexe Médiateur et relations fonctionnelles avec d'autres composants du PIC.	8
Figure 5. Positionnement relatif des motifs de liaison (grammaire) et coopérativité entre les facteurs de transcription.	12
Figure 6. Les modèles de fonctionnement des enhancers.	14
Figure 7. Interactions enhancer-promoteur.	16
Figure 8. CRM de <i>Dll</i>	18
Figure 9. Robustesse de la transcription grâce à l'interaction enhancer-enhancer et la redondance.	20
Figure 10. Robustesse de l'archipel de la régulation.	22
Figure 11. Evolution des éléments régulateurs en <i>cis</i> du gène <i>yellow</i> et implication dans la pigmentation chez différentes espèces de drosophiles.	24
Figure 12. L'expression de <i>bab1</i> dans l'abdomen est régulée par deux CREs (éléments <i>cis</i> -régulateurs).	26
Figure 13. Modèle pour le fonctionnement et l'évolution du « Switch » génétique dimorphique.	28
Figure 14. Schéma montrant la structure de différents insulateurs de la drosophile et de vertébrés.	30
Figure 15. CTCF régule les interactions entre enhancer-promoteur pour contrôler le type de latence des virus Epstein-Barr.	30
Figure 16. Structure de certains domaines créés par l'interaction entre les insulateurs CTCF dans des cellules souches embryonnaires de souris.	32
Figure 17. Les mécanismes indirects de coopérativité entre les facteurs de transcription.	38
Figure 18. Structure du nucléosome.	44
Figure 19. Mécanismes de régulation transcriptionnelle par la méthylation de l'ADN.	46
Figure 20. Les variants d'histones: structure et fonction.	54
Figure 21. Représentation schématique des 4 grandes familles d'enzyme de remodelage ATP-dépendant.	56
Figure 22. Les différents modes d'action des complexes de remodelage ATP-dépendants.	56
Figure 23. Organisation du complexe Bithorax.	58
Figure 24. Antenne sauvage de <i>Drosophila melanogaster</i>	64
Figure 25. Patte sauvage de <i>Drosophila melanogaster</i>	64
Figure 26. Cycle de vie de <i>Drosophila melanogaster</i>	64
Figure 28. Formation des territoires et des primordia au sein du disque œil-antenne.	66
Figure 29. Origine embryonnaire des disques imaginaux de pattes et d'aile.	68
Figure 30. Origine embryonnaire du disque œil-antenne.	68
Figure 31. Patte et antenne sont deux structures homologues.	70
Figure 32. Vue d'ensemble du développement de la patte.	72
Figure 33. Relations génétiques et patrons d'expression des acteurs de l'axe P/D.	74
Figure 34. Modèles en gradient versus en cascade.	76
Figure 35. Signalisation EGFR dans les segments tarsaux.	78
Figure 36. Profil d'expression de <i>rn</i> et répression progressive de l'expression de <i>Bar</i> dans <i>ta3</i> par <i>rn</i>	80
Figure 37. Modèle d'interactions génétiques dans le patterning du tarse.	82
Figure 38. Interactions fonctionnelles entre les gènes impliqués dans le développement du t5 et du prétarse.	82
Figure 39. L'expression de <i>Dl</i> et des cibles de Notch définissent plusieurs domaines à travers les « vrais » segments de la patte.	88
Figure 40. Modèle de la formation des jonctions dans les tarses de la patte.	90
Figure 41. <i>Dll</i> , <i>hth</i> et <i>ss</i> sont essentiels pour l'identité antenne.	92
Figure 42. Expression génique différentielle antenne/patte.	98
Figure 43. Domaines d'expression des principaux marqueurs de l'antenne le long de l'axe P/D.	98
Figure 44. Le locus <i>bab</i> se compose de deux gènes.	100
Figure 45. <i>Bab1</i> et <i>Bab2</i> possèdent un domaine BTB, un domaine Psq et un motif AT-hook-like.	100
Figure 46. L'expression de <i>bab2</i> est dynamique au cours du développement de l'antenne et de la patte.	102
Figure 47. Le locus <i>bab</i> est requis pour la formation des jonctions dans la partie distale de la patte et de l'antenne.	102

Figure 48. Expression de <i>bab2</i> au cours du développement cardiaque.	104
Figure 49. L'expression de <i>bab2</i> dépend de <i>Dll</i> dans les disques de patte, d'antenne et d'aile.	108
Figure 50. L'expression de <i>bab2</i> dépend en partie de <i>ss</i> dans l'antenne et dans la patte.	110
Figure 51. Expression de <i>bab2</i> dans un disque de patte mutant pour <i>rn</i>	110
Figure 52. Interactions régulant l'expression des gènes <i>bab</i> le long de l'axe PD de l'antenne et de la patte. ...	112
Figure 53. <i>bow1</i> est requis pour développement normal des segments tarsaux.	114
Figure 54. Modèle de régulation des gènes <i>bab</i> par <i>bow1</i> dans les parties proximale et distale du tarse.	116
Figure 55. Répression de l'expression de <i>bab2</i> par l'activité de <i>dac</i> , <i>sal</i> et <i>hth</i> dans l'antenne.	118
Figure 56. Stratégie de clonage adoptée pour la construction de la banque de FTs.	180
Figure 57. Stratégie de clonage de la séquence d'ADN cible à tester ('Bait').	182
Figure 58. Test d'auto-activation pour les souches qui ont intégré au moins une copie du gène rapporteur <i>his3</i>	184
Figure 59. Exemple de crible simple hybride obtenu avec la séquence LAE.	186
Figure 60. La boîte CR2 est nécessaire à l'activation de l'anneau distal, dans les tissus de pattes et d'antenne	190
Figure 61. La région intensité 3' contient des régions activatrice et répressive.	192
Figure 62. Les sites potentiels de liaison de Bowl sont requis pour réprimer l'activité du LAE en dehors du domaine d'expression normal de <i>bab2</i> , proximale pour les disques de pattes et d'antenne, ainsi que distalement et dans les régions inter-anneaux pour ceux de pattes.	194
Figure 63. Autorégulation négative des gènes <i>bab</i>	198
Figure 64. Model du fonctionnement de l'enhancer LAE dirigeant l'expression du gène <i>bab2</i> dans les tissus de pattes et l'antenne.	218

Préambule

Le développement de la drosophile est un processus complexe au cours duquel les cellules vont se déterminer et se différencier progressivement de manière coordonnée pour aboutir à la mise en place des différents tissus, organes et structures de l'adulte. Plusieurs approches de génétique moléculaire, fondées principalement sur l'identification et l'étude de mutants, ont révélé les mécanismes majeurs régissant le développement, faisant intervenir différentes cascades de régulation génique. Toutefois, en raison du grand nombre d'acteurs identifiés et de la complexité de leurs interactions, il devient de plus en plus difficile de déterminer avec exactitude les acteurs directement impliqués dans la régulation d'un gène donné. Afin de mieux comprendre (i) les mécanismes moléculaires régissant le contrôle de l'expression des gènes impliqués dans le développement et (ii) les réseaux génétiques s'établissant en réponse à des processus de signalisation cellulaire, de plus en plus d'études s'intéressent aujourd'hui aux éléments *cis*-régulateurs des gènes. C'est dans ce cadre que s'inscrit cette thèse consacrée à l'étude du "contrôle de l'expression transcriptionnelle des gènes *bric-à-brac* (*bab1/2*) lors de la morphogenèse des appendices chez le modèle drosophile".

Les appendices de la drosophile adulte se développent à partir des disques imaginaux et peuvent être classés en deux catégories : (i) les appendices dorsaux, non segmentés, regroupant l'aile et l'haltère (ii) les appendices ventraux, segmentés, regroupant les pièces buccales, l'antenne, les pattes et les pièces génitales (génitalias).

Au cours de ma thèse je me suis intéressé à l'expression des gènes *bab* dans les disques imaginaux de pattes et de l'antenne, deux appendices ventraux homologues qui auraient évolué à partir d'une même structure ancestrale (Snodgrass, 1935) et présentant plusieurs homologies aussi bien sur le plan moléculaire que structural. Ainsi mon projet comportait deux parties consistant, (1) en la détermination (1) du ou des module(s) régulateur(s) en *cis* (CRMs) requis pour l'expression des gènes *bab1/2* dans les disques imaginaux de pattes et/ou d'antenne, et (2) en l'identification de facteurs de transcription spécifiques (FTs) s'associant directement en *trans* à ce ou ces CRM(s).

I-Introduction

Partie I : Régulation de la transcription

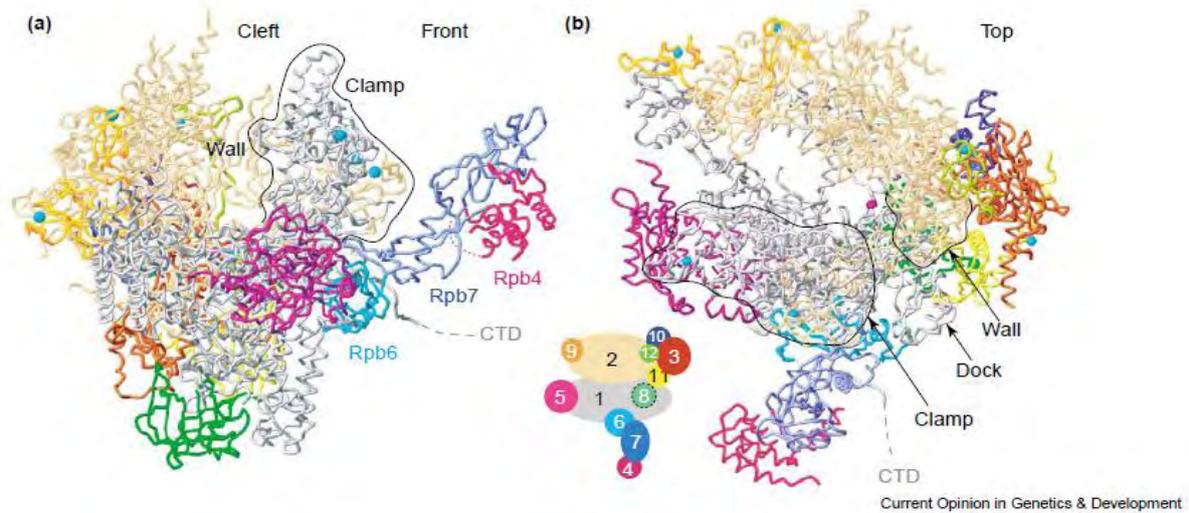


Figure 1. Structure de la PolIII. Vue de face (a) et de dessus (b) de la PolIII. Le schéma du centre représente l'espace et la position occupés par les sous-unités Rpb1 à Rpb12 au sein de la PolIII, le code couleur étant le même pour les trois représentations. Les principales structures de la PolIII sont représentées, à savoir la pince (clamp) le mur (wall), le sillon et le CTD. Les deux sous-unités centrales Rbp1 et Rbp2 forment le sillon chargé positivement dans lequel s'insère l'ADN simple brin. La pince permet d'ouvrir et de fermer le sillon. (D'après Cramer, 2004)

1. LA TRANSCRIPTION

Au cours du développement des eucaryotes supérieurs l'activité des gènes doit être finement modulée. Cette modulation peut s'effectuer à plusieurs niveaux, à savoir la transcription, la maturation et la stabilité des ARNm, la traduction, les modifications post traductionnelles et enfin la stabilité de la protéine. Néanmoins le contrôle de la transcription reste l'une des étapes les plus cruciales.

La transcription est le processus par lequel l'information génétique (ADN) est transcrite en ARN par les ARN polymérase (ARN pol). Chez les eucaryotes, il existe trois ARN pol nucléaires différentes. Chacune de ces ARN pol est responsable de la transcription d'un type d'ARN : l'ARN pol I transcrit les ARN ribosomiques (ARNr) 18S et 28S ; l'ARN PolII transcrit les ARN messagers (ARNm); enfin, l'ARN pol III transcrit l'ARN ribosomique 5S, l'ARN de transfert (ARNt) ainsi que des petits ARN (ARNsn, ARNno et pARNi) (Lee et Young, 2000 ; Thomas et Chiang, 2006).

1.1. STRUCTURE DE L'ARN POLYMERASE II (POLII)

L'ARN polymérase II (PolII) est l'unité catalytique de synthèse de l'ARNm et elle est composée de 12 sous-unités nommées Rpb1-12. Ces dernières sont désignées selon un ordre décroissant de la masse moléculaire (Figure 1) et sont hautement conservées de la levure à l'homme. En plus de leur conservation inter-espèces, un bon nombre des sous-unités de la PolII sont partagées (Rbp5, 6, 8, 10 et 12) ou possèdent des homologues (Rbp1, 2, 3 et 11) avec l'ARN pol I (PolI) et l'ARN pol III (PolIII) (Thomas et Chiang 2006).

La plus grande sous-unité de la PolIII eucaryote, Rpb1, se différencie de ses homologues de la PolI et PolIII par le domaine CTD (« carboxy-terminal domain ») localisé à son extrémité C-terminale. Ce dernier est constitué par une répétition de l'heptapeptide Tyr1-Ser2-Pro3-Thr4-Ser5-Pro6-Ser7 (Corden et al., 1985) dont le nombre varie d'une espèce à l'autre (26 chez la levure *Saccharomyces cerevisiae*, 44 chez la drosophile et 52 chez les mammifères). Le CTD est essentiel à l'activité de la PolIII *in vivo* et sa délétion est létale chez la levure, la drosophile et la souris (Allison et al., 1988). Les heptapeptides répétés du CTD comportent plusieurs sites pouvant être phosphorylés et permettent, selon leur état de phosphorylation, de moduler l'activité de la PolIII pendant les différentes étapes de la transcription.

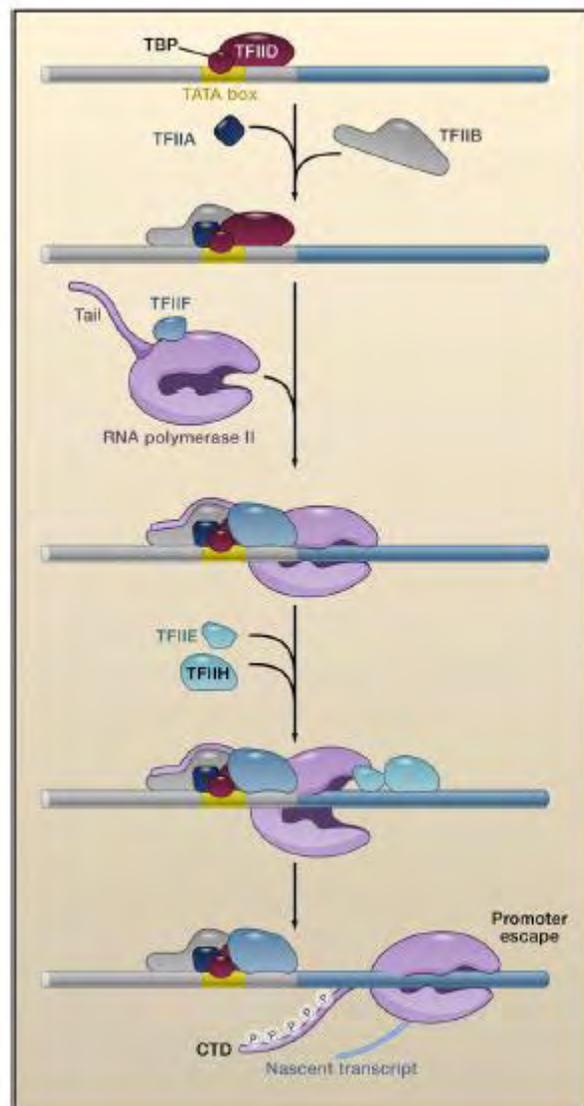


Figure 2. Etapes de l'initiation de la transcription PolII. L'initiation de la transcription débute par la fixation du facteur de transcription général TFIID, sur des éléments spécifiques au niveau du promoteur, y compris TATA (via TBP). Cela conduit au recrutement d'autres facteurs de transcription généraux, y compris TFIIA, TFIIB, TFIIF, et, finalement, l'ARN polymérase II (PolII). La liaison du complexe de pré-initiation (PIC) est initialement instable, étant donné que le complexe est dans une conformation fermée. Le recrutement de TFIIH conduit à la formation d'un complexe ouvert et au début de la transcription. L'efficacité de la transcription dépend de la phosphorylation de la PolII au niveau du CTD (PolII C-terminal region), qui favorise la fuite du promoteur. (Adapté de Gilmour, 2009).

1.2. LES ETAPES DE LA TRANSCRIPTION

La transcription des gènes codant pour des protéines comporte quatre étapes : la pré-initiation, l'initiation, l'élongation et enfin la terminaison.

La pré-initiation constitue la première étape dans l'expression d'un gène et implique la formation du complexe de pré-initiation (PIC) au niveau du promoteur proximal ou « core » (revue par Thomas et Chiang, 2006; Sikorski et Buratowski, 2009). Le PIC se compose de la PolII et des facteurs de transcription généraux associés dont certains (tels que TFIID et TFIIB) reconnaissent des séquences d'ADN spécifiques au sein du promoteur « core ». Parmi ces sites on trouve les boîtes TATA (reconnue par la sous-unité TBP du complexe TFIID), Inr (initiator), BRE (reconnue par TFIIB) et l'élément DPE (« downstream promoter element ») (Juven-Gershon et al., 2008) (Figure 2). Ce complexe est initialement instable étant donné que l'ADN du promoteur « core » est dans une conformation double hélice fermée. Ainsi la prochaine étape dans l'initiation consiste en la dissociation locale de l'ADN. Cette action dépend du recrutement de TFIIF, un complexe multiprotéique qui a approximativement la même taille que la PolII elle-même (Kim et al., 2000). TFIIF se lie en aval de la PolII et contient des hélicases ATP-dépendantes qui vont ouvrir l'ADN double-brin. Ainsi la PolII peut initier la synthèse d'un ARN de 20-30 nucléotides en aval de la boîte TATA. Toutefois, cette synthèse est d'abord inefficace et résulte souvent en la production d'un petit ARN composé de moins de 10 nucléotides. C'est seulement lorsque la PolII réussit à synthétiser un ARN naissant de plus de 10 nucléotides, qu'elle devient compétente (processive) pour assurer une transcription sur toute la longueur du gène. Ce phénomène dépend de la phosphorylation du domaine répété CTD de la grande sous-unité de PolII, par TFIIF et d'autres facteurs de transcription généraux tels que le complexe Médiateur (Hirose et Ohkuma, 2007; Boeing et al., 2010).

Après la synthèse d'environ 30 nucléotides d'ARN naissant, la PolII se décroche du promoteur « core » initiant ainsi l'étape d'élongation. Une partie du PIC reste alors fixée sur le promoteur et constitue un échafaudage qui permet une ré-initiation rapide de la transcription (Dieci et Sentenac, 2003). Enfin, lorsque le complexe tripartite PolII/ARN/ADN atteint le terminateur, la transcription s'achève (Kuehner et al., 2011) et l'ARN pré-messager est libéré.

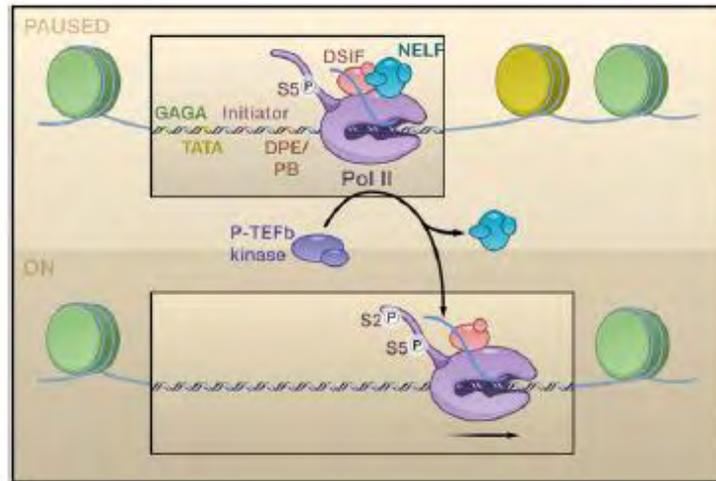


Figure 3. PolII en pause et sa libération du promoteur proximal. Le panneau supérieur montre un promoteur sur lequel est liée la PolII en pause. La région promotrice contient des éléments de séquence qui favorisent la liaison et l'activation de la PolII, y compris GAGA, TATA, initiateur et les motifs DPE/PB (downstream promoter element/pause button). La PolII est généralement en pause juste en aval de la région DPE. Elle « s'échappe » du promoteur (promoter escape) et présente une phosphorylation de la sérine 5 (Ser5) dans le domaine C-terminal (CTD) de Rpb1. DSIF (5,6-dichloro-1-b-D-ribofuranoxybenzimidazole sensitivity-inducing factor) et NELF (negative elongation factor) aident à arrêter la PolII en se liant au transcrit naissant (typiquement 30-50 nucléotides de longueur). Le recrutement de P-TEFb (positive transcription elongation factor b) provoque la libération de NELF et la phosphorylation de la Ser2 du CTD, ce qui libère la PolII et relance la transcription. (D'après Levine, 2011).

1.3. « PAUSED » POLII

Des études réalisées sur des cellules souches embryonnaires (ES) de souris (Bernstein et al., 2006; Guenther et al., 2007; Bilodeau et al., 2009) ont révélé la présence simultanée dans plusieurs régions du génome de deux isoformes d'histone H3 : une forme triméthylée sur la lysine 4 (H3K4me3) et une forme triméthylée sur la lysine 27 (H3K27me3). Ces dernières sont des marques généralement associées respectivement à l'activation et à la répression transcriptionnelle (Bernstein et al., 2006; Levine, 2011). Plusieurs gènes inactifs sont enrichis par ces deux types d'histones modifiées H3K4me3 et H3K27me3, indiquant à la fois des gènes actifs et non actifs. De façon surprenante la PolII est recrutée sur ces gènes, pourtant non transcrits, observation qui a permis l'émergence de la notion de promoteurs préparés à la transcription (« paused promoters ») (Fisher et Fisher 2011). Ainsi, les gènes contenant les deux marques bivalentes sont considérés comme réprimés, mais prêts pour une activation rapide dès lors que les signaux activateurs appropriés deviennent disponibles. Un tel état de la répression pourrait être important pour la différenciation ordonnée de différents types de cellules dérivant de cellules progénitrices multipotentes.

Des essais de liaison menés sur le génome entier suggèrent que 10 à 30% des gènes inactifs présentent une fixation de la PolII sur leurs régions promotrices et ce quel que soit le type cellulaire (Levine, 2011). Chez la drosophile, les gènes inactifs ayant des promoteurs associés à la PolII sont nettement surreprésentés parmi ceux contrôlant le développement (Levine, 2011). De nombreux gènes *Hox*, des déterminants de tissus (comme *tinman*, un gène impliqué dans le développement du cœur), et des composants des voies de signalisation cellulaire (par exemple, les gènes codant pour le récepteur du facteur de croissance de type FGF) fixent la PolII depuis le stade de l'embryon précoce, alors qu'ils sont activés plus tardivement au cours du développement (Zeitlinger et al., 2007; Muse et al., 2007; Chopra et al., 2009a).

Des études récentes suggèrent que la PolII s'arrête souvent en aval de l'élément DPE du promoteur « core », situé entre +28 pb et +32 pb en aval du site initiateur de la transcription (Figure 3 ; Juven-Gershon et al., 2008). Un autre motif de séquence court, l'élément GAGA, qui lie le facteur TrG trithorax-like (Trl) (Tsukiyama et Wu, 1995; Okada et Hirose, 1998), a été également impliqué dans l'association de la « paused » PolII au promoteur proximal (Lee et al., 1992; Wilkins et Lis, 1997). En effet, la délétion de l'élément GAGA situé à proximité de la région 5' du promoteur du gène *hsp70* de la drosophile, altère la capacité de ce dernier à avoir une « paused » PolII au niveau de son promoteur. Toutefois,

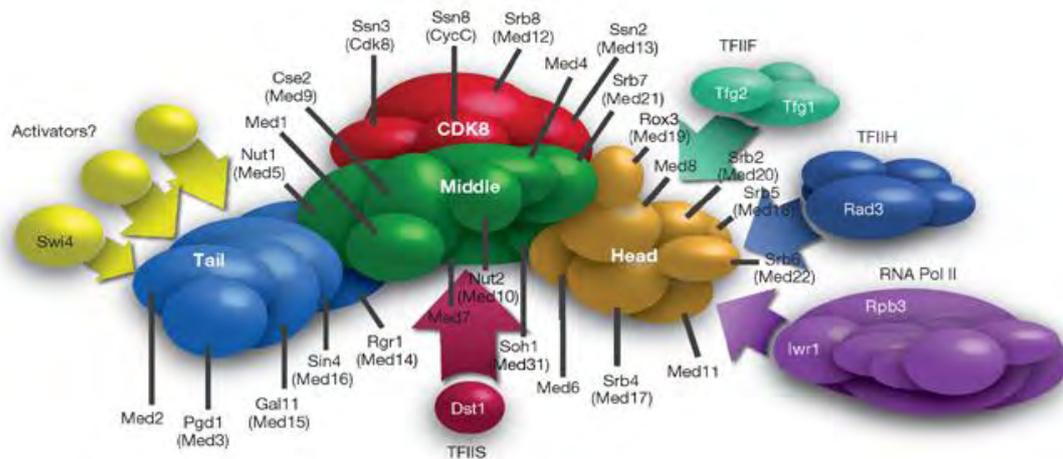


Figure 4. Structure modulaire du complexe Médiateur et relations fonctionnelles avec d'autres composants du PIC. Le complexe Médiateur présente une organisation modulaire avec des modules appelés : tête, milieu, queue et CDK8. Certaines relations fonctionnelles entre les modules du Médiateur et divers composants du PIC sont représentées (Collins et al., 2007).

il est à noter que cette délétion ne bloque pas l'activation d'un promoteur modifié de *hsp70* (Wilkins et Lis, 1997).

La libération de la PolII à partir du promoteur proximal est associée à la phosphorylation du résidu sérine-2 (Ser2P) du CTD (Boehm et al., 2003; Keogh et al., 2003). Cette libération dépend du facteur la transcription d'élongation positif b (P-TEFb) (Figure 3 ; Boehm et al., 2003), qui se compose de la kinase CDK9 et de son régulateur, la cycline T (Marshall et Price, 1995; Peng et al., 1998). CDK9 phosphoryle soit directement ou indirectement le CTD de Rpb1 au niveau des résidus Ser2. Le recrutement de CDK9 est souvent suffisant pour activer l'expression des gènes (revue de Peterlin et Price, 2006).

Il est actuellement difficile de savoir si le recrutement de la « paused » PolII est utilisé comme une stratégie de régulation en matière de développement quoique quelques cas le laissent présager. Par exemple, le gène *singled-minded (sim)* qui est le déterminant d'une région spécialisée du système nerveux central chez la drosophile, la ligne médiane ventrale (Nambu et al., 1991; Wheeler et al., 2009), présente une association « paused » PolII - promoteur dépendante du tissu considéré. En effet, le promoteur de *sim* contient une « paused » PolII dans l'ectoderme, mais pas dans le mésoderme où le gène ne s'exprime jamais.

D'autre part, il est suggéré que le positionnement de la PolII au niveau des promoteurs anticipe une activation rapide et synchrone des gènes et ceci en échappant à l'étape de recrutement de la PolII, qui peut être stochastique.

1.4. LE COMPLEXE MÉDIATEUR

Le Médiateur, est un complexe multiprotéique de grande taille, conservé chez tous les eucaryotes, jouant un rôle essentiel dans la régulation de la transcription des gènes (Boube et al., 2002; Bourbon, 2008; Sato et al., 2004). Il est composé par au moins 25 sous-unités et présente une structure modulaire, dans laquelle on distingue quatre modules (Figure 4) : la Tête, le Milieu, la Queue qui constituent le complexe « core » et le module CDK8 (Guglielmi et al., 2004). Dans les cellules de levure, une fraction du complexe Médiateur « core » est associée au module kinase, qui comprend la kinase cycline-dépendante Cdk8, cycline C et deux grandes sous-unités supplémentaires désignées Med12 et Med13. Une autre fraction du complexe médiateur « core » n'est pas attachée au module CDK8, mais elle est par contre étroitement associée à la PolII dans ce qui est parfois appelé le complexe Médiateur « holoenzyme ». La majeure partie du Médiateur « core » métazoaire et le complexe

holoenzyme comprennent une sous-unité spécifique, Med26. Enfin, une petite population de Médiateur qui contient à la fois le module kinase et Med26 peut être isolée à partir d'extraits de cellules de mammifères (Ding et al., 2009; Takahashi et al., 2011).

Le complexe Médiateur a été identifié pour la première fois dans des extraits de transcription de *S. cerevisiae* et purifié par chromatographie par Roger Kornberg et ses collègues grâce à sa capacité à permettre l'activation de la transcription PolIII par des facteurs de transcription (FTs) liant l'ADN, et ceci dans un système enzymatique reconstitué avec des facteurs généraux purifiés (Kim et al., 1994; Myers et Kornberg, 2000). Des études ultérieures menées sur le mécanisme d'action du Médiateur de levure et des organismes supérieurs ont révélé que le Médiateur favorise l'activation de la transcription par la PolIII et ceci via des interactions directes à la fois avec la PolIII et les FTs liant l'ADN (Malik et Roeder, 2010; Balamotis et al., 2009). Bien que ne se liant pas directement à l'ADN, ces propriétés permettent au MED de faciliter l'assemblage de complexes PIC fonctionnels (Baek et al., 2006; Wang et al., 2005). Plus récemment, des preuves ont été apportées sur l'implication du Médiateur dans les étapes de post-initiation de la transcription. Ces études ont mis en lumière la capacité du Médiateur (i) à contourner les activités des facteurs régulant négativement l'élongation (Cheng et al., 2012; Jishage et al., 2012), (ii) à recruter des facteurs d'élongation de la transcription et les facteurs de maturation des ARN pré-messagers (Donner et al., 2010; Huang et al., 2012) et (iii) à contrôler la phosphorylation des répétitions de l'heptapeptide du CTD de la grande sous-unité de la PolIII (Boeing et al., 2010; Donner et al., 2010). Le CTD phosphorylé joue un rôle central dans la coordination de la maturation de l'ARN pré-messager en recrutant des enzymes et des protéines essentielles pour la bonne formation de la coiffe, pour l'épissage, pour la polyadénylation des ARN pré-messagers, ainsi que pour l'exportation nucléaire appropriée et la localisation de l'ARNm mature (Buratowski, 2009; Perales et Bentley, 2009). De ce fait, le Médiateur peut participer indirectement à l'ensemble de ces processus en modulant la phosphorylation du CTD ou celle d'autres facteurs impliqués dans la transcription PolIII (ex : pTEFb). Cdk8 peut aussi phosphoryler directement des FTs conduisant à leur répression ou leur activation selon les cas. Enfin, de plus en plus de données indiquent un rôle du MED dans la coordination entre le remodelage de la chromatine et l'initiation de la transcription.

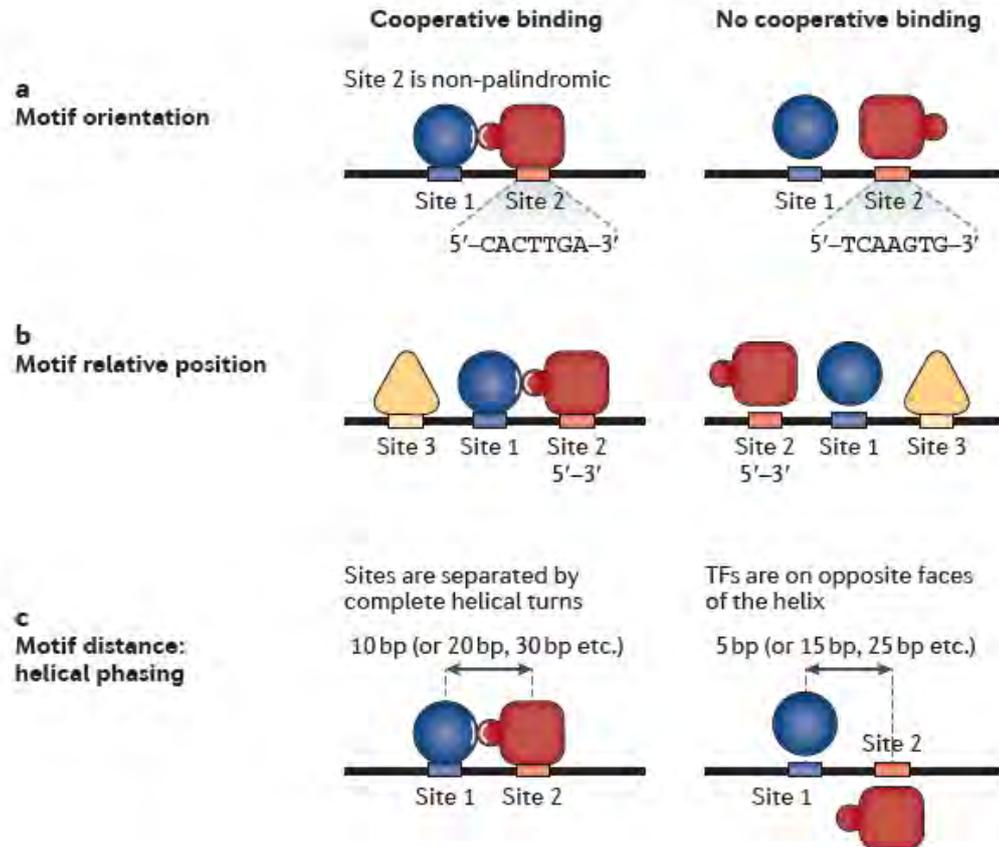


Figure 5. Positionnement relatif des motifs de liaison (grammaire) et coopérativité entre les facteurs de transcription. La colonne de gauche montre des exemples dans lesquels deux facteurs de transcription (FTs) ont une interaction protéine-protéine directe qui favorise leur liaison coopérative. **a)** Un des facteurs de transcription (en rouge) se lie à un motif non palindromique (site 2, 5' - CACTTGA-3'). Lorsque l'orientation relative du site 2 est inversée, ce FT ne peut pas être recruté sur l'ADN **b)** De la même manière, lorsque la position relative du site 2 par rapport au site 1 est modifiée, les interactions directes protéine-protéine entre les deux facteurs ne peuvent plus se produire. **c)** Il y a environ 10 paires de bases (pb) par tour d'hélice d'ADN. Si deux FTs se lient de manière coopérative à l'ADN à travers une interaction directe protéine-protéine, ils doivent généralement être placés sur la même face de l'ADN. Ceci est facilité lorsque l'espace entre les deux motifs est de ~10 pb, 20 pb ou 30 pb. Lorsque les sites 1 et 2 sont séparés par des distances telles que ~ 5 pb, 15 pb ou 25 pb, les TF seront positionnés sur des faces opposées de la double hélice de l'ADN, bloquant ainsi leur capacité à se lier directement l'un à l'autre (colonne de droite). (D'après Spitz et Furlong, 2012).

2. RÉGULATION DE LA TRANSCRIPTION

2.1. ÉLÉMENTS RÉGULATEURS EN *CIS*

L'expression des gènes est régulée par l'action appropriée de nombreux éléments régulateurs en *cis*, incluant les promoteurs proximaux (Lenhard et al., 2012), ainsi que des modules *cis*-régulateurs (CRMs) parfois localisés à de très grandes distances du site d'initiation de la transcription, tels que les 'enhancers' (Levine 2010 ; Bulger et Groudine 2011), les 'silencers' (Petrykowska et al., 2008 ; Vokes et al., 2008 ; Ayer et Benyajati 1990), les insulateurs (Gaszner et Felsenfeld 2006) et les éléments 'tethering' (Ohtsuki et al., 1998 ; Calhoun et al., 2002). Parmi cette constellation d'éléments, les enhancers et leurs facteurs de transcription spécifiques (FTs) associés ont un rôle de premier plan dans l'initiation de l'expression des gènes.

2.1.1. ENHANCERS

Les enhancers sont des segments d'ADN relativement petits, typiquement de quelques centaines de paires de bases (pb), qui servent de plates-formes pour recruter des FTs reconnaissant généralement de petites séquences d'ADN plus ou moins dégénérées ayant généralement une taille de 6 à 12 pb de longueur. Cette propriété intrinsèque suggère que des règles plus complexes, autres que la simple affinité individuelle des FTs à l'ADN, sont impliquées à la fois dans le contrôle de leur fixation aux enhancers et de leur activité.

✓ MOTIFS DE SEQUENCE ET ARCHITECTURE DES ENHANCERS

Deux propriétés des enhancers ont été étudiées en détail: la composition et le positionnement des motifs. La composition en motifs est la présence dans les enhancers de motifs de liaison pour des facteurs de transcription spécifiques qui sont essentiels à l'expression génétique dans un type cellulaire donné. Dans plusieurs cas, il est possible de prédire l'activité spatio-temporelle des éléments *cis*-régulateurs en se basant uniquement sur les données de liaison des FTs aux enhancers générés par la technique d'immunoprécipitation de la chromatine (ChIP) (Zinzen et al., 2009). Le positionnement des motifs (souvent désigné comme «grammaire» des motifs) indique l'ordre, l'orientation et l'espacement relatif des motifs de liaison des FTs sur un enhancer (Figure 5) et est souvent considéré comme une contrainte à la fixation des FTs aux éléments régulateurs (Senger et al.,

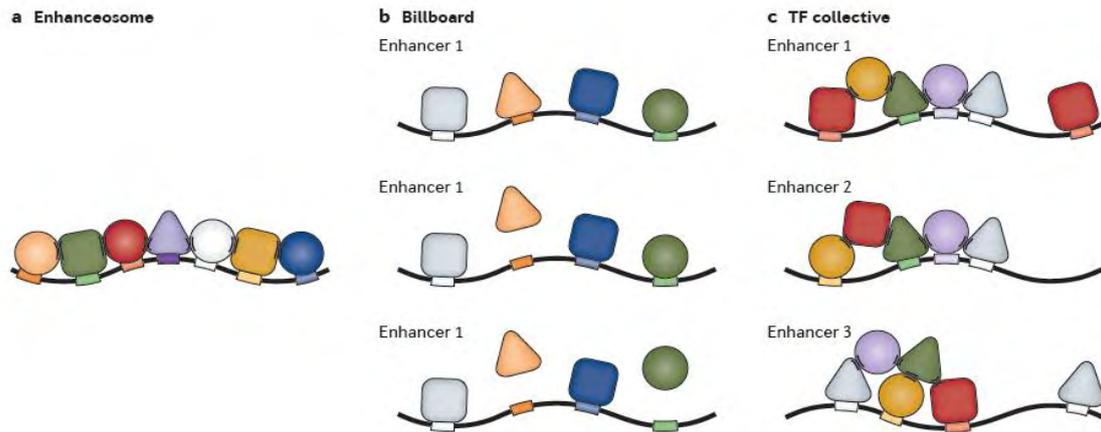


Figure 6. Les modèles de fonctionnement des enhancers. **a)** Le modèle de l' « enhanceosome » représente une situation dans laquelle tous les facteurs de transcription (FTs) qui se lient à un enhancer sont essentiels pour l'activation de l'enhancer. La composition des motifs d'ADN et leur positionnement relatif (grammaire des motifs) agissent comme un échafaudage pour recruter de manière coopérative tous les FTs, qui forment une interface protéique complexe et régulent la transcription. **b)** Le modèle « billboard ». Pour tous les enhancers, le positionnement des sites de liaison des FTs est flexible. Seul un sous-ensemble de sites de l'enhancer peut être actif à un moment donné. **c)** Le modèle dit de « collectif de FTs » représente une situation dans laquelle le même ensemble de facteurs de transcription se lient à de nombreux enhancers. Ils peuvent occuper chacun de ces enhancers d'une manière différente (avec un sous-ensemble de facteurs de transcription, ou tous les facteurs de transcription, en contact direct avec l'ADN). (D'après Spitz et. Furlong, 2012).

2004 ; Swanson et al., 2010). Typiquement, le positionnement des motifs assure que les facteurs de transcription soient positionnés de manière appropriée pour faciliter les interactions protéines-protéines et favoriser des liaisons coopératives, ainsi que le recrutement de cofacteurs et de la machinerie transcriptionnelle. L'enhanceosome (Figure 6) représente un exemple extrême de «grammaire» des motifs, car il contient un dispositif fixe de sites de liaison couvrant toute la longueur de l'enhancer (Panne et al., 2007). N'importe quel changement dans l'espacement entre les sites adjacents perturbe la fonction de l'enhanceosome, à l'exception de l'insertion de 10 paires de bases d'ADN, soit un tour d'hélice, ce qui maintient l'orientation des complexes hétérodimériques adjacents (Thanos et Maniatis 1995 ; Merika et Thanos 2001). Cependant, la plupart des enhanceurs impliqués dans le développement ont des arrangements des motifs plus flexibles que l'enhanceosome. En effet, la liaison des FTs aux éléments *cis*-régulateurs se fait pour certains de manière coopérative, alors que d'autres le font de manière additive ou indépendante. Ce mode de positionnement plus flexible des motifs a amené au modèle d'enhancer «billboard» (Kulkarni et Arnosti 2003) (Figure 6). Une étude récente menée sur cinq FTs impliqués dans le développement cardiaque de l'embryon de drosophile, a permis de proposer un nouveau modèle d'enhancer qualifié de «collectif de FTs» (Figure 6 ; Junion et al., 2012). Selon ce modèle, la fixation des FTs sur les enhanceurs se fait selon la loi de tout ou rien à savoir l'absence de l'un entraîne la perte de fixation des autres facteurs. Comme l'ensemble des enhanceurs étudiés ne partagent pas une grammaire de motifs évidente et ne présentent chacun qu'un sous ensemble de motifs de fixation pour les cinq FTs, il a été suggéré que l'association collective des FTs nécessite une coopérativité via des interactions protéine-protéine. Ce mode de régulation montre bien que les déterminants de l'activité globale d'un enhancer peuvent s'étendre au-delà d'un code linéaire basé sur la séquence pour inclure aussi les interactions protéines-protéines.

✓ INTERACTION ENHANCER-PROMOTEUR

Plusieurs preuves montrent que dès lors qu'une combinaison appropriée d'activateurs et de co-activateurs sont liés à un enhancer, ce dernier forme une boucle régulatrice avec les régions promotrices des gènes cibles (Sanyal et al., 2012). Ces boucles peuvent être visualisées en utilisant la technique de capture de la conformation chromosomique (3C) (van Berkum et dekker 2009 ; Sanyal et al., 2012), qui permet par PCR d'identifier le rapprochement physique de segments d'ADN très éloignés sur le chromosome. Les séquences impliquées dans la formation de la boucle entre l'enhancer et le promoteur ne sont pas

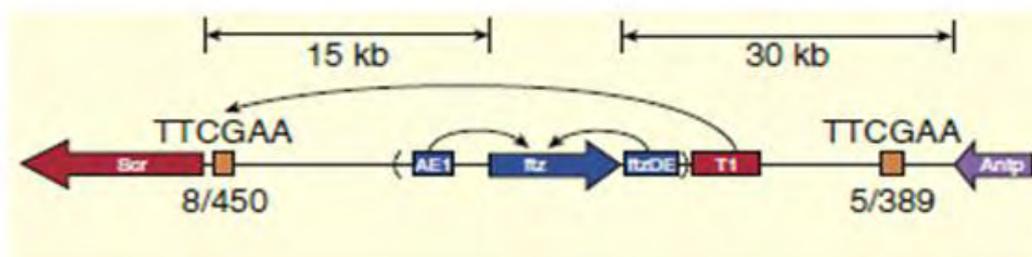


Figure 7. Interactions enhancer-promoteur. L'enhancer *T1* contourne le locus *ftz* pour activer *Scr* au niveau du complexe de gènes *Antennapedia* chez la drosophile. (D'après Levine, 2010)

forcement impliquées dans l'activation de la transcription. En effet, dans le cas de l'enhancer de *Pax2* chez *Drosophila melanogaster* la région qui est requise pour l'activation à longue distance (en permettant le rapprochement enhancer/promoteur) est sans importance lorsque l'enhancer est placé à proximité du promoteur (Swanson et al., 2010).

Dans certains cas, les enhancers peuvent ignorer un gène voisin et activer une unité de transcription plus distale. Ceci est observé au niveau du complexe Antennapedia chez la drosophile. En effet, l'enhancer *TI* situé du côté 3' de l'unité transcriptionnelle *fushi tarazu* (*ftz*) évite le promoteur de *ftz* afin d'activer le gène *Sex combs reduced* (*Scr*) plus distal (Gindhart et al., 1995; Calhoun et al., 2002 ; Calhoun et Levine, 2003). L'interaction sélective *TI*-promoteur *Scr* dépend d'un élément 'tethering' proximal situé immédiatement du côté 5' du promoteur cœur de *Scr* (Figure 7). *TI* peut activer l'expression du promoteur de *ftz* suite à l'insertion d'un élément 'tethering' de 300 pb. Ce type de spécificité enhancer-promoteur est également influencé par les éléments de séquence contenus dans les promoteurs de base. Les promoteurs contenant une boîte TATA pourraient interagir préférentiellement avec certains enhancers, alors que des promoteurs contenant un élément DPE sont activés par différents enhancers (Ohtsuki et al., 1998; Juven-Gershon et al., 2008 ; Butler et Kadonaga, 2001). Les éléments 'tethering', et la spécificité du promoteur, pourraient permettre à des enhancers agissant à des distances lointaines d'ignorer les gènes voisins et d'activer des unités de transcription distales.

✓ REGULATION MODULAIRE DES GENES DU DEVELOPPEMENT

La plupart des gènes du développement sont pléiotropiques (influencent la formation ou l'apparition de multiples traits) et présentent des promoteurs modulaires caractérisés par la présence de plusieurs enhancers, contrôlant chacun un aspect spatio-temporel spécifique de l'expression d'un gène donné (Davidson et al., 2001; Wray et al., 2003 ; Nelson et al., 2004). Un des exemples les plus concrets est la régulation du gène *Dll* au cours du développement de la patte. La dissection du locus *Dll* a révélé que l'expression de *Dll* dans les pattes dépend de plusieurs CRMs assurant chacun une partie (un sous-domaine) du profil d'expression complexe de *Dll* (Figure 8), parmi lesquels on trouve :

Dll304

Le premier signe de la formation d'appendice dans l'embryon de drosophile est l'activation de *Dll* à ~6h après la ponte (AEL; stade 11) dans des domaines circulaires comprenant ~20-30

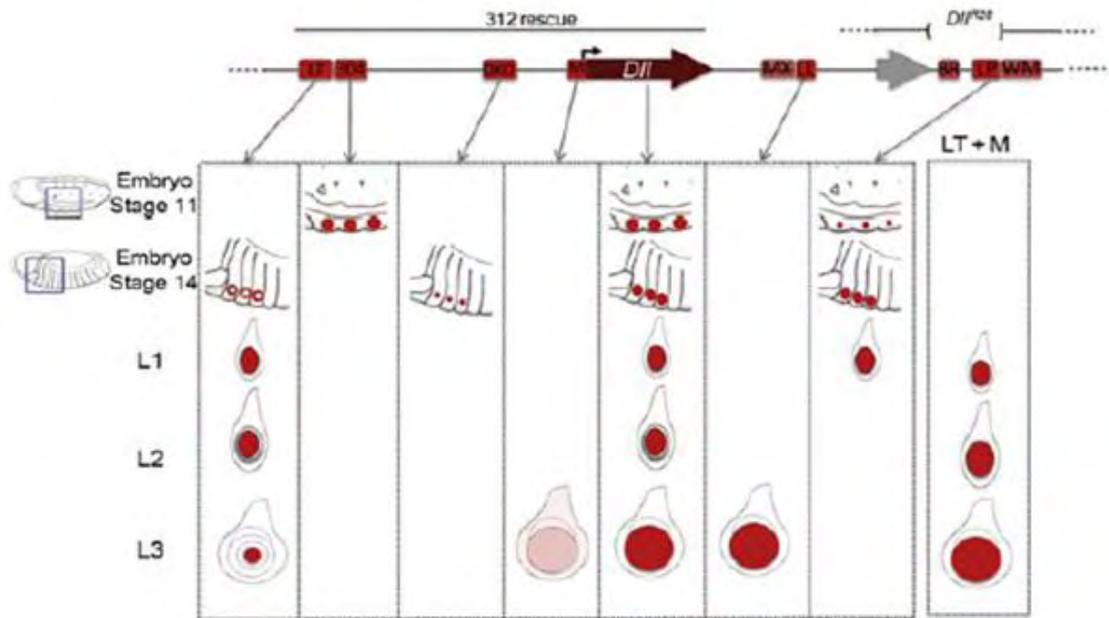


Figure 8. CRM de *Dll*. Schéma de la région génomique de *Dll* montrant les positions des CRMs identifiés (cases colorées) et de l'unité de transcription (grande flèche). Les profils d'expression dirigés par les CRMs individuels sont indiqués. Les CRMs *DllMX* et *DllWM* ne sont pas actifs au cours du développement de la patte et *DllBR* n'est actif qu'à un stade très tardif (Galindo et al., 2011; Pappu et al., 2005). La déficience *DllR28* engendre des défauts principalement au niveau des segments tarsaux. (D'après Estella et al., 2012).

cellules par héli-segment thoracique (Cohen, 1990). A ce stade, un élément régulateur précoce de *Dll* situé à 11 kb du côté 5' du site d'initiation de la transcription de *Dll*, appelé *Dll early enhancer* ou *Dll304*, est en mesure de diriger l'expression d'un gène rapporteur avec un profil similaire à *Dll* (Vachon et al., 1992). *Dll* et l'activité du CRM *Dll304* dépendent tous deux à ce stade d'un apport positif de la signalisation Wg, mais pas de la signalisation Dpp (Cohen, 1990; Cohen et al., 1993). En fait, *Dll304* est réprimé dorsalement et ventralement par les voies Dpp et EGFR, respectivement (Goto et Hayashi, 1997; Kubota et al., 2000).

DILLT

L'enhancer *DILLT* (*Leg Trigger*) commence à être actif à ~8h AEL et permet de maintenir l'expression de *Dll* quand l'activité de *Dll304* décline (Cohen et al., 1993; Estella et al., 2008; Vachon et al., 1992). Cet élément est seulement actif dans un sous-ensemble de cellules exprimant *Dll* correspondant aux 15 cellules les plus externes du domaine *Dll* dans chaque héli-segment (Cohen et al., 1993; McKay et al., 2009). L'activation de *DILLT* nécessite Dll, Sp1 (Castelli-Gair et Akam, 1995; Estella et Mann, 2010; McKay et al., 2009) ainsi que l'action directe des voies de signalisation Wg et Dpp et ceci par la fixation des facteurs de transcription situés en aval, à savoir respectivement dTCF/Pangolin (Pan) et Mothers against Dpp (Mad) (Estella et al., 2008). Toutefois, il est à noter que *DILLT* reproduit l'expression de *Dll* jusqu'au stade larvaire L2 et se restreint au centre des disques de pattes au stade larvaire L3. En outre, s'il est placé à proximité du module *cis*-régulateur *DllM* (pour maintenance) il produit une expression précise et robuste similaire à celle de *Dll* même au stade L3 (Estella et al., 2008).

DIIDKO

DILLT est actif dans seulement 15 des cellules exprimant *Dll* dans chaque héli-segment thoracique au 14^{ème} stade embryonnaire. Les autres cellules exprimant *Dll* ont une identité neuronale tel que révélé par l'expression de gènes tels *achaete* (*ac*) et *cut* (*ct*; Bolinger et Boekhoff-Falk, 2005; Cohen et Jurgens, 1989; McKay et al., 2009). Un troisième CRM de *Dll*, nommé *DIIDKO* (pour Distal-less Keilin organ) a été identifié à ~3kb en 5' du début du site d'initiation de la transcription *Dll* et est activé spécifiquement dans ces cellules *DILLT* négatives. *DIIDKO* est activé par les membres du complexe Achaete-scute (ASC) et Dll, ce qui restreint son activité aux cellules neurogènes dans les primordia des pattes (McKay et al., 2009).

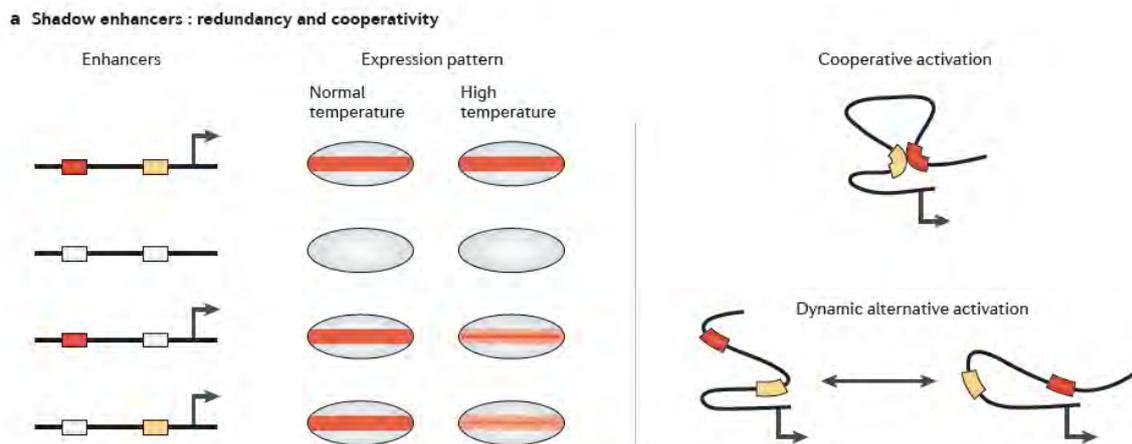


Figure 9. Robustesse de la transcription grâce à l'interaction enhancer-enhancer et la redondance. La présence de deux enhancers ayant des activités qui se chevauchent confère une robustesse à l'expression des gènes (panneau de gauche), en particulier au changement des conditions environnementales (par exemple, température élevée), ce qui établit des limites précises d'expression des gènes. Il est actuellement difficile de savoir si cette robustesse est associée à la formation d'un complexe multi-enhancer stable, ou à l'activation alternative par chaque élément séparément (représenté dans le panneau de droite), (D'après Spitz et. Furlong, 2012).

DLLP* et *DLLT

Dll304, *DLLT* et *DllDKO* sont tous situés dans une région de 14 kb du côté 5' du site d'initiation de la transcription de *Dll*. Récemment, d'autres CRMs de *Dll* situés du côté 3' de l'unité transcriptionnelle ont également été identifiés (Galindo et al., 2011; Figure 8). Parmi eux un élément, appelé *DLLP* (pour *Leg Primordium*), est actif dans un sous-ensemble de cellules exprimant *Dll* au 10^{ème} stade embryonnaire et reste actif jusqu'à la fin du premier stade larvaire. *DLLP*, comme *DLLT*, est activé par Wg et Dpp dans l'embryon, mais on ne sait pas si cette activation est directe ou pas. Un deuxième élément 3', appelé *DLLL* (pour *Leg Late*), est actif uniquement à partir de la moitié du troisième stade larvaire et dépend comme *DLLT*, de Dll pour son activité.

✓ SHADOW ENHANCER

L'analyse du génome entier d'embryons précoces de drosophile par des expériences de 'ChIP on chip', suggère que plusieurs gènes impliqués dans le développement contiennent des enhancers secondaires qui produisent des profils d'expression similaires sinon identiques à ceux générés par les enhancers primaires (Zeitlinger et al 2007 ; Hong et al 2008) (Figure 9). La distinction entre enhancers «primaires» et «secondaires» se base uniquement sur la distance physique qui les sépare du promoteur du gène ou sur l'ordre dans lequel ils ont été découverts et n'implique en aucun cas une hiérarchie ou une importance fonctionnelle.

Joung-Woo Hong et ses collaborateurs (Hong et al., 2008) ont proposé le terme "enhancers ombres" (shadow enhancers) pour désigner les enhancers secondaires localisés dans des régions lointaines de leurs gènes cibles et suggèrent que ces derniers émanent de la duplication des enhancers primaires et ceci de la même manière que la duplication et la divergence des séquences codantes. Ils suggèrent aussi que les enhancers ombres puissent acquérir de nouvelles activités régulatrices sans pour autant perturber les fonctions fondamentales des gènes contrôlant le développement présentant ainsi un potentiel évolutif important. Des études récentes suggèrent que les enhancers ombres permettent de tamponner l'expression des gènes impliqués dans le développement procurant ainsi de la robustesse face aux perturbations génétiques et environnementales, comme par exemple l'augmentation de la température (Perry et al 2010 ; Barolo et al 2012) (Figure 9).

b Robustness and regulatory archipelagos

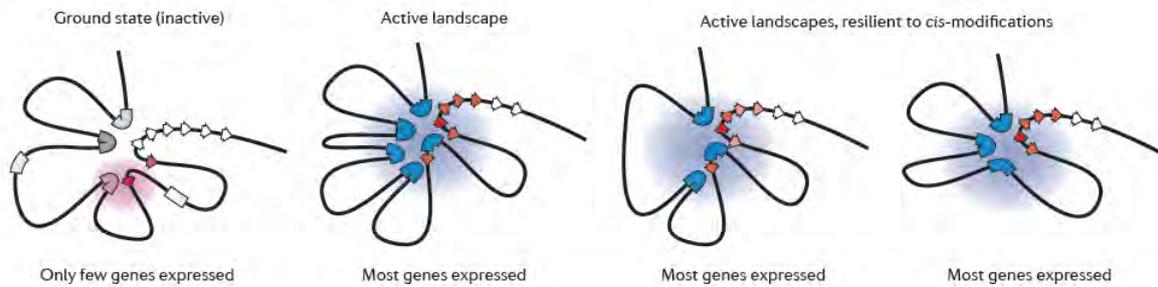


Figure 10. Robustesse de l'archipel de la régulation. L'archipel de la régulation dans le locus *HoxD* chez la souris est un exemple de coopérativité en *cis* entre de multiples éléments enhancers. Ces éléments sont répartis sur >800 kb et forment une structure complexe qui se caractérise par des interactions croisées multiples (qui sont soit simultanées soit dynamiques). Dans la partie distale des membres, de nombreux enhancers deviennent fonctionnelles et interagissent entre eux, ce qui active l'expression de plusieurs gènes dans leur voisinage. L'absence de certains éléments (par le biais de délétions ou d'autres réarrangements chromosomiques) ne peut pas être préjudiciable dans la mesure où d'autres éléments activateurs peuvent compenser. Ainsi, l'archipel semble pouvoir résister aux changements et confère une robustesse à l'expression de *HoxD* à travers un réseau dense d'interactions tridimensionnelles entre les éléments de régulation, dont chacun affiche des activités régulatrices autonomes légèrement différentes. (D'après Spitz et. Furlong, 2012).

✓ INTEGRATION DE L'ACTIVITE DE PLUSIEURS ENHANCERS

Cependant, la manière dont l'activité des différents enhanceurs est intégrée reste encore mal comprise. Sur le plan mécanistique, l'intégration pourrait être réalisée de différentes manières. Dans certains cas, le profil d'expression d'un gène est différent de la somme des activités individuelles de chaque enhanceur. Cela peut s'expliquer par une répression croisée, où les répresseurs se liant à un enhanceur restreignent l'activité de l'autre enhanceur régulant le même gène (Barolo et al., 1997b ; Perry et al., 2011b). Alternativement, l'intégration de l'activité des enhanceurs peut conduire à un effet synergique, comme c'est le cas, par exemple, des gènes *brk* (Yao et al., 2008) et *slp1* (Prazak et al., 2010) chez la drosophile. Chez la souris, plusieurs enhanceurs dirigeant l'expression du gène *HoxD* dans les bourgeons des membres au cours du développement, sont répartis sur une grande région de 800 kb en amont du gène (Spitz et al., 2003 ; Gonzalez et al., 2007 ; Montavon et al., 2011). Des expériences de capture de la conformation de la chromatine (3C) ont révélé une proximité physique omniprésente entre ces différents modules, suggérant que ce qui apparaît comme des «archipels régulateurs» dispersés peuvent se regrouper pour agir sur les promoteurs du gène *HoxD* comme une unité régulatrice unique (Montavon et al., 2011) (Figure 10), conduisant à un profil d'expression qui est plus large que la somme des activités des modules isolés.

✓ EVOLUTION DES SEQUENCES CIS-REGULATRICES

Au cours de l'évolution une grande diversité morphologique a été générée chez les animaux. Cette diversité est liée sans doute à la modulation de l'activité des gènes régulant le développement embryonnaire.

Plusieurs mécanismes génétiques seraient à l'origine de la diversité morphologique, parmi lesquels la modification de la spécificité des protéines. Toutefois les séquences codantes sont souvent soumises à des contraintes considérables, en raison de la nécessité de maintenir la structure tridimensionnelle de la protéine et de conserver sa fonction (Carroll et al., 2001). Un autre processus impliqué dans le changement morphologique serait l'évolution des séquences *cis*-régulatrices à l'origine du changement de l'expression génétique (Carroll et al., 2001; Carroll, 2005; Wittkopp, 2006). En effet, les séquences régulatrices tolèrent considérablement la variation du nombre et de la topologie des sites de liaison, ce qui leur permettent d'évoluer plus facilement que les séquences codantes (Wray et al., 2003).

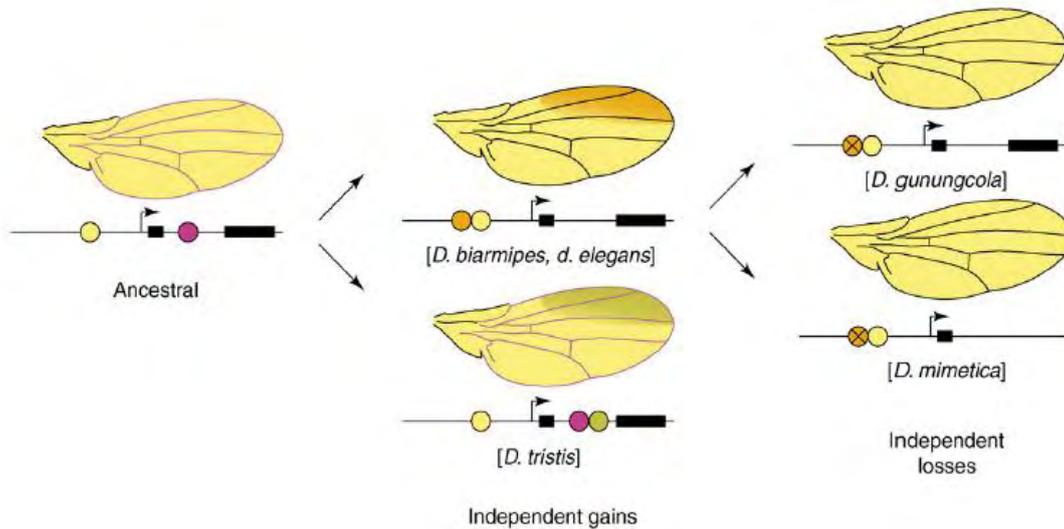


Figure 11. Evolution des éléments régulateurs en cis du gène *yellow* et implication dans la pigmentation chez différentes espèces de drosophiles. L'évolution convergente de la pigmentation en spot chez *Drosophila tristis* implique un élément cis-régulateur du gène *yellow*, « intron spot », qui est différent de celui présent chez *D. biarmipes* et *D. elegans*, l'élément «spot». L'ancêtre commun de *D. biarmipes*, *D. elegans* et *D. tristis* n'avait pas de tâche sur l'aile. L'expression de *yellow* dans la région spot a évolué deux fois: une fois dans la lignée *D. biarmipes*-*D. elegans* et une fois dans la lignée *D. tristis*, à travers deux éléments cis-régulateurs différents préexistants (l'élément « spot », est symbolisé par les cercles jaune et orange, et l'élément «intron-spot» est symbolisé par les cercles rose et vert). Le spot de pigmentation a été perdu indépendamment chez *D. gunungcola* et *D. mimetica* par inactivation de l'élément « spot ». (D'après Prud'homme, et al., 2006)

Plusieurs exemples d'évolution de séquences *cis*-régulatrices ont été rapportées dans la littérature. Je me limiterai ici aux exemples du gène *yellow* dans les ailes et des gènes *bab* dans l'abdomen chez les *Drosophilidae*.

- Evolution de l'expression du gène *yellow* dans les ailes des *Drosophilidae*

Les motifs pigmentaires des ailes de mouches drosophiles représentent un trait morphologique spécifique à l'espèce. Ces motifs dépendent principalement de la disposition des nervures (sources des précurseurs des pigments) et du profil transcriptionnel des gènes *yellow* (*y*) et *ebony* (*e*) (Wittkopp et al., 2003 ; Wittkopp et al., 2002). *y* et *e* codent pour deux enzymes nécessaires, respectivement, à la synthèse des pigments (mélanine) noir (dopa-mélanine) et brun (tan) (N- β -alanyl-dopamine NBAD).

Chez les *Drosophilidae*, la disposition des nervures est largement conservée. Par conséquent, même les espèces dont les ailes sont non mélanisées, telles que *D. melanogaster*, possèdent une prédisposition à la mélanisation. Plusieurs articles récents montrent que la variation de la pigmentation de l'aile entre les espèces est en corrélation avec des changements de la réponse du gène *yellow* (*y*) au réseau de signalisation conservé dans différentes espèces (Gompel et al., 2005; Prud'homme et al., 2006). En effet, il a été montré que l'élément *cis*-régulateur appelé « spot » (de 675 pb), récapitulant la majeure partie de l'expression de *y* dans l'aile de *D. biarmipes* dans la partie antéro-distale de l'aile, est capable de diriger l'expression d'un gène rapporteur chez *D. melanogaster* selon un profil similaire au spot de pigmentation observé chez l'espèce d'origine.

Prud'homme et collaborateurs suggèrent que l'élément « spot » est né par « subfunctionalization » (divergence après duplication, de telle sorte que chaque copie effectue une fonction individuelle), à partir d'un élément ancestral préexistant responsable de l'expression de *y* dans l'aile. Les auteurs suggèrent aussi que la tache pigmentaire chez *D. tristis* une espèce du groupe obscura, a été acquise par un mécanisme indépendant de celui qui a servi à l'apparition du spot de pigmentation dans l'aile de *D. biarmipes* et *D. elegans* et proposent l'élément «spot intron» situé au niveau de l'intron de *y* comme l'élément *cis*-régulateur responsable de cette pigmentation (Figure 11). Enfin, ils proposent que l'élément «spot intron» ait évolué à son tour à partir d'un élément ancestral impliqué dans l'expression de *y* dans les nervures. Ainsi, on peut dire que les deux chemins moléculaires qui ont permis de diriger l'expression de *y* dans la partie antéro-distale de l'aile se sont établis sur la base d'un élément ancestral permettant l'expression dans l'aile.

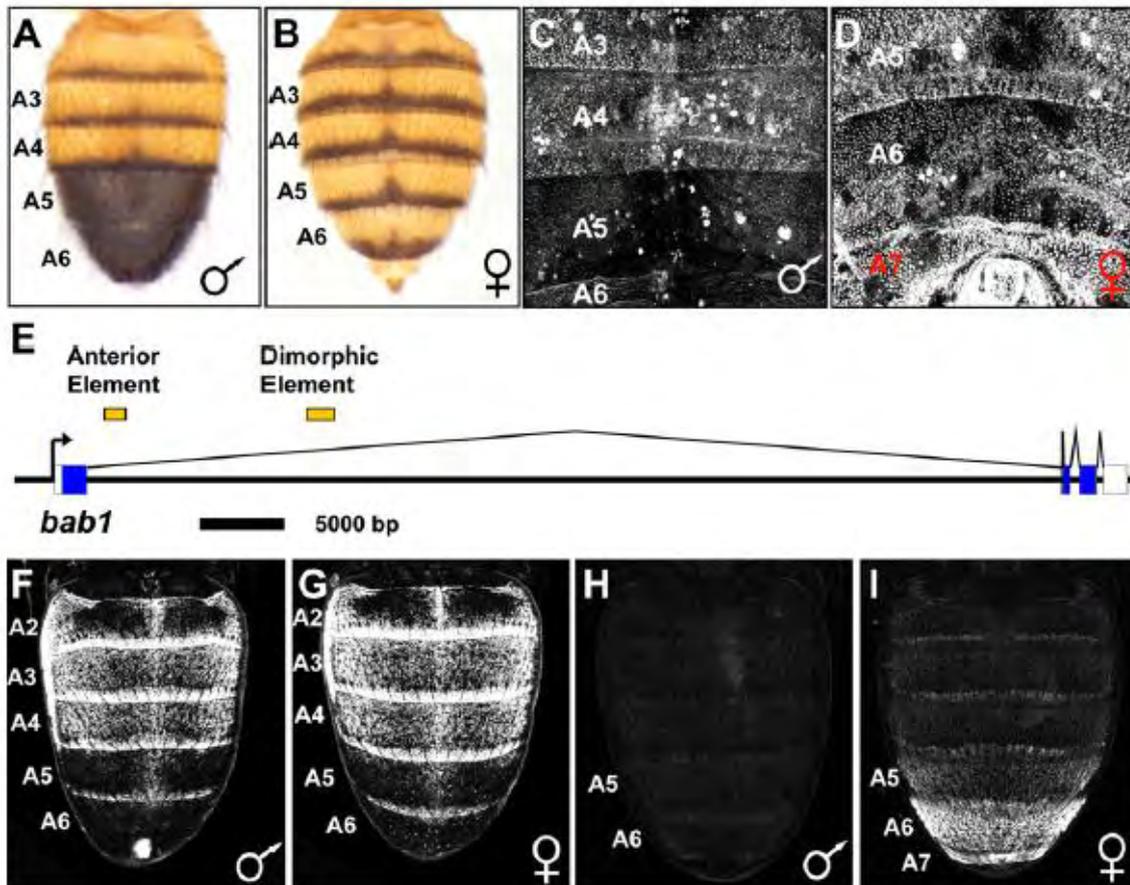


Figure 12. L'expression de *bab1* dans l'abdomen est régulée par deux CREs (éléments *cis*-régulateurs).

(A et B) Vue dorsale d'abdomens de mouches *D. melanogaster* adultes. Les segments A5 et A6 sont entièrement pigmentés chez les mâles (A). Chez les femelles, la pigmentation de ces segments est limitée à une bande postérieure (B).

(C et D) Expression de *bab1* chez des pupes mâles et femelles à 72 h après la formation de la puppe. L'expression de *bab1* chez les mâles est limitée aux segments A2-A4 (C), mais s'étend chez les femelles, aux segments A5 et A6, ainsi que dans le segment femelle spécifique A7 (D).

(E) les deux CREs, l'élément antérieur et l'élément « dimorphique », se trouvent dans le 1^{er} intron de *bab1* et régissent l'expression de Bab1 dans l'épiderme abdominal.

(F-I) expression du gène rapporteur GFP dans l'abdomen dorsal de pupes.

(F et G) L'élément antérieur dirige l'expression du gène rapporteur GFP dans les segments A2-A4 des mâles (F) et des femelles (G).

(H) L'élément dimorphique est inactif chez les mâles.

(I) L'élément dimorphique dirige l'expression du gène rapporteur dans les segments femelles A5-A7, avec des niveaux croissants de l'avant vers l'arrière. (D'après Williams et al., 2008).

- Evolution de l'expression des gènes *bab* dans l'abdomen des *Drosophilidae*

Pour comprendre le développement et l'évolution du dimorphisme sexuel l'équipe de Sean Carroll a étudié en 2008 l'évolution de la pigmentation mâle spécifique chez *Drosophila melanogaster* (Williams et al., 2008). Les mâles de cette espèce ont typiquement une pigmentation dorsale au niveau des plaques cuticulaires (tergites) sur les segments abdominaux A5 et A6 (Figure 12). En revanche, chez les femelles la pigmentation des tergites A5 et A6, est limitée à une bande postérieure, similaire à celle présente sur les tergites des segments abdominaux A2-A4 des deux sexes. Les gènes paralogues *bab1* et *bab2*, du locus *bab*, codant pour des protéines liant l'ADN (Lours et al., 2003) ont été identifiés comme des répresseurs dominants de la pigmentation postérieure (Couderc et al., 2002; Kopp et al., 2000). Tandis que les pupes femelles expriment les gènes *bab* au niveau des segments abdominaux A2-A6, l'expression des gènes *bab* chez les pupes mâles est limitée aux niveaux des segments A2-A4, ce qui permet d'avoir une forte pigmentation aux niveaux des segments A5 et A6 (Figure 12) (Kopp et al., 2000). Les analyses génétiques ont montré que le gène *Hox Abd-B* agit comme un répresseur des gènes *bab* dans les segments postérieurs et suggèrent que la répression des gènes *bab* est atténuée chez les femelles par l'activité du facteur de transcription DsxF, codé par un transcrite du locus sexe spécifique *doublesex (dsx)* (Kopp et al., 2000). Afin de comprendre les mécanismes moléculaires par lesquels l'expression des gènes *bab* est régulée et a évolué, l'équipe de Sean Carroll a identifié les CREs (*Cis-Regulatory Elements*) qui régissent l'expression des gènes *bab*. Williams et ses collaborateurs (Williams et al., 2008) ont ainsi identifié deux CREs qui gouvernent l'expression de *bab1* dans l'abdomen. Le premier élément appelé « antérieur » régule l'expression de *bab1* dans les segments A2-A4 des deux sexes, alors que le second élément appelé « dimorphique » régule son expression dans les segments postérieurs A5-A7 des femelles. Il est suggéré que l'apparition de la pigmentation dimorphique spécifique du mâle chez l'espèce *D. melanogaster* résulte d'une évolution au sein du sous-genre *Sophophora* dont l'ancêtre commun avait une expression monomorphique des gènes *bab* (Gompel et Carroll, 2003; Kopp et al, 2000) et une pigmentation abdominale monomorphique (Jeong et al, 2006.). Afin de suivre l'évolution de l'expression dimorphique de *bab1*, l'équipe de Sean Carroll a comparé les CREs des gènes *bab* impliqués dans la pigmentation abdominale aux séquences orthologues chez un membre monomorphe, extérieur au groupe *Melanogaster*, *D. willistoni*.

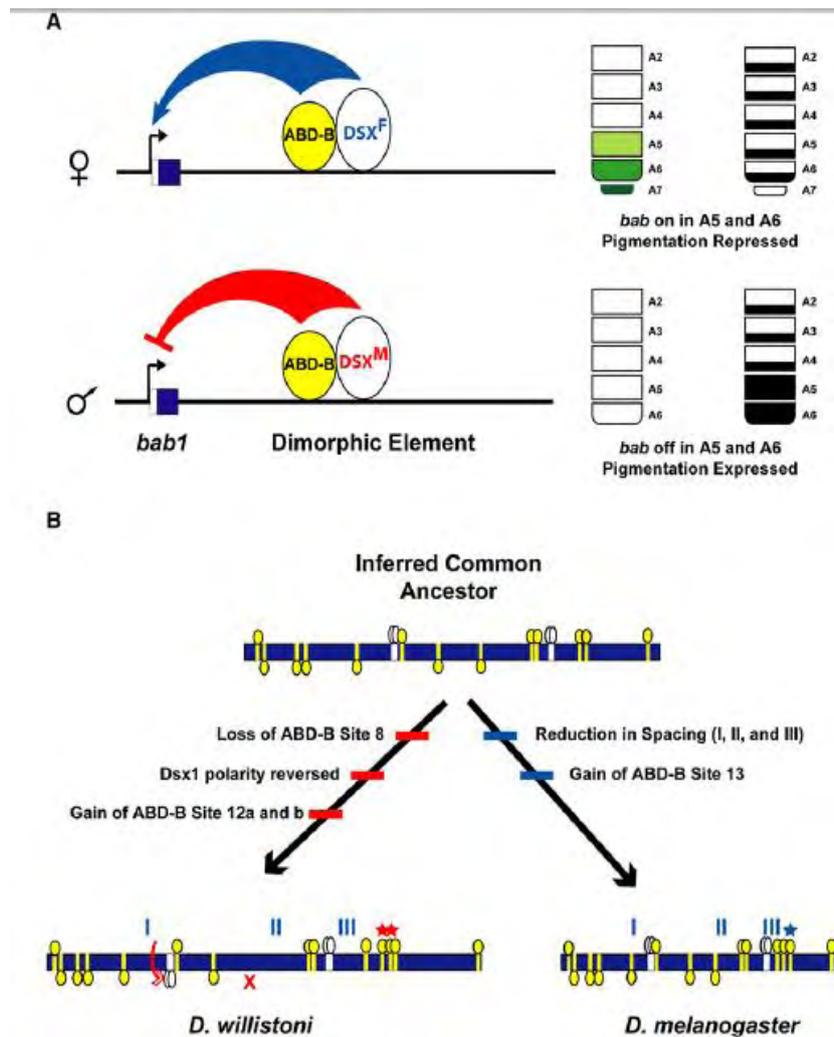


Figure 13. Modèle pour le fonctionnement et l'évolution du « Switch » génétique dimorphique. (A) L'expression de *bab1* dans les segments postérieurs abdominaux A5-A7 des femelles est médiée par les activités combinées de la protéine Hox Abd-B et de l'isoforme femelle spécifique Dsx^F. L'expression de *bab1* résulte en la répression de la pigmentation dans tous les tergites dans ces segments. Au niveau des segments A5 et A6 chez le mâle, elle est réprimée par Dsx^M, l'isoforme spécifique mâle. L'absence d'expression de *bab1* dans ces segments permet aux tergites d'être entièrement pigmentés. (B) Représentation schématique de l'évolution de l'élément dimorphique de l'ancêtre commun déduit de *D. melanogaster* et *D. willistoni*. Les cases jaunes indiquent les sites de liaison pour Abd-B et les cases blanches indiquent les sites de liaison des protéines Dsx. Les ovales jaunes et blancs représentent Abd-B et des monomères de protéines Dsx, respectivement. Le CRE commun ancestral contenait deux orthologues et treize sites de liaison pour Dsx et Abd-B, respectivement. Dans la lignée conduisant à *D. willistoni*, 8 sites de liaison à Abd-B ont été perdus, la polarité du site Dsx1 a été inversée (flèche rouge) et des sites de fixation de Abd-B 12a et 12b ont été acquis (étoiles rouges). Dans la lignée conduisant à *D. melanogaster*, l'espacement des sites de liaison a été réduit dans les régions I, II, et III, et le site Abd-B 13 a été acquis (étoile bleue). (D'après Williams et al., 2008).

Cette espèce, exprime *bab1* et *bab2* selon un schéma monomorphe dans les segments A5 et A6 chez les mâles et les femelles. La pigmentation des segments A5 et A6 chez les mâles de l'espèce *D. willistoni* est limitée comme celle des femelles (Kopp et al, 2000) à une bande postérieure. L'analyse de la région orthologue à l'élément « antérieur » chez *D. willistoni* a montré qu'elle est capable d'induire l'expression d'un gène rapporteur GFP depuis le segment A2 et jusqu'aux segments les plus postérieurs dans les deux sexes et ceci selon le même patron que le gène endogène *bab1*. Ce qui suggère que l'évolution de la pigmentation dimorphique chez *D. melanogaster* a impliqué l'évolution de la répression à la fois de *bab1* et de *bab2* dans les segments postérieurs des mâles. Ainsi, lors de l'évolution de l'espèce *D. melanogaster* la fonction de l'élément « antérieur » a été modifiée de telle sorte qu'il n'induit plus l'expression des gènes *bab* dans les segments A5 et A6 des deux sexes. L'équipe de Sean Carroll a identifié également au cours de cette étude la séquence orthologue de l'élément « dimorphique » chez *D. willistoni*. Ce dernier induit l'expression d'un gène rapporteur au niveau du segment A7 femelle chez *D. melanogaster*, mais pas dans les segments A5 et A6. Ces résultats suggèrent que l'élément actif « dimorphique » existait dans un ancêtre commun à *D. willistoni* et *D. melanogaster*. Pour tester cette idée, les auteurs ont également isolé un élément dimorphique orthologue d'une espèce éloignée, *D. virilis*, dont l'activité est également limitée au segment femelle A7. Ces résultats suggèrent que l'élément « dimorphique » de *D. melanogaster* aurait une ascendance lointaine. Par conséquent, le profil d'expression des gènes *bab* dans les segments A6 et A5 femelles de *D. melanogaster* représente une expansion de l'activité de l'élément « dimorphique ». Ainsi on peut conclure que l'évolution du dimorphisme de pigmentation a impliqué la modification fonctionnelle des deux CREs de *bab* abdominaux ancestraux.

Enfin, Les auteurs ont démontré que l'élément « dimorphique » fait partie d'un « switch » génétique qui, en combinaison avec la protéine Hox Abd-B et les activités sexe-spécifique des isoformes mâle et femelle de la protéine Dsx, dirige l'activation de manière spécifique chez la femelle et la répression spécifique des gènes *bab* chez les mâles dans les segments postérieurs (Figure 13).

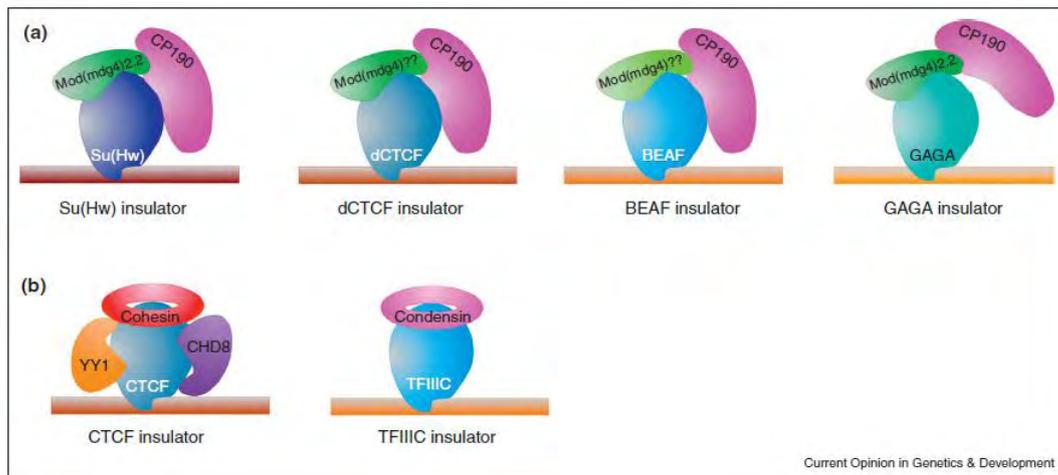


Figure 14. Schéma montrant la structure de différents insulateurs de la drosophile et de vertébrés. (a) Chaque sous-classe d’insulateurs drosophile contient une protéine de liaison différente qui peut définir la fonction spécifique de la sous-classe correspondante. Tous les insulateurs ont en commun la protéine CP190 et une isoforme de Mod(mdg4). L’insulateur gypsy/Su(Hw) contient Mod(mdg4) 2.2. Les insulateurs dCTCF et BEAF n’ont pas cette isoforme, mais contiennent une variante différent de Mod(mdg4). (b) Structure des insulateurs vertébrés CTCF et TFIIC. (D’après Yang et Corces, 2012).

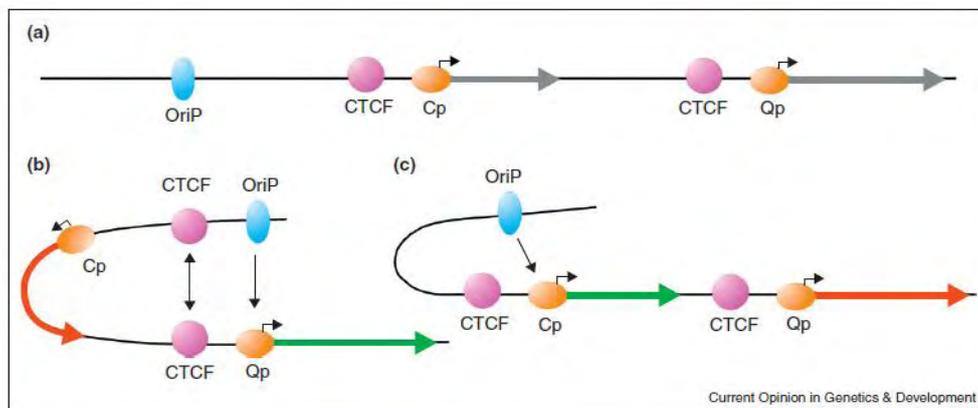


Figure 15. CTCF régule les interactions entre enhancer-promoteur pour contrôler le type de latence des virus Epstein-Barr. (a) Organisation linéaire du génome d’EBV. (b) Modèle montrant la formation d’une boucle médiée par CTCF et qui amène l’enhancer OriP en contact avec le promoteur de type I Qp dans les cellules. (c) Dans les cellules de type III, l’activateur OriP interagit directement avec le promoteur Cp. (D’après Yang et Corces, 2012).

2.1.2. LES INSULATEURS

Les insulateurs sont des complexes ADN-protéines qui peuvent médier des interactions en *cis* ou en *trans* entre les différentes régions du génome. Ils ont été identifiés initialement chez la drosophile. Depuis, plusieurs types d'insulateurs, partageant tous les protéines Mod (mdg4) et CP190, ont été caractérisés chez cette espèce (Figure 14 ; Bushey et al., 2009). Par la suite, les insulateurs ont été retrouvés chez la plupart des eucaryotes, allant des levures à l'homme (Phillips et al., 2009). Les insulateurs de *S. cerevisiae* et de *S. pombe* sont essentiellement limités aux séquences promotrices de la PolIII contenant des sites de liaison pour TFIIC (Valenzuela et al., 2009 ; Noma et al., 2006 ; Iwasaki et al., 2010). Chez les vertébrés, le facteur insulateur le plus largement étudié est CTCF, qui exige la cohésine pour son fonctionnement (Figure 14 ; Wendt et al., 2008 ; Parelho et al., 2008 ; Rubio et al., 2008). Les insulateurs ont été définis à l'origine sur la base de leur aptitude à bloquer la communication entre enhancer et promoteur ou encore leur capacité à servir de barrières contre la propagation de l'hétérochromatine dans des systèmes rapporteurs. Cependant, les résultats accumulés lors des deux dernières années suggèrent que les insulateurs participent à de nombreux processus nucléaires (Yang et Corces, 2012). En effet, en plus du blocage des interactions entre enhancers et promoteurs, les insulateurs peuvent également diriger les enhancers aux promoteurs appropriés. De plus, les insulateurs peuvent non seulement bloquer la propagation de l'hétérochromatine, mais également délimiter les frontières entre des variétés d'états épigénétiques. En outre, l'effet d'insulateurs sur la biologie du génome va au-delà de leur implication dans les processus de transcription. En effet, ils sont également impliqués dans la régulation de la recombinaison V (D) J (Yang et Corces, 2012).

✓ REGULATION DE L'INTERACTION ENHANCER PROMOTEUR

Les enhancers ont la capacité d'agir sur de longues distances pour activer l'expression d'un gène donné. Cette propriété donne à l'enhancer une grande flexibilité fonctionnelle, étant capable d'activer les promoteurs de gènes non adjacents sur des distances allant jusqu'à 1 Mb. Cependant, les mêmes propriétés qui sont au cœur de la fonction enhancer sont également un problème potentiel. Comment font les enhancers pour trouver leurs promoteurs cibles et comment leur activité se limite t'elle à un gène spécifique?

La capacité d'un insulateur à rapprocher des séquences éloignées les unes des autres dans le génome linéaire sous-entend un rôle dans le ciblage des enhancers aux promoteurs spécifiques. Un exemple intéressant est l'utilisation de CTCF pour la sélection de promoteur

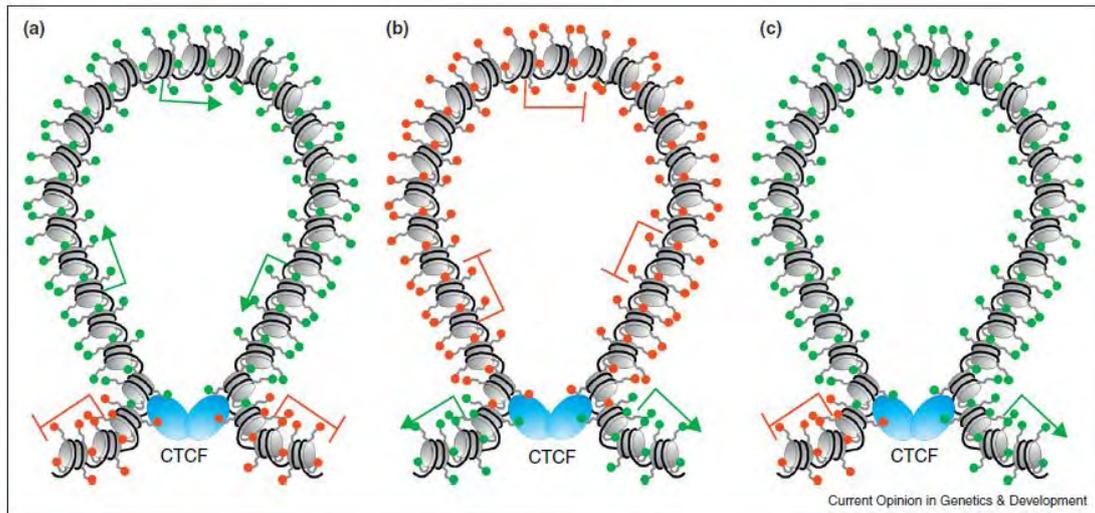


Figure 16. Structure de certains domaines créés par l'interaction entre les insulateurs CTCF dans des cellules souches embryonnaires de souris. Les gènes qui sont activement transcrits sont représentés par une flèche verte et les gènes réduits au silence par une ligne rouge; les nucléosomes et les queues d'histones sont représentés en gris. Les modifications activatrices des histones sont indiquées par des sphères vertes et les modifications répressives par des sphères rouges. L'ADN est représenté en noir et CTCF sous forme d'ovales bleus. (a) CTCF forme une boucle pour séparer un domaine contenant des modifications des histones activatrices et des gènes transcrits, des marques répressives et des gènes silencieux. (b) CTCF forme une boucle pour séparer un domaine contenant des modifications répressives des histones et des gènes silencieux, des marques activatrices et des gènes transcrits. (c) CTCF forme une boucle contenant des nucléosomes enrichis en H3K4 et H3K4 mono, diméthylée et triméthylée aux extrémités de la boucle, tandis que les modifications activatrices de la transcription, H3K36me3, et répressives, H3K27me3, sont observées à l'extérieur des boucles sur les côtés opposés. (D'après Yang et Corces, 2012).

et le contrôle du cycle de latence du virus d'Epstein-Barr (Figure 15). Dans le génome de ce virus, l'enhancer OriP est partagé par deux promoteurs en aval, Cp et Qp. Cp détermine l'expression des gènes de type III pour le cycle de latence III tandis que Qp détermine le programme d'expression des gènes de type I pour le cycle de latence I (Tempera et al., 2010 ; Chau et al., 2006 ; Tempera et al., 2011). CTCF se lie à la fois en amont de Cp et Qp. Toutefois par défaut OriP sélectionne et active le promoteur à proximité : Cp. Cette activation est indépendante de CTCF (Tempera et al., 2010 ; Chau et al., 2006 ; Tempera et al., 2011). Les interactions entre OriP et le promoteur Qp dépendent par contre de CTCF. En effet, des mutations dans l'un des sites de liaison CTCF perturbent l'interaction entre OriP et Qp et conduisent à la réactivation de transcription Cp. Ainsi, la structure 3D de la chromatine contribue à la sélection du promoteur, probablement en augmentant la capacité de Qp à concurrencer Cp (Tempera et al., 2011). Une étude réalisée sur le génome entier de cellules ES de souris (HandoKo et al 2011) confirme le rôle de CTCF. En effet, ce dernier est impliqué dans respectivement 1480 et 336 interactions intra- et inter-chromosomiques. En utilisant p300 comme marqueur de l'emplacement des séquences enhancers, il s'est avéré que CTCF forme des boucles qui permettent d'amener les enhancers à proximité immédiate des promoteurs qu'elle régule dans les cellules ES de souris.

✓ MISE EN PLACE DE DOMAINES CHROMATINIENS

Les boucles formées par la liaison entre les sites insulateurs ne déterminent pas seulement les interactions entre les séquences régulatrices mais peuvent également définir différents domaines chromatinien caractérisés par la présence de modifications spécifiques au niveau des histones. Une étude de la distribution de CTCF à l'échelle du génome a mis en évidence un sous-ensemble de sites CTCF localisés au niveau des frontières entre les domaines de la chromatine active (activatrice) et répressive, marqués respectivement par l'histone H2A acétylée en Lys5 (H2AK5Ac) et l'histone H3 triméthylée en Lys27 (H3K27me3) (Cuddapah et al., 2009). Etant donné que ces régions sont différentes entre les cellules HeLa et les cellules T CD4+, les auteurs ont suggéré que la délimitation des frontières entre les domaines par CTCF permet de promouvoir une régulation génétique de manière spécifique d'une lignée.

L'alignement des boucles médiées par CTCF avec l'emplacement des modifications des histones a révélé cinq classes de boucles différentes, dont certaines délimitent des

domaines chromatinien ayant des caractéristiques épigénétiques différentes (Figure 16). Les boucles de catégorie I sont enrichies en modifications actives des histones comme H3K4me1, les H3K4me2 et H3K36me3, tandis que les marques répressives H3K9me3, H3K20me3 et H3K27me3 sont absentes à l'intérieur mais présentes à l'extérieur des boucles. Les boucles de catégorie II montrent une distribution opposée des modifications des histones, avec H3K9me3, H3K20me3 et H3K27me3 à l'intérieur des boucles, indiquant la formation de domaines de la chromatine répressive délimitée par la présence de sites CTCF. Les boucles de catégorie III contiennent des nucléosomes H3K4 enrichis en mono-, di et triméthylation à la frontière des boucles, tandis que des modifications activatrices (H3K36me3) et répressives (H3K27me3) sont observées en dehors des boucles sur des côtés opposés (Cuddapah et al., 2009). Comme décrit précédemment, les insulateurs établissent des interactions intra et inter-chromosomique, afin d'obtenir une réponse fonctionnelle spécifique. De ce fait, il est possible que la structure tridimensionnelle du génome détermine sa fonction. Dans ce cas, il est probable que les insulateurs façonnent la structure tridimensionnelle de la fibre chromatinienne de manière cellule spécifique, permettant ainsi une expression différentielle des gènes au cours du développement. La structure 3D médiée par les insulateurs pourrait porter l'information épigénétique qui devra être maintenue lors de la mitose. L'examen de ces hypothèses demeure l'un des principaux défis des recherches futures dans le domaine.

2.2. ÉLÉMENTS RÉGULATEURS EN *TRANS*

La transcription des gènes dépend principalement de la fixation des facteurs de transcription spécifiques (FTs) au niveau des séquences *cis*-régulatrices. Ces facteurs appelés aussi *trans*-régulateurs sont des protéines présentant au moins deux domaines : un domaine de fixation à l'ADN et un domaine effecteur qui peut être activateur ou répresseur de la transcription.

Plusieurs familles de facteurs de transcription ont été définies sur la base de la structure des domaines de liaison à l'ADN. Les plus importantes étant les familles de protéines à doigts de zinc, Hélice-Tour-Hélice et Hélice-Boucle-Hélice. Ces derniers lient le génome via des domaines de liaison à l'ADN conservés appelés respectivement : domaines à doigts de zinc, homéodomains et hélice-boucle-hélice (Vaquerizas et al., 2009). Chacune de ces classes de FTs contient plusieurs membres, par exemple, chez l'homme, il y a plus de 650 protéines à doigts de zinc, ~ 250 protéines à homéodomaine et ~ 80 protéines hélice-boucle-hélice (Vaquerizas et al., 2009).

2.2.1. ACTIVATION DE LA TRANSCRIPTION

Les FTs peuvent activer la transcription d'un gène en interagissant avec la machinerie transcriptionnelle basale ou encore en permettant la liaison coopérative d'autres FTs via des interactions directes protéine-protéine ou encore indirectes en recrutant directement ou indirectement des enzymes modifiant la structure chromatinienne.

✓ INTERACTION AVEC LA MACHINERIE TRANSCRIPTIONNELLE BASALE

Les FTs jouent un rôle important dans l'initiation de la transcription en conditionnant l'interaction entre les composants de la machinerie transcriptionnelle et l'ADN (Thomas et Chiang 2006). L'interaction des FTs avec la machinerie transcriptionnelle concerne particulièrement les composantes du complexe de pré-initiation (PIC) et s'effectue via leurs domaines effecteurs appelés *trans*-activateurs (TADs). Par exemple le TAD de E2F1, établit des contacts directs avec les facteurs de transcription généraux, y compris TATA-binding protein (TBP), et les « facteurs associés à TBP » (TAFs), TFIIA, TFIIB et TFIID suite à sa liaison à des sites situés à la fois à proximité et loin des promoteurs cœurs (Cujec et al., 1997 ; Zhu et al., 1994). Les TADs peuvent également recruter et interagir avec les composantes du complexe Médiateur, un complexe multiprotéique impliqué dans l'activation d'un grand nombre de gènes (Lee et Young 2000).

Les domaines TADs eucaryotes sont généralement classés en fonction de leur composition en acides aminés. Les TAD peuvent être riches en résidus d'acides aminés acides (par exemple, E2F1 et p53), en résidus de glutamine (par exemple Oct1, Oct2, Sp1) ou en résidus de proline (par exemple Ap-2 et CTF/NF1). De façon plus appropriée, les TAD eucaryotes ont été regroupés sur la base de leur fonction en deux groupes : ceux qui stimulent l'initiation versus ceux qui stimulent l'élongation (Blau et al., 1996). Des études antérieures suggèrent que de nombreux activateurs agissent principalement au niveau de l'initiation de la transcription. Toutefois, le contact entre un TAD et la machinerie transcriptionnelle générale peut stimuler la transcription non seulement en stabilisant le complexe PIC, mais également en améliorant l'élongation (Fuda et al., 2009, Rahl et al., 2010 ; Selth et al., 2010). Par exemple, en plus de stimuler l'initiation de la transcription, le domaine TAD de c-Myc favorise également l'élongation de la transcription par le recrutement de P-TEFb (positive transcription elongation factor b, qui est composé de CycT1 et CDK9). D'une manière intéressante, le TAD de c-Myc augmente également la maturation de l'ARNm, le recrutement de polysomes et le taux de traduction. Ces augmentations résultent de la phosphorylation du

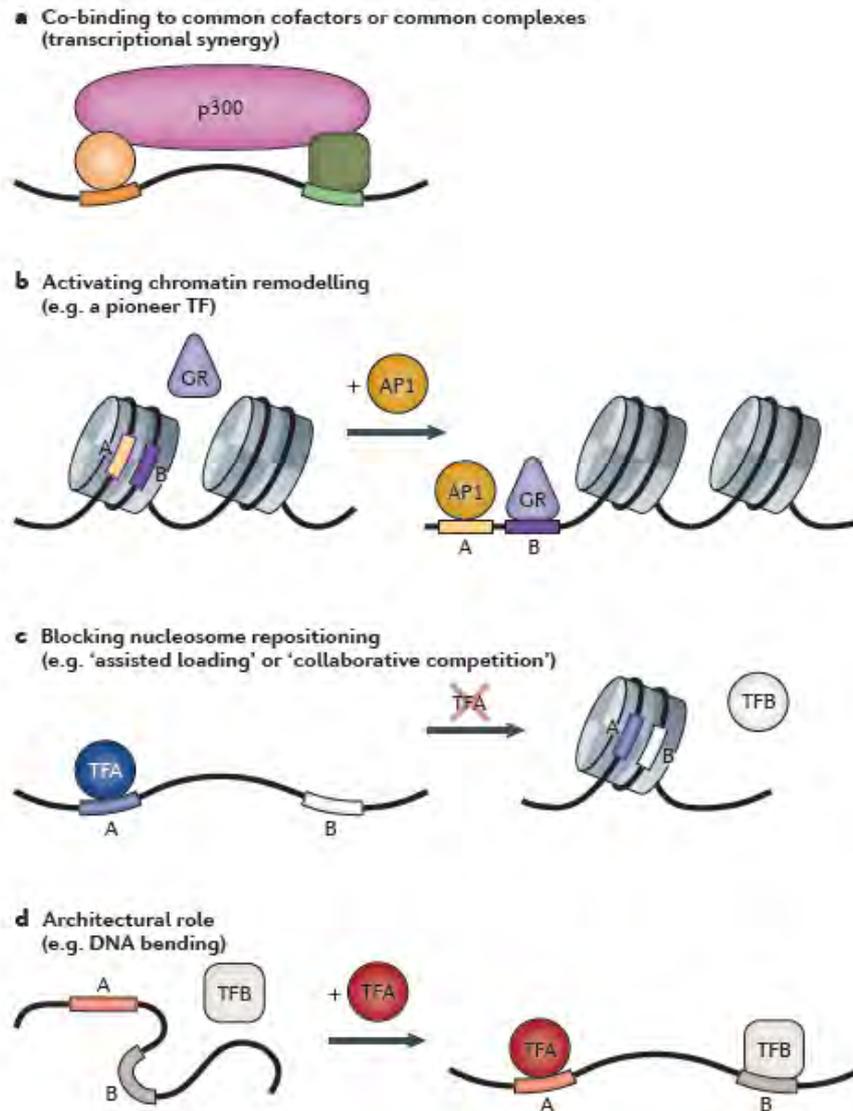


Figure 17. Les mécanismes indirects de coopérativité entre les facteurs de transcription. Il existe de nombreux mécanismes indirects par lesquels les facteurs de transcription (FTs) peuvent agir coopérativement pour réguler les enhanceurs. a) Deux ou plusieurs FTs lient ensemble un même élément enhanceur et recrutent un cofacteur commun (par exemple, p300), ou encore différents composants d'un complexe multiprotéique (par exemple, les complexes Médiateur ou SAGA), ce qui peut conduire à une augmentation nette de l'affinité de chaque FT pour leur site de liaison (ou une augmentation du temps de fixation des facteurs de transcription à l'enhancer). b) Certains facteurs de transcription peuvent agir en coopération, en activant le remodelage de la chromatine. Après liaison du FT AP1 au site 1, le nucléosome est activement repositionné, ce qui expose le site de liaison GRE (site 2) et permet au facteur GR de se lier. AP1 agit ainsi comme un FT pionnier et assiste la fixation de GR, sans qu'il y ait pour autant d'interaction directe entre les deux FTs. c) En restant lié à un site donné, une FT (FT A sur la figure) peut empêcher le repositionnement d'un nucléosome et peut donc faciliter la liaison d'un autre facteur (FT B sur la figure) à un site voisin qui, autrement, pourrait devenir inaccessible. Ceci peut constituer une méthode passive d'amorçage d'enhancer. d) Certains facteurs de transcription (par exemple, HMG1) peuvent induire une flexion locale de l'ADN, ce qui peut augmenter l'affinité des autres FTs pour les sites de l'enhancer, ou peut favoriser le recrutement et la formation du complexe. (D'après Spitz et. Furlong, 2012).

CTD de la PolII par c-Myc, via la stimulation de Cdk9 et la phosphorylation des résidus Ser2 des heptades.

✓ COOPERATIVITE ENTRE LES FTS.

La liaison coopérative à l'ADN est souvent considérée comme le résultat d'interactions directes protéine-protéine entre FTs qui se lient à des sites adjacents sur l'ADN, augmentant ainsi leur affinité respective au site de liaison (Johnson et al., 1981 ; Oehler et al., 2006). Cependant, d'autres modes plus indirects de coopérativité entre FTs existent. La coopérativité transcriptionnelle entre la protéine Gal4 de levure et le récepteur aux glucocorticoïdes de rat (GR) (Lin et al., 1990) est une manifestation importante de cette coopérativité alternative. En effet, ces deux facteurs proviennent de deux organismes différents et par conséquent il est improbable que leur coopérativité passe par des interactions directes protéine-protéine, mais plutôt via des interactions avec les composants communs conservés de la machinerie transcriptionnelle. Une telle synergie transcriptionnelle est probablement fréquente étant donné que de nombreux facteurs de transcription interagissent avec des co-activateurs communs (tels que la famille p300-CREB-binding protein (CBP) (Merika et al., 1998) (Figure 17) ou des co-répresseurs (comme Groucho), ainsi qu'avec des membres des complexes Médiateur, SAGA et TFIID.

La coopérativité transcriptionnelle passe également par le recrutement de co-activateurs distincts, qui à leur tour fonctionnent de manière synergique pour activer l'expression des gènes. Par exemple, l'activateur A pourrait recruter CBP, qui assure l'acétylation des histones, alors que l'activateur B pourrait recruter SWI/SNF, qui remodèle la chromatine par déplacement des nucléosomes (Malik et Roeder 2005 ; Plashne et Gann 2001). Un autre mode de liaison coopérative a été décrit, selon lequel un facteur de transcription dit « pionnier » se lie à des enhanceurs inaccessibles à d'autres facteurs et peut recruter des complexes de remodelage de la chromatine qui conduisent à un repositionnement des nucléosomes pour faciliter la liaison d'autres FTs au niveau de sites voisins (Schwabish et Stuhl 2007 ; Shuey et Parker 1986). Par exemple, la liaison du facteur de transcription AP1 facilite le chargement des GR par le recrutement de BRG1 (un homologue de Brahma), qui maintient la chromatine dans un état ouvert (Biddie et al., 2011) (Figure 17b). La simple fixation d'un FT pionnier sur un enhanceur est généralement insuffisante à la formation d'un complexe d'activation (Ghisletti et al., 2010). En revanche, cette fixation joue un rôle essentiel en induisant une série de modifications qui vont permettre par la suite à l'enhancer de recruter

des FTs complémentaires et déclencher ainsi l'activation transcriptionnelle à un stade ultérieur du développement. La liaison d'un FT peut également déclencher une flexion locale de l'ADN, favorisant ainsi la liaison d'un FT voisin sur l'enhancer (Panne et al., 2007; Falvo et al., 1995) (Figure 17d). Ces mécanismes peuvent contribuer à transformer les interactions de faibles affinités avec l'ADN en des liaisons robustes. Ces mécanismes suggèrent qu'une proportion importante des coopérativités fonctionnelles qui sont détectées entre les FTs liant le même enhancer, peuvent se produire par des effets indirects plutôt que par la formation de complexes stables. Enfin la coopérativité entre deux FTs au niveau d'un enhancer donné peut se produire également par compétition sur le même site de fixation (Figure 17c). Contrairement à ce à quoi on peut s'attendre, la compétition entre les FTs ne résulte pas en l'extinction de l'activité de l'un ou de l'autre mais au contraire à une augmentation du temps de liaison de chaque FTs au site de fixation. Voss et col (2011) explique ce phénomène par le fait que l'échange rapide des FTs empêche le repositionnement des nucléosomes.

La plupart de ces mécanismes de synergie activatrice décrits ci-dessus ne sont pas exclusifs. Il est donc possible, par exemple, que deux activateurs se lient de façon coopérative à des sites et recrutent de manière coordonnée un ou plusieurs co-activateurs.

2.2.2. RÉPRESSION DE LA TRANSCRIPTION

Trois mécanismes principaux de répression de la transcription chez les eucaryotes ont été décrits : (i) interaction directe entre les FTs répresseurs et la machinerie de transcription basale; (ii) compétition entre FTs activateurs et répresseurs sur des sites de fixation communs ou chevauchants; (iii) et enfin recrutement de cofacteurs. Concernant le premier mécanisme il est très peu documenté. Néanmoins, la plupart de ces études se basent sur des mécanismes définis *in vitro*, ce qui laisse douter de leur pertinence *in vivo* (Austin et Biggin, 1995). Le mécanisme de répression par compétition quant à lui, a fait ces preuves à plusieurs reprises. Toutefois, la généralisation de ce modèle est incertaine. En effet, la plupart des sites de liaison des répresseurs ne chevauchent pas directement les sites activateurs (Arnosti et al., 1996). Il semble donc que l'activité des facteurs répresseurs s'appuie plutôt sur les interactions médiées par des corépresseurs. Ces derniers peuvent agir seuls ou en médiant l'assemblage d'un complexe formé par plusieurs sous-unités.

Les complexes corépresseurs peuvent présenter une large gamme d'activités, y compris les activités de types histone-désacétylases (HDAC), méthyl transférase ou histone déméthylase. Un autre groupe de cofacteurs conservés sont les complexes de remodelage des nucléosomes SWI / SNF, qui servent généralement à l'activation, mais dans certains cas, sont

recrutés par des facteurs de transcription pour induire la répression des gènes et peuvent donc être considérés comme un type spécial de corépresseurs. Le complexe Mi2-NURD corépresseur combine une activité de remodelage du nucléosome et une activité histone déacétylase (Burke et Baniahmad, 2000 ; Perissi et al., 2010). En plus de la modification des nucléosomes, il a été montré que les corépresseurs interfèrent directement avec les co-activateurs tels que l'histone acétyltransférase CBP/p300 ou le complexe Médiateur (Gromoller et Lehming, 2000 ; Kim et al., 2005). Ces derniers exemples semblent être des exceptions visant des cibles non histones, mais en général, les enzymes de modification des histones sont capables de modifier les FTs avec les mêmes modifications post-transcriptionnelles que celles qui s'exercent au niveau des queues d'histones.

L'un des complexes corépresseurs le mieux étudié est le complexe multiprotéique Sin3, un corépresseur conservé chez la levure, les plantes et les animaux. La protéine Sin3 s'associe avec les protéines histones désacétylases HDAC1 et HDAC2 (Silverstein et Ekwall, 2005) et la désacétylation semble être le mécanisme clé de sa fonction répressive quand il est recruté par les FTs tels que les récepteurs nucléaires aux hormones, p53 et Elk1. Le complexe Sin3 peut être recruté directement par des facteurs de transcription spécifiques ou bien indirectement par l'interaction avec d'autres complexes corépresseurs. Par exemple, NCoR et SMRT, qui sont à leur tour recrutés par des FTs de type récepteurs hormonaux nucléaires (Heinzel et al., 1997 ; Nagy et al., 1997). Outre la désacétylation, le complexe Sin3 peut être également associé à d'autres activités enzymatiques, y compris de remodelage de la chromatine par SWI / SNF, des protéines glycosylases, des histones méthylases et des ADN méthylases.

Un autre facteur largement étudié est le corépresseur Groucho, un membre de la famille WD40 (Gasperowicz et Otto, 2005; Liu et Karmarkar, 2008), qui forme des complexes avec des histones désacétylases. Recruté par les FTs Hairy chez la drosophile et HES chez les vertébrés, le corépresseur Groucho est un acteur important dans la voie de signalisation Notch et dans la neurogenèse (Liu et Karmarkar, 2008). La protéine 'C-terminal binding protein' (CtBP) est également connue pour son rôle de corépresseur. La caractérisation biochimique du complexe associé à CtBP a identifié une série d'enzymes modifiant la chromatine, y compris des histones désacétylases, histones méthyltransférases et histones déméthylases (Shi et al., 2003). CtBP module des processus cellulaires importants, y compris la différenciation, l'apoptose et la tumorigenèse (Chinnadurai, 2003).

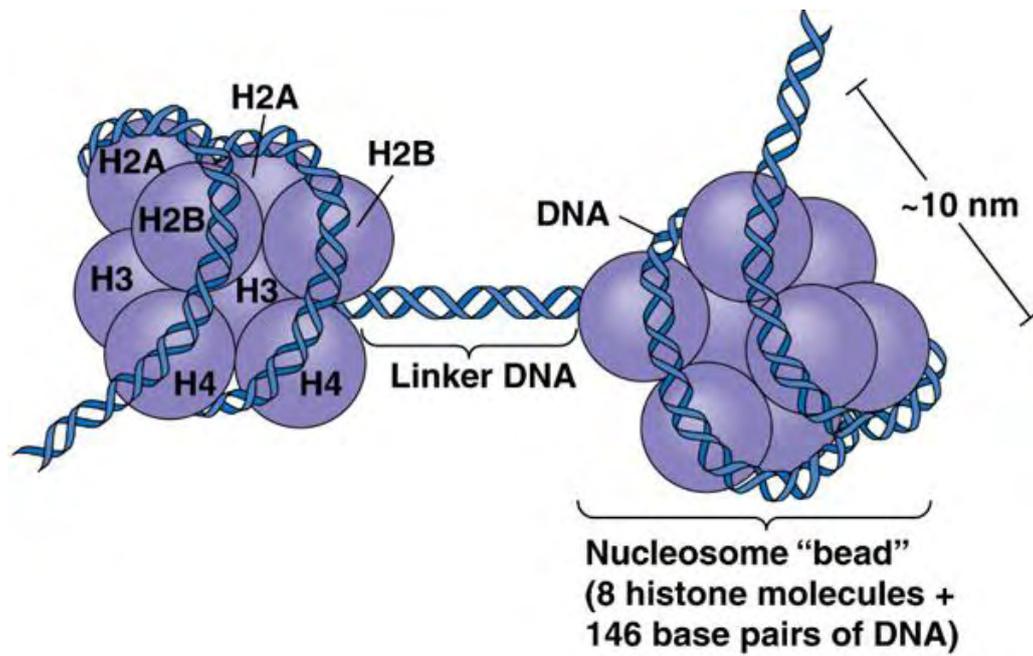


Figure 18. Structure du nucléosome. Chaque nucléosome comporte deux histones H3, H4, H2A et H2B autour desquelles s'enroule 146 pb d'ADN

2.2.3. MODULATION DE LA SPECIFICITE DE SEQUENCE RECONNUE PAR LES FTS PAR DES COFACTEURS.

La liaison coopérative de deux FTs résulte en une augmentation nette de l'affinité des deux facteurs pour leurs motifs, alors que leur propre spécificité de séquence reste la même. Deux études récentes font état d'un mécanisme alternatif par lequel l'interaction protéine-protéine avec un cofacteur peut modifier la spécificité de séquence d'ADN reconnue par un FT. Chez la levure, l'interaction avec le cofacteur Met28 permet à l'hétérodimère centromere-binding factor (Cbf1)–Met4 de reconnaître un motif plus étendu, par rapport au motif reconnu par Cbf1-Met4 isolément (Siggers et al., 2011). Les membres d'une autre famille de facteurs de transcription, les protéines Hox, contenant un homéodomaine, affichent des spécificités de liaison très similaires l'une à l'autre, tant *in vitro* (Noyes et al 2008) qu'*in vivo* (Carr et Biggin, 1999), et pourtant, leur inactivation engendre des phénotypes très différents *in vivo*. Les interactions entre Extradenticle (Exd), Homothorax (Hth) et différentes protéines Hox de *D. melanogaster* ont été récemment montrées comme capables d'induire des modifications subtiles, mais potentiellement importantes, de la spécificité de liaison à l'ADN des protéines Hox *in vitro* (Slattery et al., 2011). Ces différences de spécificité de séquence induites par les cofacteurs Exd et Hth pourraient contribuer à certaines des fonctions distinctes et spécifiques des protéines Hox sur différents gènes cibles *in vivo*.

3. LA CHROMATINE

Le terme chromatine utilisé pour la première fois par Flemming en 1880 désigne le matériel génétique se trouvant sous forme compacte au niveau du noyau. L'unité fondamentale de la chromatine eucaryote est le nucléosome. Ce dernier est constitué par l'enroulement de 147 paires de bases d'ADN autour d'un octamère d'histones composé par deux copies de H2A, H2B, H3 et H4 (Figure 18 ; Luger et al. 1997).

La chromatine se présente sous forme deux états (i) l'euchromatine riche en gènes actifs et en histones acétylées et (ii) l'hétérochromatine qui constitue la chromatine mise en silence par répression épigénétique. Cette dernière est riche en séquences répétées et pauvre en gènes. Les régions d'hétérochromatine se regroupent à la périphérie du compartiment nucléaire créant ainsi un environnement riche en protéines répressives. Par contre, l'euchromatine occupe la partie interne du nucléoplasme. Cette organisation nucléaire est importante pour la régulation de l'expression des gènes.

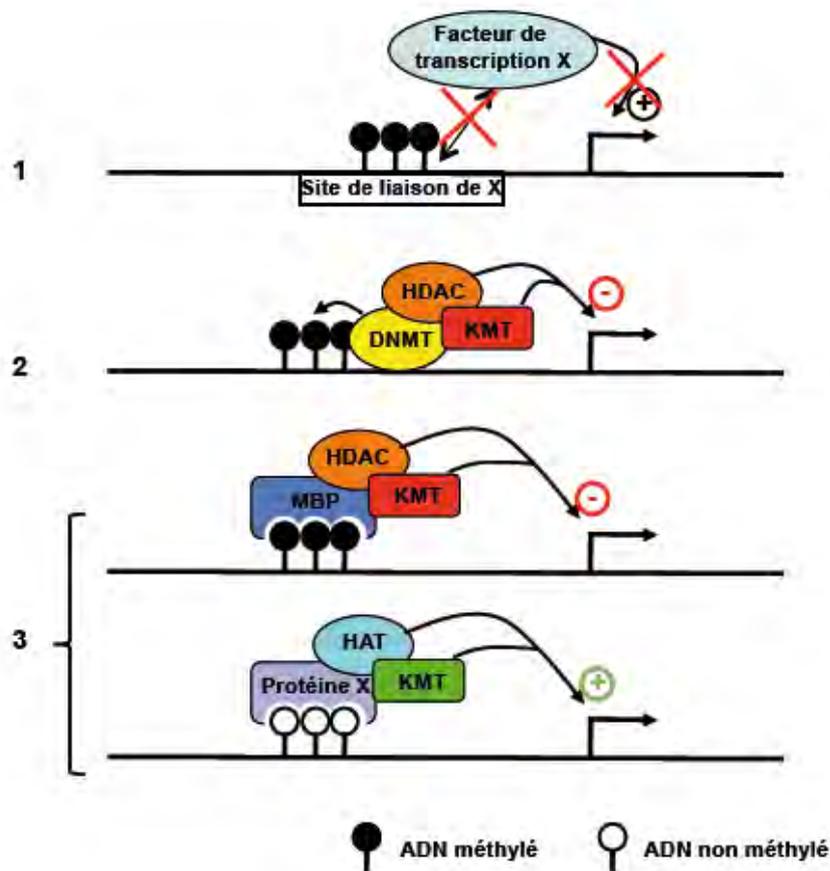


Figure 19. Mécanismes de régulation transcriptionnelle par la méthylation de l'ADN. 1 : La méthylation de l'ADN au niveau du site de liaison d'un facteur de transcription peut empêcher le recrutement de ce dernier. 2 : Les DNMTs recrutent directement des enzymes de modifications pour établir une structure non permissive à la transcription. 3 : La reconnaissance des CpG méthylés et déméthylés par des protéines spécifiques permet le recrutement d'enzymes de modifications induisant respectivement la répression ou l'activation de la transcription

3.1. REGULATION DE LA STRUCTURE CHROMATINIENNE

L'état de la chromatine peut être modulé par différents mécanismes, notamment la méthylation de l'ADN, l'incorporation de variants d'histones, l'intervention de complexes de remodelage ATP-dépendant ou encore les modifications post-traductionnelles des histones.

3.1.1. MÉTHYLATION DE L'ADN

La méthylation de l'ADN désigne l'ajout, par des ADN méthyltransférases (DNMTs), de groupements méthyles aux bases de l'ADN. Cette modification épigénétique est à la fois réversible et transmissible au cours des divisions cellulaires mitotiques et méiotiques. La méthylation de l'ADN est présente dans la plupart des grands groupes d'organismes. En revanche, les profils de méthylation de différents génomes varient considérablement tant par la nature des cibles de la méthylation, leur contexte de séquences, que par leur répartition au sein du génome. Chez les eucaryotes, la méthylation s'effectue principalement sur les cytosines (en N4 et C5) (Goll et Bestor, 2005 ; Buryanov et Shevchuk, 2005). Au sein du règne animal, cette marque se restreint majoritairement aux cytosines présentes dans des séquences symétriques 5'-CpG-3', appelées doublets CpG. Quelques cas rares de méthylation de dinucléotides CpT ou CpA existent, mais leur fréquence reste négligeable (Ramsahoye *et al.*, 2000; Lyko *et al.*, 2000; Malone *et al.*, 2001).

La mouche du fruit *Drosophila melanogaster*, a été longtemps considérée comme dépourvue de toute méthylation au niveau de l'ADN, jusqu'à ce que des études montrent que l'ADN génomique méthylé est très répandu au cours du développement embryonnaire précoce. Toutefois le niveau de méthylation global chez la drosophile est inférieur à celui des vertébrés et des plantes. En effet, il a été montré, par chromatographie à haute sensibilité, que seulement ~ 0,05 à 0,1%, des cytosines totales, sont méthylées et que ces dernières sont localisées majoritairement dans des dinucléotides CpT (Lyko *et al.*, 2000; Gowher *et al.*, 2000). Le système putatif de méthylation de l'ADN chez la drosophile contient une dDNMT2 méthyltransférase (Hung *et al.*, 1999; Tweedie *et al.*, 1999) et une protéine appelée dMBD2/3. La protéine dDNMT2 montre une homologie significative avec les méthyltransférases de mammifères de la famille DNMT2. En revanche, la protéine dMBD2/3 code pour une protéine n'ayant pas d'homologies avec les protéines mammifères MBD2 et MBD3. Ces dernières reconnaissent les dinucléotides CpG méthylés et permettent le recrutement d'enzymes de modification de la chromatine comme des histones méthyltransférases et des histones désacétylases conduisant ainsi à la mise en silence du gène

concerné. Une répression similaire peut être obtenue par les DNMTs qui en méthylant l'ADN peuvent recruter aussi les enzymes de modification de la chromatine. Enfin la répression de la transcription peut être assurée en empêchant les facteurs de transcriptions de se lier à leurs sites de fixation (Figure 19).

Ainsi on peut dire que la méthylation de l'ADN représente une composante importante de la régulation transcriptionnelle en permettant une expression spatio-temporelle correcte de certains gènes spécifiques.

3.1.2. MODIFICATIONS POST-TRADUCTIONNELLES DES HISTONES

Les histones peuvent subir plusieurs modifications post-traductionnelles (MPTs), telles que l'acétylation, la méthylation, l'ubiquitination, la phosphorylation, la sumoylation ou encore l'ADP-ribosylation, les deux premières étant les plus étudiées. Ces modifications interviennent principalement au niveau des queues N-terminales des histones, qui émergent de la structure de base du nucléosome, et contribuent à la régulation des gènes par l'intermédiaire d'actions en *cis* ou en *trans* (Bonasia et Reinberg 2010). Les MPTs peuvent être aussi localisées soit au niveau du domaine globulaire (comme la méthylation de la lysine 79 de H3, ou H3K79me) soit de l'extrémité C-terminale (par exemple l'ubiquitination de H2A sur la lysine 119).

Les mécanismes en *cis* modifient les interactions intra- et inter-nucléosomes, et conduisent à des changements de la structure chromatinienne et de son état de compaction. Un exemple classique est l'acétylation des résidus lysines: lorsque la lysine 16 de l'histone H4 chargée positivement est acétylée (H4K16ac), elle devient électrostatiquement neutre. Ainsi, l'interaction entre la chaîne latérale de la lysine et l'ADN chargé négativement est affaiblie et, par conséquent, la structure de la chromatine devient relâchée (Shogren-Knaak et al., 2006). En revanche, les mécanismes de régulation en *trans* servent de signaux pour le recrutement de facteurs de lecture/écriture/effacement spécifiques, qui sont impliqués dans le remodelage de la chromatine, l'activation des gènes, ou d'autres processus biologiques dynamiques (Taverna et al., 2007 ; Kouzarides et al., 2007). Les mécanismes de régulation *cis* et *trans* œuvrent ensemble et contribuent à la modulation et la régulation de la transcription, la réplication, la condensation des chromosomes, ainsi que d'autres fonctions biologiques.

✓ METHYLATION

La méthylation des histones est une MPT qui peut s'effectuer soit sur les résidus de lysine (K) par les lysines méthyltransférases (KMTs) soit sur les résidus d'arginine (R) par les arginine méthyltransférases (RMTs). Les résidus de lysine peuvent être mono-, di-, ou triméthylés au groupe ϵ -amino (ϵ -N-méthyl-Lys), tandis que les résidus d'arginine peuvent être mono-ou di-méthylés dans différentes configurations, y compris ω -NG arginine monométhylée, ω -NG / NG-diméthyl arginine asymétrique et ω -NG / N'G-diméthyl arginine symétrique. La famille des protéines arginine méthyltransférases (PRMT) des mammifères est composée de neuf membres classés en trois groupes : les PRMTs de type I (PRMT1, 2, 3, 4, 6 et 8) catalysent la diméthylation asymétrique, les PRMTs de type II (PRMT5 et 7) sont responsables de la diméthylation symétrique et enfin les PRMTs de type III, de la monométhylation de certains résidus (PRMT7). Les histones ne sont pas les seuls substrats des RMTs, ce qui les distingue des KMTs. Les RMTs peuvent en effet catalyser la méthylation de nombreuses protéines impliquées dans la transduction du signal, la prolifération cellulaire, la réparation de l'ADN et la maturation des ARNs (Bedford et Richard 2005). Par exemple, la méthylation de l'arginine de l'histone acétyltransférase CBP par la transférase CARM1 (*coactivator associated arginine methyl-transferase 1*) contribue à ses effets co-activateurs.

Presque toutes les KMTs sont constituées d'un domaine SET (*su(var)3-9, enhancer-of-zeste, trithorax*) conservé. Cependant, ce domaine n'est pas suffisant pour assurer l'activité catalytique de l'enzyme : deux régions adjacentes riches en résidus de cystéine, appelées pre-SET et post-SET, sont aussi nécessaires (Chaib et al., 2011).

La relation entre la méthylation des histones et l'état transcriptionnel est complexe et dépend du résidu modifié (Kouzarides 2002; Berger 2007; Kouzarides 2007; Bannister et Kouzarides 2011) ainsi que du degré de méthylation. Par exemple, la monométhylation est généralement associée à l'activation de la transcription, tandis que la triméthylation des résidus de lysine est souvent liée à la répression. La méthylation des résidus d'arginine (Litt et al., 2009) et de lysine (Martin et Zhang 2005) sont impliquées dans plusieurs fonctions biologiques telles que la régulation de la transcription, le remodelage de la chromatine, la réparation de l'ADN, l'épissage de l'ARN et la transduction du signal. La méthylation des histones n'influence pas la structure du nucléosome comme le fait l'acétylation. En revanche, ce type de modification permet le recrutement de facteurs impliqués dans la transcription au niveau des nucléosomes. Ces facteurs reconnaissent les marques spécifiques de la méthylation

via des domaines de reconnaissance spécifiques, comme les domaines TUDOR, les chromodomains, les domaines WD40 ou encore les domaines PHD. La déméthylation des histones est catalysée par amine oxydation (LSD1, *lysine specific demethylase 1*) ou par hydroxylation par une protéine à domaine Jumonji (JmJ)-C des histones diméthylases (HDM). La seule histone déméthylase de résidus d'arginine identifiée à ce jour est JMJD6, qui déméthyle l'histone H3 au niveau de l'arginine 2 (H3R2) et l'histone H4 au niveau de l'arginine 3 (H4R3) (Chang et al., 2007).

✓ ACETYLATION

L' ϵ -N-acétylation des résidus de lysine dans les histones a été rapportée en 1964 (Allfrey et al., 1964). A cette époque, on s'attendait à ce que les histones soient impliquées dans la régulation de l'activité de l'ADN et la découverte de l'acétylation des histones a renforcé la théorie fondamentale de leur rôle dans la régulation transcriptionnelle. L' ϵ -N-acétylation est principalement impliquée dans l'activation des gènes, du fait que l'ajout du groupement acétylé neutralise la charge positive des résidus de lysine affaiblissant ainsi les interactions des histones avec l'ADN. En outre, l'acétylation est impliquée dans la réparation, la réplication et la condensation de l'ADN. L' ϵ -N-acétylation de résidus de lysine est catalysée par les histones acétyl-transférases (HATs) et la suppression du groupement acétylé est médiée par des histone-désacétylases (HDACs).

Les HATs sont divisées en 2 classes majeures : celles de type A, et celles de type B. Les HATs de type A sont classées en 3 groupes en fonction de leurs homologues de séquences et de structures, constituant les familles GNAT (pour Gcn5 related N-AcetylTransferase), MYST et CBP/p300. Elles peuvent acétyler les histones nucléosomales et/ou libres sur de nombreux résidus. Les HATs de type B sont quant à elles majoritairement cytoplasmiques et de ce fait acétylent principalement les histones libres. Elles sont par exemple responsables de l'acétylation de H4K5 et H4K12, deux marques importantes pour l'incorporation des histones sur les brins d'ADN néosynthétisés après la réplication.

Concernant les HDACs, on distingue 4 classes. La classe I comprend les HDACs 1, 2, 3 et 8, la classe II regroupe les HDACs 4, 5, 6, 7, 9 et 10, tandis que la classe IV ne comporte qu'un membre, HDAC 11. Les HDACs de classe III correspondent à la famille des sirtuines. L'acétylation joue un rôle crucial dans le dialogue entre les modifications. Par exemple, la mutation du H3K14, un site connu d'acétylation, entraîne la perte de triméthylation spécifique de H3K4, mais pas de sa mono-ou di-méthylation (Nakanishi et Sanderson, 2008).

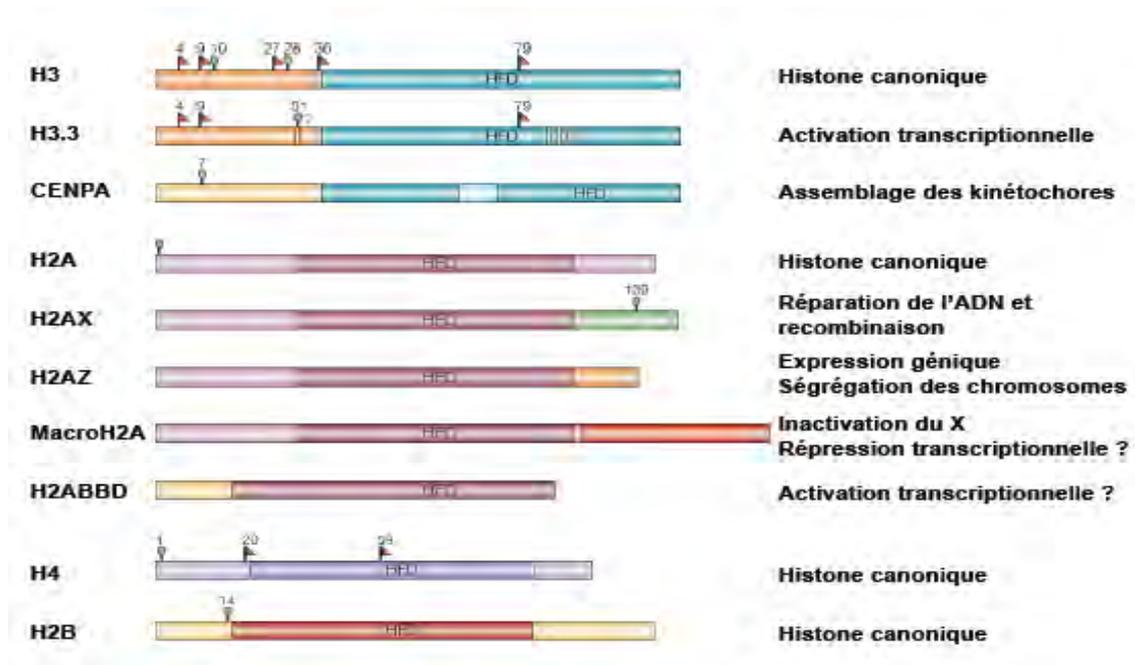


Figure 20. Les variants d'histones: structure et fonction. Représentation schématique des différents variants d'histones. Les principaux sites de phosphorylation (en vert) et de méthylation (en rouge) sont représentés. Les résidus qui diffèrent entre H3 et H3.3 sont en jaune dans la séquence de ce dernier. Dans CENPA, l'extrémité N-terminale spécifique est représentée en jaune, et la région permettant le ciblage au centromère en bleu clair. Les extrémités C-terminales non conservées entre les variants de H2A sont notifiées par des couleurs différentes, et l'extrémité N-terminale de H2ABBD, qui lui est propre, est en jaune pâle. Les fonctions proposées pour les différents variants sont listées sur la droite. (D'après Sarma et Reinberg, 2005).

✓ PHOSPHORYLATION

Les histones peuvent être phosphorylées sur des résidus de sérine (S), de thréonine (T) et de tyrosine (Y) et leur niveau de phosphorylation dépend de l'action antagoniste des protéines kinases et des protéines phosphatases. Plusieurs études ont rapporté un dialogue entre la phosphorylation et l'acétylation ou la méthylation des histones, résultant souvent en l'activation transcriptionnelle. Ceci est le cas de la phosphorylation de H3S10 qui est corrélée à l'acétylation de H3K9 et H3K14 et à une activation transcriptionnelle (Lau et Cheung 2011 ; Simboeck et al., 2010). De même, il a été rapporté que la phosphorylation de H3T11 active la transcription et ceci en stimulant l'activité catalytique de la déméthylase JMJD2C, qui en retirant la marque répressive H3K9me_{2/3}, induit la transcription des gènes cibles du récepteur aux androgènes (Metzger et al. 2008; Cerutti et Casas-Mollano 2009).

✓ BIOTINYLATION

La biotinylation des histones a été détectée dans plusieurs variants d'histones, y compris H2A, H3 et H4. Cette MPT est censée être impliquée dans la répression transcriptionnelle (Kothapalli et al., 2005 ; Filenko et al., 2011), en modifiant la structure des nucléosomes (Filenko et al., 2011). La biotinylation est catalysée par la holocarboxylase synthétase, alors que la biotinidase est impliquée dans débiotinylation. Plusieurs résidus d'acides aminés dans la queue d'histone H4 ont été caractérisés comme cibles de ces enzymes (Camporeale et al., 2004).

3.1.3. VARIANTS D'HISTONES

Chez les eucaryotes, l'agencement des nucléosomes de la chromatine s'établit généralement durant la réplication par la déposition des histones canoniques (H2A, H2B, H3.1/H3.2, et H4) synthétisées au cours de la phase S. Toutefois, plusieurs autres isotypes d'histones peuvent être incorporés dans les nucléosomes, dans certaines zones génomiques et sous certaines conditions, et ceci indépendamment de la réplication (Orsi et al., 2009). Par exemple, dans des cellules de mammifères le traditionnel H3 est remplacé dans les nucléosomes de la chromatine centromérique par cinq isotypes différents nommés H3.1, H3.2, H3.3, H3t et CENH3/CENP-A. Toutes les histones canoniques, à l'exception de H4, présentent des variants (Figure 20) qui nécessitent pour leur déposition sur le nucléosome l'action des complexes de remodelages dépendants d'ATP et/ou l'intervention de protéines chaperons.

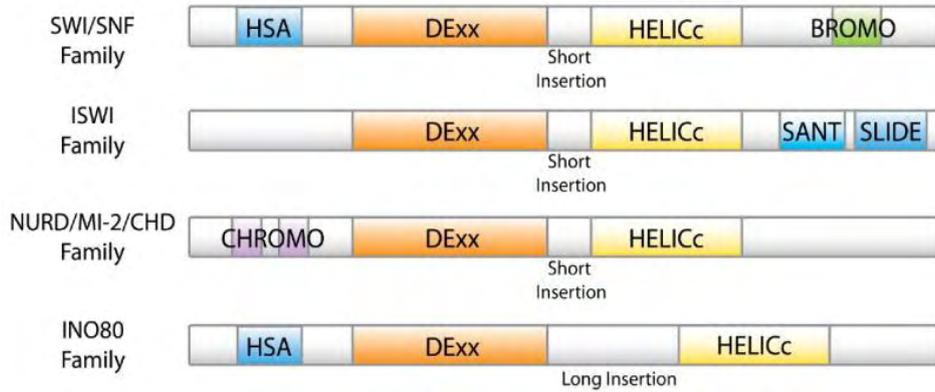


Figure 21. Représentation schématique des 4 grandes familles d’enzyme de remodelage ATP-dépendant. La position des différents domaines fonctionnels est indiquée. HSA: domaine de liaison à l’actine, SANT et SLIDE: domaines importants pour la liaison aux histones, DExx et HELICc: activité catalytique, BROMO: reconnaissance des lysines acétylées, CHROMO: reconnaissance des lysines méthylées (D’après Tang et al., 2010).

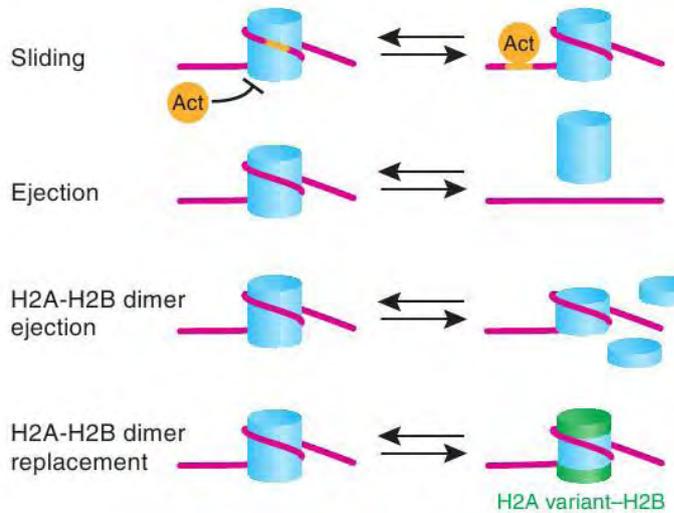


Figure 22. Les différents modes d’action des complexes de remodelage ATP-dépendants.

La séquence des variants d'histones peut différer de celle des histones canoniques au niveau de la queue N-terminale (macroH2A), du domaine globulaire (H2ABbd) ou seulement de quelques acides aminés (H3.3). L'intégration des variants d'histones dans les nucléosomes permet à la chromatine d'acquiescer une structure spécifique et/ou d'affecter l'activité transcriptionnelle (Sarma et Reinberg 2005; Campos et Reinberg 2009). Par exemple, l'incorporation du variant H2AZ est associée à une compaction de la chromatine (Campos et Reinberg 2009; Gamble et Kraus 2010) et par conséquent à une répression de la transcription.

3.1.4. REMODELAGE DE LA CHROMATINE

Les métazoaires présentent un génome condensé et inaccessible d'où la nécessité de maintenir la chromatine sous un état ouvert facilitant ainsi l'accès des différentes machineries cellulaires impliquées entre autres dans la transcription ou encore la réplication de l'ADN. Ce maintien dans un état ouvert fait intervenir des complexes de remodelage et de modification de la chromatine. Il existe quatre familles de complexes de remodelage classées sur la base de la séquence et de la structure de la sous-unité ATPase et qui sont conservées au cours de l'évolution (Figure 21) (Clapier et Cairns 2009). Elles utilisent l'énergie libérée par hydrolyse de l'ATP pour modifier l'interaction entre les histones et l'ADN. Chaque domaine ATPase est formé par deux sous domaines : DExx et HELICc.

Ces complexes de remodelage agissent soit en faisant glisser le nucléosome, soit en l'éjectant entièrement ou partiellement (dimère H2A/H2B) ou bien en échangeant un variant d'histone par un autre (Figure 22). Dans les 3 premiers cas, ceci permet à d'éventuels facteurs liant l'ADN d'y avoir accès.

La famille SWI/SNF : appelée Brahma chez la drosophile, elle contient un bromodomaine et un domaine HSA (helicase –SANT) et permet de faire glisser ou bien d'éjecter le nucléosome (Morhmann et Verrijzer 2005).

La famille ISWI : Ces complexes ont été initialement identifiés chez la drosophile, qui ne possède qu'une seule ATPase de ce type (Iswi) (Bouazoune et Brehm 2006). Les membres de cette famille présentent à leur extrémité C-terminale deux domaines SANT et SLIDE (SANT-like ISWI) (Corona et Tamkun, 2004). Elle comprend les complexes CHRAC, ACF et NURF. Les deux premiers permettent l'assemblage de la chromatine et la répression de la transcription, le dernier provoque au contraire une désorganisation de la chromatine et favorise par conséquent la transcription.

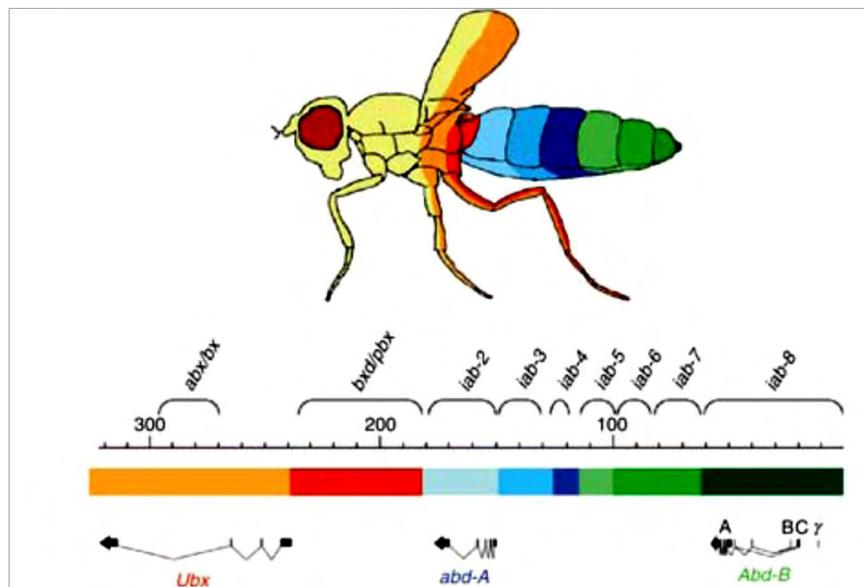


Figure 23. Organisation du complexe Bithorax. Ce locus comprend trois gènes clés du développement de la drosophile. Les régions contrôlées par les différents éléments régulateurs sont représentées en couleur (Daprès Maeda et Karch 2009).

La famille CHD : Les membres de cette famille présentent, à l'extrémité N-terminale de la sous-unité catalytique, 2 chromodomaines organisés en tandem (Marfella et Imbalzano, 2007) qui lui permettent de reconnaître des résidus méthylés ainsi qu'un domaine de liaison à l'ADN. Ces enzymes peuvent être associées soit à l'activation soit à la répression, comme c'est le cas du complexe Mi-2/NuRD (Nucleosome Remodeling histone Deacetylase).

La famille INO80 : Du point de vue structural, cette famille se distingue par la présence d'une longue insertion entre les deux domaines DExx et HELICc de l'ATPase. Chez la levure, l'homologue de p400 (Swr1) est responsable de l'incorporation du variant d'histone H2AZ, une fonction qui semble conservée chez les mammifères pour les membres de cette famille (Morrison et Shen, 2009). Ces complexes interviennent pour favoriser la transcription mais aussi pour la réparation de l'ADN.

Tous ces complexes de modification des histones et de remodelage de la chromatine agissent en étroite coopération. En effet, l'expression d'un gène peut nécessiter la déstabilisation des nucléosomes par un complexe de remodelage pour permettre l'accès à une histone acétyltransférase (Cosma *et al.*, 1999), alors qu'un autre nécessitera à l'inverse d'abord l'acétylation des histones et la liaison de la PolIII pour permettre le recrutement d'un complexe de remodelage (Agalioti *et al.*, 2000).

3.2. GROUPES POLYCOMB ET TRITHORAX

Le développement de la drosophile est régulé entre autres par les membres de deux complexes protéiques clés, appartenant aux groupes Polycomb (PcG) et Trithorax (TrxG). Ils agissent en modifiant la chromatine après fixation sur les séquences PRE (PcG Response Element) et TRE (TrxG Response Element), respectivement, impliqués dans la maintenance de la répression (PRE) ou de l'activation (TRE) des gènes établies lors de l'embryogenèse. Ce mode de régulation a été illustré durant les années 90 suite à l'étude de la régulation des gènes homéotiques localisés au niveau du complexe Bithorax (BX-C) chez la drosophile (Figure 23) (Busturia *et al.*, 1989 ; Chiang *et al.*, 1995 ; Simon *et al.*, 1993). En effet, au cours de ces études deux types d'éléments *cis*-régulateurs ont été révélés : des éléments initiateurs, actifs pendant les trois premières heures du développement et des éléments de maintenance qui permettent de maintenir l'historique transcriptionnel des gènes après la disparition des signaux initiateurs, et ce jusqu'au stade adulte (Busturia *et al.*, 1989 ; Chiang *et al.*, 1995 ; Simon *et al.*, 1993).

3.2.1. ACTIVATION PAR LES PROTEINES TRXG.

Les protéines TrxG appartiennent à différentes classes. La première comprend des protéines possédant un domaine SET dont la protéine Trx elle-même qui lie l'élément *Fab-7* et maintient l'activation d'*Ubx* au sein du complexe Bx-C (Dejardin & Cavalli 2005). Les gènes *Ash1* (*absent, small or homeotic discs 1*) et *trx* (*trithorax*) codent pour des méthyltransférases qui triméthylent H3K4, empêchant ainsi la mise en silence de ces régions par les protéines PcG (Klymenko & Muller 2004).

La seconde classe inclut des composants des complexes de remodelage de la chromatine SWI/SNF et NURF dont le recrutement est facilité en présence de la marque H3K4me3 (Wysocka *et al.*, 2006). Ces complexes facilitent le remodelage des nucléosomes aux promoteurs et le long des gènes pour permettre la liaison de la PolII ainsi que l'élongation. Ainsi, dans ce groupe de protéines, des modificateurs d'histones et des complexes de remodelage collaborent pour maintenir l'activation des gènes.

3.2.2. REPRESSION PAR LES PROTEINES PCG

Les protéines PcG appartiennent à 3 complexes. Le complexe PRC2 est responsable de la di- et triméthylation de la lysine 27 de l'histone H3 qui est une marque reconnue par le complexe PRC1 et dont la liaison favorise la compaction de la chromatine par un mécanisme qui reste à déterminer (Cao & Zhang 2004). De plus, PRC1 est responsable de la mono-ubiquitination de la lysine 119 de H2A qui corrèle, comme H3K27me3, avec les régions transcriptionnellement réprimées. Le troisième complexe est PhoRC qui possède un domaine de liaison à l'ADN et semble important pour cibler le recrutement de PRC2 sur les PRE (Wang *et al.* 2004). Ces protéines collaborent pour maintenir la chromatine sous forme compacte et réprimer la transcription et ce pendant de nombreux cycles cellulaires.

PARTIE II :

**Développement des pattes et de
l'antenne**

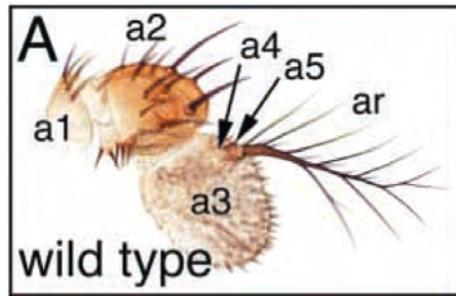


Figure 24. Antenne sauvage de *Drosophila melanogaster*. (Dong et al., 2002).

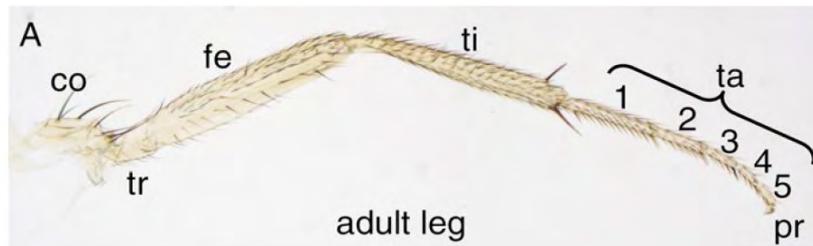


Figure 25. Patte sauvage de *Drosophila melanogaster*. (Kojima, 2004).

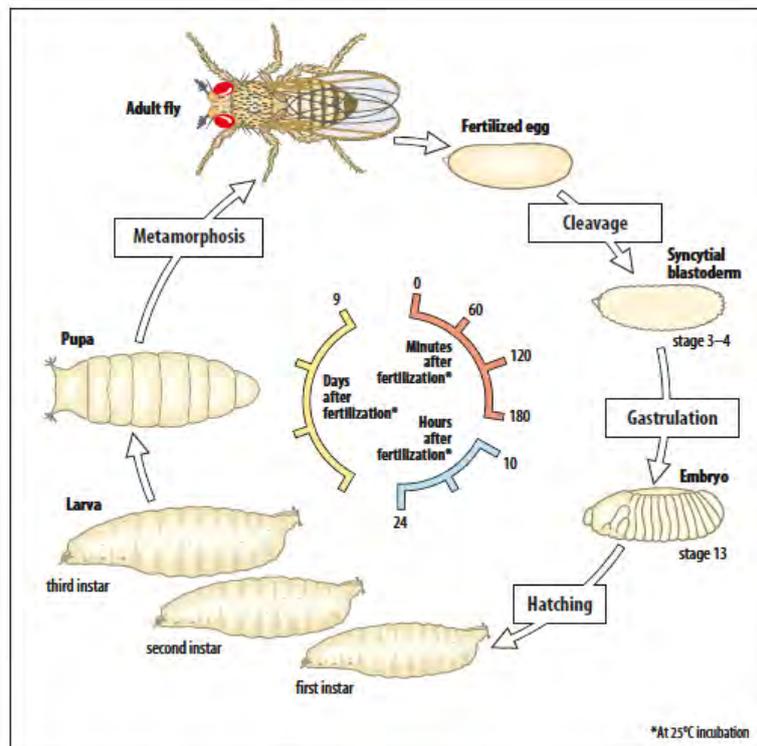


Figure 26. Cycle de vie de *Drosophila melanogaster*.

4. DEVELOPPEMENT DE LA PATTE ET DE L'ANTENNE

4.1. STRUCTURE DE LA PATTE ET DE L'ANTENNE

Chez la drosophile, les antennes constituent des structures sensorielles qui hébergent des organes d'audition, d'olfaction et de perception tactile et sont par conséquent cruciales pour l'insecte afin qu'il puisse interagir avec son environnement (Chapman, 1998 ; Snodgrass, 1935). Toutes les antennes des insectes possèdent trois segments primaires, le scape, le pédicelle et le flagellum. Chez *Drosophila melanogaster* le segment scape correspond à a1, le pédicelle à a2 et le flagellum aux quatre segments a3, a4, a5 et arista (Figure 24). Comme les segments primaires, les segments flagellaires sont séparés par des jonctions ectodermiques, mais dépourvus de muscles individuels (Snodgrass, 1928).

Les pattes adultes de la drosophile contiennent comme l'antenne des organes sensoriels qui servent à la perception du milieu. Toutefois, leur principale fonction est de permettre la locomotion et le toilettage de la mouche. La drosophile possède trois paires de pattes appelées prothoraciques (T1), mésothoraciques (T2) et métathoraciques (T3) qui présentent une organisation générale commune. Chaque patte est formée du proximal vers le distal de dix segments : la coxa, le trochanter, le fémur, le tibia, cinq segments tarsaux (t1-t5 qui forment le tarse) et le prétarse portant deux griffes (Figure 25). Comme pour l'antenne, les segments de la patte sont séparés par des structures flexibles appelées jonctions.

4.2. DISQUES IMAGINAUX

La drosophile est un insecte holométabole (à métamorphose complète) dont le cycle vital comporte un stade œuf, trois stades larvaires mobiles séparés par des mues et un stade nymphal immobile, la puppe, qui conduit après métamorphose au stade adulte (Figure 26). Au cours de la métamorphose presque toutes les structures larvaires se désintègrent et sont remplacées par les structures adultes. Ces dernières se développent à partir des cellules imaginales qui prolifèrent tout au long du stade larvaire sans pour autant contribuer à la vie larvaire. Parmi les cellules imaginales les mieux caractérisées se trouvent les disques imaginaux à l'origine des structures cuticulaires adultes : tête, thorax et appendices. Chez la drosophile, il y a 19 disques imaginaux (9 paires et un disque génital unique). Ces derniers portent le nom de l'appendice qui en résulte (Figure 27) et se répartissent au niveau de la tête (disque labial, clypeolabral, oeil-antenne), des segments thoraciques (disques de pattes, d'aile et d'haltère) et de la queue (disque génital) (Cohen, 1993; Fristrom et Fristrom 1993).

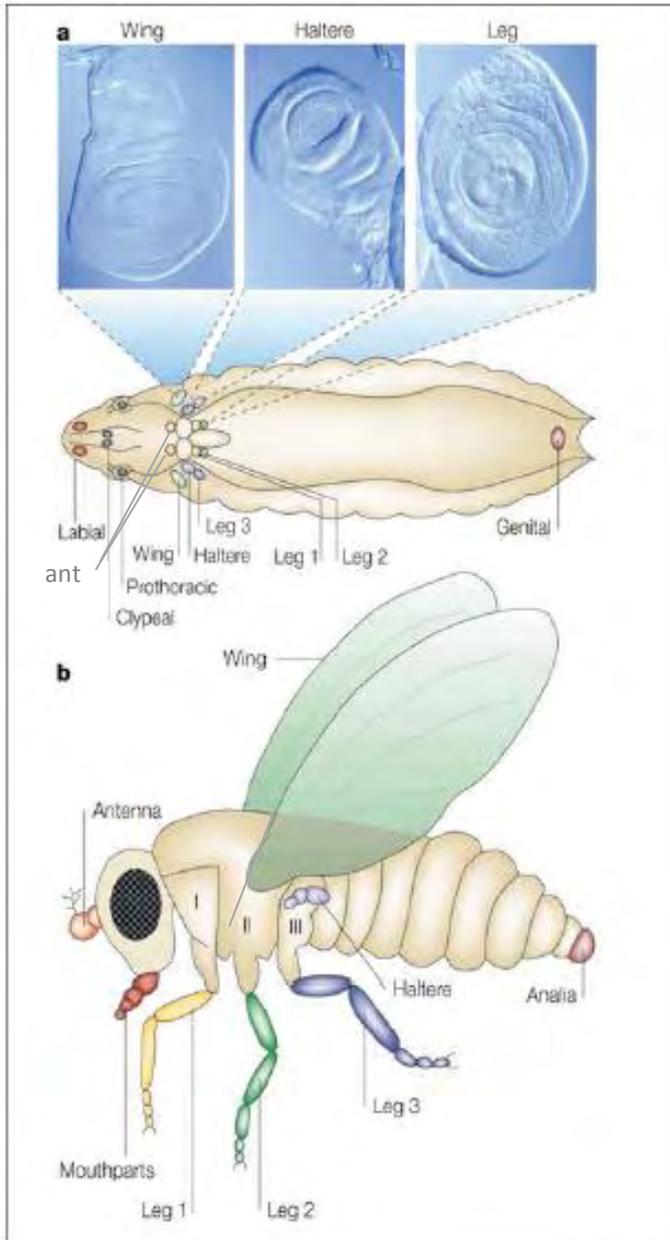


Figure 27. Origine larvaire des appendices de la drosophile adulte: les disques imaginaux. (a) Schématisation de la position des disques imaginaux dans une larve de drosophile adulte où les divers appendices sont représentés avec un code couleur correspondant aux divers disques imaginaux dont ils dérivent. (D'après Morata G., 2001)

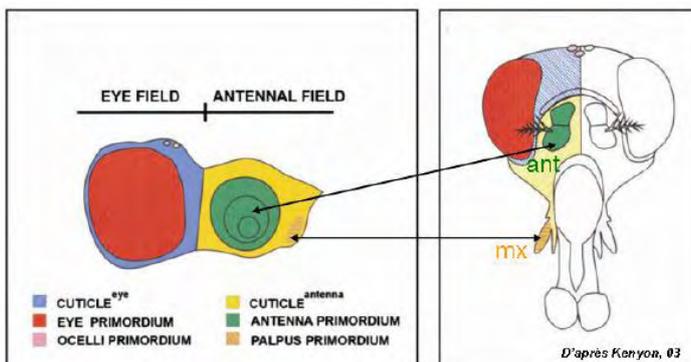


Figure 28. Formation des territoires et des primordia au sein du disque œil-antenne. Chaque territoire regroupe différents primordia qui forment la quasi-totalité de la tête de l'adulte (D'après Kenyon, 2003).

Les disques imaginaux sont des îlots de cellules embryonnaires d'origine ectodermique déterminées mais indifférenciées, qui se forment par invagination au cours de l'embryogenèse et présentent une structure de sacs aplatis, constitués d'une seule couche de cellules épithéliales. Au cours des stades larvaires, ces structures prolifèrent puis sous l'influence de l'ecdysone (hormone de mue) s'évaginent (se retournent comme des doigts de gants) et s'engagent au cours de la métamorphose dans la formation des appendices de l'imago (Baena-Lopez et al., 2005; Bryant, 1970). Ainsi dans le cas des pattes et de l'antenne les régions centrales des disques vont donner les segments distaux des appendices et les régions périphériques donneront les segments proximaux (Kojima, 2004). Les disques imaginaux sont des exemples classiques de champs morphogénétiques. En effet, leur transplantation dans des larves de même âge n'altère pas leur capacité à former les organes correspondants (Bergantiños et al., 2011). Chaque disque imaginal correspond à un champ morphogénétique d'où émerge un primordium qui donnera naissance à un organe particulier, à l'exception du disque œil-antenne qui est formé de plusieurs champs morphogénétiques à l'origine de plusieurs organes morphologiquement et fonctionnellement distincts (Figure 28). En effet, ce disque va donner naissance à quatre organes sensoriels distincts (œil, ocelle, antenne et palpe maxillaire) ainsi qu'à la plus grande partie de la capsule céphalique (Haynie et Bryant, 1986).

Dans ce qui suit, je me limiterai au développement des disques de pattes et d'antenne (respectivement à l'origine des trois types de pattes et de l'antenne).

4.2.1. ORIGINE EMBRYONNAIRE DES DISQUES IMAGINAUX DE PATTES

A sa naissance le disque de patte ne compte que 10 à 30 cellules (Cohen 1993). Par prolifération ce nombre va atteindre plus de 10000 cellules durant les deuxième et troisième stades larvaires. Deux types de cellules commencent à être reconnaissables : les cellules squameuses qui constituent la membrane péripodiale et contribuent au développement de la paroi du corps et les cellules en colonnes plus larges qui constituent le disque épithélial proprement dit. C'est ce dernier qui va se différencier en patte (Fristrom et Fristrom, 1993). Au milieu du stade embryonnaire, quand les deux protéines à homéodomaine Homothorax (Hth) et Extradenticle (Exd) hétérodimérisent et sont transloquées dans presque tous les noyaux des cellules des segments thoraciques (Rieckhof et al., 1997; Pai et al., 1998; Kurant et al., 1998; Aspland et White, 1997; Gonzalez-Crespo et al., 1998), de petits groupes de cellules dans la région ventro-latérale des segments thoraciques commencent à exprimer *Dll*

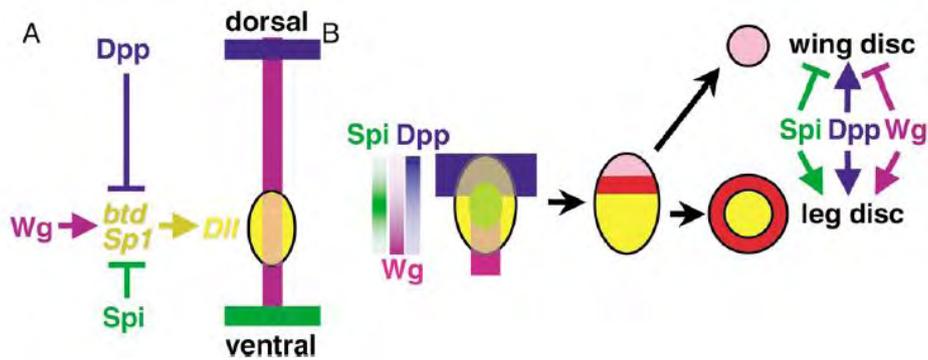


Figure 29. Origine embryonnaire des disques imaginaux de pattes et d'aile. (A) par la combinaison de l'activité activatrice de Wg et répressive de Dpp et des ligands de EGFR, *btd/Spl* sont exprimées dans un petit groupe de cellules à la région ventro-latérale du thorax chez l'embryon, et induisent l'expression de *Dll*. (B) Peu de temps après l'induction du domaine *Dll*, l'expression ou la production de Dpp, Wg et des ligands de EGFR sont modifiées (à gauche) et subdivisent le domaine d'expression de *Dll* en trois régions en fonction de leurs activités. Les cellules les plus dorsales forment le disque aile, et les cellules les plus ventrales et médiales forment le disque de patte (droite). Les cellules les plus ventrales continuent à exprimer *Dll* mais pas *hth/esg* et constituent le domaine distal du disque de patte, tandis que les cellules médianes ayant perdu l'expression de *Dll* constituent la région proximale. (Kojima, 2004).

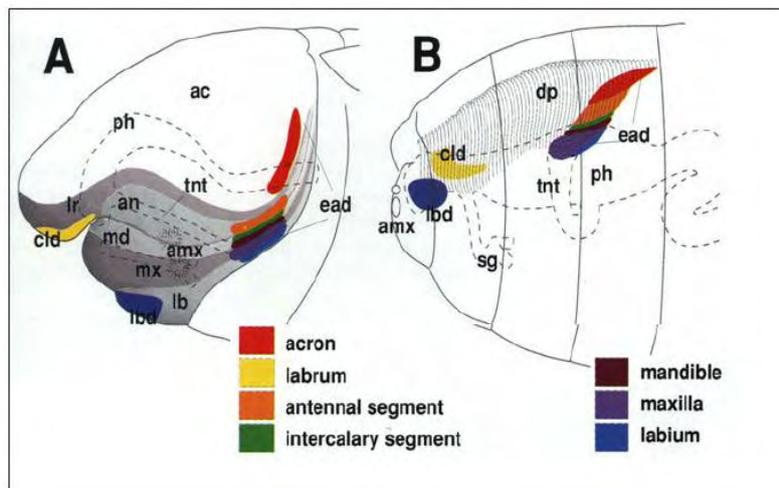


Figure 30. Origine embryonnaire du disque œil-antenne. Les cellules du disque œil antenne (ead) seraient issues des segments labial (lb), maxillaire (mx), mandibulaire (md), antennaire (an), labral (lr) et de l'acron (ac). Les deux autres disques qui participent à la formation de la tête, les disques labial (lbd) et clypéolabral (cld) proviennent respectivement des segments labial et labral.

codant pour un facteur de transcription à homéodomaine (Figure 29A; Cohen, 1990; Gonzalez-Crespo et al., 1998). A ce stade, *wg* est exprimé selon une bande de cellules tout le long des frontières AP de l'ectoderme embryonnaire, *dpp* est exprimé sous forme d'une bande dorsale et les ligands de la voie EGFR (*s*) sont produits à la ligne médiane ventrale ('ventral midline') (Cohen et al 1993; Golembo et al 1996, 1999; Goto et Hayashi, 1997). L'expression initiale de *Dll* est induite par la signalisation Wg, alors qu'elle est respectivement réprimée dorsalement et ventralement par les activités des voies Dpp et EGFR (Cohen, 1990; Goto et Hayashi, 1997). Peu de temps après l'induction de *Dll*, l'expression de *dpp* et la production des ligand(s) de la voie EGFR se produisent respectivement dans les régions dorsales et médio-ventrales ('mid-ventral') des domaines exprimant *Dll* (Figure 29B) (Kubota et al., 2000).

L'activité antagoniste des voies EGFR et Dpp vont déterminer trois populations de cellules dans le domaine primaire *Dll* (Figure 29B): les cellules situées dans la partie dorsale destinées à former les disques d'aile et d'haltère; les cellules ventrales qui formeront la partie distale de la patte adulte, et celles situées dans la zone médiane qui formeront sa partie proximale (Cohen et al., 1993; Kubota et al., 2000). A ce stade, l'expression de *wg* se rétracte ventralement et son expression au sein du domaine *Dll* se limite aux cellules ventrales (Cohen et al., 1993; Kubota et al., 2003). Wg inhibe la formation des disques d'aile et d'haltère, alors qu'il favorise la formation du disque de patte (Kubota et al., 2003). Ainsi, le domaine primaire *Dll* contient à la fois des cellules précurseurs des appendices dorsaux (aile et haltère) et ventraux (pattes).

En fin de stade embryonnaire, les cellules précurseurs des disques d'aile et d'haltère migrent dorsalement, cessent d'exprimer *Dll*, et commencent l'expression de gènes nécessaires à leur formation, comme *vestigial (vg)*, *snail (sna)* et *escargot (esg)* (Cohen et al., 1993; Williams et al., 1991; Fuse et al. 1996; Kim et al. 1996; Goto et Hayashi, 1997). Les cellules distales des pattes continuent à exprimer *Dll* mais perdent l'expression de *hth* et d'Exd nucléaire. A l'inverse, les cellules proximales de la patte exprimant *hth* et ayant Exd nucléaire, cessent d'exprimer *Dll* et entourent les cellules exprimant *Dll* (Gonzalez-Crespo et al 1998; Kubota et al., 2003). L'expression d'*esg* apparaît également dans les cellules proximales (Goto et Hayashi, 1997). Il s'agit là du premier stade, où l'on peut reconnaître une subdivision PD au niveau du disque de patte.

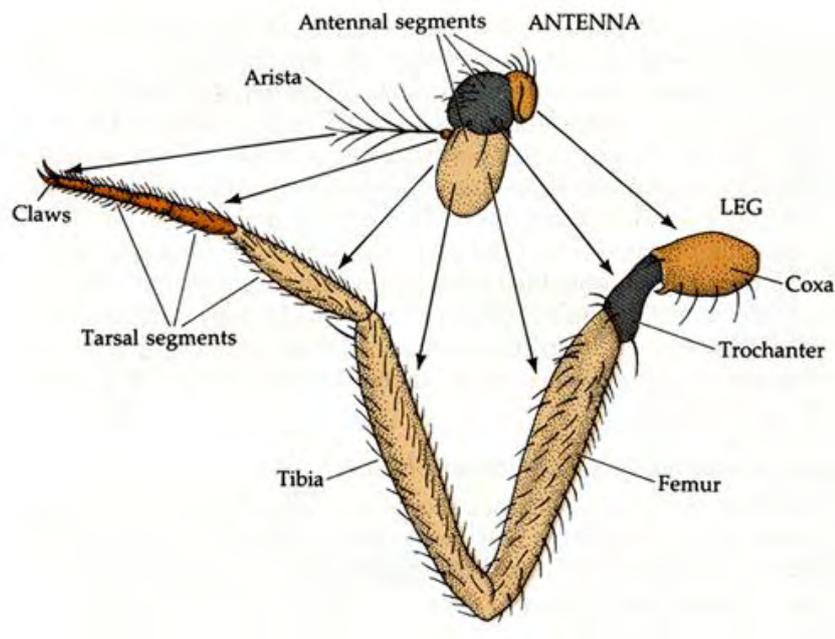


Figure 31. Patte et antenne sont deux structures homologues. Chez le mutant *Antennapedia*, l'antenne se transforme en patte T2. Les flèches indiquent la correspondance entre les segments d'antenne et de patte. (D'après Postlethwait et Schneiderman, 1971).

4.2.2. ORIGINE EMBRYONNAIRE DU DISQUE IMAGINAL DE L'ANTENNE

Dans une larve nouvellement éclos, le disque œil-antenne se situe à l'extrémité postérieure de la poche dorsale (Postlethwait, 1978) et se compose d'environ 70 cellules provenant de plusieurs segments embryonnaires de la tête et de l'acron (Figure 30; Jurgens et Hartenstein, 1993). Toutes les cellules du disque œil-antenne expriment les gènes de type *Pax6 eyeless (ey)* et *twin-of-eyeless (toy)* durant le premier stade larvaire (L1). Cependant, durant le stade larvaire L2, seuls les deux tiers postérieurs du disque composite œil-antenne expriment les deux gènes *Pax6*, alors que le tiers antérieur exprime *cut (ct)*. Les domaines *ct* et *Pax6* au stade L2 correspondent respectivement au disque d'antenne et au disque d'œil (Kenyon et al., 2003). Le disque œil-antenne donne lieu à quatre organes différents (œil, antenne, ocelle, et palpe) et à la cuticule entourant la tête (Figure 28). La partie œil du disque composite comprend les primordia œil et ocelles ainsi qu'une partie du primordium de la capsule céphalique, tandis que le disque d'antenne comprend les primordium antenne, palpe maxillaire (Mx) ainsi que le reste du primordia de la capsule céphalique (Haynie et Bryant, 1986; Morata et Lawrence, 1979; Postlethwait et Schneiderman, 1971).

4.3. MISE EN PLACE DES AXES DANS LES DISQUES DE PATTES ET D'ANTENNE

Chez la drosophile, l'antenne, les pièces buccales, les pattes et les organes génitaux sont considérés comme des appendices ventraux homologues (Cohen, 1993). Cela signifie que malgré leurs structures et leurs fonctions très différentes, il a été proposé qu'ils se soient développés à partir d'une même structure de base « ground state ». En agissant sur cet état de base, c'est l'expression de gènes sélecteurs spécifiques à un segment donné qui définit sa morphologie propre (Casares et Mann, 2001; Duncan et al., 1998; Estrada et Sanchez-Herrero, 2001; Joulia et al., 2006). Ainsi il a été proposé que les segments antennaires les plus proximaux a1 et a2 soient respectivement homologues au coxa (cx) et au trochanter (tr) d'une patte ; le troisième segment antennaire (a3) correspond à la fois au fémur (fe), au tibia et au premier segment tarsal (t1) ; l'arista (ar) et sa base (segments a4 et a5) sont homologues à la griffe (« claw ») et aux tarsi t2-5 (Figure 31) (Postlethwait et Schneiderman 1971; Angelini et Kaufman, 2005). Les développements de la patte et de l'antenne présentent plusieurs similarités, notamment au cours de la mise en place des axes antéro-postérieur A/P,

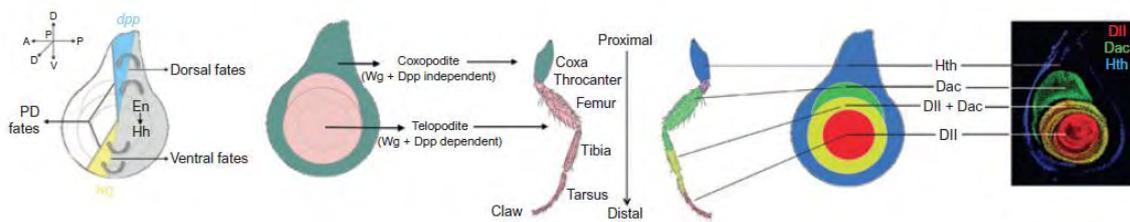


Figure 32. Vue d'ensemble du développement de la patte. Sur la gauche sont représentées les relations entre *en*, *hh*, *wg*, et *dpp* et la définition du télépodite (domaine dépendant de *hh*, *wg* et *dpp*) et le coxopodite (domaine indépendant de *hh*, *wg*, et *dpp*). Sur la droite on voit la relation entre les trois domaines primaires d'expression des gènes PD : *hth*, *dac* et *Dll*. (Estella et al., 2012)

dorso-ventral (D/V) et proximo-distal (P/D) (Angelini et Koufman, 2005). Toutefois plusieurs gènes et interactions antenne-spécifiques ont été identifiés, comme par exemple l'expression continue de *spineless* (*ss*) qui spécifie l'identité antennaire (Emmons et al., 2007).

4.3.1. MISE EN PLACE DE L'AXE ANTEROPOSTERIEUR (A/P)

Dès leur formation, les primordia de pattes et d'antenne sont subdivisés en compartiments cellulaires antérieur et postérieur (Morata, 2001). Le compartiment postérieur est défini par l'expression du gène à homeobox *engrailed* (*en*) qui se comporte comme un gène sélecteur en réprimant les gènes spécifiques du compartiment antérieur, tels que, *cubitus interruptus* (*ci*), *patched* (*ptc*), *decapentaplegic* (*dpp*) et *wingless* (*wg*) (Sanicola et al., 1995 ; Tabata et al., 1995 ; Zecca et al., 1995). En revanche le compartiment antérieur est défini par l'activité du morphogène *hedgehog* (*hh*), qui est induit par *en* dans le compartiment postérieur. La diffusion de la protéine Hh dans le compartiment antérieur va induire dans ce dernier l'expression de molécules de signalisation codées par *decapentaplegic* (*dpp*) et *wingless* (*wg*), qui vont permettre par la suite de définir l'axe D/V des disques (Figure 32) (Basler et Struhl, 1994; Campbell et al., 1993; Diaz-Benjumea et al., 1994).

4.3.2. MISE EN PLACE DE L'AXE DORSO-VENTRAL (D/V)

En réponse à la signalisation Hh, les deux morphogènes *dpp* et *wg* sont induits, par la forme activée de Ci (Ci155), respectivement dans les cellules antérieures dorsales et ventrales qui se trouvent à proximité de la frontière antéropostérieure (Aza-Blanc et al., 1997; Chen et Struhl, 1998 ; Basler et Struhl 1994). Cette expression localisée en bande pour *wg* et *dpp* (Figure 32) est due à la répression mutuelle qui s'exerce entre ces deux morphogènes (Jiang et Struhl, 1996; Johnston et Schubiger, 1996; Theisen et al., 1996). En effet, lorsque la signalisation Wg (ou Dpp) est abolie dans les disques de pattes ou d'antenne, l'expression de *dpp* (ou *wg*) s'étend tout le long de la frontière A/P (Theisen et al., 1996). Cet antagonisme mutuel a été confirmé par des expériences de gain de fonction : la surexpression de Wg dans un disque de patte conduit à la répression de *dpp* et inversement (Jiang et Struhl, 1996; Theisen et al., 1996 ; Penton et Hoffmann, 1996 ; Johnston et Schubiger, 1996). L'antagonisme entre Wg et Dpp a été noté aussi au niveau de la régulation de gènes cibles spécifiques, ce qui permet une meilleure régionalisation D/V des disques. Par exemple, *optomotor-blind* (*omb*) est activé par Dpp et réprimé par Wg, alors que *H15* est activé par Wg et réprimé par Dpp (Brook et Cohen, 1996).

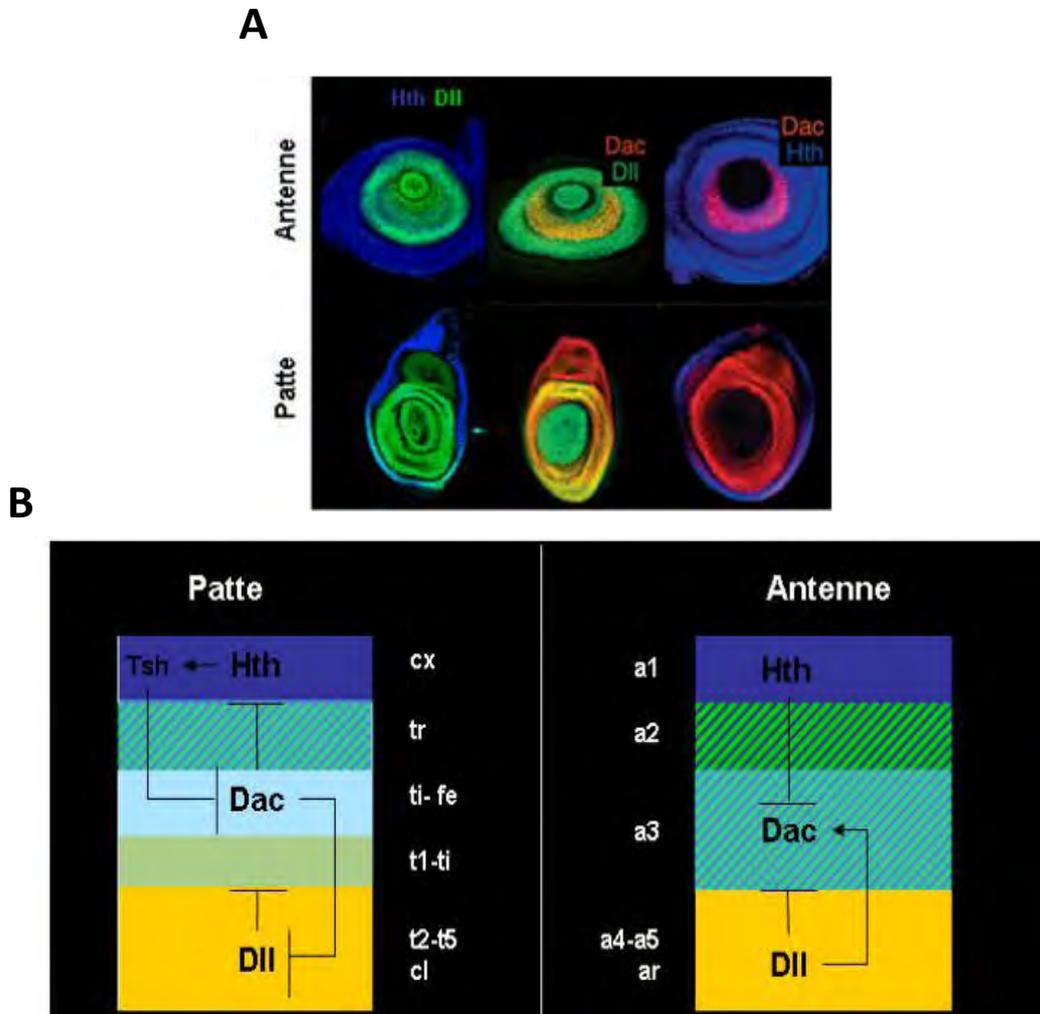


Figure 33. Relations génétiques et patrons d’expression des acteurs de l’axe P/D. (A) Expression des protéines Dll, Hth et Dac dans les disques de pattes et d’antenne. (B) Relations génétiques entre *dac*, *hth* et *Dll* dans les tissus de pattes et d’antenne (D’après Dong et al, 2000 (A), Dong et al., 2001 (B))

4.3.3. MISE EN PLACE DE L'AXE PROXIMO-DISTAL (P/D)

✓ SUBDIVISION PRÉCOCE: HTH/DAC/DLL

Le développement des appendices requiert l'établissement d'un troisième axe, l'axe proximo-distal (P/D). D'une manière intéressante, l'intersection des signaux à partir des deux axes perpendiculaires A/P et D/V est suffisante pour définir ce troisième axe (Meinhardt et al., 1983). L'axe P/D des pattes et de l'antenne est établi par l'activité combinée des signalisations Wg et Dpp, qui se croisent au centre du primordium des membres (Figure 32). Ensemble Wg et Dpp vont ainsi induire l'expression de *Distal-less (Dll)*, codant une protéine à homéodomaine nécessaire au développement de toutes les structures distales des appendices (Cohen et al., 1989; Diaz-Benjumea et al., 1994), et *dachshund (dac)*, codant une protéine nucléaire nécessaire pour les segments intermédiaires (fémur et tibia) des pattes et le segment a5 de l'antenne. Ainsi, au début du troisième stade larvaire, le primordium de la patte se trouve subdivisé en au moins trois régions exprimant les protéines Hth dans la partie proximale, Dac dans la partie médiale et Dll dans la partie distale (Figure 33). Le gène *dac* commence à être exprimé à la fin du deuxième stade larvaire (Abu-Shaar et Mann, 1998; Galindo et Couso, 2000). A ce stade du développement, les protéines Dac et Dll ne sont pas co-localisées. Par la suite, à partir du milieu du stade L3, une région exprimant à la fois les deux gènes apparaît permettant ainsi la définition de trois territoires distincts: Dac+Dll-/ Dac+Dll+/ Dac-Dll+ correspondant respectivement aux régions: fémur, tibia/ tibia, tarse 1/ tarse 2-5 et la griffe. Il est à noter qu'un peu plus tardivement au cours du développement, un anneau supplémentaire exprimant Dll apparaît dans les disques de pattes au niveau de la région à l'origine du trochanter.

Ainsi à la fin du troisième stade larvaire la patte présente l'organisation suivante:

Hth +, Dac -, Dll -: coxa.

Hth +, Dac +, Dll -: trochanter.

Hth -, Dac +, Dll -: fémur et première moitié du tibia.

Hth -, Dac +, Dll +: deuxième moitié du tibia et tarse 1.

Hth -, Dac -, Dll +: tarse 2-5 et griffes.

Dans le disque d'antenne en revanche la disposition des domaines d'expression de Hth, Dac et Dll diffèrent fortement de celles établies dans les disques de pattes où ces trois domaines se chevauchent partiellement. En effet, au sein du disque d'antenne, l'expression de *Dll* s'étend plus proximale et inclut le domaine d'expression de *dac* alors qu'au sein d'un disque de patte, l'expression de ces deux gènes est principalement exclusive.

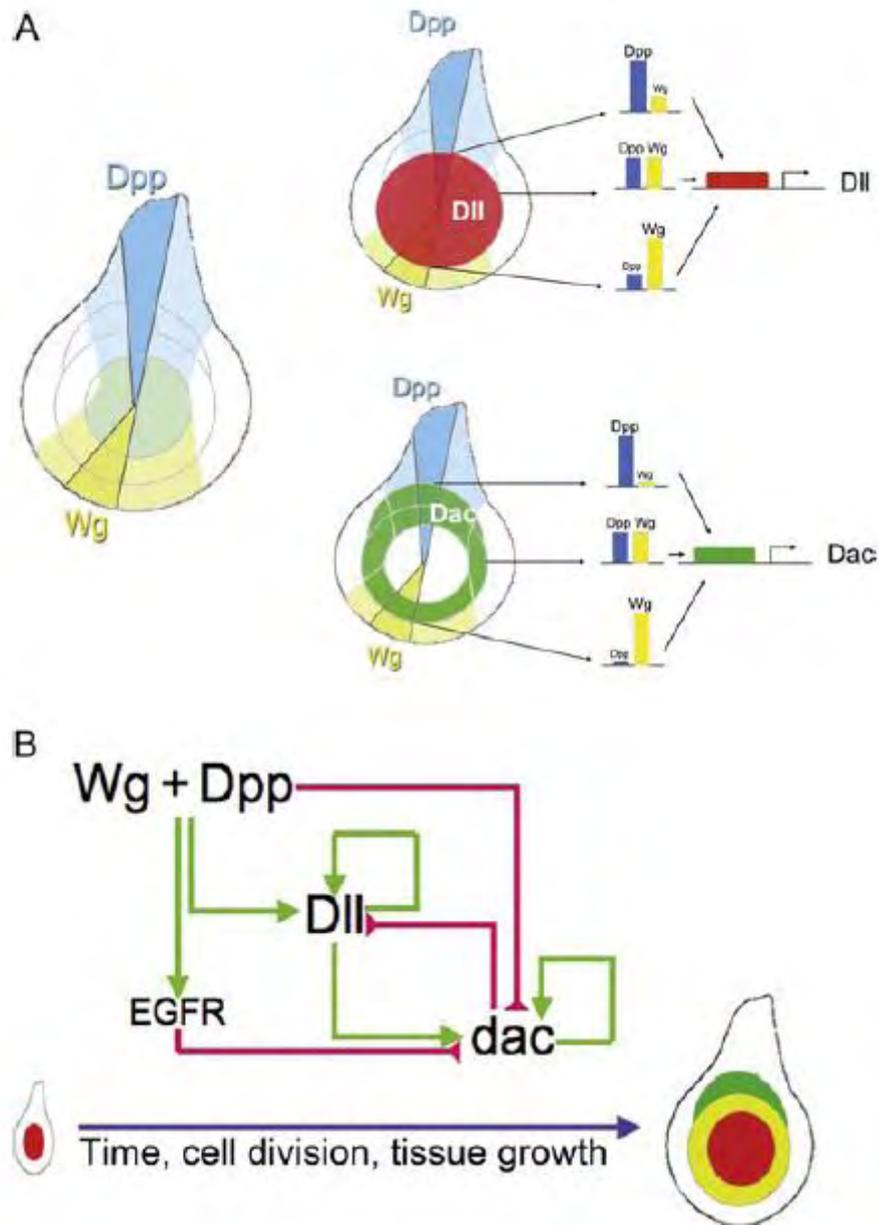


Figure 34. Modèles en gradient versus en cascade. (A) Modèle en gradient. Selon ce modèle, suivant la position de la cellule dans les disques, les CRMs *Dll* et *dac* doivent interpréter des niveaux de signalisation de Dpp et Wg très différents. (B) Modèle en cascade. Selon ce modèle les voies Dpp et Wg sont nécessaires seulement pour initier la formation de l'axe PD en activant *Dll* et les ligands de la voie EGFR. *Dll* à son tour active *dac*, et par la suite *Dll* et *dac* maintiennent leur expression d'une manière indépendante de Wg et Dpp. L'activité EGFR maintient la répression *dac*, tandis que Wg et Dpp répriment *dac* dans la partie centrale du disque de patte au début du développement. (Estella et al., 2012).

De même, *hth* s'exprime sur un territoire plus large dans l'antenne que dans la patte en développement. En effet, *hth* s'exprime sur tout le disque antennaire durant le premier et le deuxième stades larvaires et ne se rétracte que partiellement au stade larvaire L3, de la partie la plus distale du disque à l'origine des segments a4, a5 et arista (Emmons et al., 2007). En revanche, durant les stades larvaires L2 et L3 l'expression de *hth* est restreinte aux cellules les plus proximales du disque de patte (Abu-Shaar et Mann, 1998; Wu et Cohen, 1999; Casares et Mann, 2001). Il est à noter que *hth* est plus faiblement exprimé dans le segment a3 que dans les segments a1 et a2, à cause de la répression exercée par *ss* dont l'expression s'étend du segment a3 aux segments les plus distaux (Emmons et al., 2006). Ainsi, à la fin du stade L3 les domaines d'expression des protéines Hth, Dac et Dll apparaissent suffisants pour la délimitation des différents segments de l'antenne (du proximal vers le distal) (Figure 33):

Hth +, Dac -, Dll -: a1.

Hth +, Dac -, Dll +: a2.

Hth +, Dac +, Dll +: a3.

Hth -, Dac -, Dll +: a4, a5 et arista.

Dans l'antenne, Dac s'exprime dans un domaine restreint comparé à la patte, sous forme de croissant au niveau du primordium du segment a3. La perte de la fonction *dac* dans les tissus d'antenne n'entraîne qu'un phénotype subtil se limitant à la fusion du segment a5 avec l'arista (Mardon et al., 1994; Lecuit et Cohen, 1997; Dong et al., 2002).

✓ REGULATION DE L'EXPRESSION DE *DAC* ET *DLL* PAR LES VOIES WG ET DPP

Selon Lecuit et Cohen (1997), l'expression de *dac* et *Dll* (dans le centre des disques de pattes) le long de l'axe PD dépend des niveaux de signalisation de Wg et Dpp qu'une cellule perçoit: de fortes concentrations de Wg et Dpp activent *Dll* et répriment *dac* dans le centre du disque de patte ; des niveaux intermédiaires activent *dac* mais pas *Dll* dans la région médiane du disque ; et de faibles niveaux de ces morphogènes n'activent ni *dac* ni *Dll* (Figure 34A). Bien que très attrayant, le modèle du gradient est difficile à imaginer à l'échelle moléculaire, où les effets des voies de signalisation Wg et Dpp doivent converger vers les CRMs de *Dll* et *dac* (Figure 34A). Le modèle est également difficile à concilier avec l'observation que les disques double mutants pour *dpp* et *brk*, qui ne disposent pas d'un gradient de Dpp, ont un axe PD (Estella et Mann, 2008). La dissection des CRMs de *Dll* et de *dac* a fourni des indications sur la façon dont ces gènes répondent aux signalisations Wg et Dpp.

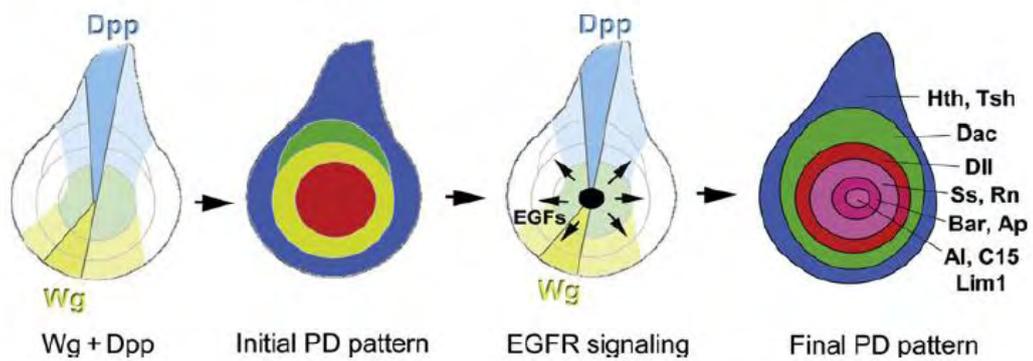


Figure 35. Signalisation EGFR dans les segments tarsaux. Après que les premiers domaines d'expression de l'axe PD primaire soient mis en place, les ligands de la voie EGFR sont produits au centre du disque de patte et vont activer l'expression d'une série de gènes cibles secondaires (*C15/cll*, *al*, *Bar*, *rn*, etc..) dans les primordia des segments du tarse et du pré-tarse. (Estella et al., 2012).

Comme décrit précédemment (page 19), le CRM *Dllt* (CRM précoce activé par Wg et Dpp) est le premier activé dans la progéniture des cellules du télépodite à la fin de l'embryogenèse. Bien que *Dllt* reste actif durant le développement larvaire, au troisième stade larvaire son activité est restreinte aux cellules qui se trouvent au centre du disque, à proximité de l'endroit où les domaines d'expression de Wg et Dpp se croisent, ce qui représente un sous-ensemble du domaine d'expression de *Dll* à ce stade (Figure 8). En outre, contrairement à *Dll*, *Dllt* continue à dépendre des signaux Wg et Dpp (Estella et al., 2008).

La dissection moléculaire d'un CRM du gène *dac*, *dacRE* (pour Ring Enhancer), qui récapitule la majorité du patron d'expression médial du gène, suggère que *dac* n'a pas non plus besoin des gradients de Wg et Dpp pour son activation dans la partie médiane de la patte (Giorgianni et Mann, 2011). L'élimination de la plupart des sites de liaison présumés de Pan et de Mad n'a pas ou très peu d'effets sur l'activité *DacRE* au troisième stade larvaire dans les disques de pattes, ce qui suggère que cet élément n'intègre pas les signaux intermédiaires de Wg et Dpp. Au lieu d'être activé par Wg et Dpp, *DacRE* est directement activé par *Dll*. En effet, d'après des expériences de traçage cellulaire, le domaine *Dac* dérive des cellules exprimant *Dllt* (McKay et al., 2009). En outre, la répression de *DacRE* dans l'extrémité distale de la patte par Wg et Dpp est transitoire et est maintenue par d'autres facteurs de transcription exprimés par la suite dans le développement. En résumé, en opposition avec le modèle du gradient, ces résultats suggèrent que la phase de formation de l'axe PD dépend d'une cascade génétique, dans laquelle Wg et Dpp activent *Dll*, et en retour *Dll* active *dac* (Figure 34B).

✓ ETABLISSEMENT PROGRESSIF DES REGIONS TARSE ET PRETARSE (VIA L'ACTION DES FTS RN, BAR, AL, C15/CLL ET DLIM1..).

L'intercalation des segments tarsaux se fait de manière progressive (Natori et al., 2012) et nécessite l'activation de la voie EGFR qui va médier trois effets, d'après les travaux de Galindo et collaborateurs (2002) et Campbell (2002) (Figure 35) : l'activation d'*aristaless* (*al*) et de *dlim1* au centre, l'activation de *Bar* (*BarH1/BarH2* collectivement désignés *Bar*; Higashijima et al., 1992) dans les cellules voisines, et la répression de *bab* et de *rn* dans la région centrale (à l'origine du t5 et du pré-tarse). D'après Galindo et collaborateurs (2002) l'activation de la voie EGFR dépend des voies de signalisation Wg et Dpp qui vont permettre la sécrétion de Vein (Vn), une protéine neuregulin-like et un ligand du récepteur EGFR, au

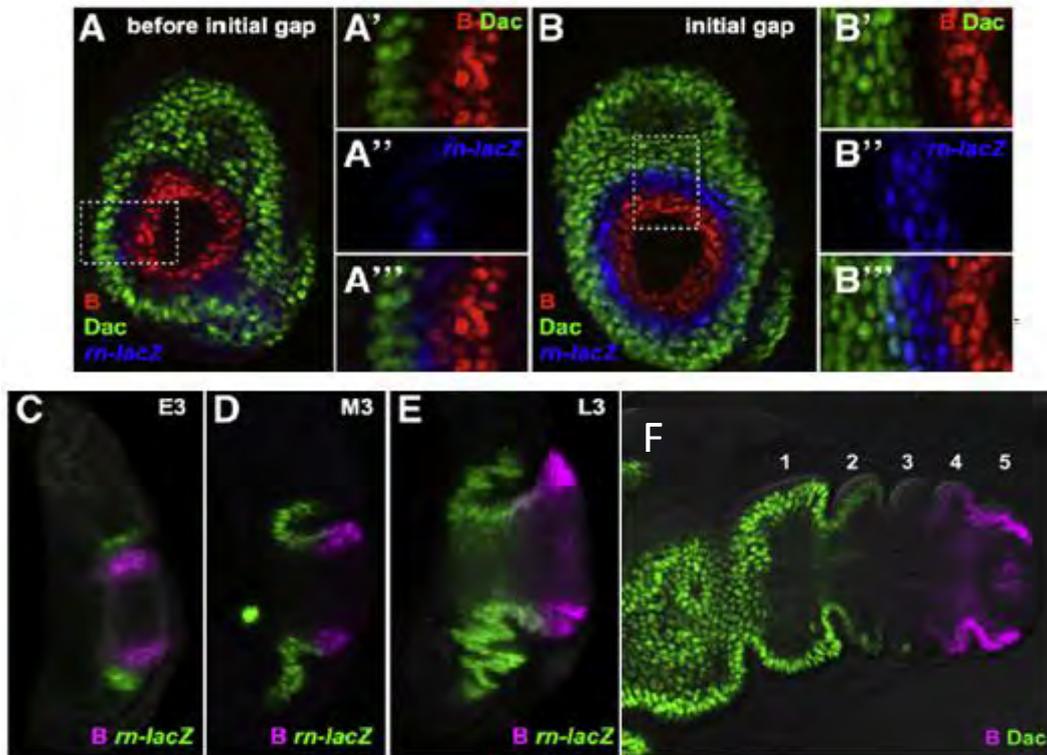


Figure 36. Profil d'expression de *rn* et répression progressive de l'expression de *Bar* dans *ta3* par *rn*. (A-F) Disques de pattes à différents stades. (A) L'expression de *rn* commence dans la région « gap » qui apparaît entre les domaines d'expression de *Bar* et *dac* et s'élargit à la fois distalement et proximale. (B-C) L'expression de *rn* chevauche celle de *dac* mais reste globalement exclusive de celle de *Bar*. (D, E) À la fin du troisième stade larvaire, l'expression de *rn* chevauche celle de *Bar*. (F) Expression de *Bar* et de *dac* dans les disques de pattes au stade pré-pupal.

centre des disques de pattes (Schnepp et al., 1996), aux alentours de 72 h après la ponte de l'œuf (APO). Ainsi l'activation de la voie EGFR va se traduire par l'activation du transducteur de signal EGFR-Ras, mitogen-activated protein kinase (MAPK) (Schweitzer et al., 1997) et ceci à partir de 80 heures et jusqu'à 96 heures APO. En utilisant des mutations thermosensibles Galindo et collaborateurs (2002) ont montré aussi que les signalisations Wg et Dpp sont nécessaires au développement de l'axe PD des pattes jusqu'à 84 heures APO, au-delà, la perte de l'expression de *wg* ou de *dpp* n'a plus d'effet sur l'organisation de l'axe PD et engendre seulement des défauts du côté ventral ou dorsal, respectivement.

Au stade L3 précoce l'expression de *Bar* débute dans la partie centrale des disques de pattes puis se résout rapidement en anneau qui jouxte proximale le domaine d'expression de *dac* et distalement celui d'*al* et de *C15/Clowless (Cll)*. Puis, avant que le pli central ne se forme (« central fold »), un repliement caractéristique du disque de patte utilisé comme repère temporel par Kojima et collaborateurs (2000), une région « gap » exprimant *rn* dépourvue de protéine Bar et chevauchant légèrement avec la protéine Dac apparaît entre les domaines d'expression de *Bar* et de *dac* (Kojima et al., 2000). Par la suite la région « gap » s'élargit et le domaine d'expression de *rn* s'étend distalement sur le territoire d'expression de *Bar* et le réprime (Figure 36). *rn* s'exprime en anneaux transitoires entre 84 à 96 heures APO dans les disques de pattes et durant tout le stade larvaire dans le disque d'antenne (St Pierre et al., 2002). En utilisant le gène rapporteur *rn-lacZ* (avec expression persistante de *lacZ*) il a été montré que l'anneau de cellules exprimant *rn* dans les disques de pattes donne lieu à la majeure partie du tarse (St Pierre et al., 2002). *rn* est crucial pour le développement du tarse et sa mutation engendre la fusion de tous les segments tarsaux. Il est également requis pour le développement des segments homologues dans l'antenne. En effet, on note chez les mutants *rn*- la perte du segment a4 et une fusion du segment a5 et de l'arista. L'expression de *rn* est induite par *ss* qui s'exprime également de manière transitoire dans les disques de pattes et permanente dans le disque d'antenne. Il est à noter que la perte de fonction de *ss* donne un phénotype qui se rapproche de celui obtenu avec une perte de fonction *rn*. Au même titre que *rn* et *ss*, les gènes *bab* permettent aussi de spécifier la région médiane du tarse, leur fonction est décrite dans le paragraphe 6.

A la fin du troisième stade larvaire, une forte expression de *Bar* est induite dans la région distale du domaine *Bar* alors que le gène *apterous (ap)*, codant pour un facteur de

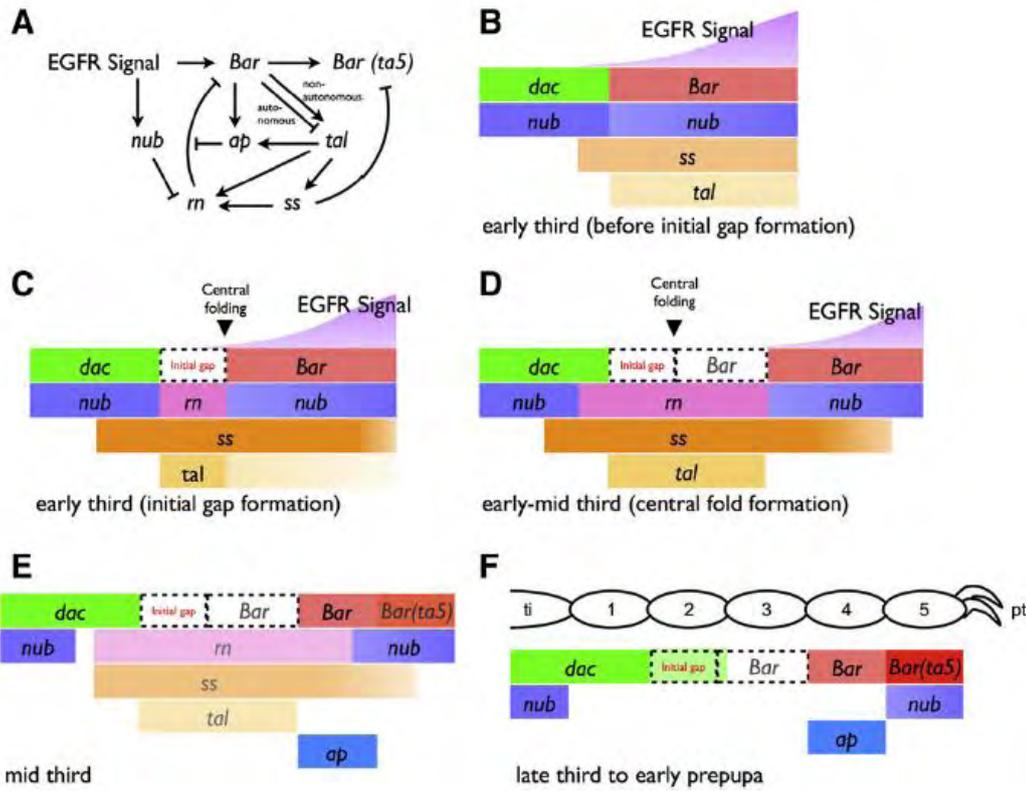


Figure 37. Modèle d'interactions génétiques dans le patterning du tarse. (A) Interactions génétiques entre les gènes de segmentation des tarse. Les flèches et les barres en T indiquent respectivement l'activation et la répression. (B-F) Profils d'expression des gènes d'intérêt pendant la structuration du tarse. Les rectangles en pointillés étiquetés comme «initial Gap» (écart initial) en rouge et «Bar» en gris indiquent respectivement la région «Gap» initiale entre les domaines d'expression de *Bar* et de *dac*, et la région proximale du «central fold» (pli central) qui perd l'expression de *Bar*.

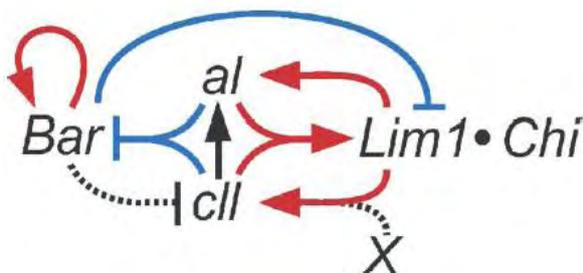


Figure 38. Interactions fonctionnelles entre les gènes impliqués dans le développement du t5 et du pré-tarse. Les flèches en rouge indiquent une activation et celles en bleu une répression. Les flèches en trait interrompu représentent des hypothèses non confirmées. (Kojima et al., 2005)

transcription à homéodomaine de type LIM, est exprimé dans la région proximale du domaine *Bar* (Kojima et al., 2000; Pueyo et al., 2000). *ap* qui commence à être exprimé au milieu du troisième stade larvaire permet de maintenir l'expression de *Bar* dans la future région t4 où *rn* est exprimé.

Récemment Natori et collaborateurs (2012) ont constaté que le gène *nubbin* (*nub*), qui code pour un facteur de transcription à homéodomaine et à domaine POU, est exprimé sur un large territoire y compris la région tarse au début du troisième stade larvaire. Les auteurs ont montré que l'expression de *rn* qui apparaît entre les domaines *Dac* et *Bar* provient de la perte d'expression de *nub* qui a une activité répressive sur l'expression de *rn* (Figure 37). En outre, il a été démontré que la séparation entre les domaines d'expression de *Bar* et *Dac* est régulée par *tarsal-less* (*tal*) qui induit l'expression de *ss* et *rn* et la croissance cellulaire. *tal*, connu aussi sous le nom de *polished rice* (*pri*), code pour quatre petits peptides agissant d'une manière cellule non-autonome (Galindo et al., 2007; Kondo et al., 2007). Récemment, *tal* a été reporté comme régulateur du clivage du FT à doigt de zinc Shavenbaby, impliqué dans la formation des denticules et la segmentation des tarse (Kondo et al., 2010; Pueyo et Couso, 2011). Il a été démontré également que *tal* est nécessaire à l'expression d'*ap* dans la future région ta4. Les pattes mutantes pour *tal* perdent presque tous les segments tarsaux à l'exception du ta5 (Pueyo et Couso, 2008). Ainsi, *tal* semble jouer un rôle fondamental dans la structuration de la région médiane du tarse (Pueyo et Couso, 2008).

Les domaines d'expression d'*al* et de *Bar* se chevauchent partiellement au troisième stade larvaire précoce, mais deviennent exclusifs ultérieurement. *Bar* atténue l'expression d'*al* par la répression de *dLim1* qui code pour une protéine à homéodomaine LIM-(LIM-HD) qui est nécessaire à une forte expression d'*al* (Kojima et al, 2000; Pueyo et al, 2000; Tsuji et al, 2000). En revanche *Al* seule n'est pas suffisant pour la répression de *Bar* et nécessite l'action conjointe de *C15/Cll* (Kojima et al., 2005), un homologue de *Hox11/Tlx* de vertébrés. *cll* (appelé aussi *C15*; Dear et Rabbitts, 1994) est exprimé spécifiquement dans le futur prétarse et est essentiel pour sa mise en place. L'expression maximale de *cll* nécessite *dLim1* et *Chip*, qui code pour un co-facteur de *dLim1* (Lilly et al., 1999; Morcillo et al., 1997; Pueyo et Couso, 2004). Les analyses biochimiques indiquent qu'*Al* et *Cll* forment un complexe qui se lie à des séquences spécifiques d'ADN, qui ne peuvent pas être reconnues par *Al* ou *Cll* seuls. Les régulations mutuelles entre *Bar*, *al*, *cll* et *dLim1* (Figure 38) (Campbell, 2005; Kojima et al., 2000, 2005; Pueyo et al., 2000; Tsuji et al., 2000) vont permettre d'établir précisément la frontière entre tarse5 et le prétarse. Finalement, au stade pré-pupal, la

subdivision de la région du tarse en cinq sous-régions correspondant à ta1-5 peut être distinguée par l'expression forte de *Bar* (ta5), l'expression modérée de *Bar* (ta4), pas d'expression de *Bar* ni de *dac* (ta3), l'expression faible de *dac* (ta2) et l'expression forte de *dac* (ta1) (Figure 36). La future région prétarse est aussi marquée par l'expression *d'al*, *cll* et *dLim1* (Campbell, 2002; Galindo et al., 2002). Il est à noter que l'expression de *dac*, *rn*, *Bar*, *al* et *cll* dans l'antenne présente la même disposition que celle décrite pour la patte. Toutefois certaines interactions ont été étudiées uniquement dans la patte et n'ont pas été examinées dans l'antenne.

4.4. FORMATION DES JONCTIONS ENTRE LES SEGMENTS DES PATTES

Les jonctions sont des structures flexibles reliant les différents segments de la patte. Selon leurs structures et leurs fonctions (Bishop et al., 1999 ; Fristrom et Fristrom, 1993; Mirth et Akam, 2002; Tajiri et al., 2010) on peut distinguer: (i) les vraies jonctions, qui séparent les segments coxa, trochanter, fémur, tibia, tarse et prétarse, caractérisées par la présence d'attachements musculaires et une diversité morphologique propre à chaque segment ; et (ii) les jonctions des segments tarsaux, caractérisées par une structure identique et l'absence d'attachements musculaires.

La différenciation des jonctions est un processus complexe qui implique des changements de la forme cellulaire, de l'adhérence entre les cellules et de la distribution de l'actine filamenteuse et des protéines de la matrice extracellulaire (Fristrom et Fristrom 1993; Mirth et Akam 2002). Sur le plan moléculaire la formation des jonctions nécessite l'activation de la voie Notch (Celis et al., 1998; Bishop et al., 1999; Rauskolb et Irvine, 1999) dans les tissus présomptifs des jonctions. Ces derniers se trouvent normalement au niveau des cellules les plus distales de chaque segment et sont distalement adjacents aux domaines d'expression des ligands de Notch : Delta (Dl) et Serrate (Ser). La perte de fonction de *Notch* ou de ses ligands conduit à la perte des jonctions et à la fusion des segments de la patte. Le signal Notch est également suffisant pour induire la formation des jonctions. En effet, des jonctions ectopiques peuvent être induites lorsque Notch ou ses ligands sont activés de manière ectopique (de Celis et al., 1998; Bishop et al., 1999; Rauskolb et Irvine 1999).

4.4.1. ACTIVATION DE LA VOIE NOTCH

La voie signalisation Notch (N) est une voie de communication intercellulaire très conservée, qui est activée suite à l'association des ligands transmembranaires de la famille DSL (Delta (Dl), Serrate (Ser) et LAG-2, à la partie extracellulaire du récepteur Notch

(Bray, 2006; Fleming et al., 1997). Lors de la liaison avec ses ligands, le récepteur N subit consécutivement deux clivages protéolytiques par les complexes ADAM-métalloprotéase et γ -sécrétase, et libère le domaine N intracellulaire (N^{icd}) (Bray, 2006; Fortini, 2009). Par la suite, le N^{icd} transloque dans le noyau où il se lie au complexe transcriptionnel CSL (CBF1/RBP-1 κ chez les mammifères, Suppressor of Hairless (Su(H)) chez la drosophile et le xénope, Lag-1 chez *C. elegans*) et active la transcription de gènes cibles (Bailey et Posakony, 1995; Bray, 2006; Lecourtois et Schweisguth, 1995).

4.4.2. EFFECTEURS DE LA VOIE NOTCH

Plusieurs gènes codant pour des facteurs de transcription sont connus comme médiateurs de la fonction de la voie Notch au cours du développement des pattes parmi lesquelles les gènes de la famille d'*odd-skipped* (*odd*) : *odd*, *bowl*, *sister of odd and bowl* (*sob*) et *drumstick* (Figure 39). Ces derniers codent pour des protéines présentant une forte similitude entre leurs domaines à doigts de zinc (Coulter et al., 1990; Hart et al., 1996; Wang et Coulter, 1996; Green et al., 2002) et sont co-exprimés dans les cellules formant les jonctions, sauf celles de la région du tarse (Rauskolb et Irvine 1999; de Celis Ibeas et Bray 2003; Hao et al., 2003). Bien qu'aucun de leurs mutants simples ne présente de défauts apparents dans la formation des joints dans lesquels ils sont exprimés, (probablement à cause d'une redondance fonctionnelle (de Celis Ibeas et Bray 2003; Hao et al., 2003), les études d'expression ectopique ont montré qu'ils sont capables d'induire des jonctions ainsi que des changements dans la forme des cellules et la distribution de l'actine (Hao et al., 2003), comme c'est le cas lors de l'activation ectopique de Notch (de Celis et al., 1998; Bishop et al., 1999; Rauskolb et Irvine 1999). Par conséquent, les gènes de la famille *odd* sont considérés comme des médiateurs de l'activité de la voie Notch dans la formation des jonctions à l'extérieur de la région du tarse. En plus de la famille *odd* plusieurs autres gènes sont des effecteurs de la voie Notch : *Activator protein-2* (*dAp-2*), *Enhancer of Split* (*E(spl)*), *disconnected* (*disco*) et *big brain* (*bib*) (Bishop et al., 1999; de Celis et al., 1998; Kerber et al., 2001). Dans le cas du gène *Ap-2*, la mutation entraîne un raccourcissement sévère des pattes avec des défauts de formation de jonctions. Cependant, *dAp-2* est incapable d'induire des jonctions en absence de l'activité Notch quand il est exprimé de manière ectopique, montrant qu'il a besoin d'un ou plusieurs autre (s) gène(s) cible(s) de Notch pour fonctionner (Kerber et al., 2001 ; monge et al., 2001).

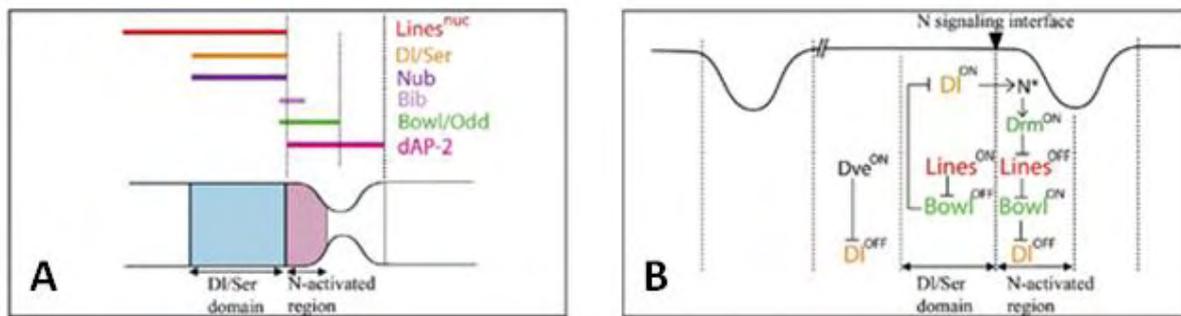


Figure 39. L'expression de *Dl* et des cibles de Notch définissent plusieurs domaines à travers les « vrais » segments de la patte. (A) Schéma représentant l'expression de *Dl* et des cibles de Notch au niveau de l'articulation tibia/tarse d'une patte en éversion. (B) Modèle décrivant les interactions qui régulent l'expression de *Dl* dans tous les vrais segments de la patte. *lines* antagonise *bowl* pour maintenir l'expression de *Dl* dans les cellules proximales. Ensemble, *lines* et *bowl* agissent comme un « Switch » binaire pour maintenir une interface stable de signalisation Notch à l'extrémité distale de chaque segment de la patte. L'expression de *Dl*, *dAP-2* et *bib* sont détectés dans tous les segments de la patte, tandis que celle de *bowl*, *odd* et *nub* n'est détectée uniquement que dans les vrais segments. (Greenberg et Hatini, 2009)

4.4.3. ROLE DE NOTCH DANS LA CROISSANCE DES PATTES

En plus de son rôle dans la formation des jonctions, Notch est aussi impliqué dans la croissance. En effet la perte de l'activité de la voie Notch conduit à une croissance réduite de la patte, et l'activation ectopique résulte en des excroissances de tissus de pattes (de Celis et al 1998; Bishop et al., 1999; Rauskolb et Irvine 1999). L'analyse clonale a indiqué que l'activité Notch est nécessaire de manière cellule non autonome dans les cellules formant les jonctions (de Celis et al., 1998). Il est donc suggéré que l'activité Notch contrôle également la croissance des segments de la patte en régulant une sorte de facteur sécrété induisant la croissance cellulaire et / ou empêchant la survie cellulaire.

4.4.4. ETABLISSEMENT DE LA FRONTIERE AU NIVEAU DES JONCTIONS

Le maintien de la répartition spatiale asymétrique entre l'expression du ligand et l'activation N est crucial pour la segmentation de la patte. En effet, l'expression ectopique des ligands dans la région qui répond à N réprime la formation des jonctions en diminuant de manière post-transcriptionnelle la signalisation N (Bishop et al., 1999; Rauskolb et Irvine, 1999). La distribution différentielle, entre le ligand dans les cellules signalantes et les récepteurs dans les cellules réceptrices, est accomplie par une boucle rétroactive négative par la signalisation N qui réprime l'expression du ligand dans les cellules réceptrices (Artavanis-Tsakonas et al., 1999; Fortini, 2009). Réciproquement, les cellules exprimant le ligand perdent leur propre capacité à répondre à Notch, par une répression transcriptionnelle ou post-transcriptionnelle de *N* (Bécam et al., 2010; Fortini, 2009; Miller et al., 2009). Ainsi, des boucles de rétroaction amplifient et renforcent la distribution différentielle du ligand et du récepteur. Les mécanismes de contrôle de la boucle de rétroaction de signalisation N diffèrent en fonction du contexte de développement (Heitzler et al., 1996; Huppert et al., 1997).

✓ DANS LES 'VRAIES' JONCTIONS

Il a été démontré que les limites nettes de signalisation N dans les vraies jonctions sont maintenues par une boucle de rétroaction négative entre la voie de signalisation N et la cassette de gènes '*odd-skipped (odd)-related drumstick (drm)-lines-bowl*' (Figure 39) (Greenberg et Hatini, 2009). Dans les cellules exprimant *Dl*, la protéine Lines est active et nucléaire conduisant à la déstabilisation de la protéine Bowl et à sa dégradation ultérieure (Greenberg et Hatini, 2009; Hatini et al., 2005). Dans les cellules adjacentes à la jonction la

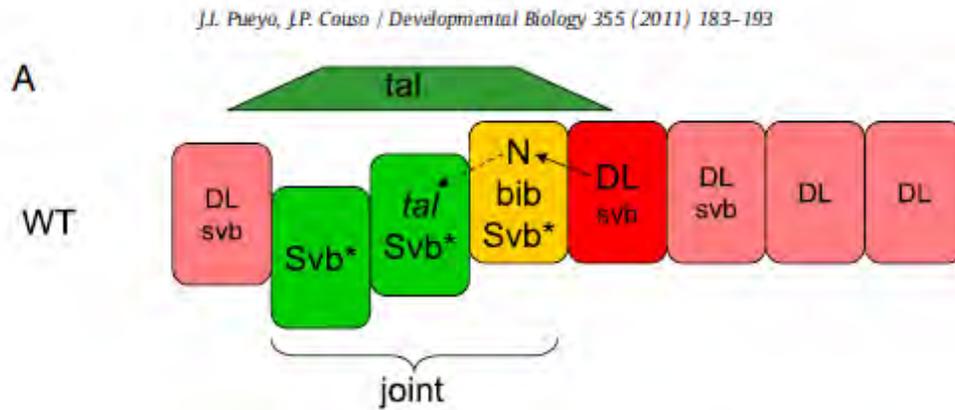


Figure 40. Modèle de la formation des jonctions dans les tarse de la patte. La représentation schématique montre le mécanisme médié par Tal, permettant la formation de la frontière de signalisation N dans la partie distale des segments du tarse dans une patte nymphale (les segments distaux du tarse sont représentés vers la gauche et les proximaux vers la droite). Une rétroaction négative entre la signalisation N et Tal régule la formation et l'entretien de la frontière de signalisation N. Les jonctions ne se forment que si a) l'expression de *tal* et *svb* se chevauchent (vert) et b) et si ce chevauchement Tal-Svb limite à son tour ou chevauche un haut niveau d'expression de *Dl* (rouge, chevauchement en jaune). Ces chevauchements conduisent à la génération d'une frontière de signalisation DI+/DI-. A la fin du troisième stade larvaire, l'expression de *Dl* (rouge) et *svb* se chevauchent légèrement dans la partie distale du segment. Au début du stade puppe, les cellules ont des niveaux élevés de *Dl* et activent la signalisation N dans les cellules adjacentes (flèche noire). La signalisation N active (directement ou indirectement) l'expression du gène *tal* dans la partie la plus distale des jonctions (flèche noire en pointillés). La signalisation Tal non autonome (vert) déclenche l'activation post-transcriptionnelle du facteur de transcription Svb (*Svb**) dans la région présumptive de la jonction. Par la suite, cette activation de Svb médiée par Tal résulte en la répression directe ou indirecte de la transcription de *Dl* dans les cellules de la jonction présumptive (jaune), générant une forte frontière de signalisation DI+/DI- qui conduit à l'activation de *bib*. La perte de la fonction *tal* ou de celle de *svb* résulte en la perte de cette frontière, et, par conséquent, des jonctions. (Pueyo et Couso, 2011).

signalisation N active l'expression de *drm*, la protéine Drm se liant alors à Lines et empêche ainsi la dégradation de Bowl. Par conséquent, Bowl peut s'accumuler dans le noyau où elle réprime l'expression de *Dl* (Greenberg et Hatini, 2009). Ce mécanisme élégant renforce la frontière de signalisation DI+/DI- et assure son maintien le long du développement de la jonction. Contrairement aux cellules les plus distales, la voie Notch n'est pas activée dans les cellules exprimant les ligands de Notch ni dans les cellules proximale-ment adjacentes. Le gène *fringe* (*fng*) qui est nécessaire au développement de la patte et qui code pour un modulateur de l'activité Notch (Irvine et Wieschaus 1994; Panin et al., 1997), a été proposé comme responsable de l'activation asymétrique de Notch le long l'axe PD dans chaque segment (de Celis et al 1998; Bishop et al., 1999; Rauskolb et Irvine 1999).

✓ DANS LES JONCTIONS TARSALES

La cascade Drm-Lines-Bowl utilisée pour le maintien des frontières au niveau des vraies jonctions ne peut expliquer la formation des jonctions tarsales puisque *lines* semble être exprimé dans toutes les cellules du tarse. Une étude récente menée par Pueyo et Couso (2011) a montré que le gène non canonique *tarsal-less* (*tal*), est exprimé dans les régions activées par N dans le tarse et est nécessaire pour le développement des jonctions inter-tarses au cours du développement nymphal. Cette fonction de *tal* est à la fois temporellement et fonctionnellement distincte de celle médiée par *tal* pour la formation des tarses au milieu du troisième stade larvaire. Dans la fonction nymphale, la signalisation N active l'expression de *tal* et réciproquement le peptide Tal rétrocontrôle N par la répression de la transcription de *Dl* dans les jonctions du tarse générant une frontière nette de signalisation (Figure 40). Cette répression Tal-induite de *Dl* est médiée par l'activation post-transcriptionnelle du facteur de transcription Shavenbaby (Svb), d'une manière similaire à celle qui a été décrite dans l'embryon. Ainsi, une boucle rétroactive négative impliquant Tal régule la formation et le maintien d'une frontière DI+/DI- dans les segments tarsaux.

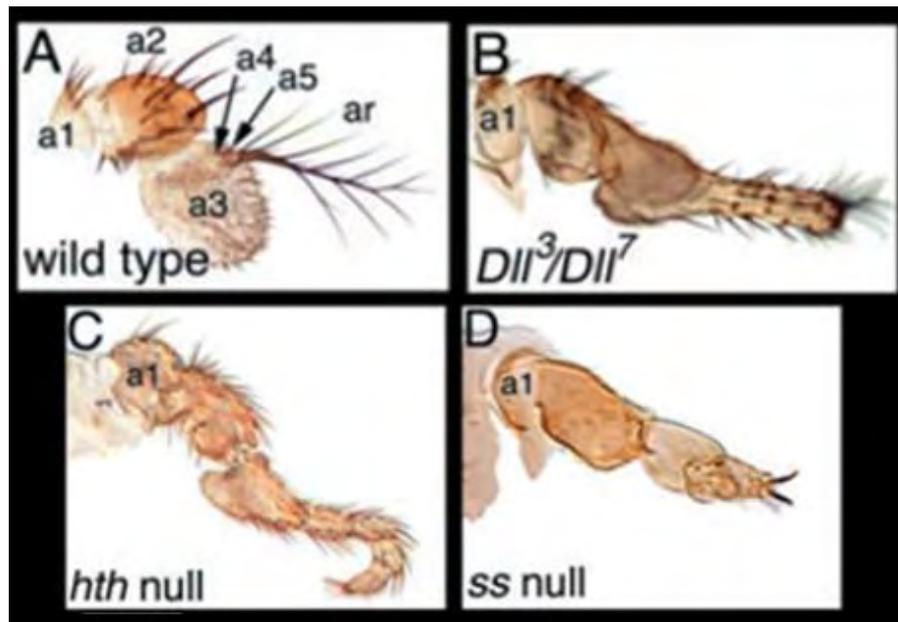


Figure 41. *Dll*, *hth* et *ss* sont essentiels pour l'identité antenne. Une combinaison allélique hypomorphique de *Dll* et les mutations nulles de *hth* et *ss* transforment toutes l'antenne en patte. (D'après Dong et al., 2000).

5. IDENTITE ANTENNE VERSUS PATTE

Chez la drosophile plusieurs mutations provoquent des transformations de l'antenne en patte. Les plus connues étant des mutations gain de fonctions dominantes du gène *Hox Antennapedia (Antp)* permettant à l'antenne de se développer comme une patte T2 complète. Les études menées par Struhl (1981, 1982) ont révélé que ces mutations provoquent l'expression ectopique d'*Antp* dans l'antenne. En revanche, il a montré au cours de ces mêmes études que la perte de fonction d'*Antp* a l'effet inverse, entraînant la transformation de structures de patte T2 en tissus antennaire, sans avoir pour autant un effet sur le développement de l'antenne. L'ensemble de ces résultats ont amené Struhl à proposer qu'*Antp* est normalement exprimé dans la patte T2, mais pas dans l'antenne, et que sa fonction est de réprimer l'activation des gènes antenne-spécifiques dans la patte T2. Au niveau des paires de pattes T1 et T3 l'identité est conférée respectivement par la présence des gènes homéotiques *Sex combs reduced (Scr)* et *Ultrabithorax (Ubx)*. En revanche aucun gène *Hox* n'est exprimé dans le disque d'antenne (Kaufman et al., 1990) ce qui suggère que l'identité antennaire serait donc définie par l'absence de fonction homéotique en général.

Il est maintenant clair que la majeure partie de l'identité de l'antenne est spécifiée par l'action combinée des protéines Dll et Hth (et sans doute Exd) (Casares et Mann, 1998; Dong et al., 2000) qui sont largement co-exprimées dans l'antenne. En effet, la perte de fonction d'*hth* ou d'*exd* entraîne une transformation cellulaire autonome des structures médiales et distales de l'antenne en structures médiales et distales de patte (Figure 41) (Casares et Mann, 1998; Gonzalez-Crespo et Morata, 1995; Pai et al., 1998). De plus, plusieurs combinaisons d'allèles hypomorphes de *Distal-less* transforment l'antenne en patte sans causer pour autant des tronçures proximo-distales (Dong et al., 2000). Il est intéressant de noter que l'activité de *Dll* dans la spécification de l'identité de l'antenne est génétiquement séparable de la fonction de patterning proximo-distal.

Les expressions de *Dll* et *hth* sont indépendantes l'une de l'autre. Cependant, l'activation des gènes cibles spécifiques impliqués dans l'identité antennaire requiert généralement l'expression combinée des protéines Hth et Dll (Duncan et al., 1998; Dong et al., 2000, 2002; Emmons et al., 2007). Parmi ces gènes citons *spineless (ss)*, *dan/danr*, *cut (ct)*, *sal/salr* et *atonal (ato)*.

5.1. *SPINELESS (SS)*

La plupart des fonctions d'identité de Hth et Dll dans la partie distale de l'antenne sont exécutées par le gène cible *ss* (Dong et al., 2002; Emmons et al., 2007). La protéine Ss, contenant un motif bHLH (domaine de fixation à l'ADN) et un motif PAS (motif intervenant dans les liaisons protéine-protéine), est un homologue des protéines murines et humaines AHR (Aryl Hydrocarbon Receptor) connues également sous le terme de récepteur à la dioxine (Duncan et al., 1998 ; Schmidt et Bradfield, 1996). Comme l'activation et la localisation nucléaire de la protéine AHR nécessite sa liaison à son ligand (la dioxine), il est suggéré que Ss puisse également se lier à un ligand.

L'expression de *ss* est détectée dans la partie distale du disque d'antenne dès la fin du second stade larvaire et sa limite proximale correspond précisément à la frontière entre les segments a2 et a3 au troisième stade larvaire. Ce facteur s'exprime également dans les disques de pattes mais de manière transitoire durant le deuxième stade larvaire dans la région tarsale t1 (partie distale)-t4. Au vu des phénotypes perte et gain de fonction de ce gène, *ss* jouerait un rôle crucial dans l'identité de l'antenne distale. En effet, la perte de fonction de *ss* transforme la partie distale de l'antenne en patte (Figure 41) et l'expression ectopique de ce gène est suffisante pour transformer les parties distales des pattes en tissus antennaires (Duncan et al., 1998). De plus, les transformations antenne en patte induites par diverses mutations (*Antp*, *Scr*, *Ubx*, *abd-A*, *Abd-B*, *Dll*) sont toutes corrélées à la répression de *ss* dans l'antenne, témoignant de sa fonction comme un déterminant de l'identité antenne (Duncan et al., 1998). L'expression de *ss* dépend de Dll et de Hth. En effet, la réduction de l'activité de *Dll* (hypomorphe viable) est associée à une perte de l'expression de *ss* dans l'antenne et dans les tarsi. De plus, l'expression de *ss* est abolie dans les clones somatiques *hth*- (Chu et al., 2002). Il est donc admis que *ss* code un facteur « réalisateur » intervenant dans la spécification de l'identité antennaire localisé en aval de la combinatoire Dll/Hth.

Afin de comprendre la régulation de *ss* dans l'antenne Emmons et collaborateurs (2007) ont cherché à identifier les éléments *cis*-régulateurs responsables de l'expression de *ss* dans les embryons et les disques imaginaux. Pour ce faire, les auteurs ont testé l'expression du gène rapporteur *lacZ* sous le contrôle de plusieurs fragments d'ADN, couvrant la majeure partie du gène *ss*. Suite à cette analyse, plusieurs enhanceurs reproduisant presque tous les aspects de l'expression de *ss* ont pu être identifiés, dont au minimum trois dirigent

l'expression dans le disque d'antenne. Emmons et collaborateurs (2007) ont choisi de travailler avec le fragment B6.9, étant donné que c'est le seul fragment qui dirige l'expression de manière spécifique dans l'antenne tout au long du développement. Par la suite, les auteurs ont testé l'activité de B6.9 dans des clones somatiques *hth*- et *Dll*- et ont montré qu'il est positivement régulé par *Dll* et *hth*. En revanche, *Dll* est nécessaire en permanence pour l'activation, alors que *hth* est nécessaire uniquement au début du développement. De plus, l'expression ectopique de *Dll* est suffisante pour activer cet enhancer antenne-spécifique, même en l'absence d'*hth*, ce qui suggère que *Dll* est son principal activateur.

Les auteurs ont également testé la régulation de *B6.9* par *cut* et *Antp* et ont montré que ces deux gènes régulent négativement son expression. De manière intéressante, *Antp* réprime l'expression de *B6.9*, même dans des clones induits longtemps après que *hth* ne soit plus requis, ce qui suggère que *ss* est une cible directe d'*Antp*, responsable de la transformation de la région distale de l'antenne en patte. Cette répression directe de *ss* indépendante d'*hth* est en accord avec le fait que des clones *Antp*- induits tardivement peuvent toujours transformer les parties distales de l'antenne en tissus de pattes alors que des clones *hth* ne le font plus. De par la constatation que les clones *Antp*- dans la patte montrent parfois une expression ectopique de *hth* dans la partie distale, Casares et Mann (1998) ont proposé que la fonction première d'*Antp* est de réprimer *hth* dans la partie distale de la patte, ce qui empêche alors l'activation des gènes cibles antennaires. Bien que ce point de vue soit largement accepté, il n'a pas fait l'objet de tests. Pour répondre à cette question, Duncan et collaborateurs (2010) ont étudié une sous-région de *B6.9* de 62 pb qui dirige l'expression spécifiquement dans le segment antennaire *a3*. Les auteurs ont pu mettre en évidence que cet enhancer répond de la même façon que le fragment *B6.9* entier vis-à-vis de *Dll*, *Hth* et *Exd*. De plus ils ont montré par gels retards que ces facteurs lient directement le fragment *a3*. De manière intéressante, la liaison de *Dll* montre une forte coopérativité avec *Hth* et *Exd*, ce qui indique que ces protéines se lient en tant que complexe. La liaison tripartite *Dll/Hth/Exd* suggère que *Dll* se comporte comme une protéine Hox en précisant l'identité de l'antenne. Étonnamment, *Antp* interagit aussi directement avec l'enhancer *a3*. De plus, *Antp* se lie en collaboration avec *Hth* et *Exd*, et antagonise la fixation de *Dll* au moins en partie, par compétition sur le site de fixation.

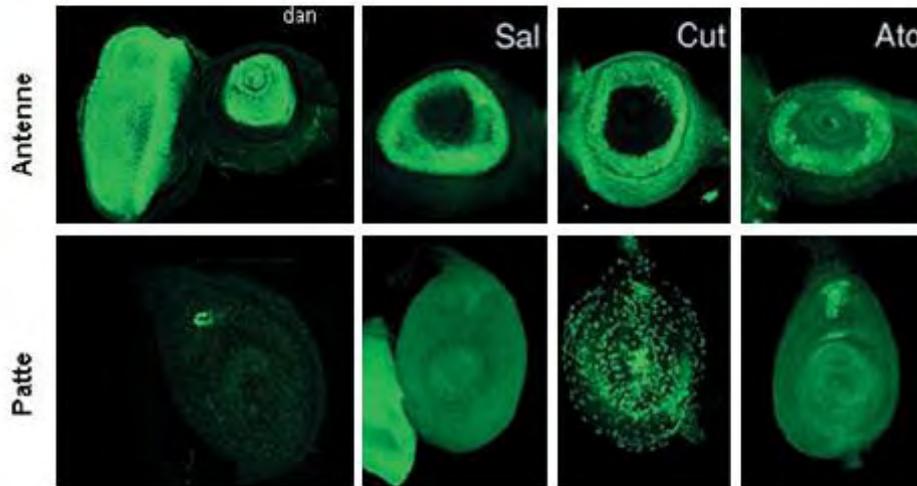


Figure 42. Expression génique différentielle antenne/patte. *sal*, *cut*, *ato* sont induits dans l’antenne indépendamment de *ss* tandis que *dan* est une cible de *ss*. (d’après Dong et al., 2002 et Cohen et al., 2003)

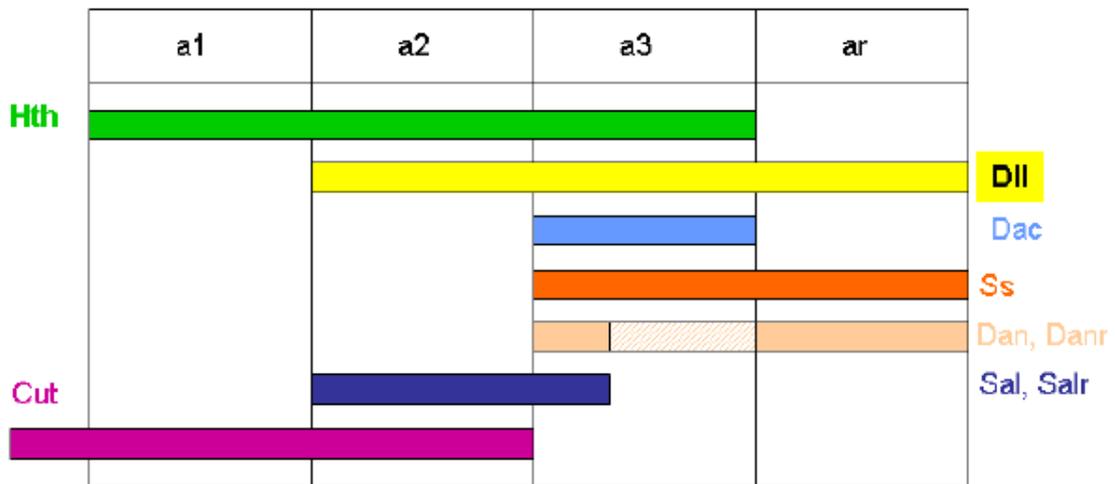


Figure 43. Domaines d’expression des principaux marqueurs de l’antenne le long de l’axe P/D.

Le fait qu'*Antp* interagisse directement avec l'enhancer *a3* a amené les auteurs à réexaminer le rôle d'*Antp* dans le développement des pattes. Ils ont constaté que l'enhancer *a3* est parfois activé dans les clones *Antp*- dans un disque de patte. De plus, une partie des cellules *Antp*- dans laquelle l'enhancer *a3* a été activé commencent à exprimer *ss*, *cut*, et d'autres marqueurs antennaires, indiquant une transformation en antenne. Cependant, l'activation de l'enhancer *a3* ne se produit que dans un anneau étroit de cellules dans la région proximale de la patte co-exprimant *Dll*, *hth*, et Exd nucléaire (Wu et Cohen, 1999). Sachant que l'expression de *hth* et *Dll* dans l'anneau proximal est inchangée dans les clones *Antp*-, les auteurs proposent que le rôle d'*Antp* ne soit pas de bloquer le développement antennaire dans la patte par la répression de *hth*, comme cela a été proposé mais plutôt de réprimer directement les gènes antennaires dans le disque imaginal de patte qui autrement seraient activés par l'expression combinée des FTs *Dll*, *Hth* et *Exd* dans l'anneau proximal.

5.2. *DISTAL ANTENNA (DAN)* ET *DISTAL ANTENNA RELATED (DANR)*

dan et *danr* sont deux gènes paralogues codant pour des protéines nucléaires, ayant un domaine de liaison à l'ADN de type « pipsqueak » (Emerald et al., 2003). *dan* et *danr* s'expriment dans l'antenne selon le même profil que *ss*, et jouxtent proximale le domaine d'expression de *cut* au niveau de la frontière a2/a3 (Emerald et al., 2003). Par contre dans les disques de pattes l'expression de *dan* et *danr* se limite à des petits groupes de cellules (Figures 42 et 43). Dans le disque d'antenne, *dan* et *danr* sont régulés positivement par *ss* et négativement par *cut* (Emerald et al., 2003). *dan* et *danr* participent à l'acquisition de l'identité antennaire dans la partie la plus distale. En effet, leur expression ectopique dans les disques de pattes peut activer l'expression des gènes *ss*, *hth* et *sal*, et transforme la griffe en arista.

5.3. *CUT (CT)*

Le gène *cut* (*ct*) code un facteur de transcription à homéodomaine (Blochlinger et al., 1993). Il s'exprime dans le disque d'antenne au niveau des segments a1 et a2 (Figure 42) (Blochlinger et al., 1993; Chu et al., 2002) alors que dans les disques de pattes il ne s'exprime que dans quelques petits groupes de cellules réparties sur l'ensemble des disques imaginaires de pattes (Blochlinger et al., 1993; Chu et al., 2002). L'expression de *cut* dans l'antenne dépend de *hth* mais pas de *Dll* et non plus de *ss*.

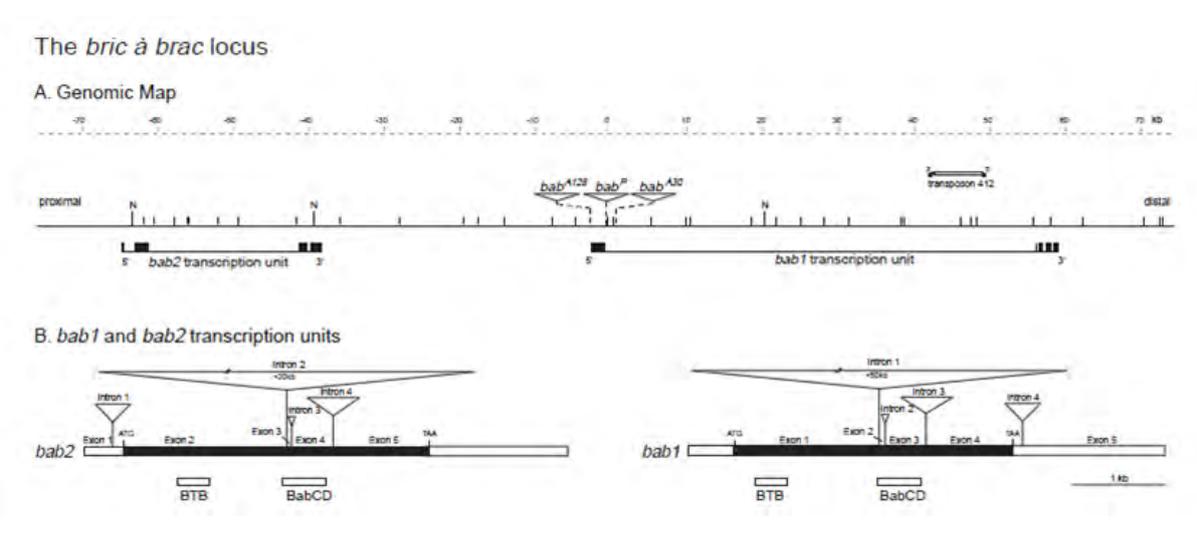


Figure 44. Le locus *bab* se compose de deux gènes. (A) Carte moléculaire. Les sites d'insertion des éléments P *babA128*, *babP* et *babA30* sont indiqués. Le point d'insertion *babP* est défini comme 0 sur la carte, le locus entier couvre environ 150 kb d'ADN. (B) Structure des unités transcriptionnelles *bab1* et *bab2*. Les exons (boîtes noires) et les positions et les tailles des introns sont affichés. Les régions codant pour les domaines BTB et BabCD sont également indiquées. (Couderc et al., 2002)

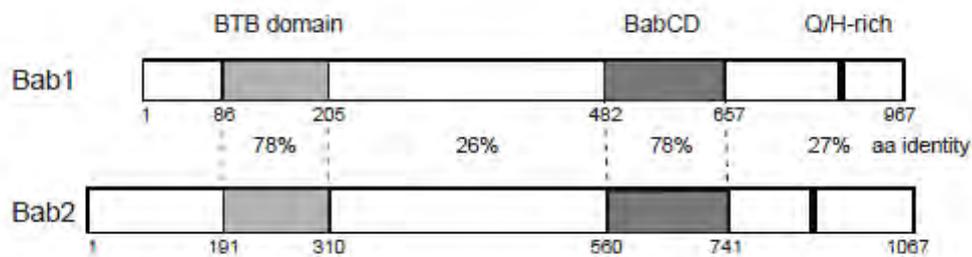


Figure 45. Bab1 et Bab2 possèdent un domaine BTB, un domaine Psq et un motif AT-hook-like. (A) La similarité de séquence entre Bab1 et Bab2 est limitée au domaine N-terminal BTB et au domaine C-terminal BabCD. L'homologie de séquence est indiquée en pour cent (Couderc et al., 2002).

5.4. LES GÈNES *SPALT (SAL)/ SPALT-RELATED (SALR)*

sal/salr sont deux gènes orthologues, fortement apparentés structurellement et fonctionnellement, qui codent pour des facteurs de transcription à doigts de zinc. Ils s'expriment dans l'antenne de manière similaire sous forme d'un anneau dans une région qui correspond au futur a2 (Figure 42) (Barrio et al., 1999 ; Chu et al., 2002). En revanche, ils ne sont quasiment pas exprimés dans les disques de pattes.

5.5. *ATONAL (ATO)*

Le gène *atonal (ato)* code un facteur de transcription à domaine bHLH. Dans le disque d'antenne, il s'exprime sous forme d'un anneau au niveau du segment a2 et dans les disques de pattes dans de petits groupes de cellules localisées dans la partie proximale dorsale (Figure 42) (Chu et al., 2002; Jarman et al., 1995). *ato* ainsi que les deux gènes *sal* et de *salr* sont requis pour la formation de l'organe de Johnston (JO) un organe d'audition (Carlson, 1996; Eberl, 1999; Sayeed et Benzer, 1996 ; Jarman et al., 1995) qui se trouve au niveau de la jonction a2/a3 et permet de percevoir les vibrations sonores en provenance de l'arista. En effet, la perte de ces gènes entraîne une réduction du segment a2 et la fusion de ce dernier avec le segment a3 ce qui engendre un problème d'audition chez la mouche (Chu et al., 2002).

6. LES GÈNES *BAB*

Les gènes *bab1* et *bab2* sont situés au niveau du locus *bab*, qui s'étend sur 148 kb, et les unités de transcription couvrent respectivement environ 60 kb et 25 kb de l'ADN génomique (Figure 44). Ils se trouvent sur le chromosome 3 au niveau de la bande polytène 61F1-2 (Godt et al., 1993 ; Couderc et al., 2002) et codent respectivement pour les protéines nucléaires Bab1 (967 aa) et Bab2 (1067aa). Ces gènes ont été initialement appelés *bab* et *BTB-II* (Godt et al., 1993 ; Zollman et al., 1994) avant d'être renommés respectivement *bab1* et *bab2* et considérés comme étant des gènes paralogues. Les protéines Bab1 et Bab2 présentent des similarités uniquement au niveau du domaine BabCD « Bab conserved Domain » (Figure 45) et du domaine évolutivement conservé BTB/POZ (Broad domain, Tramtrac, Bab). Ce dernier est partagé par un grand nombre de produits de gènes impliqués dans le développement et participe à des interactions protéines-protéines (Zollman et al., 1994), y compris dans l'homodimérisation de Bab1 *in vitro* (Chen et al., 1995). Le domaine BabCD quant à lui contient deux motifs connus pour leur capacité à lier l'ADN, le domaine Psq et le motif AT-hook. Le domaine Psq formé

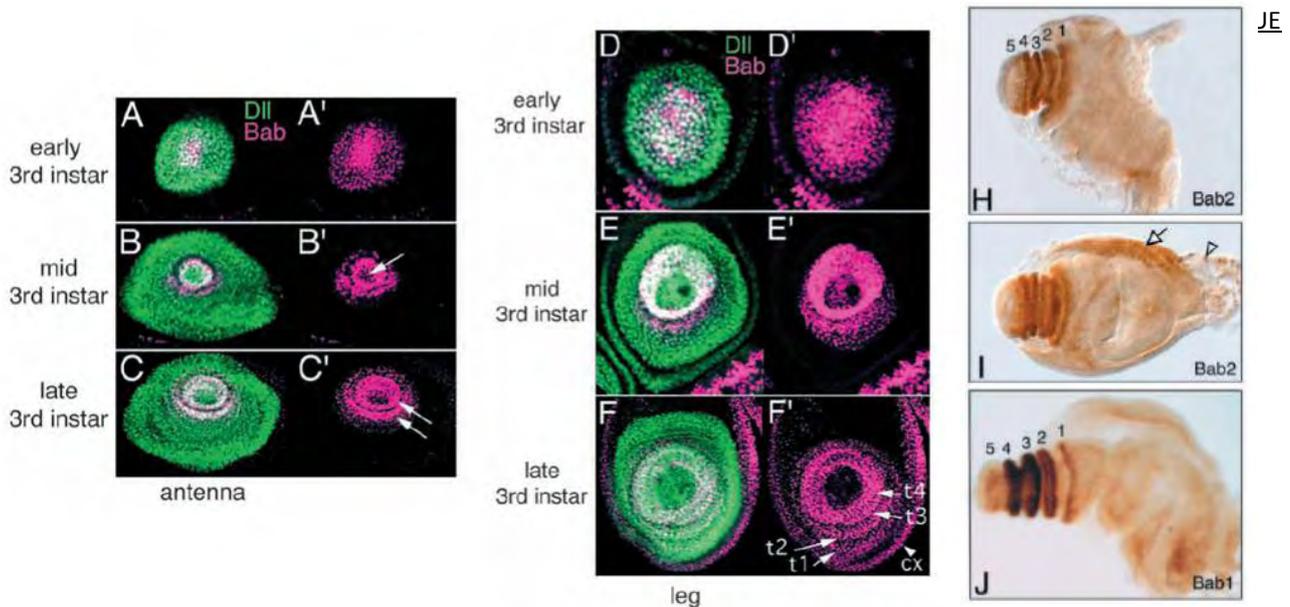


Figure 46. L'expression de *bab2* est dynamique au cours du développement de l'antenne et de la patte. Au troisième stade larvaire précoce, *bab2* est exprimé dans presque tout le domaine d'expression de *Dll*, dans l'antenne (A, A') et dans la patte (D, D'). Au milieu du troisième stade, l'expression de *bab2* est perdue dans la partie la plus distale de l'antenne (flèche B') et deux anneaux cellulaires exprimant *bab2* émergent (B, B'). Simultanément, l'expression de *bab2* est perdue aussi dans la partie la plus distale de la patte (E, E'). À la fin du troisième stade, *bab2* s'exprime selon deux anneaux concentriques dans l'antenne (flèches C et C') et selon quatre anneaux concentriques dans la région tarse de la patte (flèches F'). *bab2* est également exprimé dans la primordium coxa (cx), où il ne se chevauche pas avec *Dll*. (D'après Chu et al., 2002). Comparaison des profils d'expression de Bab1 et Bab2 dans les primordia imaginaux. (H-J) Dans les disques imaginaux de pattes, les protéines Bab se trouvent dans une distribution graduelle dans le primordium du tarse avec le plus haut niveau d'expression au niveau du t4 et t3, plus faible en t2 et encore plus faible dans la région distale de t1 (les chiffres indiquent les primordia des segments du tarse). Contrairement à Bab1 (J), Bab2 se trouve également dans la région proximale de t5 (H, I), dans la membrane péripodiale (flèche en I), et dans la périphérie du disque (flèche I). (D'après Chu et al., 2002 et Couderc et al., 2002)

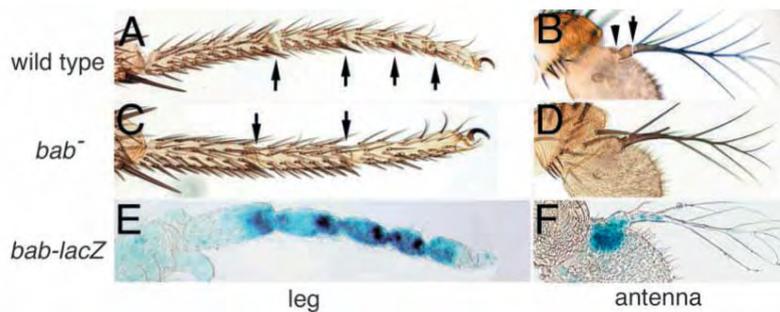


Figure 47. Le locus *bab* est requis pour la formation des jonctions dans la partie distale de la patte et de l'antenne. Structure sauvage d'une patte de drosophile (A), et de l'antenne (B). Les jonctions du tarse sont indiquées par des flèches. Les jonctions a4/a5 et a5/arista dans l'antenne sont indiquées par deux flèches. (C) les jonctions du tarse (flèches) sont fusionnées ou absentes dans des pattes *bab⁻* (Godt et al., 1993). (D) Les jonctions a4/a5 et a5/arista sont absentes dans l'antenne *bab⁻* (Godt et al., 1993). (E, F) Les profils d'expression d'un rapporteur *bab-lacZ* pendant le stade puppe sont compatibles avec les défauts des mutants *bab* au niveau des pattes et de l'antenne. (d'après Chu et al., 2002).

par 48 acides aminés, présente 97% d'homologie entre Bab1 et Bab2, constituant ainsi la région la plus similaire au niveau du domaine BabCD. Les deux protéines Bab contiennent aussi un court motif très riche en résidus de glutamine et d'histidine, qui pourrait participer à la transactivation (Courey et Tjian, 1988; Gerber et al., 1994). L'architecture des domaines de Bab1 et Bab2 associé à leur localisation nucléaire (Godt et al., 1993; Godt et Laski, 1995), suggère que ces protéines sont des régulateurs transcriptionnels (Reeves et Nissen 1990; Lehmann et al., 1998; Couderc et al., 2002; Lours et al., 2003).

6.1. EXPRESSION ET FONCTIONS DES GENES *BAB*

Durant l'embryogenèse, seule l'expression de *bab2* est détectable. Toutefois il a été suggéré que le locus *bab* n'a pas une fonction essentielle dans l'embryogenèse vu que chez des embryons mutants pour les deux gènes *bab* le développement ne semble pas être affecté (Zollman et al., 1994). Par la suite tout au long de la phase du développement post-embryonnaire les deux gènes *bab1* et *bab2* sont généralement co-exprimés, avec *bab1* souvent exprimé dans un sous-domaine de *bab2*. Les deux gènes *bab* sont impliqués dans plusieurs processus développementaux : le développement des pattes et de l'antenne, la formation des filaments terminaux dans l'ovaire ainsi que la pigmentation abdominale (voir page 26-27 evolution CRE *bab*) (Couderc et al., 2002). Selon les tissus, *bab1* et *bab2* peuvent agir de manière synergique, redondante ou bien distincte. Cependant, le rôle de *bab2* semble prédominant.

Dans l'ovaire, *bab1* s'exprime exclusivement dans les filaments terminaux, alors que *bab2* s'exprime fortement dans les filaments terminaux et plus faiblement dans les cellules apicales de l'ovaire (Couderc et al., 2002). Les mouches qui portent des mutations qui affectent les deux gènes *bab* montrent des défauts de formation au niveau des filaments terminaux, des cellules apicales et du primordium basal (stalk), se traduisant par des femelles stériles et des ovaires avec seulement quelques ovarioles rudimentaires, tandis que les mutations qui affectent un seul gène *bab* résultent en de plus faibles phénotypes (Godt et Laski 1995; Couderc et al., 2002). Les deux gènes *bab* sont également exprimés dans les disques génitaux femelle et mâle à l'origine des structures internes et externes des appareils génitaux, des tergites A8, A9 et des plaques anales (Bryant, 1978; Casares et al., 1997).

Dans les disques imaginaires de pattes et d'antenne, les gènes *bab1* et *bab2* s'expriment selon des patrons similaires. Leur expression débute à la fin du deuxième stade larvaire au niveau des régions distales des disques de pattes et d'antenne. Durant le troisième stade larvaire, ils se définissent en anneaux concentriques : deux anneaux dans le

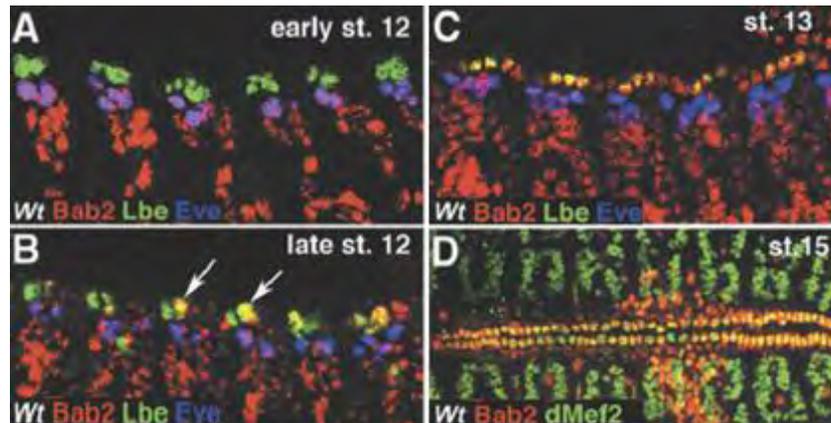


Figure 48. Expression de *bab2* au cours du développement cardiaque. (A-D) expression de Bab2 dans les précurseurs cardiaques et le vaisseau dorsal. (A) Au stade précoce 12 Bab2 est exprimée au niveau des précurseurs cardiaques Eve positifs, mais pas dans les Lb positifs. (B) À la fin du stade 12, une faible expression de Bab2 commence à apparaître dans les cardioblastes y compris les cellules Lb positives (flèches). (C) Au stade 13, Bab2 est exprimée dans toutes les cellules cardiaques. (D) Après fusion des ébauches cardiaques, au stade 15, la protéine Bab2 est particulièrement bien visible au niveau du cœur. (Junion et al., 2007)

disque d'antenne et quatre anneaux dans les disques de pattes (Figure 46) (Chu et al., 2002). Cette expression va par la suite se maintenir tout au long du stade puppe et va se localiser à proximité des jonctions inter-tarse dans les pattes et des jonctions a4/a5 et a5/arista dans l'antenne. Au niveau des disques de pattes, les gènes *bab* (surtout *bab1*) présentent une expression graduée, élevée au niveau des segments tarsaux distaux ts3-ts4, et plus faibles dans les segments proximaux ts2 et surtout ts1 (Figure 46). Il est à noter aussi que contrairement à Bab1, l'expression de Bab2 n'est pas limitée à une région sub-distale des disques de pattes, mais se retrouve également dans la membrane péripodiale, et dans la périphérie du disque imaginal de patte qui est à l'origine de certaines structures du thorax (Figure 46). Des expériences de perte de fonction des gènes *bab* ont montré que ces derniers sont impliqués dans la formation des jonctions inter-tarses dans la patte et des jonctions a4/a5 et a5/arista dans l'antenne (Figure 47). En revanche le rôle de *bab2* est prédominant dans les pattes (Couderc et al., 2002). En effet, un simple mutant *bab1* est sans effet, alors qu'un simple mutant *bab2* engendre un phénotype patte ce qui montre encore une fois le rôle prédominant de *bab2* vis-à-vis de *bab1*. Par ailleurs, les gènes *bab* jouent aussi un rôle dans la spécification des soies au niveau des tarse. En effet, chez des mutant *bab* (hypomorphes ou nuls), on note une transformation de certaines soies distales en soies de type proximal. Ceci se traduit par l'apparition de rangées de soies transversales ('transverse bristle rows') et de peignes sexuels (structure spécifique du tarse le plus proximal t1 des pattes mâles antérieures, formée d'une seule rangée d'environ dix à douze soies) au niveau des tarse les plus distaux des pattes T1 mâles (Godt et al., 1993). Il a été rapporté que l'apparition de peignes sexuels ectopiques chez les mutants *bab* est due à la dé-répression de *Sex combs reduced* (*Scr*) au niveau des tarse distaux (Randsholt et Santamaria 2008). Ceci suggère que la concentration des protéines Bab spécifie l'identité des soies dans les segments du tarse d'une manière dose-dépendante. Il est à noter que le gène Hox *Scr* est strictement requis pour le développement des peignes sexuels et que les pattes dépourvue de *Scr* se développent comme des pattes T2 (Struhl, 1982; Rogers et al., 1997; Shroff et al., 2007).

Au cours d'une étude menée par Junion et collaborateurs (2007), il a été montré que le gène *bab2* joue aussi un rôle dans la morphogenèse cardiaque durant le stade embryonnaire et ceci en réprimant le facteur à homéodomaine Lb (codé par *ladybird* (*lb*)) au stade 12 embryonnaire au niveau des cellules péri-cardiaques exprimant *eve* (Figure 48). Il a été montré aussi que Lb régule négativement *bab2* dans les cellules cardiaques pendant le stade 12 et positivement

à partir de la fin du stade 12 permettant ainsi à l'expression de *bab2* de s'étendre à tous les précurseurs cardiaques. Au cours de cette étude et moyennant la technique ChEST ('ChIP-enriched in silico target') le CRM reproduisant l'expression de *bab2* dans les tissus cardiaques à partir du stade 13 a été également identifié au niveau de la région inter-génique *bab1/bab2* (à la position I: 1146918–1147918). Il est à noter qu'en plus de l'expression dans les cellules cardiaques, *bab2* s'exprime aussi au niveau des cellules des vaisseaux dorsaux à partir du stade 12.

Afin d'analyser les contraintes évolutives qui s'exercent sur le locus *bab* Bickel et collaborateurs (2008) ont comparé la séquence du locus *bab* (148 kb) dans 94 lignées de *D. melanogaster* collectées au même endroit (verger à Wolfskill) mais sur plusieurs années. Les auteurs ont trouvé que deux régions non codantes semblent être sous pression sélective. La première se trouve du côté 3' du gène *bab2* dans une région qui n'a pas de rôle fonctionnel connu et la deuxième se trouve au niveau de *bab1* dans la région qui contient le CRE contrôlant l'expression spécifique de *bab* dans l'épiderme abdominal femelle (Williams et al., 2008). De plus, la région codante pour *bab1* semble être sous une contrainte fonctionnelle plus forte que la séquence codante pour *bab2*. Ce qui suggère que les deux paralogues évoluent sous des contraintes sélectives différentes au sein d'une même population en milieu naturel.

Bien que les domaines d'expression de *bab1* et *bab2* se chevauchent largement et qu'ils aient des rôles partiellement redondants dans le développement (Kopp et al., 2000; Couderc et al., 2002), ces deux gènes ont été maintenus dupliqués dans l'ensemble des génomes des drosophilidae (Ryan et al., 2009) ce qui laisse penser qu'ils présentent des différences subtiles au niveau de leurs fonctions. La plupart des structures dépendant des gènes *bab* sont impliquées dans le dimorphisme sexuel, notamment les gonades (Sahut-Barnola et al., 1995), les peignes sexuels sur les pattes avant des mâles (Godt et al., 1993; Barmina et Kopp 2007; Randsholt et Santamaria 2008), les soies abdominales ventrales (Kopp et al., 2000), et la pigmentation abdominale dorsale (Kopp et al., 2000; Williams et al., 2008). Ceci pourrait suggérer que la sélection sexuelle aurait agi sur le locus *bab*. En outre, il a été démontré que les peignes sexuels sont importants pour le succès d'accouplement chez les mâles (Ng et Kopp 2008).

L'analyse fonctionnelle détaillée du locus *bab* a révélé des éléments *cis*-régulateurs (CRE) (Williams et al., 2008) impliqués dans l'expression des gènes *bab* dans l'épiderme

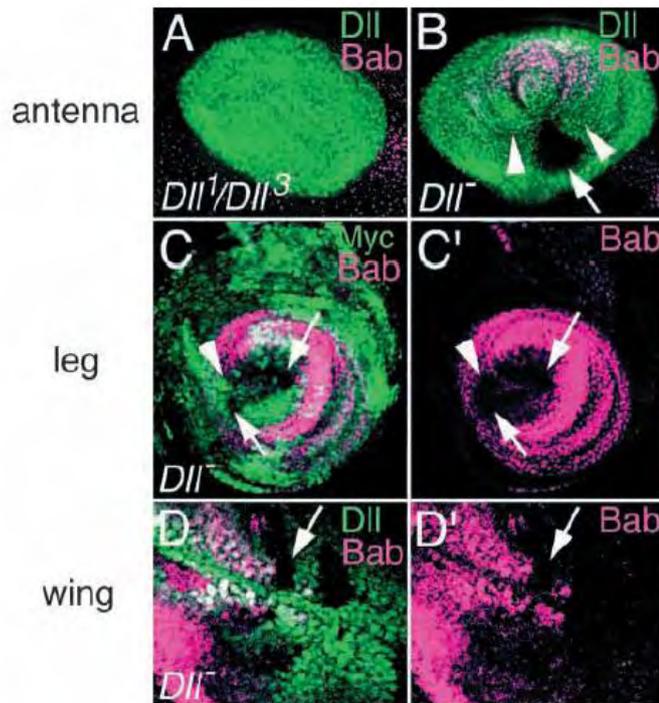


Figure 49. L'expression de *bab2* dépend de *Dll* dans les disques de patte, d'antenne et d'aile. (A) Bab2 n'est pas détectée de manière distale dans des disques antennaires hypomorphes *Dll1/Dll3*. (B) Bab2 n'est pas détectée dans (flèche) ou autour (pointes de flèches) d'un clone *Dll-* dans le disque d'antenne. (C, C') Bab2 n'est également pas détectée à l'intérieur (flèches) ou autour de clones *Dll-* dans un disque de patte. Ces clones ont été identifiés par l'absence d'expression du marqueur Myc. (D et D') Bab2 est perdue de manière cellulaire autonome dans un clone *Dll-* dans un disque d'aile.

abdominal nymphal (grand intron de *bab1*), les pattes (région inter-génique entre *bab1* et *bab2*), et les oenocytes (grand intron de *bab2*). Étonnamment, seul un élément de régulation a été identifié pour ces tissus qui expriment à la fois *bab1* et *bab2*, ce qui suggère que les deux gènes paralogues seraient contrôlés par les mêmes CREs. Ceci n'exclue pas l'existence d'autres éléments régulateurs spécifiques qui moduleraient l'expression de chaque gène d'une manière plus subtile.

6.2. REGULATION DE L'EXPRESSION DES GENES *BAB* DANS LA PATTE ET L'ANTENNE

L'étude de la cascade de signalisation impliquée dans le développement des pattes et de l'antenne a révélé que l'activation des gènes *bab* dépend de l'activité de *Dll* et *ss* alors que la répression passerait par l'activité de *dac* et de *bowl* auxquelles se rajoutent celles de *sal* et d'*hth* dans l'antenne en développement. Néanmoins, on ne sait pas si les facteurs de transcription codés par ces gènes régulent directement l'expression des gènes *bab*.

6.2.1. ACTIVATEUR DE L'EXPRESSION DES GENES *BAB*

✓ DLL

Chez la drosophile *Dll* est impliqué dans le développement des parties distales des divers appendices (Panganiban, 2000), entres autres les pattes et l'antenne. En effet, la perte de fonction de *Dll* entraîne la déléition des segments a2, a3 et de l'arista dans l'antenne et de tous les segments de la patte à l'exception du coxa (Cohen et al., 1989; Dong et al., 2000). L'une des propriétés de *Dll* est sa capacité à induire de manière ectopique dans certains tissus un centre organisateur exprimant *Wg* et *Dpp* qui assure la mise en place d'un nouvel axe P/D, ainsi l'apparition d'antennes ectopiques sur la tête et de pattes sur les ailes (Gorfinkiel et al., 1997) a pu être observée. De plus, l'expression ectopique de *Dll* dans la partie proximale des disques de pattes ou d'antenne entraîne la duplication de ces structures.

Dll est strictement requis pour l'expression des gènes *bab* dans les disques de pattes et d'antenne. En effet, *Bab2* n'est plus détectée ni dans des mutants *Dll* hypomorphes (Campbell et Tomlinson, 1998), ni dans des clones *Dll*- (Figure 49) aussi bien dans les disques de pattes que d'antenne. De manière intéressante, bien que l'expression de *bab2* soit perdue de manière cellulaire autonome dans des clones *Dll*- (Chu et al., 2002), *Bab2* est également perdue de manière cellulaire non-autonome sous certaines conditions. Dans l'antenne, par exemple, l'expression de *bab2* est aussi perdue dans des cellules entourant les clones *Dll*- lorsque ces

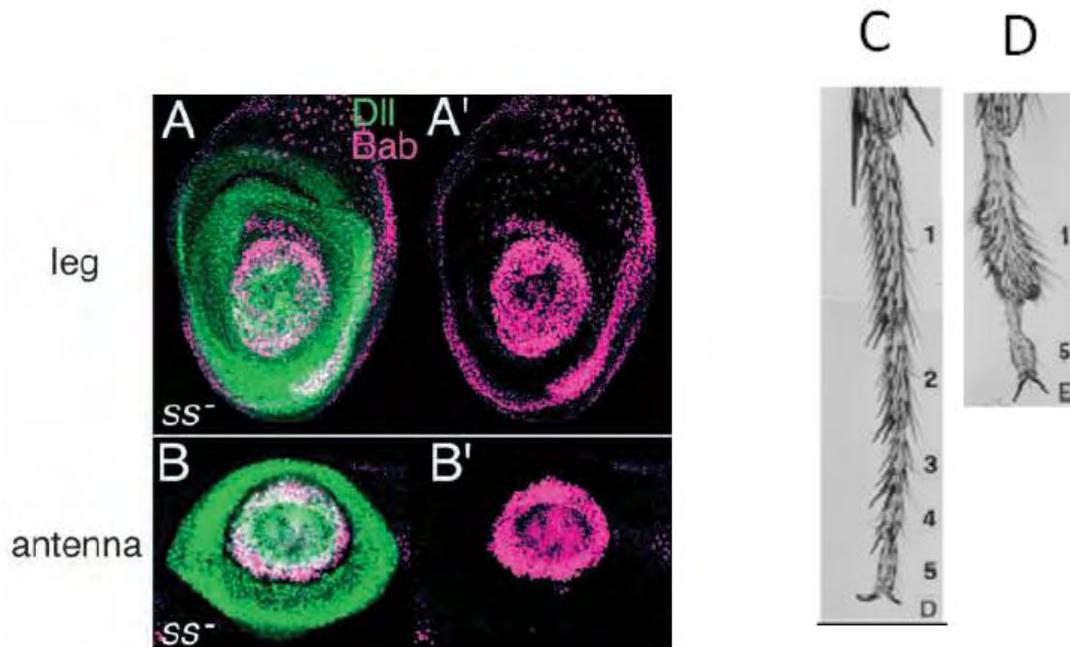


Figure 50. L'expression de *bab2* dépend en partie de *ss* dans l'antenne et dans la patte. (A, A') l'expression des anneaux proximaux de Bab2 est perdue dans une patte *ss*⁻, tandis qu'un anneau distal d'expression de Bab2 est maintenu. L'expression de Bab2 dans la coxa est insensible à la mutation *ss*. (B, B') L'expression de Bab2 est partiellement perdue dans l'antenne *ss*⁻, seul un anneau reste. Étant donné que l'antenne est transformée en patte, cet anneau est susceptible de représenter l'anneau distal de Bab2 dans une patte *ss*⁻. L'expression de *bab2* est partiellement dé-réprimée distalement à la fois dans les disques de patte et d'antenne *ss*⁻. L'expression de *Dll* semble normale dans ces disques (vert en A et B). (C) Partie distale d'une patte sauvage. (D) Partie distale d'une patte *ss*⁻. A l'exception du segment t1, les segments t2- t4 du tarse sont déléétés.

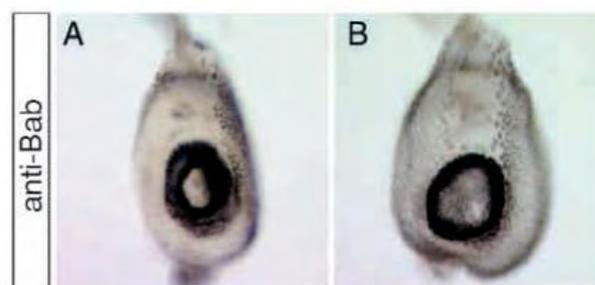


Figure 51. Expression de *bab2* dans un disque de patte mutant pour *rn*. (A, B) Expression de *bab2* dans des disques de patte de type sauvage (A) et *rn16/rn20* (B) au milieu du troisième stade larvaire. (St Pierre et al., 2002).

clones se trouvent près de la limite des domaines d'expression normale des gènes *bab* et de *dac*. Dong et collaborateurs (2002) proposent que ces phénomènes soient dus à l'activation cellulaire non-autonome de *dac*, un répresseur des gènes *bab*, autour des clones *Dll*⁻. Dans les pattes, les clones *Dll*⁻ engendrent aussi une perte d'expression non autonome d'expression de *bab2* autour des clones. Toutefois, dans ce cas il semble que l'expression de *Dll* soit également perdue de façon non autonome. Ainsi la perte d'expression cellule non autonome de *bab2* dans la patte est probablement due à la perte de l'activateur *Dll* même si cela n'exclue pas l'existence d'un mécanisme de répression dépendant de *Dac* comme dans l'antenne.

Dans l'aile, l'expression de *bab2* recouvre partiellement celle de *Dll* dans le 'wing pouch'. Bien que l'expression de *bab2* ne soit pas visiblement réduite dans des disques d'ailes *Dll* hypomorphes, l'expression de *bab2* est perdue dans certains clones *Dll*⁻ dans le 'wing pouch' (Figure 49). Ainsi pour conclure *Dll* est requis pour l'expression de *bab2* dans les disques de pattes et d'antenne et dans certaines cellules du 'Wing pouch'.

✓ SS

La perte de fonction *ss* se traduit par la perte d'une partie du segment de t1 et des segments t2, t3 et t4 de la patte, y compris leurs jonctions. Ainsi, dans des pattes *ss*⁻, il ne reste qu'une seule jonction distale positionnée entre t1 et t5 (Figure 50 ; Duncan et al., 1998). Conformément au phénotype adulte, à la fin du troisième stade larvaire un seul anneau d'expression de *bab2* est observé dans les disques de pattes *ss*⁻ (Figure 50). Un double marquage des disques de pattes *ss*⁻ à la fois par l'anticorps Bar qui marque les segments t4 et t5 (Kojima et al., 2000) et l'anticorps Bab2, indique que l'anneau d'expression de *bab2* restant correspond à l'anneau le plus distal (Chu et al., 2002). Ceci montre que les anneaux proximaux d'expression de *bab2* dépendent de *ss*, alors que l'anneau distal d'expression de *bab2* est indépendante de *ss* (Figure 52). De la même façon, les disques antennaires *ss*⁻, présentent un seul anneau distal d'expression de *bab2*. Ceci est en accord avec la présence d'une seule jonction distale, entre le segment a3 et l'arista transformées chez les adultes *ss*⁻ (Duncan et al., 1998). Cependant, en raison de la transformation de l'antenne en patte chez les mutant *ss*⁻, il est difficile d'évaluer dans quelle mesure *ss* est requis pour la formation de l'antenne. D'après St pierre et collaborateurs (2002) l'expression de *bab2* semble normale dans des disques de pattes *rn*⁻. Ce qui suggère que l'activation de *bab2* par *ss* ne nécessite pas

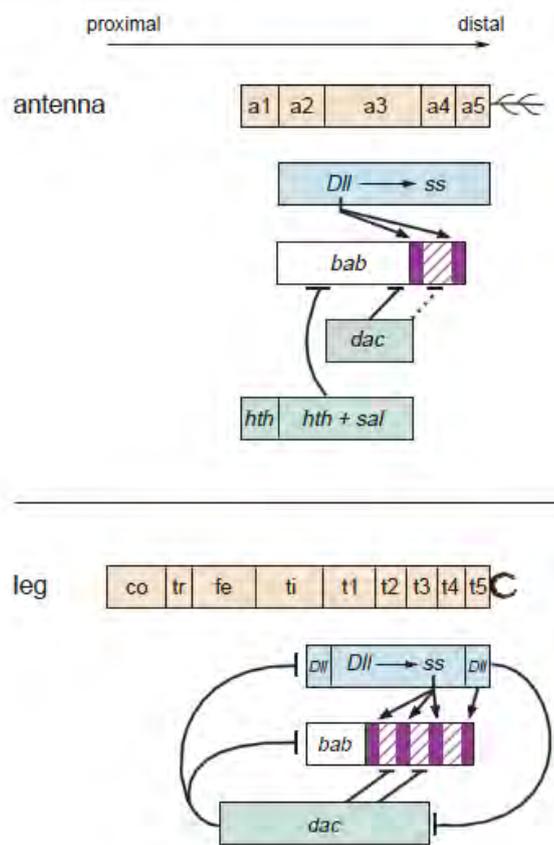


Figure 52. Interactions régulant l'expression des gènes *bab* le long de l'axe PD de l'antenne et de la patte. *bab2* est activé par *Dll* aussi bien dans les disques de pattes que d'antenne. L'expression de *bab2* est *ss* dépendante dans les trois premiers segments tarsaux (t1-3), mais pas dans le tarse t4 de la patte. Le rôle de *ss* dans l'expression de *bab2* dans l'antenne n'a pu être évalué, car une antenne *ss-* se transforme en patte. La limite d'expression proximale de *bab2* est régulée par *dac* dans les pattes et dans l'antenne, avec également une implication de *sal* et *hth* dans la répression de *bab2* spécifiquement dans l'antenne. *dac* intervient également pour moduler les niveaux d'expression inter-anneaux à la fois dans les disques de pattes et d'antenne. (Chu et al., 2002).

l'activité de *rn* qui se trouve en aval de *ss* dans la cascade de signalisation de la patte (Figure 51).

6.2.2. REPRESSEURS DE L'EXPRESSION DES GENES *BAB*

Bien que les gènes *bab* soient exprimés dans la majeure partie du domaine *Dll* au début du troisième stade, à la fin du troisième stade larvaire leur expression se limitent à une sous-partie du domaine *Dll*. Ceci suggère qu'il existe des répresseurs des gènes *bab* qui sont exprimés dans la partie complémentaire du domaine *Dll*. Deux FTs candidats ont été proposés comme répresseurs de *bab* dans les pattes et l'antenne : *Dac* dans la partie proximale et *Bowl* au niveau des parties proximale et distale (segment t1). Deux autres répresseurs spécifiques antenne ont été également proposés à savoir *Sal* et *Hth*. Toutefois aucune action directe de ces régulateurs n'a été établie. Enfin, la voie EGF a été proposée comme capable d'inhiber l'expression des gènes *bab* dans la partie distale des disques de pattes mais le(s) facteur(s) impliqué(s) dans cette répression n'a (ont) pas été déterminé(s).

✓ *DAC*

L'expression de *dac* dans les disques de pattes commence à la fin du deuxième stade larvaire et chevauche légèrement à ce stade, avec le domaine d'expression de *Dll*. Cependant, à la fin du troisième stade larvaire, l'expression de *dac* s'étend davantage sur la région proximale du domaine de *Dll*, et correspond au primordium des futurs tibia, tarse1 et tarse2 (Lecuit et Cohen, 1997). La protéine *Dac* chevauche ainsi l'expression de *bab2*, dans les anneaux t1 et t2. En revanche, dans le disque d'antenne les domaines d'expression de *dac* et des gènes *bab* sont principalement exclusifs. S'appuyant sur le fait que l'expression de *bab2* est induite dans des clones *dac*- proximaux et que l'expression ectopique de *Dac* réprime *bab2* dans les disques de pattes et d'antenne, Chu et collaborateurs (2002) ont proposé *Dac* comme un répresseur proximal des gènes *bab* (Figure 52). Toutefois, Il est à noter que l'expression de *bab2* est aussi induite dans certaines cellules qui se trouvent à proximité des clones *dac*-, le mécanisme responsable de cette activation n'étant pas encore élucidé.

Etant donné que *Dll* est également activé dans les clones *dac*- qui présentent une forte expression des gènes *bab* et que *Dll* est réprimé par l'expression ectopique de *Dac* (Abu-Shaar et Mann, 1998; Dong et al., 2001), il est possible que la répression des gènes *bab* par *dac* dans la patte soit médiée en partie par la réduction de l'activité de *Dll*. Néanmoins, le fait que les protéines *Dll*, *Dac* et *Bab2* soient co-exprimées dans t1 et t2 au troisième stade

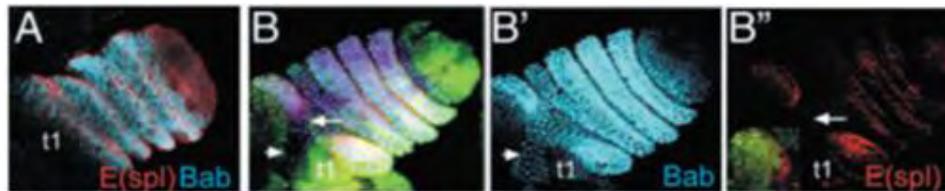


Figure 53. *bowl* est requis pour développement normal des segments tarsaux. Expression de *bab2* dans la région du tarse d'un disque de patte sauvage au stade pupé (A) et un disque de patte contenant un clone mutant *bowl2* (B) ; les cellules mutantes sont marquées par l'absence de GFP, vert). L'expression de Bab2 s'étend proximale dans les cellules mutantes (tête de flèche, B, B') et l'expression d'*E(spl)mβ-CD2* (cible de la voie Notch) est perturbée (flèche, B, B'').

larvaire tardif, indique que *dac* est insuffisant pour réprimer *Dll* ou *bab2*. En plus de fixer la limite proximale du domaine *bab* dans la patte, *dac* joue également un rôle dans la modulation de l'expression inter-anneaux des gènes *bab*. En effet, dans les disques de pattes qui sont fortement hypomorphes pour *dac*, l'expression de *bab2* inter-anneaux n'est plus modulée et *bab2* s'exprime uniformément à des niveaux élevés dans t1-t3. Le même phénomène est observé au niveau du disque d'antenne (Chu et al., 2002). L'expression uniforme de Bab2 dans les pattes *dac*- est corrélée à des phénotypes similaires à la perte de la fonction de *bab2*, à savoir la perte des segments t1-t3 tarse (Chu et al., 2002). Ainsi, il semble qu'une alternance entre des niveaux élevé et faible des gènes *bab* soit critique à la formation des jonctions. De manière similaire la perte de fonction *dac* dans l'antenne entraîne la perte des jonctions a4/a5 et a5/arista. Ainsi pour conclure *dac* permet de fixer la limite proximale d'expression des gènes *bab* dans les disques de patte et d'antenne (répression au niveau du a3) et joue le rôle de modulateur de l'expression inter-anneaux (Chu et al., 2002).

✓ BOWL

Comme le gène *bowl* est essentiel dès les premiers stades du développement (un embryon mutant pour *bowl* n'écloie pas; Wang et Coulter, 1996), il est difficile de déterminer l'effet d'une perte de fonction totale de *bowl* sur le développement des pattes et de l'antenne. Toutefois quand un clone mutant *bowl* couvre la majeure partie distale de la patte, cette dernière est sévèrement tronquée, avec peu ou pas de segmentation/jonction évidente. Inversement, les clones qui se limitent à la partie centrale du tarse n'aboutissent pas souvent à un phénotype détectable. L'effet observé avec *bowl* sur le développement des tarsi est également observé avec le gène étroitement lié *odd* (Wang et Coulter, 1996; Rauskolb et Irvine, 1999) qui s'exprime selon un profil similaire à *bowl* dans les disques de pattes. *bowl* et *odd* s'expriment depuis le début de troisième stade larvaire à proximité distale des domaines d'expression des ligands de Notch, Serrate et Delta. Toutefois, à partir de la fin du troisième stade les expressions de *bowl* et d'*odd* ne sont pas détectées à proximité des nouveaux domaines d'expression de Serrate et Delta qui émergent au niveau des inter-tarsi.

L'analyse de l'expression de *bab2* dans des clones *bowl*- dans les disques de pattes a montré que *bab2* est induit dans la région proximale au domaine d'expression normal de *bab2* ainsi qu'au niveau distal dans la région qui correspond au futur t5 et où la protéine Bab2 est normalement faible voire absente (Figure 53). La dé-répression de *bab2* dans les régions proximale et distale s'accompagne respectivement d'une réduction de l'expression de *dac* et

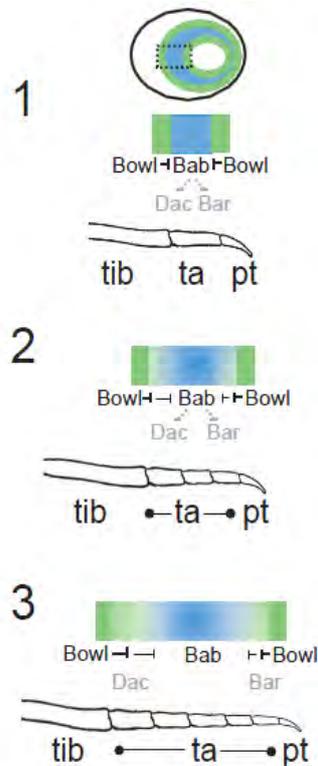


Figure 54. Modèle de régulation des gènes *bab* par *bowl* dans les parties proximale et distale du tarse. (1) stade précoce du développement du tarse. Les pointillés délimitent la région montrée au-dessous ainsi aux stades ultérieurs. Au sein de cette région (rectangle), le domaine d'expression de *bab2* est uniforme et flanqué par celui de *bowl*. *bowl* inhibe *bab2*. *bab2* (ou une autre cible de *bowl*) réprime l'expression de *dac* et de *Bar* (pointillés gris). Au-dessous, est montrée la patte distale hypothétique, le tarse est constitué d'un seul segment. Il est prédit que la différenciation des membres s'arrête à ce stade chez les arthropodes qui possèdent un tarse non segmenté. (2) stade moyen de développement des torses. La taille du territoire du tarse (rectangle) a augmenté. L'expression de *bab2* s'élargit au fur et à mesure que l'expression de *bowl* diminue. (3) stade tardif de structuration du tarse. La longueur du tarse (rectangle) a encore augmenté. Le domaine d'expression de *bab2* a atteint la taille maximale, comme celui de *bowl* décroît encore, *bab2* ne peut plus être induit et l'expression de *dac* et de *Bar* est augmentée dans t1 et t5, respectivement (en haut).

de *Bar*. Les clones *bowl* altèrent aussi l'expression de *E(spl)mb* et entraîne dans certains cas la perte totale de son expression ce qui suggère que l'effet de la perte de fonction *bowl* passe par la modification de l'activation de la voie Notch dans la patte (Figure 53). L'expression de la protéine Bowl de manière ectopique dans le domaine *Dll* réduit fortement l'expression de *bab2* et entraîne la progression des domaines *dac* et *Bar* vers le centre de la région tarse dans les disques de pattes. Cette expression ectopique s'accompagne également d'une extension des phénotypes t1 et t5 respectivement dans les tarsi distaux et proximaux chez l'adulte. Il est à noter que l'inhibition de *bab2* par l'activité de *bowl* entraîne un phénotype semblable à une perte de fonction *bab* avec l'apparition de peignes sexuels au niveau des tarsi les plus distaux. L'analyse de l'expression de la protéine Bowl produite par l'expression ectopique de *bowl* montre que des niveaux élevés de la protéine sont accumulés uniquement dans les régions d'expression de *bowl* (de Celis Ibeas et Bray 2003). Ceci est dû certainement à la dégradation de la protéine à cause de la protéine Lines nucléaire (Greenberg et Hatini, 2009).

Les gènes *bowl* et *odd* sont exprimés aux frontières du domaine *bab2* durant les premiers stades. A ~ 76-80 heures de développement, un premier anneau d'expression de *bowl* et d'*odd*, large de 2 à 3 cellules, entoure les cellules exprimant *bab2*. A ce stade la plupart des cellules exprimant Bab2 expriment également BarH1. Dans la région proximale d'expression de *bab2* seulement un anneau large de 1 à 2 cellules exprime Bab2 et non BarH1. Par la suite, un autre anneau d'expression de *bowl* et d'*odd* apparaît au centre du domaine *bab2*, à la frontière avec le pré-tarse et Bab2 est rapidement perdue à l'intérieur de cet anneau. Le domaine d'expression de *bab2* est désormais flanqué à la fois proximale et distalement par celui exprimant *bowl/odd*. Aux stades ultérieurs, les écarts apparaissent entre le domaine *bab2* et les anneaux *bowl/odd*. L'expression de *bab2* devient plus progressive, avec une diminution des niveaux sur les bords de son domaine d'expression. Les profils d'expression de *bowl* et d'*odd* et les phénotypes associés à leurs mutations suggèrent que les cellules à l'intérieur des régions t1/t2/t3 et t5 du tarse contiennent les protéines Bowl/Odd à 76-86 heures. de Celis Ibeas et Bray (2003) proposent que l'expression de *bowl/odd* soit peu à peu perdue à partir des cellules du tarse au cours de sa croissance (Figure 54), donnant lieu à un gradient temporel des protéines Bowl/Odd (expression prolongée dans t1, période plus courte d'expression dans t3/t2).

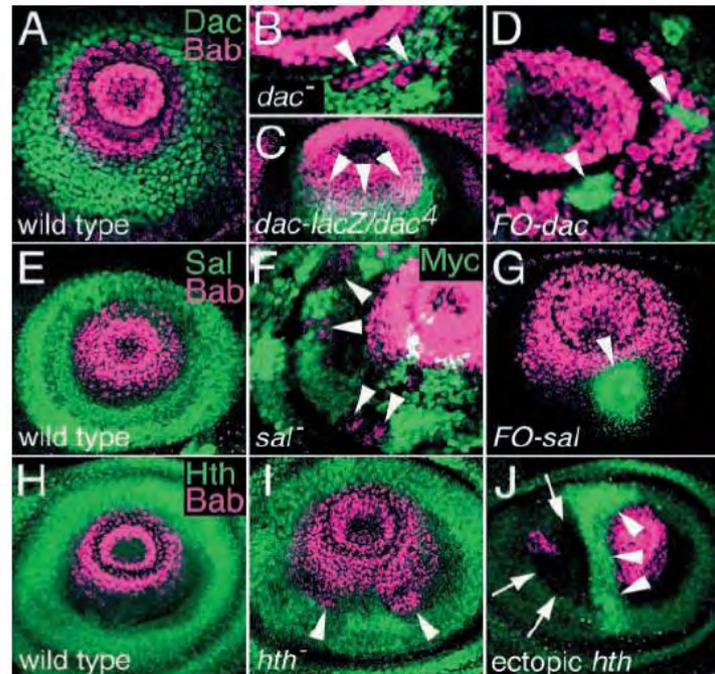


Figure 55. Répression de l'expression de *bab2* par l'activité de *dac*, *sal* et *hth* dans l'antenne. (A) Domaines d'expressions de *dac* et *bab2* dans un disque d'antenne sauvage. (B) *bab2* est déréprimé dans les clones *dac*- en a3. (C) dans un disque *dac*- l'expression de *bab2* s'étend proximale-ment dans a3. Notez que le chevauchement présent dans C est absent dans A. (D) *bab2* est réprimé par dans les clones exprimant Dac ectopiquement. (E) profil d'expression sauvage de *bab2* et *sal*. (F) *bab2* est déréprimé dans les clones *sal*- dans a2. Les clones sont marqués par l'absence du marqueur Myc. (G) *bab2* est réprimé par l'expression ectopique de Sal dans les clones 'flipout'. (H) expressions de *bab2* et *hth* dans le type sauvage. (I) *bab2* est activé dans les clones *hth*-. (J) *bab2* est réprimé par l'expression ectopique de Hth produite dans le domaine *dpp*. (Chu et al., 2002)

✓ VOIE EGFR

Selon les travaux de Galindo et collaborateurs (2002), le blocage de la signalisation médiée par le ligand Vn par l'expression ectopique de l'antagoniste sécrété de la voie EGFR, *argos*, dans le domaine *Dll* produit, en plus des défauts structuraux chez l'adulte (absence du pré-tarse à l'extrémité de la patte et absence ou malformation du t5), une perte de l'expression du marqueur de la pointe distale *dLim1* (Pueyo et al., 2000) et la dé-répression des gènes *bab* dans le centre du disque (Galindo et al., 2002). Des résultats similaires ont été obtenus en utilisant une version dominante négative de la protéine EGFR. De plus, l'expression ectopique d'une version activée d'EGFR dans les pattes, altère la structure des tarsi, abolit l'expression des gènes *bab* et entraîne une expression ectopique de *dlim1*, d'*al* et de *Bar*. Ces expériences montrent bien que la signalisation EGFR réprime l'expression des gènes *bab* dans le centre des disques de pattes. Néanmoins on ne sait si cette répression se fait de manière directe à travers les effecteurs de la voie (facteurs de type ETS tels Pnt/Yan) ou de manière indirecte via l'activation de l'expression de gènes codant des FTs tels *dLim1*, *cfl* ou *al* qui pourraient médier directement cette répression.

✓ SAL ET HTH REPRESSEURS ANTENNE SPECIFIQUE DES GENES *BAB*

Dans l'antenne, *dac* mais aussi *sal* et *hth* sont impliqués dans la répression proximale des gènes *bab* (Figure 55). L'activité de *sal* réprime *bab2* uniquement au niveau du segment a2. En effet, seuls les clones induits au niveau de la région a2 peuvent entraîner la dé-répression de *bab2*. En revanche *bab2* est souvent dé-réprimé dans des clones *hth*- au niveau des régions a2 et a3. Il est à noter que la perte de *hth* dans les clones entraîne la perte d'expression de *sal* et l'expression ectopique de *dac* dans les segments a2 et a3 (Dong et al., 2000, Dong et al., 2001). Ainsi, dans les clones *hth*- au niveau des segments a2 et a3, *bab2* et son répresseur *Dac* sont souvent co-exprimés. Cela confirme que *Dac* est insuffisant pour la répression de *bab2* comme décrit précédemment. Ainsi il semble que *dac* nécessite *hth* (ou le produit d'un gène activé par *Hth*) pour réprimer *bab2* dans l'antenne.

Résultats

Les disques imaginaux de pattes et d'antenne de la drosophile fournissent un paradigme permettant de comprendre les mécanismes fondamentaux du développement qui génèrent des structures adultes complexes à partir de primordia. L'une des étapes cruciales pour le développement de la patte et de l'antenne est la formation de l'axe P/D caractérisé par la subdivision des disques imaginaux en sous-domaines exprimant des gènes *gap* comme *dac*, *bab*, *Bar*, etc.

Au cours de ma thèse, je me suis intéressé à la régulation de l'expression des deux gènes paralogues *bab1* et *bab2* dans les tissus de patte et d'antenne considérés comme deux structures homologues (Snodgrass, 1935). Les gènes *bab* sont exprimés dans les disques imaginaux de pattes et d'antenne selon des patrons d'expression similaires et sont tous les deux impliqués dans le développement normal des structures adultes qui en émergent, à savoir les trois types de pattes et l'antenne, avec un rôle prédominant pour *bab2* (Couderc et al., 2002; Godt et al., 1993). Bien que les organisations P/D des pattes et de l'antenne soient assez semblables, le profil d'expression des gènes *bab* n'est le même, à savoir selon quatre anneaux dans la région tarsale des pattes et selon deux anneaux dans les segments a4 et a5 de l'antenne, à la fin du troisième stade larvaire (Couderc et al., 2002; Godt et al., 1993).

Afin de comprendre les mécanismes moléculaires à l'origine de cette expression différentielle, mon projet de thèse a consisté à la caractérisation des séquences *cis*-régulatrices, ainsi que des facteurs de transcription directement impliqués en *trans* dans le contrôle de l'expression des gènes *bab* dans les tissus de pattes et d'antenne. Pour ce faire, je me suis basé sur l'analyse du locus *bab* réalisée par (Williams et al., 2008) qui a permis d'identifier par marche chromosomique deux fragments chevauchants s'étalant sur environ 11 kb capables de diriger l'expression d'un gène rapporteur GFP selon le même patron que celui du gène *bab2* dans les pattes aussi bien que l'antenne en développement. En découpant cette région de 11 kb, j'ai pu montrer que les 'enhancers' patte et antenne sont confondus. Par ailleurs, j'ai montré également que cet enhancer unique (que nous avons appelé LAE pour 'Leg Antenna Enhancer') fonctionne de manière analogue dans les tissus de pattes et d'antenne, ce qui conforte l'hypothèse que les deux types d'appendices dérivent d'une même structure ancestrale.

Cette section s'articule autour de deux parties : la première concerne un article en préparation, qui est consacrée à l'identification, la validation et l'analyse fonctionnelle du LAE, ainsi qu'à deux activateurs directs, Dll et Rn; la deuxième traitera de données qui seront

présentées dans un deuxième article. Ces dernières, correspondent (1) à des résultats issus d'un crible simple hybride effectué au laboratoire LSBG ('Laboratory of Systems Biology and Genetics') dirigé par le Prof. Bart Deplancke à l'EPFL (École Polytechnique Fédérale de Lausanne); (2) à la caractérisation fonctionnelle d'autres séquences cis-régulatrices spécifiques de l'activation distale dans le LAE; (3) à des données concernant la répression directe des gènes *bab* par Bowl; et enfin (4) à l'auto-régulation négative des gènes *bab*

A single enhancer governs *Drosophila bab2* expression along rings in both leg and antennal tissues

Aissette Baanannou¹, Luis Mojica-Vazquez¹, Gaylord Darras^{1\$}, Jean-Louis Couderc², David L. Cribbs¹, Muriel Boube^{1*} and Henri-Marc Bourbon^{1*}

¹Centre de Biologie du développement, UMR5547 CNRS/UPS, Université de Toulouse, 118 Route de Narbonne, 31062 Toulouse, France

²Laboratoire Génétique, Reproduction et Développement, UMR6293, CNRS/Université de Clermont Ferrand, France

\$Present address: Institute for Research in Biomedicine, Parc Científic de Barcelona, c/Baldiri Reixac 10, 08028 Barcelona, Spain

Keywords: antenna/CRM/evolution/leg/PD axis

*Corresponding authors : boube@cict.fr, bourbon@cict.fr

Running title: Direct *bab2* activation by Distal-less and Rotund

Résumé

Dans cette publication, nous nous sommes intéressés aux mécanismes moléculaires contrôlant l'expression de *bab2* dans les disques imaginaux de pattes (4 anneaux) et d'antenne (2 anneaux). Ainsi en analysant les modalités d'expression d'un gène rapporteur GFP dans des lignées transgéniques obtenues par intégration spécifique de site, ainsi qu'en réalisant des tests de sauvetage des phénotypes mutants *bab* par un BAC sauvage ou modifié par recombineering, nous avons montré que l'expression de *bab2* dans les disques de pattes et d'antenne dépend d'un seul module cis-régulateurs (CRM) de 566 paires de bases (pb), appelé par la suite LAE, pour 'Leg Antennal Enhancer'. Le LAE se trouve dans la partie centrale du locus *bab* (qui s'étend sur 140 kilobases) et présente une forte conservation parmi 21 autres espèces de drosophiles. De manière significative, nous avons montré que la région LAE de *D. virilis* est suffisante pour diriger l'expression de notre transgène GFP selon le même profil que le gène *bab2* endogène de *D. melanogaster*, révélant ainsi une conservation fonctionnelle entre Drosophilidae. Moyennant des délétions et des mutations dirigées ('Linker scanning'), nous avons identifié des motifs d'ADN agissant positivement dans tout le domaine d'expression des gènes *bab* ou dans des sous-domaines (proximaux ou distaux), et ceci de manière similaire dans les tissus de pattes et d'antenne. Enfin, par des expériences de perte et de gain de fonction *in vivo*, ainsi que des analyses de complexes ADN/protéine en 'gels retards' *in vitro*, nous avons révélé que Distal-less (Dll), protéine à homéodomaine, active directement *bab2* et l'enhancer LAE au niveau de tous les anneaux exprimant les gènes *bab*, tandis que Rotund (Rn), une protéine à doigt de zinc, est seulement nécessaire à l'activation directe dans les anneaux proximaux.

Abstract

A proper spatio-temporal expression of *bric-a-brab2* is essential for the proximo-distal (PD) patterning of the *Drosophila* leg and antenna. Here, we investigated the molecular mechanisms controlling *bab2* expression in developing distal leg (4 rings) and antennal (2 rings) tissues. Using site-targeted GFP reporter assay and BAC recombineering, we show that restricted *bab2* expression relies on a single 0.56 kb *cis*-regulatory module (CRM), which is necessary and sufficient for Bab2 function in both developing leg and antenna. Through deletion and site-directed mutagenesis approaches, we identified discrete DNA motifs acting either positively or negatively. Finally, loss- and gain-of-function genetic analyses in combination with *in-vitro* electrophoretic mobility shift assays established that the Distal-less homeodomain protein directly activates *bab2* in all rings, whereas the Rotund C2H2 zinc-finger protein is specifically required for direct activation in the proximal-most rings. This work paves the way to unravel the molecular making of concentric epithelial cell rings expressing patterning genes.

Introduction

A spatially and temporally fine-tuned transcriptional regulation of the genome is essential to ensure proper development (Levine, 2010; Levine & Tjian, 2003). Morphogenesis of the *Drosophila* adult leg along the proximo-distal (PD) axis offers a good genetic model for deciphering how patterning genes are tightly regulated, since the genetic cascades operating within the developing leg imaginal disc, the larval epithelial sheet giving rise after metamorphosis and morphogenesis to the adult appendices, are well characterized (Couso & Bishop, 1998; Estella et al, 2012; Galindo & Couso, 2000; Kojima, 2004; Morata, 2001; Natori et al, 2012). Nevertheless due to the large number of patterning genes encoding transcription factors (TF) and of the complexity of their interactions, it is more and more difficult to determine which TFs are directly implicated in the regulation of a given gene. To sort out of this complexity and for getting a full understanding of gene cross-regulations, it is crucial to characterize cis-regulatory modules (CRM) binding those developmentally-regulated TFs. A good model to answer this question is the control of *bric-a-brac* (*bab*) expression during the establishment of the leg proximo-distal axis (Couderc et al, 2002).

The *Drosophila* leg is composed of ten segments which are articulated to each other by characteristic joints (Morata, 2001). The distal portion of the adult leg, the tarsus, is divided into five segments (ts1-5), and PD tarsal patterning occurs by successive intercalations of new positional fates within the growing leg imaginal disc (Estella et al, 2012; Natori et al, 2012). The combination of a dorsal signal provided by the *BMP4* homolog *decapentaplegic* (*dpp*) with a ventral signal provided by the *Wnt* homolog *wingless* (*wg*) establishes the PD axis, in addition to organizing the dorso-ventral appendage pattern (Campbell et al, 1993; Couso et al, 1993; Couso & Bishop, 1998; Jiang & Struhl, 1996). The more *wg* and *dpp* a cell receives the more distal it becomes, the center of the leg primordium giving rise after evagination of the entire organ to the adult claws (Kojima et al, 2005). Early during this process, *wg* and *dpp* signalling regulates directly or indirectly the expression of TFs encoded by *Distal-less* (*Dll*), *dachshund* (*dac*) and *homothorax* (*hth*) PD genes (Abu-Shaar & Mann, 1998; Diaz-Benjumea et al, 1994; Lecuit & Cohen, 1997; Wu & Cohen, 1999). *Dll*, *dac* and *hth* are activated in concentric broad domains and define the distal, medial and proximal parts of the adult appendages, respectively (Estella et al, 2008; Giorgianni & Mann, 2011). Together with EGFR signalling emanating from the distal-most cells, they in turn regulate downstream TF-encoding genes, such as *spineless*, *rotund*, *bric-a-brac1/2*, *BarH1/2* and *apterous*, all expressed dynamically in smaller concentric rings in the tarsal

domain (Estella et al, 2012; Natori et al, 2012). In the antenna, unlike the leg, *Dll* and *hth* expression domains overlap, leading to the specific activation of *spalt* and *cut* (Dong et al, 2000).

The *bric-a-brac* (*bab*) locus consists of two paralogous genes, *bab1* and *bab2*, with only the latter being critically required for distal leg and antennal segmentation (Couderc et al, 2002; Godt et al, 1993). Both *bab1* and *bab2* display dynamic expression in merely similar (see below) restricted PD sub-domains and exhibit distinct expression patterns between leg and antenna (Couderc et al, 2002; Godt et al, 1993). Null *bab* mutation causes transformation toward the tarsal segment 1 (ts1) fate along the PD axis and fusion of tarsal segments 2 through 5 (ts2-5) of the leg and of the fourth and fifth segments (a4-5) and arista of the antenna (Couderc et al, 2002; Godt et al, 1993).

While initially expressed within the *Dll*-expressing distal domain in mid third-instar larvae (L3), *bab1* and *bab2* expression in late L3 consists of four and two concentric rings in the distal leg and antennal tissues, respectively, with distinct intensities (Couderc et al, 2002). Later on at pupal and adult stages a PD expression gradient within each ring is observed (Chu et al, 2002; Godt et al, 1993). This graded expression pattern is essential in defining the identity of the tarsal ts2-4 segments (Couderc et al, 2002). While antennal *bab2* expression extends more proximally than *bab1* in late L3, in the early pupal leg (i) *bab2* expression extends more distally than *bab1* in proximal ts5; (ii) the expression intensity of the two proximal rings (ts1-2) is higher for *bab2* than *bab1*; (iii) *bab2*, but not *bab1*, is also expressed in the disc periphery and the peripodial membrane (Couderc et al, 2002). Finally, no expression is detected in the center of mid and late L3 imaginal discs that gives rise to the primordia of distal ts5 and the pre-tarsus of the leg and the antennal arista (Couderc et al, 2002; Godt et al, 1993).

In both developing leg and antennal discs, *bab* expression depends on activity of the homeodomain Dll protein (Campbell & Tomlinson, 1998; Chu et al, 2002; Galindo et al, 2002). This could be mediated in part by *spineless* (*ss*), a direct transcriptional target of *Dll* (Duncan et al, 2010), which encodes a bHLH-PAS family TF (Chu et al, 2002; Duncan et al, 1998; Galindo et al, 2002). As the *ss* gene is no longer expressed in late L3 leg tissues, the requirement for *ss* activity must be transient, however (Duncan et al, 1998; Natori et al, 2012; Pueyo & Couso, 2008). Interestingly, *rotund* (*rn*), a tarsal patterning gene encoding a C2H2 zinc-finger (ZF) protein (St Pierre et al, 2002), is a target of *ss* in a mid L3 leg disc (Natori et al, 2012; Pueyo & Couso, 2008). Of note, unlike leg, *ss* and *rn* expression patterns are

maintained in developing antenna up to the adult stage (Duncan et al, 1998; St Pierre et al, 2002).

In addition to positive inputs from *Dll* and *ss*, *bab* expression is restricted proximally by *dac* activity (Chu et al, 2002) and distally by a gradient of EGFR-Ras signalling (Campbell, 2002; Galindo et al, 2002), via the activity of *C15/clawless (cll)* and *aristaless (al)*, which encode HD-containing TFs (Campbell, 2005). Lastly, graded *bab* expression distally, as well as proximally, have been linked to Notch signalling via a repressive effect of *bowl* activity in response to *lines* down-regulation (de Celis Ibeas & Bray, 2003; Greenberg & Hatini, 2009).

Starting from a systematic identification of *bab* enhancers performed by (Williams et al, 2008), here we characterized an evolutionarily-conserved 567 base pair (bp) CRM, which reproduces expression of the *bab2* endogenous gene. We show that this CRM (termed LAE, for leg and antennal enhancer) is functionally conserved between *D. melanogaster* and *D. virilis*. We then show that LAE is both necessary and sufficient *in-vivo* for proper *bab2* expression in leg and antennal discs, and for normal segmentation of the adult appendices. Using deleted constructs and site-directed mutagenesis, we show that leg and antennal *cis*-regulatory elements are intermingled. Furthermore, activation of *bab2* expression in proximal- and distal-most rings is dependent on separate DNA elements. Finally, we provide evidence that Distal-less and Rotund proteins bind directly to LAE sites to activate *bab2* expression in all and the proximal-most expressing cells, respectively. Taken together, our work indicates that a single enhancer governs *Drosophila bab2* expression along rings in both developing leg and antenna.

Results

A conserved cis-regulatory module recapitulates limb-specific bab2 expression

A systematic study of the 150-kilobase (kb) *bab* locus identified leg-specific *cis*-regulatory elements within a 11 kb region encompassing two overlapping genomic fragments (termed BP42 and BP47) localized between the *bab1* and *bab2* transcription units (Williams et al, 2008) (Fig. 1A). Both BP42 and BP47 fragments reproduce also the antennal *bab2* expression (Supplementary Fig. S1). To identify appendage-specific *bab* CRMs, we further dissected the 11-kb BP42+BP47 region, using a new GFP reporter vector including an *attB* sequence (see Methods) (Bischof et al, 2007).

Six overlapping genomic fragments (#1 to 6) covering the BP42+BP47 region (Fig. 1A) were examined for GFP expression in developing leg and antenna. Only fragments #3 and

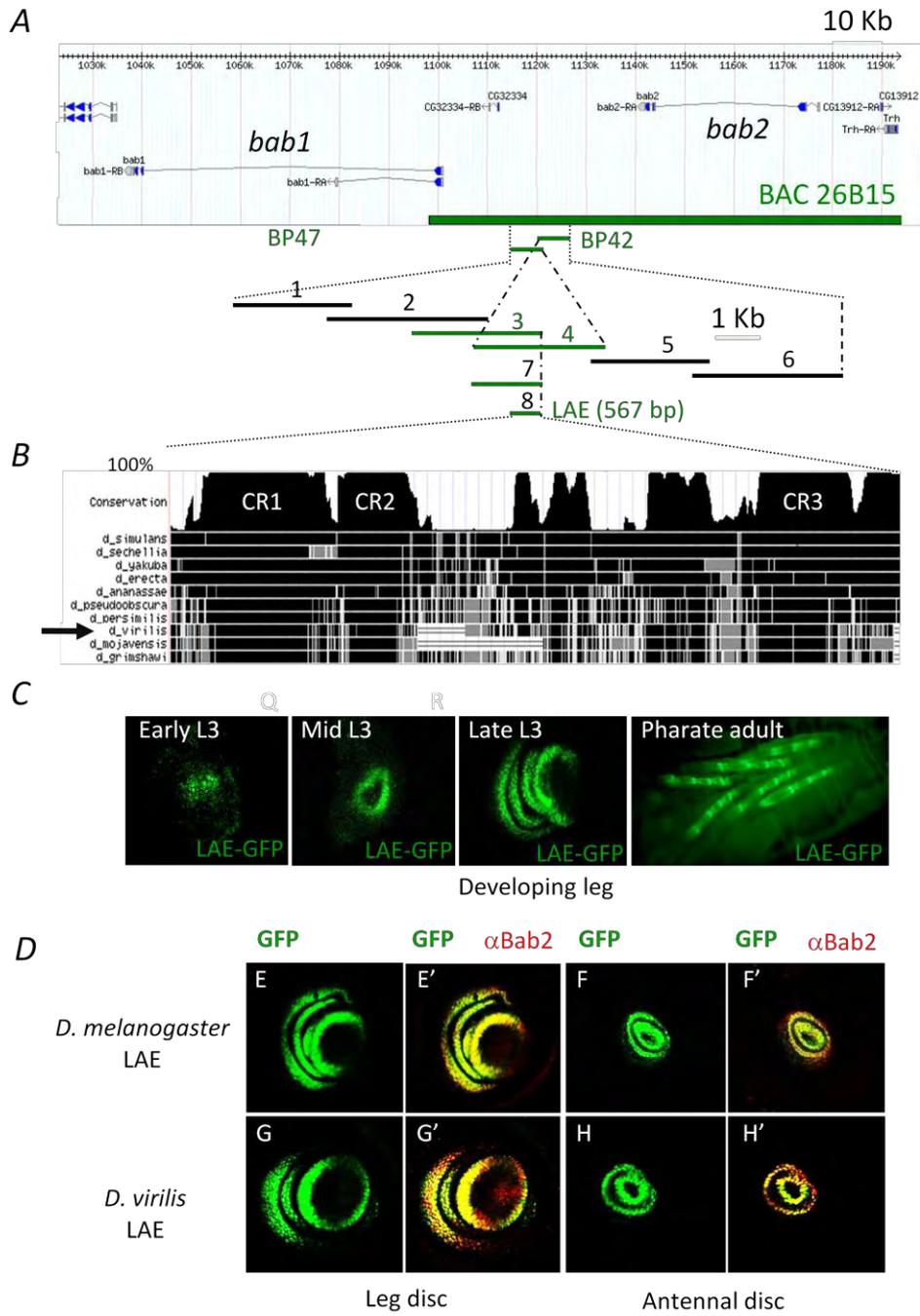


Figure 1: Identification of an evolutionarily-conserved *cis*-regulatory module (CRM) driving GFP-reporter activity in normal *bab2*-expressing leg and antennal tissues.

(A) A site-specific GFP-reporter assay identified a leg and antennal enhancer (LAE) element situated in the middle of the 150 kb *bric-a-brac* locus, encoding two paralogous genes *bab1* and *bab2*. The position of the BAC 26B15 used for phenotypic rescue experiments is shown, as well as the two overlapping genomic fragments (BP42 and BP47) that drove GFP-reporter expression in developing appendages (Williams et al, 2008). In beneath, an enlarged view of the 11-kb BP42-47 region shows the positions of the eight genomic sub-fragments (#1-8). The four fragments (#3, 4, 7 and 8) reproducing wild-type leg and antennal *bab2* expression are indicated in green.

(B) Conservation of the LAE sequence among Drosophilidae. Identical, highly- or non-conserved LAE sequences among a set of 10 *Drosophila* species, with respect to the *D. melanogaster* sequence, are depicted, respectively, by thick horizontal black, grey or white lines, whereas thin grey lines indicate gaps. The position of the *D. virilis* LAE sequence is indicated by a horizontal arrow. As determined from the UCSC genome browser (<http://genome.ucsc.edu/>), sequence conservation (from 0 to 100 %) over a 10-bp sliding window is represented above by the height of the conservation plot, with the three >20-pb-long highly-conserved CR1-3 regions indicated within (full alignment from LAE sequences of 22 Drosophilidae species is provided in Supplementary Fig. S2).

(C) The 567 bp LAE recapitulates *bab2*-expression pattern in developing leg, both spatially and temporally. The expression of the *LAE-GFP* reporter construct is shown, from early, mid and late L3 larval, up to the pharate adult stages.

(D) Functional conservation of the *D. virilis* LAE regulatory activity in developing *D. melanogaster* appendages. GFP expression (green) alone or in combination with Bab2 immunostaining (red) of leg (E-E' and G-G') and antennal (F-F' and H-H') imaginal discs, dissected from late third-instar larvae expressing the *D. melanogaster* or *D. virilis* *LAE-GFP* construct (E-F and G-H, respectively), are shown.

#4 drove strong GFP expression in leg and antennal tissues (Fig. S1) confirming that the relevant cis-regulatory elements are located within the 1.5 kb BP42+BP47-overlapping region. In support to this, a fragment (#7) containing only this 1.5 kb region was sufficient to reproduce *bab2* expression in leg and antennal tissues (Fig. S1).

Before dissecting further the 1.5 kb fragment, we examined its evolutionary conservation between 12 *Drosophila* species whose sequences were available (Clark et al, 2007). We found only three >20 bp motifs (termed CR1-3) that are highly conserved (Figs. 1B and S2). We therefore tested the activity of a 567-bp fragment (#8) encompassing the CR1-3 motifs, and found that its activity faithfully recapitulated *bab2* expression in both developing leg and antenna (Fig. 1C-D). We then tested the equivalent 0.7 kb *D. virilis* region and found that its activity was similar to that of the *D. melanogaster* LAE (Fig. 1D, compare E-F to G-H). This result pointed to the functional importance of the conserved CR1-3 motifs. Taken together these data indicate that the evolutionarily-conserved 567-bp region recapitulates limb-specific *bab2* expression. Given that the 567-bp region contains all the regulatory information for timely expression in the correct distal domain of the leg and antennal tissues, with properly positioned rings and shaped PD gradients, we therefore termed it LAE, for leg and antennal enhancer.

The LAE is necessary and sufficient for bab2 expression and function in-vivo

To define whether the LAE is also necessary to drive *bab* expression *in-vivo*, we used a *P[acman]* BAC construct (26B15) (Venken et al, 2009), including the *bab2* transcription unit and the LAE sequence (Fig. 1A). Adults homozygous for the *bab*^{ARO7} null allele display shortened tarsi with segmental joint fusions (Fig. 2A-B). A single copy of the intact 26B15 BAC construct was sufficient to restore normal Bab2 protein expression in both developing leg and antenna (Fig. 2H and M) as well as to rescue *bab* mutant phenotypes (Fig. 2C). In contrast, no appendage-specific Bab2 expression and no phenotypic rescue were observed when a LAE-deleted version of 26B15 BAC was analyzed (Fig. 2D, I and N), concluding that the LAE CRM is strictly required for *bab2* expression and function in developing limbs.

To know whether the LAE is not only necessary but also sufficient to drive normal Bab2 expression, we tested the ability of a LAE-driven *bab2* cDNA construct (LAE-Bab2cDNA) to rescue *bab* mutant phenotypes. Both adult phenotypes and Bab2 expression defects of *bab2* null adults were rescued by the *LAE-Bab2cDNA* construct (Fig. 2E, J and O), demonstrating that the LAE is both essential and sufficient for limb-specific *bab2* expression and function.

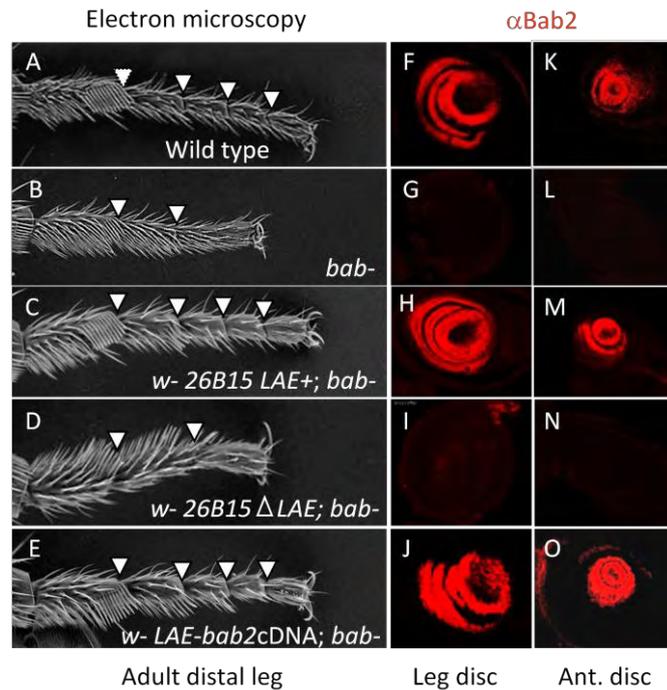


Figure 2: The LAE is necessary and sufficient for wild-type *bab2* expression and function *in-vivo*.

(A-E) Electron microscope views of adult distal legs from wild-type (A), *bab*^{AR07} (B), *w*¹¹¹⁸; *26B15-LAE+*; *bab*^{AR07} (C), *w*¹¹¹⁸; *26B15 LAE*; *bab*^{AR07} (D) and *w*¹¹¹⁸ *LAE-bab2cDNA*, *bab*^{AR07} (E), are shown. Wild-type tarsal joints are indicated by arrowheads. (F-O) Bab2 immunostaining (red) of leg (F-J) and antennal (K-O) imaginal discs, dissected from wild-type (F and K), *bab*^{AR07} (G and L), *w*¹¹¹⁸ *26B15-LAE+*; *bab*^{AR07} (H and M), *w*¹¹¹⁸ *26B15 LAE*; *bab*^{AR07} (I and N) or *w*¹¹¹⁸ *LAE-bab2cDNA*; *bab*^{AR07} (J and O) third-instar larvae, are shown. The *bab*^{AR07} allele (Couderc et al, 2002) corresponds to a large deletion removing the entire *bab1-2* intergenic region that contains the LAE, and most of the *bab2* transcription unit (unpublished). Both the *26B15LAE+* PacMan and *LAE-bab2cDNA* constructs rescue the tarsal fusion phenotypes and restore the wild-type leg and antennal Bab2-expression patterns.

Functional dissection of the LAE CRM: shared leg and antennal regulatory information

To functionally dissect the LAE CRM, we tested series of 3' and 5'-truncated constructs (Fig. 3). Deletion of 89 bp at the 3'-end (*F3* construct), removing the CR3 motif, only led to a slightly decreased GFP expression (72 and 77% of signal strength, compared to the entire LAE, in leg and antennal tissues, respectively) without affecting signal specificity, i.e. timely expression in the correct cells (Fig. 3, B-B' and I-I'). Deletion of 148 additional bp (*S5* construct) reduced further the level of expression, again in both tissues and without specificity defects (35 and 32%, in leg and antennal tissues, respectively) (Fig. 3, C-C' and J-J'). Deletion of 94 additional bp at the 3'-end (*S1* construct), now removing the CR2 sequence, led to a further drastic reduction of the signal strength, particularly in antennal tissues (3%), with a nearly-complete loss of GFP expression in the distal-most *bab2*-expressing cells in either leg or antennal tissues (arrows in Fig. 3, D-D' and K-K'). These data suggested (i) that the 94 bp CR2-encompassing region is required for the making of the distal-most *bab2*-expressing ring, and (ii) that the 137 bp LAE 5'-end region, including the large CR1 sequence (68 bp), is sufficient for proximal ring expression, albeit with low signal strength.

Conversely, a 100-bp 5' deletion removing the CR1 sequence (*F1* construct) led to complete loss of GFP expression, in both leg and antennal discs (Fig. 3, F-F' and M-M'). This indicates a key regulatory function for CR1 in all *bab2*-expressing cells. Nevertheless, the isolated 68-bp CR1 fragment (*B1* construct) did not drive either leg- or antennal-specific GFP expression (Fig. 3, G-G' and N-N). These data indicate that CR1 is critical but is not sufficient for GFP-reporter activity in both developing appendages. Given that the CR2-containing *F1* construct deleted for CR1 did not allowed GFP expression (above), we conclude that CR2 is also not sufficient for LAE activity in the distal-most *bab2*-expressing ring.

The 32 bp region at the LAE 5' end is a poorly conserved sequence (Fig. S2) and its deletion (*S3* construct) affected GFP reporter activity only partially: the expression levels of the two proximal-most leg rings (spanning tarsi ts1-2 and ts2-3) and the proximal antennal ring (spanning a3-4 segments) were both decreased (brackets in Fig. 3, E-E' and L-L'), albeit without specificity defects.

Taken together these data indicate that (i) the entire LAE is required to recapitulate normal *bab2* expression patterns, with shared leg and antennal regulatory elements; (ii) the 337 bp 3'-part is only required for signal strength; and finally (iii) the 230 bp 5'-part is critical for signal specificity, that is timely expression in the correct cells, with a key role of the CR1

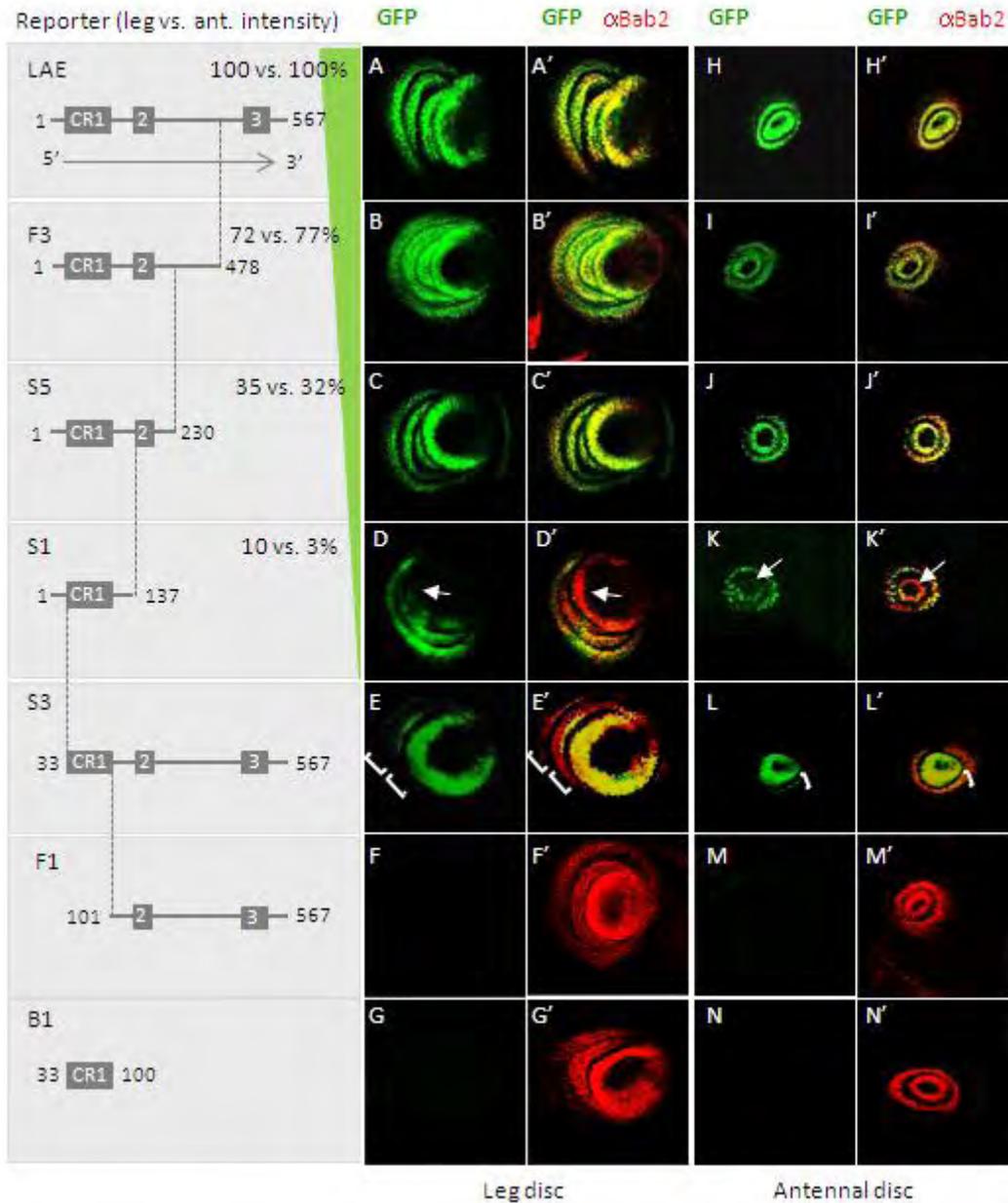


Figure 3: The LAE contains separate regions required for spatial expression and signal intensity level, via intermingled leg and antennal regulatory elements.

The name and relative position (with respect to the entire LAE) of each tested genomic fragments are shown on the left side. The CR1-3 conserved sequences are indicated by grey boxes. The relative levels of GFP expression (Methods) observed in the different constructs, in leg versus antennal tissues, are indicated in % with respect to the entire LAE. GFP expression (green) alone or in combination with Bab2 immunostaining (red) of leg (A-G) and antennal (H-N) imaginal discs from dissected late third-instar transgenic larvae, are shown on the right side for each construct. The *S1* (D-D' and K-K') and *S3* (E-E' and L-L') constructs, both in the leg and antennal tissues, displayed decreased GFP expression in the distal- or proximal-most *bab2*-expressing rings, respectively, visualized by the red-orange signals (see arrowheads and brackets, respectively) instead of yellow colour normally seen in the merged images. The CR1 sequence alone (B1) and the CR2-3 region (F1) did not drive limb GFP expression.

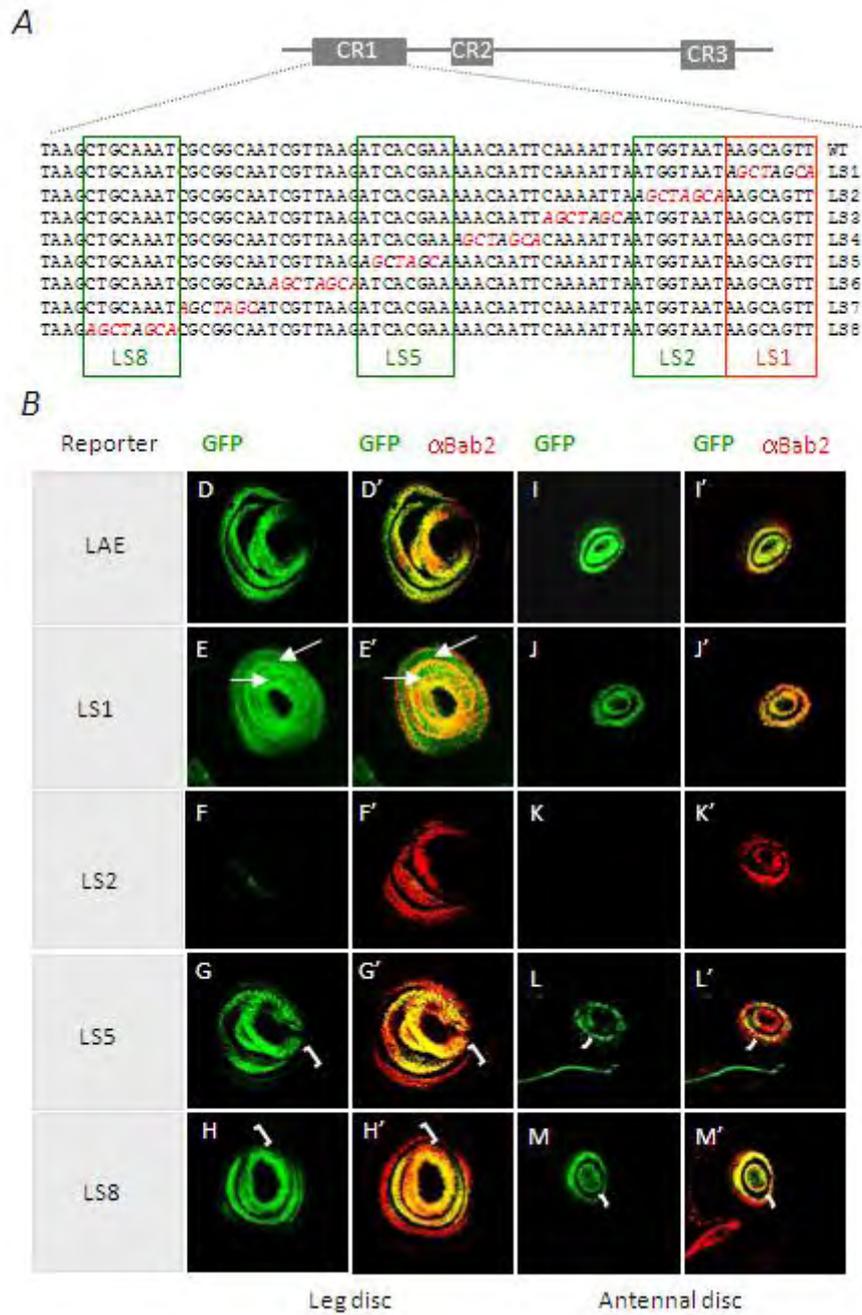


Figure 4: The key CR1 sequence includes both separate activating and repressive regulatory information.

(A) Linker scanning (LS) mutagenesis of the 68 bp CR1 conserved sequence. The sequences of the wild-type and of the eight mutated (LS1-8) CR1 regions (within the LAE) are shown, with the names of the *LS* constructs indicated on the right side. The linker element is italicized within the mutated CR1 sequences, and nucleotide substitutions are in red. The most critical regulatory elements are framed in green (activation) or red (repression).

(B) Leg (D-H) and (I-M) antennal imaginal discs from dissected late third-instar larvae, expressing a given LS-mutated construct. GFP expression (green) alone and in combination with Bab2 immunostaining (red), for the four reporter constructs displaying abnormal expression patterns, are shown. LS2-driven GFP reporter expression was strongly reduced, particularly in distal *bab2*-expressing cells, in both developing leg and antenna (F-F' and K-K'). Conversely, the *LS1-GFP* construct displayed inter-ring de-repression in leg tissues (E-E'), as determined by the green (arrows) instead of yellow colour normally seen in the merged images. Finally, the *LS8* and to a lesser extent *LS5* constructs exhibited strong GFP decrease in the proximal-most *bab2*-expressing tarsal and antennal cells (G-H and L-M) as determined by the red staining (see brackets) instead of the yellow colour normally seen on the merge.

sequence, the 32 bp 5'-end and CR2 sequence being required only quantitatively for normal expression levels of the proximal- and distal-most rings, respectively.

The key CR1 sequence includes separate activating and repressive regulatory information

Because it is essential for limb-specific expression, we decided to dissect the 68-bp-long CR1 sequence by substituting each 8 bp by a linker sequence (see Methods) (Fig. 4A). Significantly, all eight CR1-mutated LAE constructs (*LS1-8*) affected GFP expression either qualitatively or/and quantitatively (Figs. 4B and S3), and a half (*LS1*, 2, 5 and 8) displayed strong defects. *LS1* is the only mutated construct that led to strong up-regulation in inter-ring cells in the developing leg (Fig. 4B, arrows in E-E'). *LS2* showed the strongest quantitative defects, with almost a complete loss of GFP expression in both leg and antennal tissues (Fig. 4B, F-F' and K-K'). The *LS5* construct showed decreased GFP expression (30%), with a pronounced effect in the proximal-most ring in both appendages (Fig. 4B, brackets in G-G' and L-L'). The *LS8* construct displayed decreased GFP expression specifically in the proximal-most *bab2*-expressing ring, again in both leg and antennal tissues (Fig. 4B, brackets in H-H' and M-M'). Of note (i) the *LS4* and *LS6* constructs displayed a partial GFP intensity decrease in the proximal-most ring, particularly marked in leg tissues; (ii) *LS7* showed a partial inter-ring de-repression in leg tissues; (ii) whereas *LS3* and *LS6* exhibited a slight GFP intensity decrease specifically in antennal tissues (50% each) (Fig. S3).

Altogether these data indicate tightly associated leg and antennal CR1 regulatory information, with partially redundant functions. Of note, the key repressive (*LS1*) and activating (*LS2*) elements are juxtaposed, and the proximally-acting *LS8* activating element is adjacent to the 32-bp-most 5' sequence that is specifically required for expression in the proximal-most *bab2*-expressing cells (*S3* construct shown in Fig. 3), suggesting that together they constitute a proximal-ring activation functional unit.

Distal-less is the master direct bab2 activator in developing appendages

Having identified key *cis*-regulatory DNA elements, we next sought for trans-acting factors directly binding to the LAE. Here, we focused on positive *bab2* regulators. We noticed that the key positively-acting *LS2* motif is embedded within an A/T-rich sequence which includes three TAAT or ATTA motifs (AAAATTAATGGTAATAA), representing as many potential homeodomain (HD) binding sites. Given that *Dll* protein is cell-autonomously required for *bab2* expression (Campbell & Tomlinson, 1998; Chu et al, 2002; Galindo et al, 2002), we examined whether the *LAE-GFP* reporter is sensitive to *Dll* activity, using a clonal analysis.

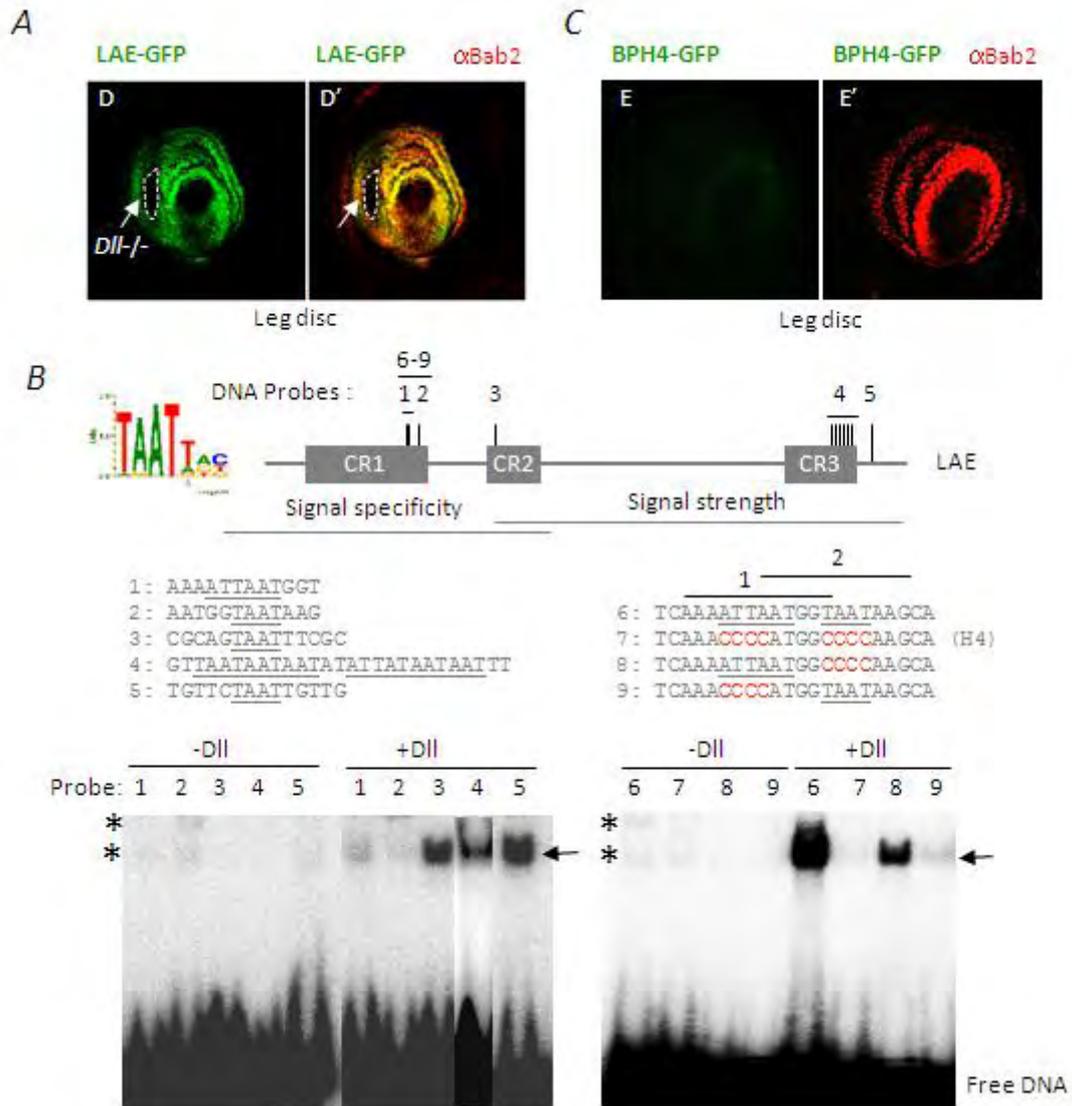


Figure 5: The Dll homeodomain protein positively regulates the LAE activity in all *bab2*-expressing cells, through direct binding to at least key CR1 AT-rich sequences.

(A) *Dll* activity is critically required for *LAE-GFP* expression in developing limbs. GFP expression (green) alone or in combination with Bab2 immunostaining (red) of leg imaginal discs dissected from late third-instar larvae carrying *Dll* null mutant clones (D-D'), are shown. A large *Dll*^{SA1} mutant clone (circled by dashed white line) is indicated by a white arrow in both panels. Both *bab2* (red) and *LAE-GFP* reporter (green) expressions were cell-autonomously lost in the *Dll* mutant cells.

(B) *In-vitro* Dll binds to TAAT motifs located throughout the LAE sequence. The sequence LOGO of the consensus Dll-binding site (mainly composed of a TAAT motif), as determined from selection in bacteria (Noyes et al, 2008), is shown in the left side. The positions of 11 TAAT/ATTA sequences within the LAE are indicated by vertical lines, and the corresponding probes used in the EMSA experiments indicated above. All but one (probe #5) sites are situated within the CR1-3 conserved regions. The LAE sequences (in grey) included in each labelled double-stranded DNA probe (see Methods) are indicated in beneath, the mutated nucleotides being written in red. PhosphorImager detection of DNA-protein complexes separated by electrophoresis on native polyacrylamide gels is shown in the lower part. *In-vitro* translated Dll was added or not, as indicated above the lanes, the numbers corresponding to the tested probe (#1 to 9). The positions of non- or Dll-specific shifted DNA-protein complexes are indicated by a horizontal arrow or by asterisks, respectively. The extended probes #6-9 included the overlapping probes #1-2. In contrast to the two singly-mutated probes (#8-9), the high-affinity Dll-specific complex detected with the wild-type probe (#6) was no longer detected with the doubly-mutated probe (#7).

(C) The triple TAAT/ATTA CR1 sequence is critical for LAE activity *in-vivo*. GFP (green) expression alone (D) and in combination with Bab2 (red) immunostaining (D') of a leg imaginal disc dissected from a late third-instar larva expressing the *H4-GFP* reporter construct, a *LAE-GFP* derivative mutated on the three CR1 TAAT/ATTA motifs (identical to EMSA probe #7). No GFP expression could be detected.

Similarly to endogenous *bab2*, GFP expression was completely lost in mitotic mutant clones for *Dll^{SA1}*, a null allele (Fig. 5A, C-C'). This indicates that Dll is cell-autonomously required for LAE regulatory activity, and therefore suggest a direct binding to the TAAT/ATTA sites present in the key CR1 motif.

In addition to the 3 present in CR1, the entire LAE comprises 8 additional putative HD-binding sites (Fig. 5B), of which 7 are clustered within the two other large evolutionarily-conserved regions (Figs. 5B and S2). Whereas one of those additional TAAT/ATTA motifs is included in the distal-specific CR2 region, suggesting a functional relevance for LAE-driven expression, the others might have at least a partial redundant function in expression intensity level. To test whether Dll is able to bind *in-vitro* to the TAAT/ATTA-containing LAE sequences, we used electrophoretic mobility shift assay (EMSA). As expected, all DNA fragments including one to six of the 11 TAAT/ATTA motifs present in the LAE, were capable of binding *in-vitro* translated Dll, albeit with distinct affinities (Fig. 5B). Unexpectedly, given the key role *in-vivo* of the LS2 region, each TAAT or ATTA site in CR1 showed rather low *in-vitro* affinity (fragments #1-2), when tested individually. We therefore tested whether Dll bind with a higher efficiency a larger fragment (TCAAATTAATGGTAATAAGCA, the AT-rich sequence covered by fragments #1-2 is underlined) including the three TAAT/ATTA motifs present in CR1. We found that Dll strongly bound this extended CR1 fragment (#6) (Fig. 5B), and stable interaction was strictly dependent upon the three TAAT/ATTA motifs. Singly-mutated (TAAT vs. CCCC) fragments bound Dll with much lower affinity, while the mutation of the double overlapping TAAT/ATTA sites (ATTAAT vs. CCCCAT; fragment #9) showing the strongest effect. These results indicate that Dll binds directly to the TAAT/ATTA-rich sequence, AAAATTAATGGTAATAA. Altogether, the clonal analysis and EMSA strongly suggest that Dll directly regulates *bab2* through the CR1 sequence in LAE.

To further confirm that CR1 mediates *in-vivo* regulation by Dll, we mutated all three TAAT/ATTA motifs within the *LAE-GFP* reporter (*H4* construct). *H4-GFP* expression was no longer detected in the leg (Fig. 5C, D-D') or the antennal tissues, indicating that HD-binding sites in CR1 are crucial for LAE activity *in-vivo*.

Altogether, *in-vitro* and *in-vivo* data provide strong evidence that Dll directly activates *bab2* transcription in developing limbs, through at least binding to the conserved HD-binding sites in the CR1 regulatory sequence.

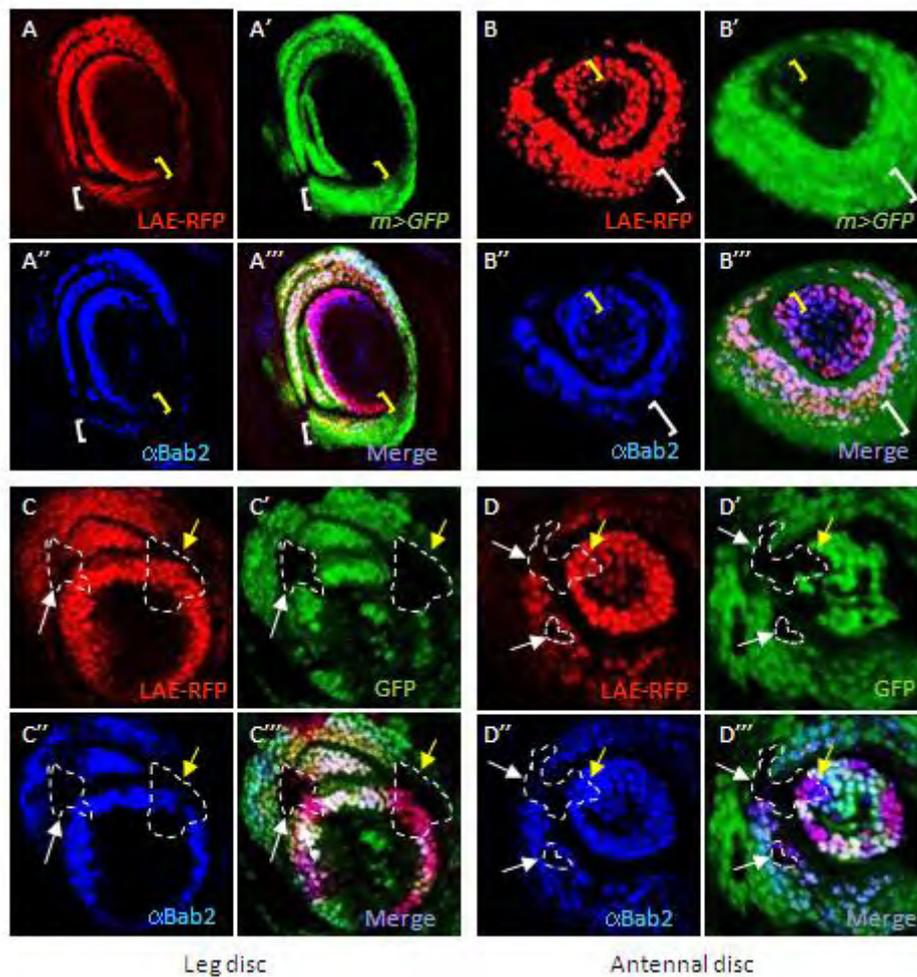


Figure 6: *rotund* is expressed in the proximal-most *bab2*-expressing cells and is required for *bab2* and *LAE-GFP* activation, in both leg and antenna.

Leg (A and C) and antennal (B and D) imaginal discs dissected from late third-instar larvae, either wild-type (A-B) or carrying *rn* null mutant clones (C-D). GFP (green) expression alone, RFP (red) expression alone and Bab2 immunostaining (blue) alone as well as the merge of all three, are shown. *rn* expression (A''-A''' and B''-B''') was monitored by combining *rn*-Gal4 and UAS-GFP constructs (*rn*>GFP). LAE regulatory activity (A-D) was monitored with a *LAE-RFP* (red) reporter (methods). *rn*>GFP expression is not detected in the distal-most *bab2*-expressing cells (yellow brackets), and extends more proximally than *bab2* expression (white brackets). In mosaic tissues, *rn*¹⁶ mutant clones (some are circled by dashed lines) did not express GFP (i.e. black areas). Note that only proximally-located mutant cells (white arrows), and not those located in the distal-most ring (yellow arrows), displayed strong *bab2* (and *LAE-RFP*) expression decrease, in either leg or antennal tissues.

Rotund contributes to bab2 activation in proximal leg and antennal rings

rotund (*rn*), which encodes a Zinc-Finger TF, is dynamically expressed in developing limbs, in patterns partly overlapping that of *bab2* (Godt et al, 1993; Natori et al, 2012; St Pierre et al, 2002). At the late L3 stage *rn-Gal4* reporter expression is only expressed in the proximal-most *bab2*-expressing tissues, and extends more proximally than *bab2*, particularly in antennal tissues (Fig. 6A-B; see yellow and white brackets, for distal versus proximal *bab2*-only and *rn*-only expressing cells, respectively.). To analyse the role of *rn* on *bab2* and LAE-driven (using a *LAE-RFP* construct, see Methods) expression, we generated mitotic clones, using the *rn*¹⁶ null allele. Strong reduction of endogenous *bab2* (in blue) and *LAE-RFP* reporter (in red) expression were detected, in a cell-autonomous manner, but only in those clones overlapping the proximal-most *bab2*-expressing rings, in either the leg or the antennal discs (Fig. 6C-D). This indicates that *rn* activity on *bab2* is restricted and contributes to *bab2* and LAE-driven expression in t1-3 leg and a3-a4 antennal tissues.

Rotund directly activates bab2 through the LAE T-rich 5'end region

Next, we examined whether the Rn protein directly controls *bab2* expression via the LAE in developing limb tissues. As detailed above (Fig. 4C), deletion analysis showed that a 32-bp regulatory element at the LAE 5'end was specifically required for GFP reporter expression in the proximal-most tarsal and antennal *bab2*-expressing ring(s). To further confirm their functional requirement, we deleted these 32 bp in the context of the 230-bp *S5-GFP* construct (Fig. 3), recapitulating the timing and sites of *bab2* expression, albeit with lower level than the full size LAE (Fig. 7A, C and H). The truncated *S5* fragment (*S10*) drove strong expression in the distal-most ring but only very weakly in the proximal-most ring(s), particularly in antennal tissues (Fig. 7A, brackets in E and J). These data confirm that the 32-pb-long 5'-end LAE region is required for LAE-driven expression in proximal *bab2*-expressing cells in both tarsal and antennal tissues.

The deleted 32-pb-long region includes a 6-pb-long poly-T track (T6) embedded within a 13-bp-long T-rich (T13) sequence (TTCGTTTTTTGTT), reminiscent of binding sites for the vertebrate Rn homolog (Torrunguang et al, 2002). To test whether the Rn protein effectively binds to T13 in LAE, EMSA experiments were performed with DNA probes covering either the T13 sequence (fragment #1) in isolation or the complete 32-bp sequence (#2) (Fig. 7B). *In-vitro* translated Rn efficiently bound to both probes, the largest 32-bp fragment giving a much stronger affinity (about 5 fold) (Fig. 7B, compare lanes 1 and 2). The sequence of the 32-bp probe fully matches the consensus binding site for *Drosophila* Rn, as

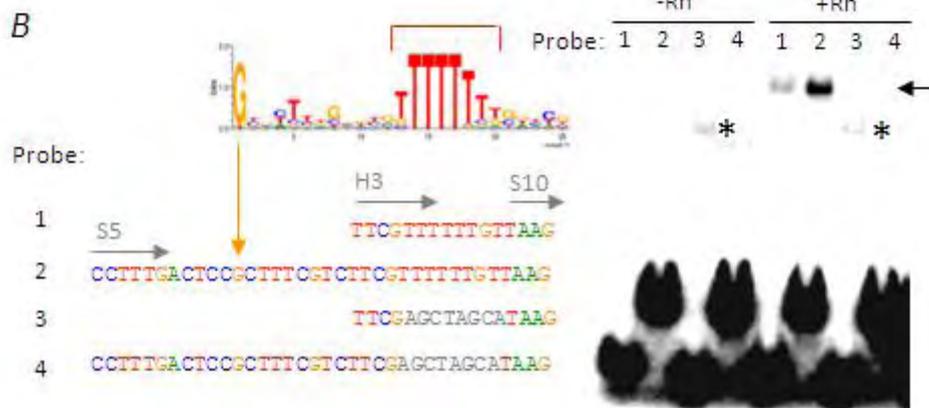
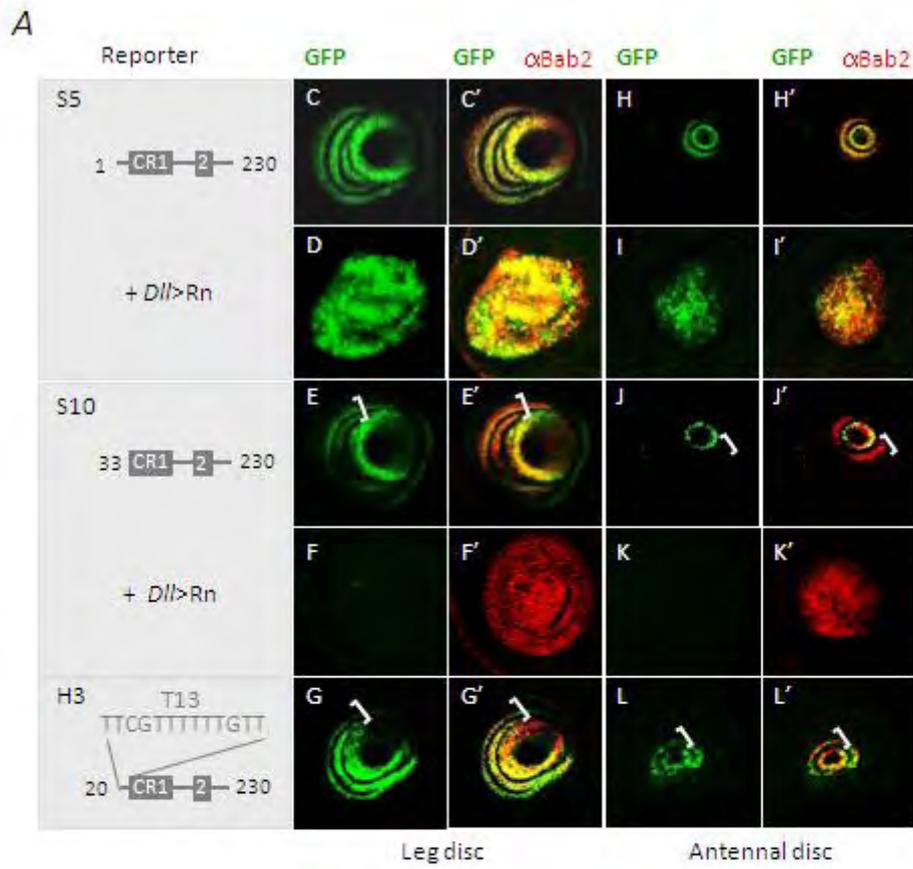


Figure 7: The Rotund Zinc-finger protein positively regulates *bab2* expression through direct binding to the T13 LAE sequence.

(A) The 32-bp T-rich LAE 5'-end sequence is necessary and sufficient for *rn*-dependent *bab2* (and GFP-reporter) activation in developing limbs. GFP (green) expression alone and in combination with Bab2 immunostaining (red) from dissected late L3 leg (C-G) and antennal (H-L) discs, expressing a given truncated derivative (indicated on the left side) of the *LAE-GFP* reporter, in the wild-type (C, E, G and H, J, L) or together (D, F and I, K) with two transgenes (*Dll-Gal4* + *UAS-Rn*) allowing over-expression of Rn under the control of the *Dll* regulatory sequences, are shown. The *S5-GFP* reporter construct drove wild-type leg (C-C') and antennal (H-H') *bab2* expression patterns. In contrast, GFP expression in the proximal-most *bab2*-expressing cells of its 5'-end truncated derivative *S10-GFP* was much weakly detected in leg or not at all in antennal tissues (brackets). The addition of the T13 sequence from the LAE 5'-end sequence to the *S10-GFP* construct (*H3*) was sufficient to increase proximal GFP expression levels (G and L, brackets), particularly in antennal tissues. When Rn was over-expressed throughout the *Dll* domain (*Dll*>Rn) endogenous *bab2* and *S5-GFP* reporter (D-I) were both up-regulated in nearly all Rn-expressing cells, in both developing appendages. In striking contrast, GFP expression in *S10* reporter tissues is not detected, indicating that the LAE 5'-end is critically required for Rn-mediated *S5-GFP* expression.

(B) Rn binds *in-vitro* to the 32-bp-long LAE 5'-end sequence. The sequence LOGO of the consensus Rn-binding site, including a T-rich track (bracket), as determined from selection in bacteria (Noyes et al, 2008), is shown in the upper right side. The LAE sequences included in the four labelled double-stranded DNA probes are indicated in beneath, with the same colour code. In probes #3-4 the wild-type T6 track was substituted by a neutral sequence shown in grey. Position of a consensus G is indicated by a vertical yellow line. The 5'-end positions of the *S5*-, *S10*- and *H3-GFP* reporter constructs tested in A, are indicated. EMSA experiments are shown on the left. *In-vitro* translated Rn was added or not as indicated, and line numbers refer to the four tested probes. The positions of non- or Rn-specific shifted DNA-protein complexes are indicated by a horizontal arrow or asterisks, respectively. Note that Rn bound the T13 fragment more strongly than the entire 32 bp LAE 5'-end fragment.

determined by recent bacterial one-hybrid binding site data (Noyes, 2012), with a preponderance of a G residue located 12 bp downstream of the LAE 5' end (see Fig. 7B). Most importantly the T6 track appeared critical for specific binding, as 32 bp and 13 bp fragments specifically mutated in the T6 track (#3-4) failed to be bound by Rn. The EMSA experiments thus indicate that the LAE T13 region constitutes a strong binding site for Rn.

To examine the functional importance of the Rn binding site *in-vivo*, we added the T13 sequence to the truncated *S10* construct to yield the *H3* construct (Fig. 7A). We found that the *H3-GFP* reporter activity was similar to that of *S5-GFP* (Fig. 7A, brackets in G and L). Even though the GFP expression level remained somewhat low in developing antenna, we conclude that the T13 sequence is sufficient for GFP-reporter activation in the proximal-most *bab2*-expressing rings in both the leg and antennal tissues.

To further confirm that T13 is essential to mediate direct regulation of *bab2* by the Rn transcription factor, we performed *rn* gain-of-function experiments, using the *S5* and *S10* reporter constructs. Ectopically-expressed Rn (*Dll>Rn*) was capable to activate *S5-GFP* and endogenous *bab2* expression throughout the *Dll*-expressing leg and antennal cells (Fig. 7A, panels D-D' and I-I'), whereas the T13-deficient *S10* construct displayed neither GFP expression in leg nor in antennal tissues (Fig. 7A, F-F' and K-K').

In addition to the clonal analyses, these data indicate (i) that Rn activity is necessary and sufficient for *bab2* and LAE-driven expression and (2) that Rn regulation of *bab2* expression is direct, via binding to the T13 sequence in the LAE.

Discussion

In this report, we characterize a 567-bp-long evolutionarily-conserved enhancer element, named LAE, situated in the middle of the 150 kb *bric-a-brac* locus, which is both necessary and sufficient to properly drive *bab2* expression in distal leg and antennal epithelial cells, from larval to adult stages. Interestingly, whereas leg and antennal regulatory signals are closely associated, full activation of the proximal- versus distal-most *bab2*-expressing ring(s) is dependent on distinct DNA elements (Fig. 8). Importantly, we provide compelling evidence that *Dll* and Rn protein directly activate *bab2* transcription in all and in the proximal-most ring(s), respectively. The *in-vivo* results are consistent with *Dll* homeodomain protein displaying *in-vitro* high affinity for a composite TAAT/ATTA-rich sequence, which is critical for LAE activity in developing distal limb tissues. Likewise, Rn binds a long T-rich sequence in LAE *in-vitro*, which is required for *bab2* up-regulation *in-vivo*.

Leg versus antennal bab2 regulation

The *Drosophila* antenna and leg are thought to be homologous structures evolved from a common ancestral appendage, as shown by leg-to-antenna or antenna-to-leg transformations caused by miss-expression of PD patterning or homeotic genes (Casares & Mann, 1998; Cummins et al, 2003; Dong et al, 2000; Duncan et al, 1998; Pai et al, 1998; Struhl, 1981). Although appendage-specific CRMs recapitulate *Dll* and *dac* expression in both developing leg and antenna (Estella et al, 2008; Giorgianni & Mann, 2011), *bab2* is the first example of a PD gene for which a single CRM is shown to be necessary and sufficient to drive a complex gene expression pattern in both appendages. In addition, none of our mutated LAE constructs specifically affected antennal or leg expression. Our findings thus support the idea that an ancestral PD genetic cascade emerged before limb diversification in insects.

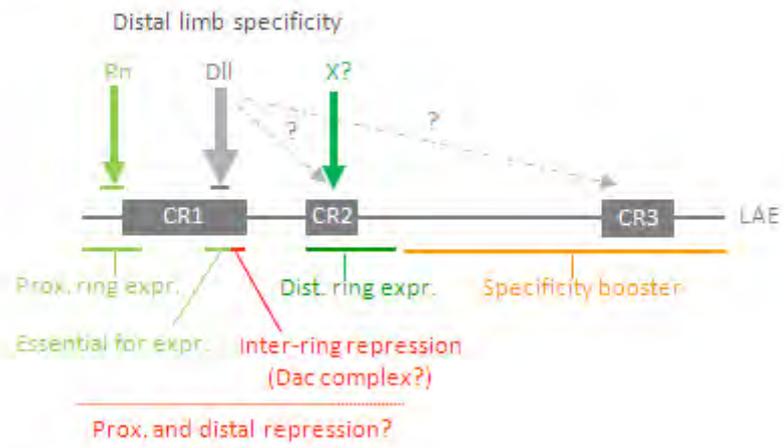
Molecular integration of both activators and repressors by a relatively-small CRM

In developing leg and antenna, *bab2* expression is dynamic and complex, going from broad distal expression at early L3 stage, distal and proximal as well as inter-ring repression at late L3 stage and beyond, giving rise to precisely positioned rings, to graded expression for each ring at pupal and adult stages (Fig. 8) (Chu et al, 2002; Couderc et al, 2002; Godt et al, 1993). Most of the specific regulation occurs through the 230-bp-long 5' part of the LAE (*S5* construct) which integrates positive inputs from both *Dll* and *Rn*. In addition, we have identified a 12 bp sequence at the CR1 3'-end (LS1 motif) required for leg inter-ring repression. The nature of the LS1-bound repressor remains to be identified, and particularly its functional link with *Dac*, a non-specific DNA binding protein known to function within a multi-protein complex in other developmental processes (Chen et al, 1997), has to be investigated. Finally, none of 16 different LAE reporter constructs displayed detectable de-repression in the distal-most territories, suggesting the existence of functionally-redundant repressive DNA elements or alternatively a competition between repressors and activators for binding the same sites. The direct implication of genetically-defined *bab2* distal repressors situated downstream the EGFR pathway, remains to be elucidated.

The Dll homeodomain protein is a distal limb-specific direct activator of bab2 expression

We provided compelling evidence that *bab2* is a direct target of the *Dll* HD-containing transcription factor, presumably acting through a set of evolutionarily-conserved TAAT/ATTA motifs, including at least the key AAATTAATGGTAAT composite sequence present in the CR1 sequence. Interestingly, similar A/T-rich *Dll* binding sites are present in

A



B

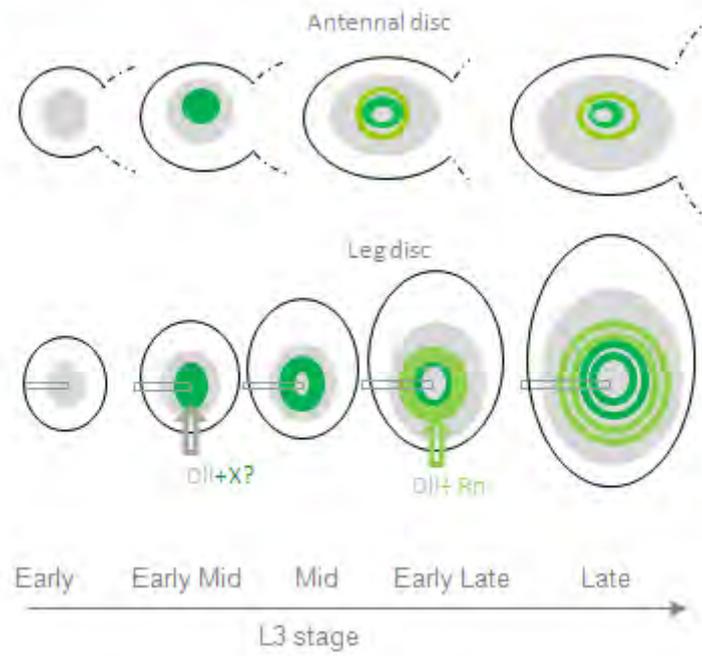


Figure 8: A model for LAE regulatory activity during distal limb development.

(A) Functional organization of the LAE. Positions of key *cis*-regulatory elements are shown in beneath the DNA stretch, with evolutionarily-conserved CR1-3 sequences shown as grey boxes. Within the 5'-half, critical for distal-limb specificity, regions required for ring activation and inter-ring repression are shown in green and red lines, respectively. Sequences responsible for proximal (prox.) and distal (dist.) repression have not been yet delineated, as indicated by a dotted line. The 3'-half, which is only required for high level of expression ("booster"), is depicted as an orange line. The binding sites for the direct *bab2* regulators Dll and Rn (in grey and light green, respectively) as well as for an unknown factor (X, in dark green) required for normal distal-most ring expression level, are indicated by vertical arrows (coloured according to B).

(B) A model for spatio-temporal regulation of *bab2* gene during the third-instar larval (L3) stage in antennal and leg imaginal discs. Onset of *bab2* expression (in dark green) occurs at the early/early-mid L3 stage as a circular domain throughout most of the Dll domain (in grey). A hypothetical distal activator (X) binding to CR2 (see A) might intervene from this stage. In response to EGFR signalling, *bab2* down-regulation in the distal-most territory occurs in mid L3, with the emergence of two rings in the antenna and a single distal ring in the leg. Only the *bab2*-expressing proximal antennal ring (in light green) requires *rn* activity. At the onset of *rn* expression in early-late L3, a second *bab2*-expressing ring (light green) emerges proximally in the leg disc. By late third-instar stage, *bab2* expression consists of two strong rings in the antenna and four concentric rings in the leg. The two proximal-most (light green) and the two distal-most (dark green) rings emerge both by tarsal growth and inter-ring down-regulation. Contrary to antenna, *rn* is no longer expressed in the late third-instar leg disc, and may no longer be required owing to the relief of *bab2* repression through decreased *dac* and possibly *bowl* activities. In the same vein, maintenance of antennal Rn (and Ss) expression may contribute to antagonize activities of *bowl*, *dac* and antennal-specific *bab2* repressors, such as *hth* and *spalt*. See main text for details.

enhancers of *ss* and *dac*, two other direct Dll target genes (Duncan et al, 2010; Giorgianni & Mann, 2011), suggesting a conserved CRM grammar for Dll-regulated genes. Although Dll is expressed throughout the larval, pupal and adult stages, onset of *bab2* expression starts at the early L3 stage, in the form of a distal domain prefiguring the distal ring (Fig. 8) (Chu et al, 2002). In this study, we identified a distal-specific *cis*-regulatory element encompassing the CR2 sequence. In addition to Dll (Fig. 5B), identification of other TF(s) binding this element will illuminate the molecular mechanisms allowing tissular specificity. Finally, the conservation of other Dll-binding sites may contribute to the expression “booster” effect of the LAE 3'-half (Fig. 8). Characterization of other TF(s) interacting with LAE 3'-half sequences may help to better understand CRM activity, both in terms of tissular specificity and expression enhancement.

A specific role for the Rn zinc-finger protein in proximal regulation

The requirement of *rotund* activity in *bab2* regulation contrast with the data reported in the literature (St Pierre et al, 2002), showing apparently normal *bab2* expression in *rn* mutant discs (St Pierre et al, 2002). Additionally, given that *rn* expression in leg ts1-3 tissues is partially dependent on *ss* activity (Natori et al, 2012), our data provide a rationale for the partial requirement for *ss* activity in normal *bab2* regulation, at the exclusion of the distal-most ring (Chu et al, 2002). For the first time, we report that ectopically-expressed Rn protein is sufficient to activate *bab2* throughout the *Dll* domain in either developing leg or antenna. Although required in both appendages, *rn* expression is only maintained in antennal tissues, in which no molecular function was previously described. In the developing leg, in addition to its role in ts1-3 growth, we propose that transiently-expressed Rn counteracts repressive activities of *dac* and/or *bowl*, both of which are dynamically expressed during the critical L3 stage (de Celis Ibeas & Bray, 2003; Natori et al, 2012). In fact, de Celis and Bray (de Celis Ibeas & Bray, 2003) anticipated that a transiently-expressed *bab2* activator should be present to relieve transient repression by *bowl*. Consistent with this view, in addition to its contribution to *rn* activation, *ss* represses *bowl* (and *dac*) expression (de Celis Ibeas & Bray, 2003; Pueyo & Couso, 2008). Given that *bowl* in turn represses *rn* expression (Greenberg & Hatini, 2009), these cross-interactions are likely to reinforce *bab2* expression in the form of rings. Given that *bowl* (and *dac*) expression have decayed in the tarsal cells which transiently expressed *ss* and *rn* at the mid third-instar stage, maintenance of *bab2* expression later on would no longer require Rn activity. This working model is attracting and deserves to be investigated by further experiments.

Conservation among Drosophilidae and beyond

The LAE is systematically located between *bab1* and *bab2* transcription units in 22 Drosophilidae species for which the entire sequence of the *bab* locus is available, suggesting that the LAE is under a strong selective pressure. Given that we have shown that the LAE is essential *in-vivo*, “shadow” enhancers (Hong et al, 2008) are unlikely to contribute to *bab2* regulation in normal culture conditions. In support to their key functions, the CR1-2 sequences are partially conserved in the tsetse fly *Glossina morsitans* (Fig. S4), which diverged from Drosophilidae about 260 million years ago (Liu et al, 2010). Further, we have shown that the LAE from *D. virilis* ensures normal regulatory functions in *D. melanogaster*. Although strongly conserved among Sophophora subgenus species (i.e. *D. melanogaster* group), the T13 Rn-binding site is poorly conserved in *D. virilis* and related Drosophila subgenus species (Fig. S4). It is formally possible that Rn has been co-opted lately in the Sophophora subgenus. Alternatively, *D. virilis* Rn functions through subgenus-specific LAE sequences. Consistent with this hypothesis, a T13-related sequence, only conserved among Drosophila subgenus species, is located in between CR1-2 (Fig. S4). Deciphering the molecular basis for Rn action in the positive regulation of *bab2* may tackle these evolutionary issues.

Material and methods

Fly stocks, culture and genetic manipulations

Drosophila lines were grown on standard yeast extract-sucrose medium. The *vasa*-PhiC31 Zh2A *attP* stock was obtained from F. Karch and was used to generate all the transgenic reporter and PacMan constructs. *rn* and *Dll* mutant clones were generated by 30 minute heat shocks at 38°C, during early first to late second-instar larvae of genotypes: *y w hsFlp; FRT82B Ub-GFP/FRT82B rn¹⁶* and *y w hsFlp; Ub-GFP FRT42D/Dll^{SA1} FRT42D. UAS-Rn¹, UAS-GFP^{S65T} and rn-Gal4* lines were obtained from the Bloomington stock center. *Dll^{EM212}-Gal4* line was provided by M. Suzanne.

Reporter constructs, mutagenesis, BAC recombineering and germline transformation

Genomic DNA fragments from the *D. melanogaster* or *D. virilis* *bab* locus were amplified using standard PCR (sequences of oligonucleotides are available upon request), cloned into pBP-S3aG, a home-made derivative of the *attB*-containing pS3aG plasmid, obtained from S. Carroll (Williams et al, 2008). Details are available upon requests. The pLAE-RFP and pLAE-*bab2* constructs were made by substituting the GFP insert of the pLAE-GFP construct

by a pH2B-RFP insert (obtained from A. Vincent) or a *bab2* full-length cDNA (from pNB-*bab2*), respectively. Site-directed mutagenesis (including linker scanning) was performed by PCR, using the overlap extension method (Ho et al, 1989). All constructs were sequence-verified. BAC recombineering and PhiC31-mediated germline transformation were done as described (Venken et al, 2006). The same 2A *attP* platform on the X chromosome was used for all constructs.

Immunofluorescence and imaging

Imaginal discs were prepared and stained using standard procedures. Confocal analyses were done with a LEICA TCS SP5 microscope. Rat anti-Bab2 antibodies (Couderc et al, 2002) were used at 1/2000. GFP fluorescence quantification was done with the ImageJ software, from 10 distinct leg and antennal discs of each genotype.

In-vitro translation and electrophoretic mobility-shift assay (EMSA)

Dll and Rn proteins were synthesized from coupled *in-vitro* transcription/translation with T7 RNA polymerase and rabbit reticulocyte lysate (TNT assay, Promega), using as DNA matrices the pET-Dll and pCS2-MycRn plasmid constructs, obtained from S. Cohen and P. Couso, respectively. EMSA experiments were done mainly as described (Gebelein et al, 2004), using Novex 6% DNA retardation gels (InvitroGene). Probes were assembled from synthetic oligonucleotides including 4 additional G bases at their 5' ends, labeled with the Klenow fragment of *E. coli* DNA polymerase I, in presence of ³²PdCTP, and purified on mini Quick spin columns (Roche). Free and shifted probes were revealed with a PhosphorImager.

Acknowledgments

We thank T. Kojima, G. Campbell, G. Boekhoff-Falk, S. Cohen, S. Carroll, P. Couso, M. Suzanne, A. Vincent, F. Laski and the Bloomington Stock Center for fly stocks and reagents, the Developmental Studies Hybridoma Bank (DSHB) at the University of Iowa for antibodies. We are grateful to A. Vincent, S. Plaza, M. Crozatier and C. Immarigeon for discussions and helpful suggestions on the manuscript or during the course of this work. We are grateful to C. Faucher and O. Bardot for contributions to this project. We also thank J. Favier and S. Bernat-Fabre for technical assistance. Lastly, we acknowledge the help of B. Ronsin in image acquisition and for the availability of the Toulouse RIO Imaging platform. This work was

supported by grants from the French governmental agency for Research (CNRS), the Paul Sabatier University (UPS) and the "Association pour la Recherche sur le Cancer" (ARC). A.B was supported by the CNRS and the Midi Pyrenees Region. L.M.V. was supported by the Mexican CONACYT.

Supplementary information

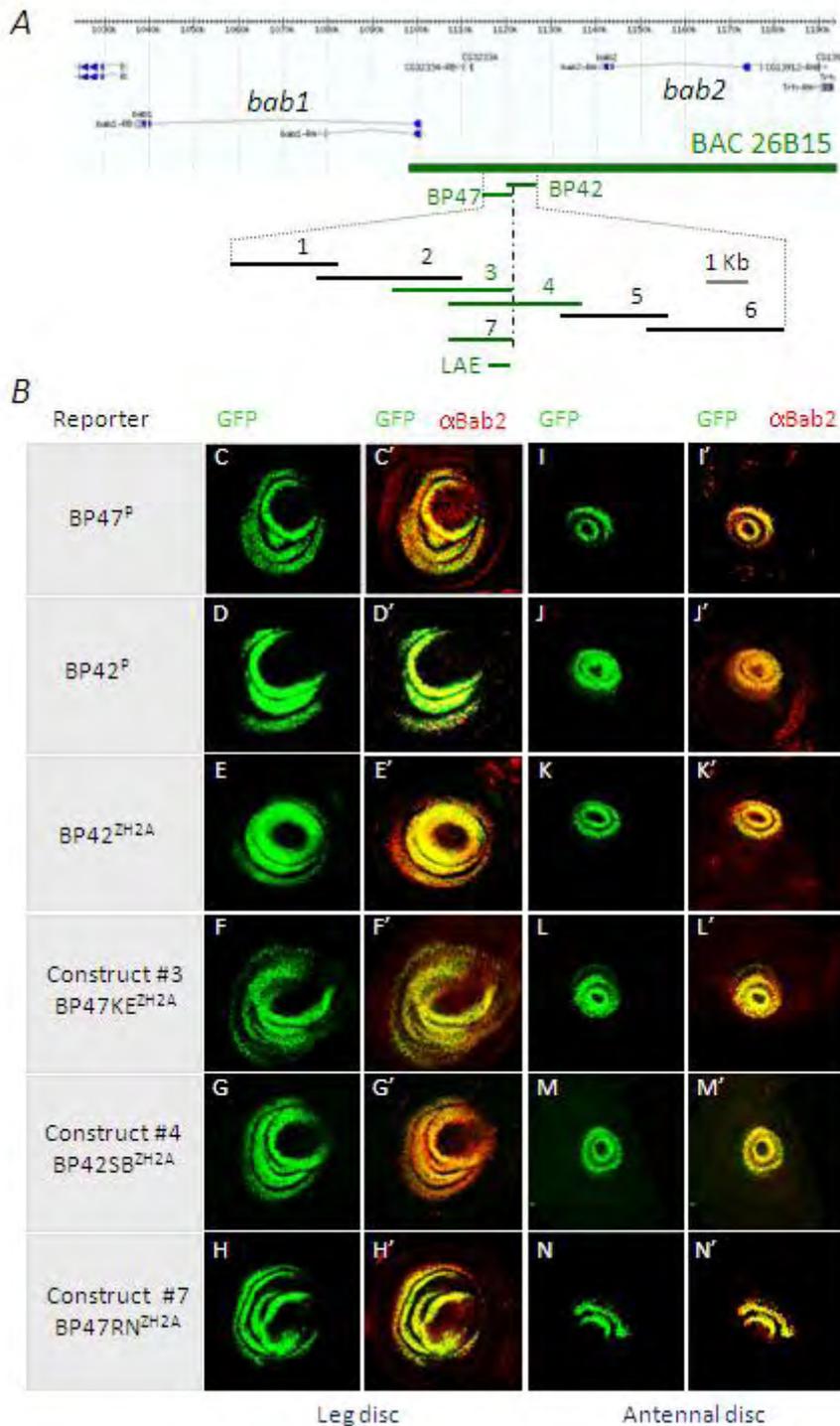


Figure S1: A site-specific GFP-reporter assay identified a single leg and antennal enhancer (LAE) element situated in the middle of the *bric-a-brac* locus.

(A) The *bab* locus is represented as shown in Fig 1A.

(B) Leg (C-H) and antennal (J-N) imaginal discs from dissected late third-instar larvae, expressing each tested GFP-reporter construct (depicted in the left side). GFP expression (green) alone and combined with Bab2 immunostaining (red) are shown for each construct. The reporter constructs encompassing the 1.5 kb-long BP47/42 overlapping region, including the 1.5 kb genomic fragment in isolation (construct #7), all recapitulate the limb-specific *bab2* expression.

CR1

```

D.melanogaster 1  CTTTGGACTCCGCTTCGCTCTCG  FTTTGTGTAAGCTGCAAAATCGGGGCAATCGTAAAGATCACGAAAAACAATCAAAAATAATGGTAATAAGCAGTTT  GC
D.simulans 1  CTTTGGACTCCGCTTCGCTCTCG  FTTTGTGTAAGCTGCAAAATCGGGGCAATCGTAAAGATCACGAAAAACAATCAAAAATAATGGTAATAAGCAGTTT  GC
D.secchelia 1  CTTTGGACTCCGCTTCGCTCTCG  FTTTGTGTAAGCTGCAAAATCGGGGCAATCGTAAAGATCACGAAAAACAATCAAAAATAATGGTAATAAGCAGTTT  GC
D.yakuba 1  CTTTGGACTCCGCTTCGCTCTCG  FTTTGTGTAAGCTGCAAAATCGGGGCAATCGTAAAGATCACGAAAAACAATCAAAAATAATGGTAATAAGCAGTTT  GC
D.erecta 1  CTTTGGACTCCGCTTCGCTCTCG  FTTTGTGTAAGCTGCAAAATCGGGGCAATCGTAAAGATCACGAAAAACAATCAAAAATAATGGTAATAAGCAGTTT  GC
D.ficusphila 1  CTTTGGACTCCGCTTCGCTCTCG  FTTTGTGTAAGCTGCAAAATCGGGGCAATCGTAAAGATCACGAAAAACAATCAAAAATAATGGTAATAAGCAGTTT  GC
D.takahashii 1  CTTTGGACTCCGCTTCGCTCTCG  FTTTGTGTAAGCTGCAAAATCGGGGCAATCGTAAAGATCACGAAAAACAATCAAAAATAATGGTAATAAGCAGTTT  GC
D.elegans 1  CTTTGGACTCCGCTTCGCTCTCG  FTTTGTGTAAGCTGCAAAATCGGGGCAATCGTAAAGATCACGAAAAACAATCAAAAATAATGGTAATAAGCAGTTT  GC
D.kikkawai 1  CTTTGGACTCCGCTTCGCTCTCG  FTTTGTGTAAGCTGCAAAATCGGGGCAATCGTAAAGATCACGAAAAACAATCAAAAATAATGGTAATAAGCAGTTT  GC
D.rhopaloo 1  CTTTGGACTCCGCTTCGCTCTCG  FTTTGTGTAAGCTGCAAAATCGGGGCAATCGTAAAGATCACGAAAAACAATCAAAAATAATGGTAATAAGCAGTTT  GC
D.biarmipes 1  CTTTGGACTCCGCTTCGCTCTCG  FTTTGTGTAAGCTGCAAAATCGGGGCAATCGTAAAGATCACGAAAAACAATCAAAAATAATGGTAATAAGCAGTTT  GC
D.eugracilis 1  CTTTGGACTCCGCTTCGCTCTCG  FTTTGTGTAAGCTGCAAAATCGGGGCAATCGTAAAGATCACGAAAAACAATCAAAAATAATGGTAATAAGCAGTTT  GC
D.bipectinata 1  CTTTGGACTCCGCTTCGCTCTCG  FTTTGTGTAAGCTGCAAAATCGGGGCAATCGTAAAGATCACGAAAAACAATCAAAAATAATGGTAATAAGCAGTTT  GC
D.ananassae 1  CTTTGGACTCCGCTTCGCTCTCG  FTTTGTGTAAGCTGCAAAATCGGGGCAATCGTAAAGATCACGAAAAACAATCAAAAATAATGGTAATAAGCAGTTT  GC
D.pseudoobscura 1  CTTTGGACTCCGCTTCGCTCTCG  FTTTGTGTAAGCTGCAAAATCGGGGCAATCGTAAAGATCACGAAAAACAATCAAAAATAATGGTAATAAGCAGTTT  GC
D.persimilis 1  CTTTGGACTCCGCTTCGCTCTCG  FTTTGTGTAAGCTGCAAAATCGGGGCAATCGTAAAGATCACGAAAAACAATCAAAAATAATGGTAATAAGCAGTTT  GC
D.miranda 1  CTTTGGACTCCGCTTCGCTCTCG  FTTTGTGTAAGCTGCAAAATCGGGGCAATCGTAAAGATCACGAAAAACAATCAAAAATAATGGTAATAAGCAGTTT  GC
D.willistoni 1  CTTTGGACTCCGCTTCGCTCTCG  FTTTGTGTAAGCTGCAAAATCGGGGCAATCGTAAAGATCACGAAAAACAATCAAAAATAATGGTAATAAGCAGTTT  GC
D.mojavensis 1  CTTTGGACTCCGCTTCGCTCTCG  FTTTGTGTAAGCTGCAAAATCGGGGCAATCGTAAAGATCACGAAAAACAATCAAAAATAATGGTAATAAGCAGTTT  GC
D.virilis 1  CTTTGGACTCCGCTTCGCTCTCG  FTTTGTGTAAGCTGCAAAATCGGGGCAATCGTAAAGATCACGAAAAACAATCAAAAATAATGGTAATAAGCAGTTT  GC
D.americana 1  CTTTGGACTCCGCTTCGCTCTCG  FTTTGTGTAAGCTGCAAAATCGGGGCAATCGTAAAGATCACGAAAAACAATCAAAAATAATGGTAATAAGCAGTTT  GC
D.grimshawi 1  TAGCCCTCCGCTTCGCTCTCG  FTTTGTGTAAGCTGCAAAATCGGGGCAATCGTAAAGATCACGAAAAACAATCAAAAATAATGGTAATAAGCAGTTT  GC
  
```

```

D.melanogaster 104  AGGCGAAG  GCCTCGGGAC
D.simulans 104  AGGCGAAG  GCCTCGGGAC
D.secchelia 104  AGGCGAAG  GCCTCGGGAC
D.yakuba 104  AGGCGAAG  GCCTCGGGAC
D.erecta 104  AGGCGAAG  GCCTCGGGAC
D.ficusphila 103  AGGCGAAG  GCCTCGGGAC
D.takahashii 103  AGGCGAAG  GCCTCGGGAC
D.elegans 103  AGGCGAAG  GCCTCGGGAC
D.kikkawai 96  CGCCGAAG  GCCTCGGGAC -AGCAGCAGGAGCAGCGT
D.rhopaloo 103  AGGCGAAG  GCCTCGGGAC
D.biarmipes 103  AGGCGAAG  GCCTCGGGAC
D.eugracilis 103  AGGCGAAG  GCCTCGGGAC
D.bipectinata 94  AGGCGAAG  GCCTCGGGAC
D.ananassae 107  AGGCGAAG  GCCTCGGGAC
D.pseudoobscura 107  AGGCGAAG  GCCTCGGGAC
D.persimilis 103  AGGCGAAG  GCCTCGGGAC
D.miranda 103  AGGCGAAG  GCCTCGGGAC
D.willistoni 107  AGGCGAAG  GCCTCGGGAC
D.mojavensis 106  AGGCGAAG  GCCTCGGGAC
D.virilis 104  AGGCGAAG  GCCTCGGGAC
D.americana 110  AGGCGAAG  GCCTCGGGAC
D.grimshawi 109  AGGCGAAG  GCCTCGGGAC
  
```

CR2

```

D.melanogaster 125  GACTCCTTCAGCGAAATTAAGTGGTACAAACAGAGCCGCTCGGCA
D.simulans 125  GACTCCTTCAGCGAAATTAAGTGGTACAAACAGAGCCGCTCGGCA
D.secchelia 105  GACTCCTTCAGCGAAATTAAGTGGTACAAACAGAGCCGCTCGGCA
D.yakuba 125  GACTCCTTCAGCGAAATTAAGTGGTACAAACAGAGCCGCTCGGCA
D.erecta 125  GACTCCTTCAGCGAAATTAAGTGGTACAAACAGAGCCGCTCGGCA
D.ficusphila 124  GACTCCTTCAGCGAAATTAAGTGGTACAAACAGAGCCGCTCGGCA
D.takahashii 124  GACTCCTTCAGCGAAATTAAGTGGTACAAACAGAGCCGCTCGGCA
D.elegans 132  GACTCCTTCAGCGAAATTAAGTGGTACAAACAGAGCCGCTCGGCA
D.kikkawai 135  GACTCCTTCAGCGAAATTAAGTGGTACAAACAGAGCCGCTCGGCA
D.rhopaloo 124  GACTCCTTCAGCGAAATTAAGTGGTACAAACAGAGCCGCTCGGCA
D.biarmipes 124  GACTCCTTCAGCGAAATTAAGTGGTACAAACAGAGCCGCTCGGCA
D.eugracilis 124  GACTCCTTCAGCGAAATTAAGTGGTACAAACAGAGCCGCTCGGCA
D.bipectinata 128  GACTCCTTCAGCGAAATTAAGTGGTACAAACAGAGCCGCTCGGCA
D.ananassae 141  GACTCCTTCAGCGAAATTAAGTGGTACAAACAGAGCCGCTCGGCA
D.pseudoobscura 144  GACTCCTTCAGCGAAATTAAGTGGTACAAACAGAGCCGCTCGGCA
D.persimilis 144  GACTCCTTCAGCGAAATTAAGTGGTACAAACAGAGCCGCTCGGCA
D.miranda 144  GACTCCTTCAGCGAAATTAAGTGGTACAAACAGAGCCGCTCGGCA
D.willistoni 166  GACTCCTTCAGCGAAATTAAGTGGTACAAACAGAGCCGCTCGGCA
D.mojavensis 193  GACTCCTTCAGCGAAATTAAGTGGTACAAACAGAGCCGCTCGGCA
D.virilis 189  GACTCCTTCAGCGAAATTAAGTGGTACAAACAGAGCCGCTCGGCA
D.americana 197  GACTCCTTCAGCGAAATTAAGTGGTACAAACAGAGCCGCTCGGCA
D.grimshawi 194  GACTCCTTCAGCGAAATTAAGTGGTACAAACAGAGCCGCTCGGCA
  
```

```

D.melanogaster 170  AGGGGCC  SCTCCCTC  GGTTCCTC  GCTTGTCCAGGTTCTCT  TCCCTC  TCCCTC  TCCCTC  TCCCTC
D.simulans 170  AGGGGCC  SCTCCCTC  GGTTCCTC  GCTTGTCCAGGTTCTCT  TCCCTC  TCCCTC  TCCCTC  TCCCTC
D.secchelia 150  AGGGGCC  SCTCCCTC  GGTTCCTC  GCTTGTCCAGGTTCTCT  TCCCTC  TCCCTC  TCCCTC  TCCCTC
D.yakuba 170  AGGGGCC  SCTCCCTC  GGTTCCTC  GCTTGTCCAGGTTCTCT  TCCCTC  TCCCTC  TCCCTC  TCCCTC
D.erecta 170  AGGGGCC  SCTCCCTC  GGTTCCTC  GCTTGTCCAGGTTCTCT  TCCCTC  TCCCTC  TCCCTC  TCCCTC
D.ficusphila 169  AGGGGCC  SCTCCCTC  GGTTCCTC  GCTTGTCCAGGTTCTCT  TCCCTC  TCCCTC  TCCCTC  TCCCTC
D.takahashii 177  AGGGGCC  SCTCCCTC  GGTTCCTC  GCTTGTCCAGGTTCTCT  TCCCTC  TCCCTC  TCCCTC  TCCCTC
D.elegans 169  AGGGGCC  SCTCCCTC  GGTTCCTC  GCTTGTCCAGGTTCTCT  TCCCTC  TCCCTC  TCCCTC  TCCCTC
D.kikkawai 180  AGGGGCC  SCTCCCTC  GGTTCCTC  GCTTGTCCAGGTTCTCT  TCCCTC  TCCCTC  TCCCTC  TCCCTC
D.rhopaloo 169  AGGGGCC  SCTCCCTC  GGTTCCTC  GCTTGTCCAGGTTCTCT  TCCCTC  TCCCTC  TCCCTC  TCCCTC
D.biarmipes 169  AGGGGCC  SCTCCCTC  GGTTCCTC  GCTTGTCCAGGTTCTCT  TCCCTC  TCCCTC  TCCCTC  TCCCTC
D.eugracilis 169  AGGGGCC  SCTCCCTC  GGTTCCTC  GCTTGTCCAGGTTCTCT  TCCCTC  TCCCTC  TCCCTC  TCCCTC
D.bipectinata 173  AGGGGCC  SCTCCCTC  GGTTCCTC  GCTTGTCCAGGTTCTCT  TCCCTC  TCCCTC  TCCCTC  TCCCTC
D.ananassae 186  AGGGGCC  SCTCCCTC  GGTTCCTC  GCTTGTCCAGGTTCTCT  TCCCTC  TCCCTC  TCCCTC  TCCCTC
D.pseudoobscura 189  AGGGGCC  SCTCCCTC  GGTTCCTC  GCTTGTCCAGGTTCTCT  TCCCTC  TCCCTC  TCCCTC  TCCCTC
D.persimilis 189  AGGGGCC  SCTCCCTC  GGTTCCTC  GCTTGTCCAGGTTCTCT  TCCCTC  TCCCTC  TCCCTC  TCCCTC
D.miranda 189  AGGGGCC  SCTCCCTC  GGTTCCTC  GCTTGTCCAGGTTCTCT  TCCCTC  TCCCTC  TCCCTC  TCCCTC
D.willistoni 211  AGGGGCC  SCTCCCTC  GGTTCCTC  GCTTGTCCAGGTTCTCT  TCCCTC  TCCCTC  TCCCTC  TCCCTC
D.mojavensis 285  AAAAGTACTTG  AGGGGCC  SCTCCCTC  FTGC  GGTTCCTC  GCTTGTCCAGGTTCTCT  TCCCTC  TCCCTC  TCCCTC  TCCCTC
D.virilis 270  AGAAGTACTTG  AGGGGCC  SCTCCCTC  FTGC  GGTTCCTC  GCTTGTCCAGGTTCTCT  TCCCTC  TCCCTC  TCCCTC  TCCCTC
D.americana 278  AGAAGTACTTG  AGGGGCC  SCTCCCTC  FTGC  GGTTCCTC  GCTTGTCCAGGTTCTCT  TCCCTC  TCCCTC  TCCCTC  TCCCTC
D.grimshawi 297  TGAAGTACTTG  AGGGGCC  SCTCCCTC  FTGC  GGTTCCTC  GCTTGTCCAGGTTCTCT  TCCCTC  TCCCTC  TCCCTC  TCCCTC
  
```

```

D.melanogaster 235  CACGATCC  TCCGCTCC  TGGCCCTG  TCC  CCCCAGAAAGAGCA  AAC  CAACAGTAGTCCCAGTC  TGGGAT  TGT
D.simulans 218  CACGATCC  TCCGCTCC  TGGCCCTG  TCC  CCCCAGAAAGAGCA  AAC  CAACAGTAGTCCCAGTC  TGGGAT  TGT
D.secchelia 197  CACGATCC  TCCGCTCC  TGGCCCTG  TCC  CCCCAGAAAGAGCA  AAC  CAACAGTAGTCCCAGTC  TGGGAT  TGT
D.yakuba 244  CACGATCC  TCCGCTCC  TGGCCCTG  TCC  CCCCAGAAAGAGCA  AAC  CAACAGTAGTCCCAGTC  TGGGAT  TGT
D.erecta 252  CACGATCC  TCCGCTCC  TGGCCCTG  TCC  CCCCAGAAAGAGCA  AAC  CAACAGTAGTCCCAGTC  TGGGAT  TGT
D.ficusphila 216  CACGATCC  TCCGCTCC  TGGCCCTG  TCC  CCCCAGAAAGAGCA  AAC  CAACAGTAGTCCCAGTC  TGGGAT  TGT
D.takahashii 224  CACGATCC  TCCGCTCC  TGGCCCTG  TCC  CCCCAGAAAGAGCA  AAC  CAACAGTAGTCCCAGTC  TGGGAT  TGT
D.elegans 234  CACGATCC  TCCGCTCC  TGGCCCTG  TCC  CCCCAGAAAGAGCA  AAC  CAACAGTAGTCCCAGTC  TGGGAT  TGT
D.kikkawai 222  CACGATCC  TCCGCTCC  TGGCCCTG  TCC  CCCCAGAAAGAGCA  AAC  CAACAGTAGTCCCAGTC  TGGGAT  TGT
D.rhopaloo 216  CACGATCC  TCCGCTCC  TGGCCCTG  TCC  CCCCAGAAAGAGCA  AAC  CAACAGTAGTCCCAGTC  TGGGAT  TGT
D.biarmipes 216  GACCCCTC  TCCGCTCC  TGGCCCTG  TCC  CCCCAGAAAGAGCA  AAC  CAACAGTAGTCCCAGTC  TGGGAT  TGT
D.eugracilis 216  CACGATCC  TCCGCTCC  TGGCCCTG  TCC  CCCCAGAAAGAGCA  AAC  CAACAGTAGTCCCAGTC  TGGGAT  TGT
D.bipectinata 245  GACGATCC  TCCGCTCC  TGGCCCTG  TCC  CCCCAGAAAGAGCA  AAC  CAACAGTAGTCCCAGTC  TGGGAT  TGT
D.ananassae 245  GACGATCC  TCCGCTCC  TGGCCCTG  TCC  CCCCAGAAAGAGCA  AAC  CAACAGTAGTCCCAGTC  TGGGAT  TGT
  
```


D.ficusphila 286 -----
 D.takahashii 281 -----
 D.elegans 309 -----
 D.kikkawai 276 -----
 D.rhopaloo 291 -----
 D.biarmipes 273 -----
 D.eugracilis 278 -----
 D.bipectinata 322 -----
 D.ananassae 316 -----
 D.pseudoobscura 297 -----
 D.persimilis 308 -----
 D.miranda 321 -----
 D.willistoni 316 -----
 D.mojavensis 810 GAGTACTACAGGTATTTTATTTCTTAATCATATCATTAAATATACTTAATAAACCTATATTTCCGACTCCGTCTATAAAAATTATCTTAATCTAATCACATTTTA
 D.virilis 394 -----
 D.americana 408 -----
 D.grimshawi 414 -----

D.melanogaster 305 -----
 D.simulans 286 -----
 D.secchelia 264 -----
 D.yakuba 309 -----
 D.erecta 323 -----
 D.ficusphila 286 -----
 D.takahashii 281 -----
 D.elegans 309 -----
 D.kikkawai 276 -----
 D.rhopaloo 291 -----
 D.biarmipes 273 -----
 D.eugracilis 278 -----
 D.bipectinata 322 -----
 D.ananassae 316 -----
 D.pseudoobscura 297 -----
 D.persimilis 308 -----
 D.miranda 321 -----
 D.willistoni 316 -----
 D.mojavensis 920 AGTATTTTTATATGCCTATACACCCAAATTCATCTGTAAATTTGGCATTAACTCTAATATAGCTATACCTAATATACCTATTTCCCGAGTTATGTTCTATAGAATC
 D.virilis 394 -----
 D.americana 408 -----
 D.grimshawi 414 -----

D.melanogaster 305 -----
 D.simulans 286 -----
 D.secchelia 264 -----
 D.yakuba 309 -----
 D.erecta 323 -----
 D.ficusphila 286 -----
 D.takahashii 281 -----
 D.elegans 309 -----
 D.kikkawai 276 -----
 D.rhopaloo 291 -----
 D.biarmipes 273 -----
 D.eugracilis 278 -----
 D.bipectinata 322 -----
 D.ananassae 316 -----
 D.pseudoobscura 297 -----
 D.persimilis 308 -----
 D.miranda 321 -----
 D.willistoni 316 -----
 D.mojavensis 1030 TTATCTTAAATCTAATCCTATTATTATCATATTTAAATATCTTTATAATCCTATGCATATCTGTAAATTTTGTATAAATCTAATATCTATATTTCCCGAGTTA
 D.virilis 394 -----
 D.americana 408 -----
 D.grimshawi 414 -----

D.melanogaster 305 -----
 D.simulans 286 -----
 D.secchelia 264 -----
 D.yakuba 309 -----
 D.erecta 323 -----
 D.ficusphila 286 -----
 D.takahashii 281 -----
 D.elegans 309 -----
 D.kikkawai 276 -----
 D.rhopaloo 291 -----
 D.biarmipes 273 -----
 D.eugracilis 278 -----
 D.bipectinata 322 -----
 D.ananassae 316 -----
 D.pseudoobscura 297 -----
 D.persimilis 308 -----
 D.miranda 321 -----
 D.willistoni 316 -----
 D.mojavensis 1140 GCCTTTTAAATCATACCTTTAAAGATCATCACAAATGTTGCAGAACAAATTTCTCTATAGTTTTAGTTTCTTAAGCTGAGCGGGTATTCGTAGTCATTTTGA
 D.virilis 394 -----
 D.americana 408 -----
 D.grimshawi 414 -----

D.melanogaster 305 ----- TGGGCGTTCAC TGGC ----- GGCTGGGACTCAAAACCTCGATGCCCATGCGG-
 D.simulans 286 ----- TGGGCGTTCAC TGGC ----- GGCTGGGACTCAAAACCTCGATGCCCATGCGG-
 D.secchelia 264 ----- TGGGCGTTCAC TGGC ----- GGCTGGGACTCAAAACCTCGATGCCCATGCGG-
 D.yakuba 309 ----- TGGGCGTTCAC TGGC ----- GGCTGGGACTCAAAACCTCGATGCCCATGCGG-
 D.erecta 323 ----- TGGGCGTTCAC TGGC ----- GGCTGGGACTCAAAACCTCGATGCCCATGCGG-
 D.ficusphila 286 ----- TGGGCGTTCAC TGGC ----- GGCTGGGACTCAAAACCTCGATGCCCATGCGG-
 D.takahashii 281 ----- TGGGCGTTCAC TGGC ----- GGCTGGGACTCAAAACCTCGATGCCCATGCGG-
 D.elegans 309 ----- TGGGCGTTCAC TGGC ----- GGCTGGGACTCAAAACCTCGATGCCCATGCGG-
 D.kikkawai 276 ----- TGGGCGTTCAC TGGC ----- GGCTGGGACTCAAAACCTCGATGCCCATGCGG-
 D.rhopaloo 291 ----- TGGGCGTTCAC TGGC ----- GGCTGGGACTCAAAACCTCGATGCCCATGCGG-
 D.biarmipes 273 ----- TGGGCGTTCAC TGGC ----- GGCTGGGACTCAAAACCTCGATGCCCATGCGG-
 D.eugracilis 278 ----- TGGGCGTTCAC TGGC ----- GGCTGGGACTCAAAACCTCGATGCCCATGCGG-
 D.bipectinata 322 ----- TGGGCGTTCAC TGGC ----- GGCTGGGACTCAAAACCTCGATGCCCATGCGG-
 D.ananassae 316 ----- TGGGCGTTCAC TGGC ----- GGCTGGGACTCAAAACCTCGATGCCCATGCGG-
 D.pseudoobscura 297 ----- GCATGGGGGCGTTCAC TGGC ----- GCATGGGGGCGTTCAC TGGC ----- GCATGGGGGCGTTCAC TGGC ----- GCATGGGGGCGTTCAC TGGC -----
 D.persimilis 308 ----- GCATGGGGGCGTTCAC TGGC ----- GCATGGGGGCGTTCAC TGGC ----- GCATGGGGGCGTTCAC TGGC ----- GCATGGGGGCGTTCAC TGGC -----
 D.miranda 321 ----- GCATGGGGGCGTTCAC TGGC ----- GCATGGGGGCGTTCAC TGGC ----- GCATGGGGGCGTTCAC TGGC ----- GCATGGGGGCGTTCAC TGGC -----
 D.willistoni 316 ----- TCTCACTGCCATGCTCTCCTCGTTTTTTGCTAATTTTTTTGGGGCGTTCAC TGGC ----- TCTCACTGCCATGCTCTCCTCGTTTTTTGCTAATTTTTTTGGGGCGTTCAC TGGC ----- TCTCACTGCCATGCTCTCCTCGTTTTTTGCTAATTTTTTTGGGGCGTTCAC TGGC -----
 D.mojavensis 1250 TCTCACTGCCATGCTCTCCTCGTTTTTTGCTAATTTTTTTGGGGCGTTCAC TGGC ----- TCTCACTGCCATGCTCTCCTCGTTTTTTGCTAATTTTTTTGGGGCGTTCAC TGGC ----- TCTCACTGCCATGCTCTCCTCGTTTTTTGCTAATTTTTTTGGGGCGTTCAC TGGC -----
 D.virilis 394 ----- TTTTTCGGGGCGTTCAC TGGC ----- TTTTTCGGGGCGTTCAC TGGC ----- TTTTTCGGGGCGTTCAC TGGC ----- TTTTTCGGGGCGTTCAC TGGC -----
 D.americana 408 ----- TTTTTCGGGGCGTTCAC TGGC ----- TTTTTCGGGGCGTTCAC TGGC ----- TTTTTCGGGGCGTTCAC TGGC ----- TTTTTCGGGGCGTTCAC TGGC -----

<i>D. eugracilis</i>	384	-----
<i>D. bipectinata</i>	418	-----
<i>D. ananassae</i>	412	-----
<i>D. pseudoobscura</i>	405	-----
<i>D. persimilis</i>	425	-----
<i>D. miranda</i>	423	-----
<i>D. willistoni</i>	710	CTAAAGTCTCGCAAACTCGACAACTGTGAATTTTGAAGTACAATTCCTCTTTATTTTGCAGGGTATAAGCAAGCTGAGCTGCAATTACTACTTTATTGTGTTTTTTT
<i>D. mojavensis</i>	1399	-----
<i>D. virilis</i>	531	-----
<i>D. americana</i>	536	-----
<i>D. grimshawi</i>	507	-----

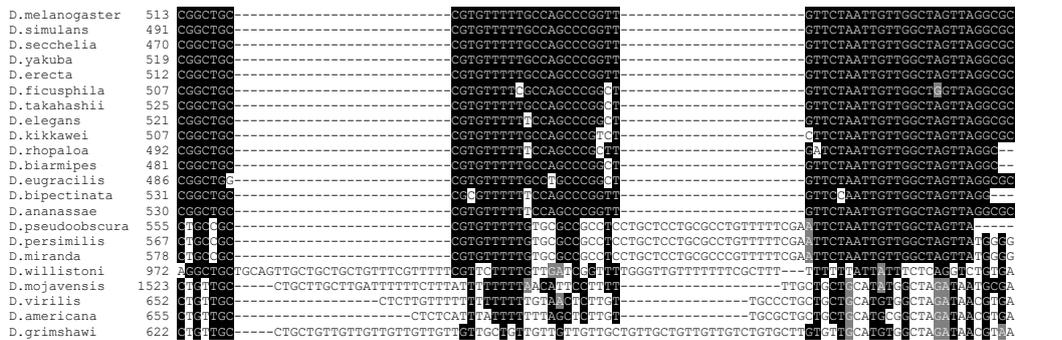
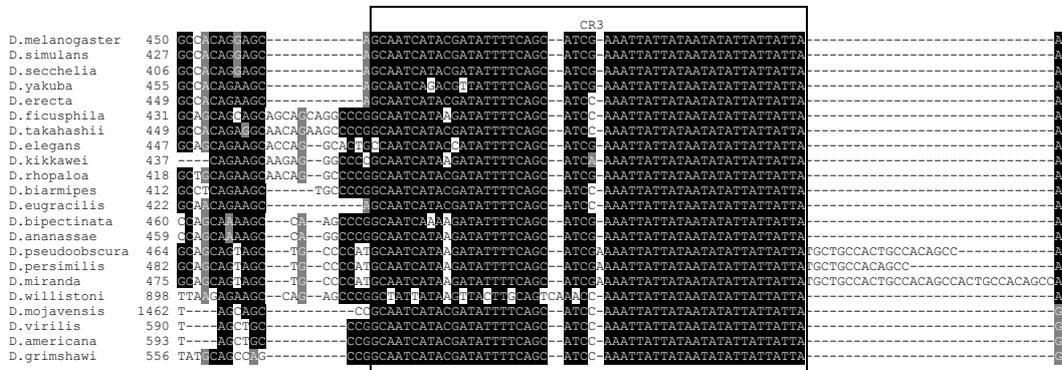
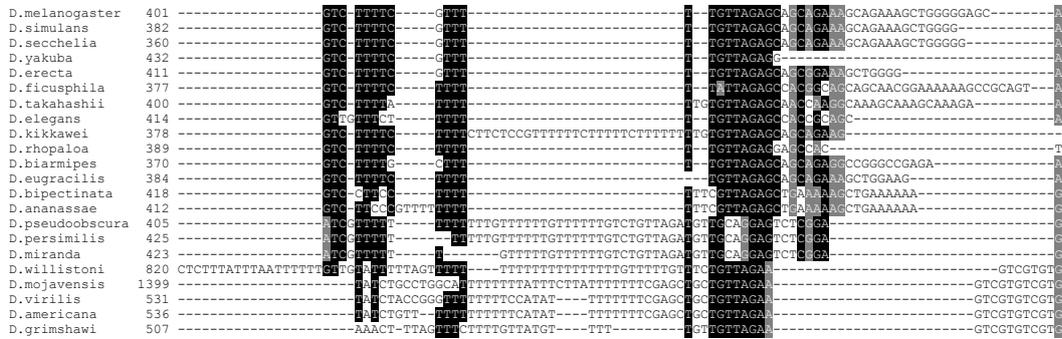


Figure S2: Sequence alignment of LAE sequences from 22 *Drosophila* species.

The genomic sequences homologous to the *D. melanogaster* LAE were recovered by Blat analysis at the UCSC Genome Browser website (<http://genome.ucsc.edu/cgi-bin/hgBlat?command=start>) and aligned with MAFFT (<http://mafft.cbrc.jp/alignment/server/>). The multiple alignment was then shaded with Boxshade (http://www.ch.embnet.org/software/BOX_form.html). The >20-bp highly-conserved regions (CR1-3) are framed. Notice that *D. mojavensis* and *D. willistoni* LAE sequences have long inserts (1-2 kb) between the CR2 and CR3 regions.

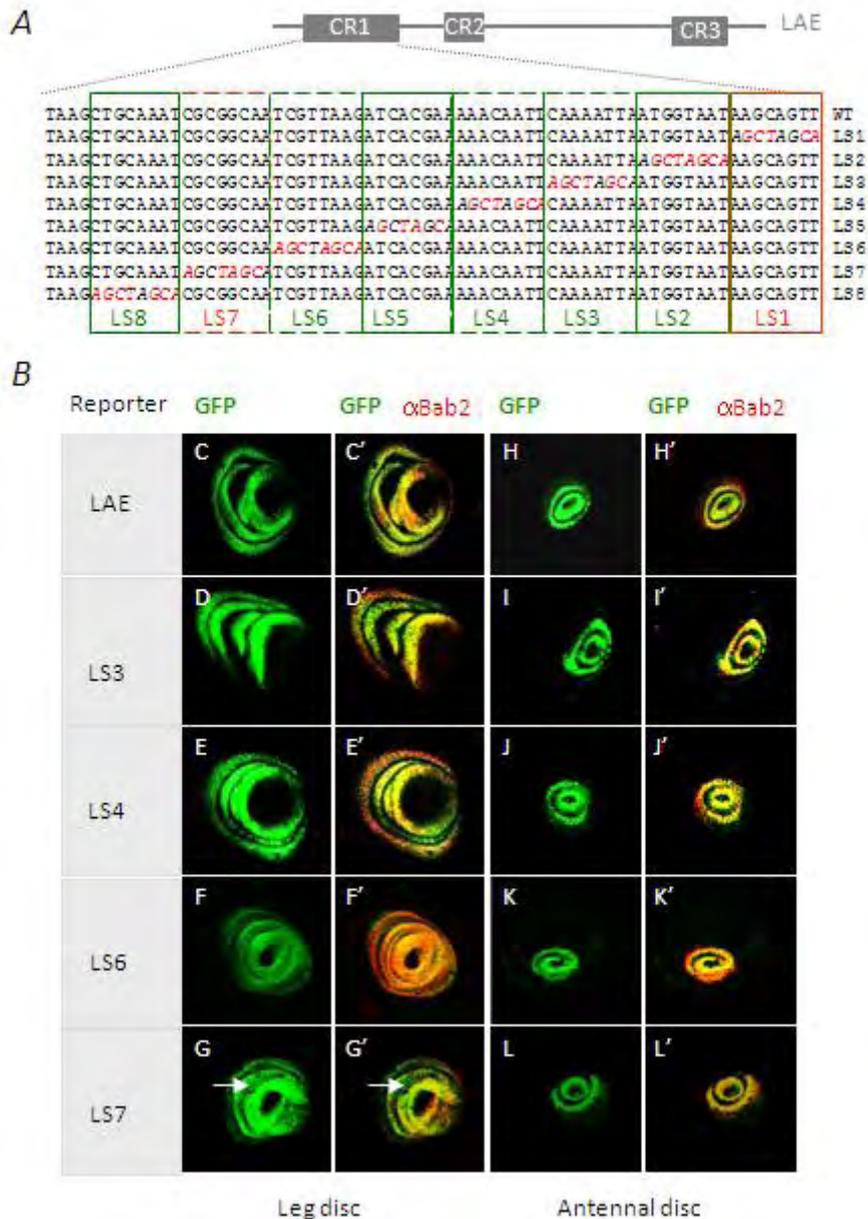


Figure S3: A linker-scanning mutagenesis of the critical CR1 region reveals functionally important motifs.

(A) The sequences of the wild-type and of each mutated CR1 region are shown in beneath the entire LAE structural organization, as determined by evolutionary conservation among *Drosophilidae* (Fig. S2). The sequence of the linker element is italicized within each LS mutant, with nucleotide substitutions depicted in red. Regulatory elements are framed in green (activation) or red (repression), dashed lines indicating partially redundant functions.

(B) Leg (C-K) and (L-T) antennal imaginal discs from dissected late third-instar larvae, expressing a given LS-mutated *LAE-GFP^{ZH2A}* construct shown in A. GFP fluorescence emission (green) in isolation and in combination with Bab2 immunostaining (red), are shown. Only four out of eight constructs displayed strongly abnormal expression patterns and/or intensities (see Fig. 4B).

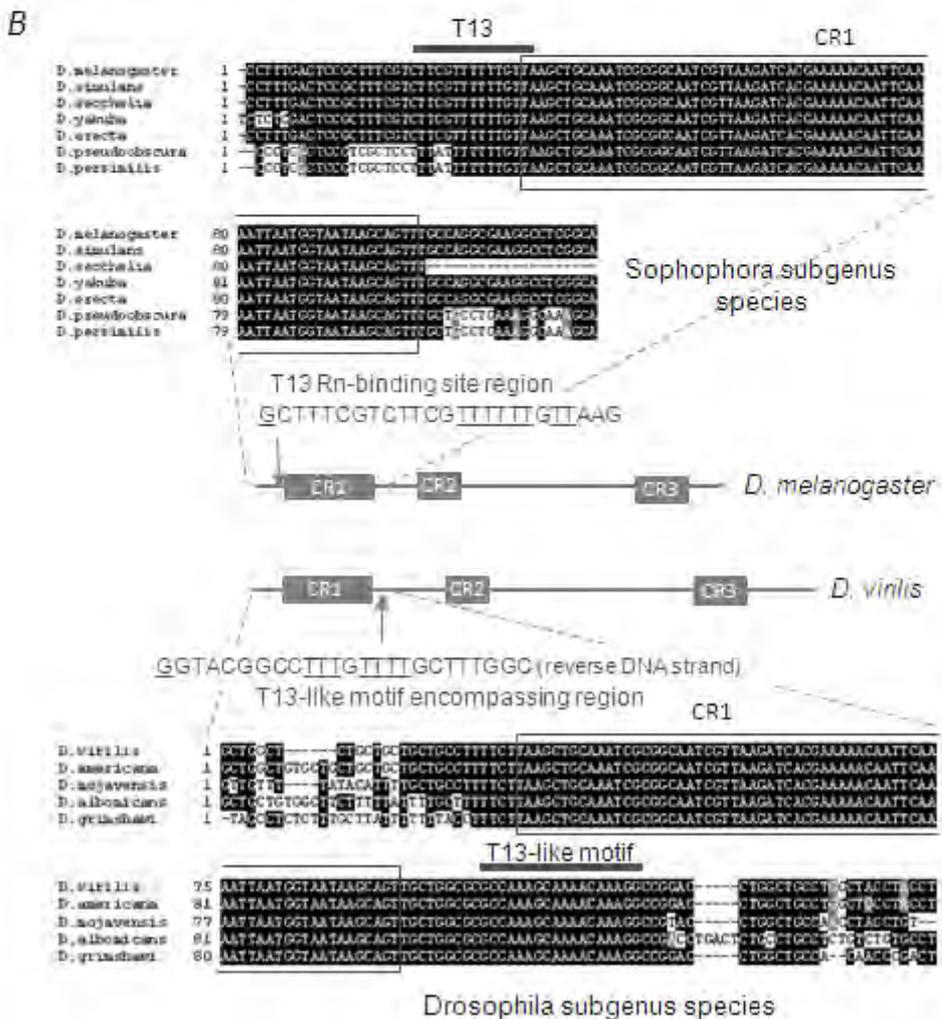
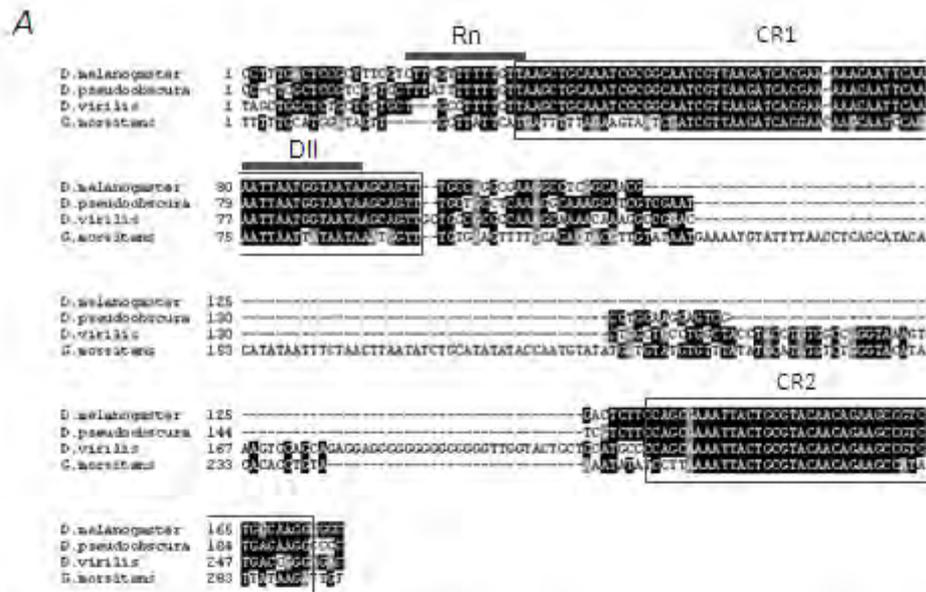


Figure S4: Evolutionary-conservation of the limb-specificity LAE sequence in the tsetse fly *Glossina morsitans* and among Drosophilae subgenus-specific portions.

(A) Sequence conservation of the limb-specificity LAE portion among Drosophilidae and in the Glossinidae *Glossina morsitans*. *G. morsitans* LAE sequence was identified through blast analyses using the Trace archive nucleotide blast server at NCBI (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>) and aligned with LAE sequences from representative Drosophilidae species, using MAFFT (<http://mafft.cbrc.jp/alignment/server/index.html>). Homology shading was made using BoxShade (http://www.ch.embnet.org/software/BOX_form.html). The CR1-2 sequences are framed and locations of the Rn and Dll binding sites are indicated above the alignment. Note that the Rn binding site is poorly conserved.

(B) Sequence conservation of the CR1-encompassing LAE portion among Sophophora and Drosophila subgenera. Sequences were aligned and processed as in A. The structural organization scheme of the entire *D. melanogaster* (Sophophora subgenus) and *D. virilis* (Drosophila subgenus) LAE sequences are shown in the middle part, with the aligned portion depicted with broken lines. The sequence and locations of the T13 and T13-like motifs are indicated. The T13 Rn-binding site of *D. melanogaster* and the T13-like sequence of *D. virilis* are well conserved among Sophophora and Drosophila subgenus species, respectively (T-rich sequences are underlined). Note that the T13-like *D. virilis* sequence (i) is located in a 3'-end extended CR1 and (ii) in an inverted orientation.

References

- Abu-Shaar M, Mann RS (1998) Generation of multiple antagonistic domains along the proximodistal axis during *Drosophila* leg development. *Development* **125**: 3821-3830
- Bischof J, Maeda RK, Hediger M, Karch F, Basler K (2007) An optimized transgenesis system for *Drosophila* using germ-line-specific phiC31 integrases. *Proc Natl Acad Sci U S A* **104**: 3312-3317
- Campbell G (2002) Distalization of the *Drosophila* leg by graded EGF-receptor activity. *Nature* **418**: 781-785
- Campbell G (2005) Regulation of gene expression in the distal region of the *Drosophila* leg by the Hox11 homolog, C15. *Dev Biol* **278**: 607-618
- Campbell G, Tomlinson A (1998) The roles of the homeobox genes *aristaless* and *Distal-less* in patterning the legs and wings of *Drosophila*. *Development* **125**: 4483-4493
- Campbell G, Weaver T, Tomlinson A (1993) Axis specification in the developing *Drosophila* appendage: the role of *wingless*, *decapentaplegic*, and the homeobox gene *aristaless*. *Cell* **74**: 1113-1123
- Casares F, Mann RS (1998) Control of antennal versus leg development in *Drosophila* [see comments]. *Nature* **392**: 723-726
- Chen R, Amoui M, Zhang Z, Mardon G (1997) *Dachshund* and *eyes absent* proteins form a complex and function synergistically to induce ectopic eye development in *Drosophila*. *Cell* **91**: 893-903.
- Chu J, Dong PD, Panganiban G (2002) Limb type-specific regulation of *bric a brac* contributes to morphological diversity. *Development* **129**: 695-704.
- Clark AG, Eisen MB, Smith DR, Bergman CM, Oliver B, Markow TA, Kaufman TC, Kellis M, Gelbart W, Iyer VN, Pollard DA, Sackton TB, Larracuenta AM, Singh ND, Abad JP, Abt DN, Adryan B, Aguade M, Akashi H, Anderson WW, Aquadro CF, Ardell DH, Arguello R, Artieri CG, Barbash DA, Barker D, Barsanti P, Batterham P, Batzoglou S, Begun D, Bhutkar A, Blanco E, Bosak SA, Bradley RK, Brand AD, Brent MR, Brooks AN, Brown RH, Butlin RK, Caggese C, Calvi BR, Bernardo de Carvalho A, Caspi A, Castrezana S, Celniker SE, Chang JL, Chapple C, Chatterji S, Chinwalla A, Civetta A, Clifton SW, Comeron JM, Costello JC, Coyne JA, Daub J, David RG, Delcher AL, Delehaunty K, Do CB, Ebling H, Edwards K, Eickbush T, Evans JD, Filipowski A, Findeiss S, Freyhult E, Fulton L, Fulton R, Garcia AC, Gardiner A, Garfield DA, Garvin BE, Gibson G, Gilbert D, Gnerre S, Godfrey J,

- Good R, Gotea V, Gravely B, Greenberg AJ, Griffiths-Jones S, Gross S, Guigo R, Gustafson EA, Haerty W, Hahn MW, Halligan DL, Halpern AL, Halter GM, Han MV, Heger A, Hillier L, Hinrichs AS, Holmes I, Hoskins RA, Hubisz MJ, Hultmark D, Huntley MA, Jaffe DB, Jagadeeshan S, .. et al. (2007) Evolution of genes and genomes on the Drosophila phylogeny. *Nature* **450**: 203-218.
- Couderc JL, Godt D, Zollman S, Chen J, Li M, Tiong S, Cramton SE, Sahut-Barnola I, Laski FA (2002) The bric a brac locus consists of two paralogous genes encoding BTB/POZ domain proteins and acts as a homeotic and morphogenetic regulator of imaginal development in Drosophila. *Development* **129**: 2419-2433.
- Couso JP, Bate M, Martínez-Arias A (1993) A wingless-dependent polar coordinate system in Drosophila imaginal discs. *Science* **259**: 484-489
- Couso JP, Bishop SA (1998) Proximo-distal development in the legs of Drosophila. *Int J Dev Biol* **42**: 345-352
- Cummins M, Pueyo JI, Greig SA, Couso JP (2003) Comparative analysis of leg and antenna development in wild-type and homeotic Drosophila melanogaster. *Dev Genes Evol* **213**: 319-327
- de Celis Ibeas JM, Bray SJ (2003) Bowl is required downstream of Notch for elaboration of distal limb patterning. *Development* **130**: 5943-5952
- Diaz-Benjumea FJ, Cohen B, Cohen SM (1994) Cell interaction between compartments establishes the proximal-distal axis of Drosophila legs. *Nature* **372**: 175-179
- Dong PD, Chu J, Panganiban G (2000) Coexpression of the homeobox genes Distal-less and homothorax determines Drosophila antennal identity. *Development* **127**: 209-216
- Duncan D, Kiefel P, Duncan I (2010) Control of the spineless antennal enhancer: direct repression of antennal target genes by Antennapedia. *Dev Biol* **347**: 82-91
- Duncan DM, Burgess EA, Duncan I (1998) Control of distal antennal identity and tarsal development in Drosophila by spineless-aristopedia, a homolog of the mammalian dioxin receptor. *Genes Dev* **12**: 1290-1303
- Estella C, McKay DJ, Mann RS (2008) Molecular integration of wingless, decapentaplegic, and autoregulatory inputs into Distalless during Drosophila leg development. *Dev Cell* **14**: 86-96
- Estella C, Voutev R, Mann RS (2012) A dynamic network of morphogens and transcription factors patterns the fly leg. *Curr Top Dev Biol* **98**: 173-198

Galindo MI, Bishop SA, Greig S, Couso JP (2002) Leg patterning driven by proximal-distal interactions and EGFR signaling. *Science* **297**: 256-259

Galindo MI, Couso JP (2000) Intercalation of cell fates during tarsal development in *Drosophila*. *Bioessays* **22**: 777-780

Gebelein B, McKay DJ, Mann RS (2004) Direct integration of Hox and segmentation gene inputs during *Drosophila* development. *Nature* **431**: 653-659

Giorgianni MW, Mann RS (2011) Establishment of medial fates along the proximodistal axis of the *Drosophila* leg through direct activation of dachshund by Distalless. *Dev Cell* **20**: 455-468

Godt D, Couderc JL, Cramton SE, Laski FA (1993) Pattern formation in the limbs of *Drosophila*: bric a brac is expressed in both a gradient and a wave-like pattern and is required for specification and proper segmentation of the tarsus. *Development* **119**: 799-812.

Greenberg L, Hatini V (2009) Essential roles for lines in mediating leg and antennal proximodistal patterning and generating a stable Notch signaling interface at segment borders. *Dev Biol* **330**: 93-104

Ho SN, Hunt HD, Horton RM, Pullen JK, Pease LR (1989) Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene* **77**: 51-59

Hong JW, Hendrix DA, Levine MS (2008) Shadow enhancers as a source of evolutionary novelty. *Science* **321**: 1314

Jiang J, Struhl G (1996) Complementary and mutually exclusive activities of decapentaplegic and wingless organize axial patterning during *Drosophila* leg development. *Cell* **86**: 401-409

Kojima T (2004) The mechanism of *Drosophila* leg development along the proximodistal axis. *Dev Growth Differ* **46**: 115-129.

Kojima T, Tsuji T, Saigo K (2005) A concerted action of a paired-type homeobox gene, aristaless, and a homolog of Hox11/tlx homeobox gene, clawless, is essential for the distal tip development of the *Drosophila* leg. *Dev Biol* **279**: 434-445

Lecuit T, Cohen SM (1997) Proximal-distal axis formation in the *Drosophila* leg. *Nature* **388**: 139-145

Levine M (2010) Transcriptional enhancers in animal development and evolution. *Curr Biol* **20**: R754-763

Levine M, Tjian R (2003) Transcription regulation and animal diversity. *Nature* **424**: 147-151.

- Liu R, Lehane S, He X, Lehane M, Hertz-Fowler C, Berriman M, Pickett JA, Field LM, Zhou JJ (2010) Characterisations of odorant-binding proteins in the tsetse fly *Glossina morsitans morsitans*. *Cell Mol Life Sci* **67**: 919-929
- Morata G (2001) How *Drosophila* appendages develop. *Nat Rev Mol Cell Biol* **2**: 89-97.
- Natori K, Tajiri R, Furukawa S, Kojima T (2012) Progressive tarsal patterning in the *Drosophila* by temporally dynamic regulation of transcription factor genes. *Dev Biol* **361**: 450-462
- Noyes MB (2012) Analysis of specific protein-DNA interactions by bacterial one-hybrid assay. *Methods Mol Biol* **786**: 79-95
- Noyes MB, Meng X, Wakabayashi A, Sinha S, Brodsky MH, Wolfe SA (2008) A systematic characterization of factors that regulate *Drosophila* segmentation via a bacterial one-hybrid system. *Nucleic Acids Res* **36**: 2547-2560
- Pai CY, Kuo TS, Jaw TJ, Kurant E, Chen CT, Bessarab DA, Salzberg A, Sun YH (1998) The Homothorax homeoprotein activates the nuclear localization of another homeoprotein, extradenticle, and suppresses eye development in *Drosophila*. *Genes Dev* **12**: 435-446
- Pueyo JI, Couso JP (2008) The 11-aminoacid long Tarsal-less peptides trigger a cell signal in *Drosophila* leg development. *Dev Biol* **324**: 192-201
- St Pierre SE, Galindo MI, Couso JP, Thor S (2002) Control of *Drosophila* imaginal disc development by rotund and roughened eye: differentially expressed transcripts of the same gene encoding functionally distinct zinc finger proteins. *Development* **129**: 1273-1281.
- Struhl G (1981) A homoeotic mutation transforming leg to antenna in *Drosophila*. *Nature* **292**: 635-638
- Torrunguang K, Alvarez M, Shah R, Onyia JE, Rhodes SJ, Bidwell JP (2002) DNA binding and gene activation properties of the Nmp4 nuclear matrix transcription factors. *J Biol Chem* **277**: 16153-16159. Epub 12002 Feb 16126.
- Venken KJ, Carlson JW, Schulze KL, Pan H, He Y, Spokony R, Wan KH, Koriabine M, de Jong PJ, White KP, Bellen HJ, Hoskins RA (2009) Versatile P[acman] BAC libraries for transgenesis studies in *Drosophila melanogaster*. *Nat Methods* **6**: 431-434
- Venken KJ, He Y, Hoskins RA, Bellen HJ (2006) P[acman]: a BAC transgenic platform for targeted insertion of large DNA fragments in *D. melanogaster*. *Science* **314**: 1747-1751

Williams TM, Selegue JE, Werner T, Gompel N, Kopp A, Carroll SB (2008) The regulation and evolution of a genetic switch controlling sexually dimorphic traits in *Drosophila*. *Cell* **134**: 610-623.

Wu J, Cohen SM (1999) Proximodistal axis formation in the *Drosophila* leg: subdivision into proximal and distal domains by Homothorax and Distal-less. *Development* **126**: 109-117

Résultats : Partie II

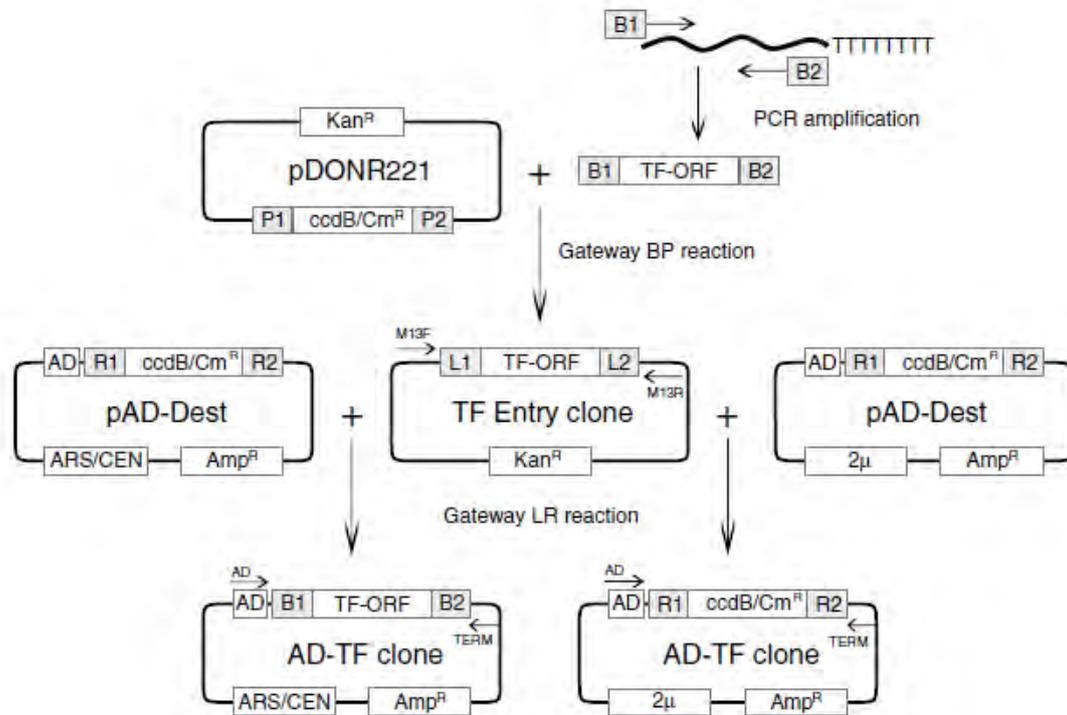


Figure 56. Stratégie de clonage adoptée pour la construction de la banque de FTs. L'ORF codant un FT donné a été amplifiée par PCR en incorporant des sites de recombinaison 'Gateway' *attB1* et *attB2* dans les amorces 5' et 3', respectivement. Le produit PCR obtenu a été cloné dans un premier temps dans le vecteur donneur pDONR221 par une réaction Gateway BP. Le vecteur d'entrée qui en résulte est combiné avec l'un ou l'autre de deux vecteurs pAD-Dest contenant soit l'origine de réplication ARS/CEN ou soit celui de 2μ, permettant d'avoir, respectivement, un nombre faible ou élevé de copies dans la levure. Le deuxième clonage est réalisé par une réaction Gateway LR. Chaque vecteur pAD-DEST recombiné code pour une protéine hybride formée par la fusion du domaine d'activation (AD) de la protéine Gal4 de levure à l'extrémité N-terminale de chacun des facteurs de transcription.

Dans cette partie seront présentés : (1) le crible simple hybride chez la levure élaboré afin de mettre en évidence les interactions entre les FTs de drosophile et le LAE ; (2) l'analyse de séquences régulatrices dans le LAE impliquées dans l'activation distale et enfin (3) les données concernant le rôle des facteurs Sp1 et Dll d'une part et Bowl, Bab et C15 d'autre part respectivement dans l'activation et la répression de *bab*.

1. CRIBLE SIMPLE HYBRIDE

1.1. PRINCIPE.

Le simple hybride est une technique de biologie moléculaire qui permet de mettre en évidence une interaction fonctionnelle entre une protéine donnée et une séquence d'ADN à tester. Cette technique se base sur le fait que les FTs possèdent une structure modulaire et que les domaines de fixation à l'ADN et les domaines effecteurs de la machinerie transcriptionnelle sont fonctionnellement indépendants et interchangeables. Ainsi, pour tester la liaison d'une protéine candidate entière (ou d'un sous-domaine particulier) à une séquence d'ADN d'intérêt, il suffit de cloner cette dernière en amont d'un gène rapporteur ayant un promoteur minimal et d'examiner l'activation de l'expression de ce gène rapporteur *in-vivo* (souvent chez la levure) ou *in-cellulo*, après transformation (ou transfection) d'un plasmide exprimant la protéine (ou un sous-domaine) à tester en fusion avec un domaine activateur donné (le plus souvent du FT de levure Gal4). Ainsi si le domaine de liaison à l'ADN à tester fixe la séquence cible, ceci se traduira par un rapprochement physique entre le promoteur minimal et le domaine activateur greffé, et par conséquent par une activation transcriptionnelle du gène rapporteur.

Afin d'exploiter au maximum le potentiel de la technique du simple hybride dans l'identification d'interactions fonctionnelles entre les facteurs de transcription et les séquences *cis*-régulatrices à tester, l'équipe du Pr. Bart Deplancke à l'EPFL (Suisse) a développé une plateforme destinée à cribler chez la levure *S. cerevisiae*, de manière systématique, standardisée et rapide, la grande majorité des FTs de *D. melanogaster* à partir d'une librairie plasmidique en cours d'achèvement. Lors de mon séjour à Lausanne, cette dernière permettait l'expression de 718 FTs de la drosophile, soit 95% des FTs prédits, fusionnés au domaine activateur (AD) de Gal4, au niveau de leur extrémité N-terminale. Pour ce faire, l'équipe de Bart Deplancke a eu recours à la technique de clonage Gateway en deux étapes (Figure 56).

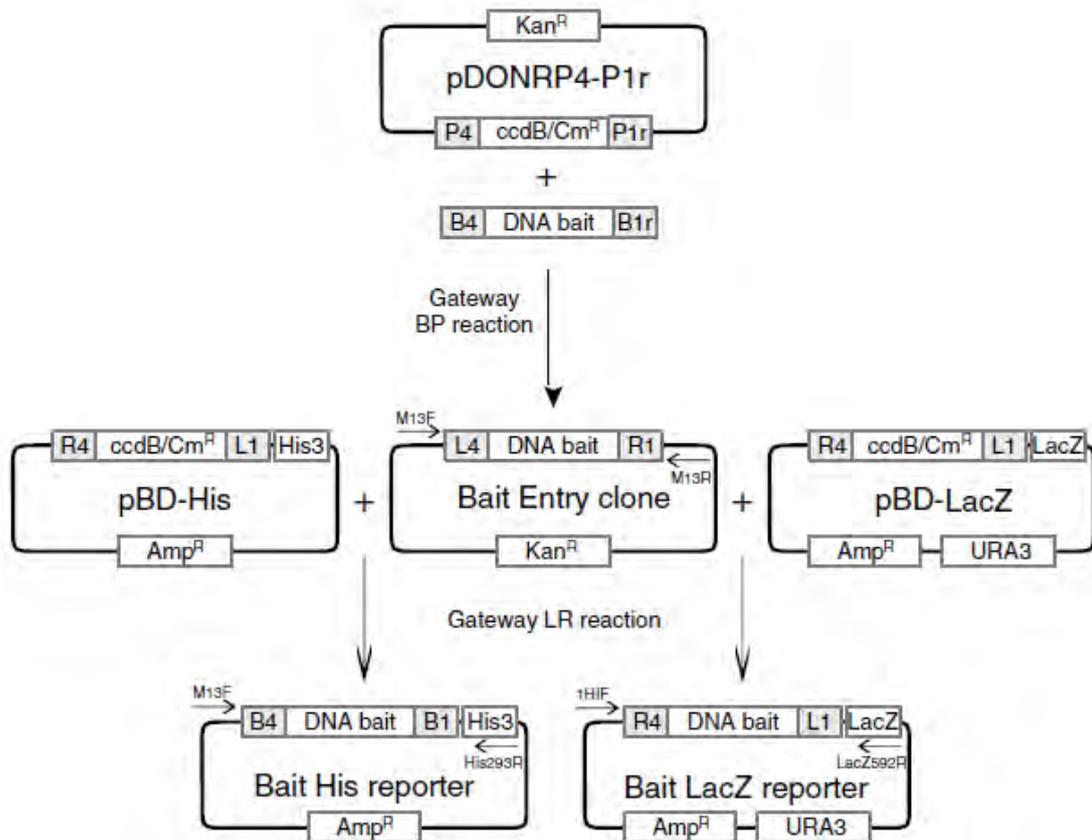


Figure 57. Stratégie de clonage de la séquence d'ADN cible à tester ('Bait'). Après amplification PCR de la séquence cible, avec addition des sites Gateway AttB4 et attB1R, respectivement aux extrémités 5' des amorces sens et antisens, une réaction BP est réalisée en présence du vecteur de clonage pDONRP4-P1R. Deux réactions Gateway LR différentes sont par la suite réalisées avec le clone d'entrée ainsi obtenu, ce qui permet de transférer la séquence cible dans le vecteur pBD-His dans une réaction ou pBD-LacZ dans l'autre.

Ainsi chacune des ORFs entières codant pour les différents FTs de la librairie a été amplifiée par PCR et clonée dans un premier temps dans le vecteur pDONR221, puis par la suite transférée du vecteur d'entrée ainsi obtenu dans deux vecteurs pAD-Dest (cf. ci-dessous) contenant l'ORF codant pour la région AD de Gal4. Comme l'expression des FTs fusionnées à AD se fait de manière transitoire (par transfection) deux variants du vecteur pAD-Dest ont été utilisés, un contenant l'origine de répllication 2μ et l'autre les séquences ARS ('autonomously replicating sequence')/CEN (centromère de la levure) ce qui permet d'avoir respectivement un nombre élevé (13 en moyenne) ou faible (1 ou 2) de copies des vecteurs à tester dans la levure.

1.2. CONSTRUCTION DES SOUCHES STABLES

Dans le but d'identifier des facteurs de liaison à l'ADN potentiellement impliqués dans la régulation des gènes *bab* dans les tissus de pattes et d'antenne, nous avons analysé par une approche simple hybride la capacité de l'enhancer LAE et le mini CRM S5 décrits précédemment dans l'article, à fixer les FTs de drosophile présents dans la librairie de l'EPFL (Figure 57). Pour ce faire, j'ai d'abord cloné les régions *cis*-régulatrices à tester, dans les vecteurs d'intégration pBD-His et pBD-LacZ, respectivement en amont des gènes rapporteurs *his3* et *lacZ*. Pour cela, les régions à tester ont été amplifiées par PCR en utilisant des amorces sens et anti-sens contenant respectivement les sites Gateway attB4 (ggggacaactttgtatagaaaagttg) et attB1R (ggggactgctttttgtacaaaacttg) aux extrémités 5'. Par la suite, j'ai intégré ces produits PCR dans le vecteur d'entrée pDONRP4-P1r par une réaction Gateway BP et enfin, après vérification de l'intégrité des inserts par séquençage, je les ai transférés dans les deux vecteurs rapporteurs pBD-His et pBD-LacZ par une réaction Gateway LR. Au final, j'ai ainsi obtenu 4 vecteurs d'intégration que j'ai appelé pLAE-His, et pLAE-LacZ ; pS5-His et pS5-LacZ, exprimant les gènes rapporteurs *his3* ou *lacZ* sous le contrôle de la séquence LAE ou S5.

Par la suite, les vecteurs rapporteurs pLAE-His et pLAE-LacZ, d'une part, et pS5-His et pS5-LacZ, d'autre part, ont été intégrés simultanément par recombinaison homologue dans le génome de la souche *S. cerevisiae* YM4271 (Hens et al., 2011). La sélection des levures recombinantes a été réalisée sur un milieu His- Ura-. En effet, comme la souche YM4271 est mutante au niveau des locus *his3* et *ura3*, les seules levures qui peuvent pousser sur ce milieu

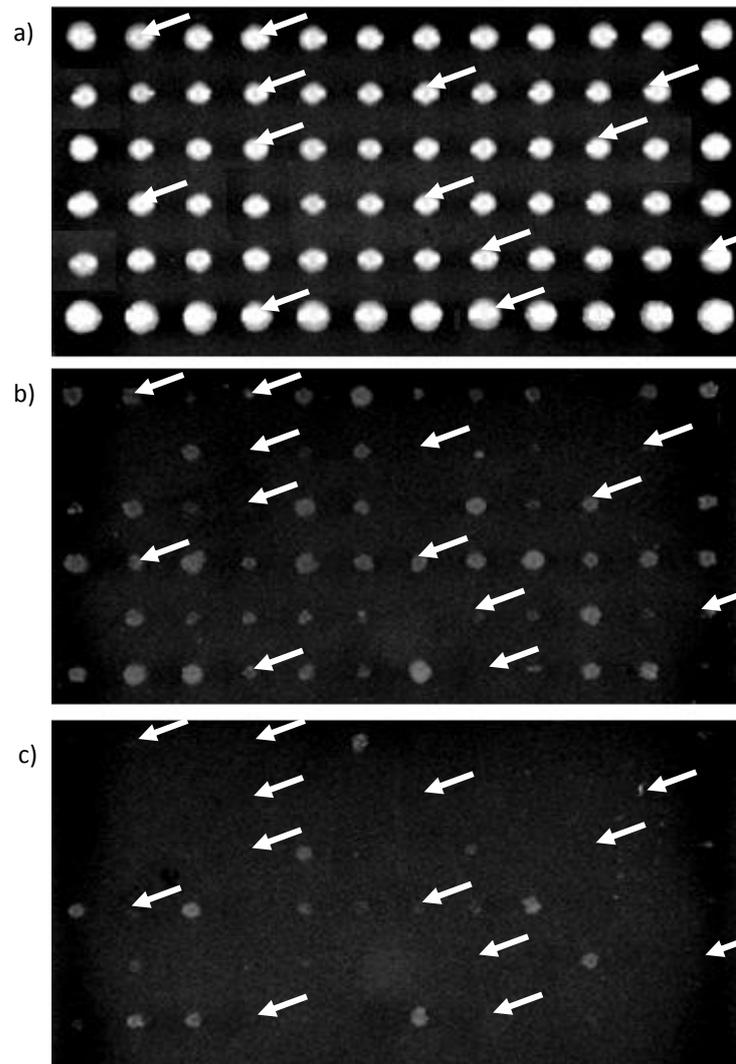


Figure 58. Test d'auto-activation pour les souches qui ont intégré au moins une copie du gène rapporteur *his3*. 12 clones individuels (alignés) pour 6 séquences cibles différentes ont été repiqués sur des boîtes contenant un milieu sélectif His- Ura- sans 3-AT, un inhibiteur de croissance, ou avec des concentrations croissantes de 3-AT variant de 10 à 100 mM. Des boîtes sans (a) ou avec 20 mM (b) ou 40mM (c) de 3-AT sont montrées ici. Les flèches indiquent les clones choisis pour la suite du crible simple hybride.

sélectif sont celles dont les mutations ont été corrigées suite à des recombinaisons homologues entre les gènes *his3* et *ura3*, se trouvant respectivement aux niveaux des vecteur pBD-His et pBD-lacZ, et les locus mutants endogènes. Il faut noter que le gène *his3* ainsi reconstitué par recombinaison homologue n'est pas corrigé totalement. En effet, le gène reconstitué se trouve alors sous le contrôle d'un promoteur minimal et de la séquence *cis*-régulatrice à tester. Toutefois ce gène *his3* présente un niveau d'expression basal suffisant pour induire la croissance cellulaire.

1.3. TEST D'AUTO ACTIVATION ET SELECTION DES SOUCHES

Afin que le crible simple hybride soit le plus efficace possible, il est nécessaire de travailler avec des souches modifiées ayant des niveaux d'expression basale de *his3* et *lacZ* les plus faibles possibles. En effet, comme le niveau d'expression du gène rapporteur *his3* détermine la vitesse de croissance de la levure sur un milieu His- et que ce dernier critère traduit la force de l'interaction entre les facteurs de transcription et les régions *cis*-régulatrices à tester, il est clair que le niveau d'expression basale de *his3* doit être le plus faible possible pour que les interactions (croissance plus importante) se détectent plus facilement. De manière similaire, le niveau d'expression du gène *lacZ* (qui code pour la β -galactosidase d'*E. coli*) doit être le plus faible possible. En fait, comme les interactions entre FTs et séquences *cis*-régulatrices sont évaluées par un test de coloration au X-gal (substrat de la β -galactosidase), le niveau de coloration basale due à l'expression de *lacZ* doit être le plus faible possible pour distinguer les clones qui ont une activité enzymatique basale de ceux dont le gène rapporteur est activé par la séquence régulatrice à tester.

L'utilisation de deux gènes rapporteurs avait pour but de croiser les résultats issus de deux systèmes rapporteurs différents, afin de réduire le nombre de faux positifs. Malheureusement, les expériences utilisant le système rapporteur *lacZ* n'ont pas montré de reproductibilité. C'est pourquoi, par la suite je me limiterai à l'analyse du gène rapporteur *his3*.

Le niveau d'expression basale, ou bruit de fond, du gène rapporteur *his3* dépend principalement du nombre de copies intégrées dans le génome de la levure. En effet, plus le nombre de copies intégrées est important plus le niveau d'expression basale est fort. Afin de sélectionner les souches ayant de faibles niveaux d'expression, et donc théoriquement un

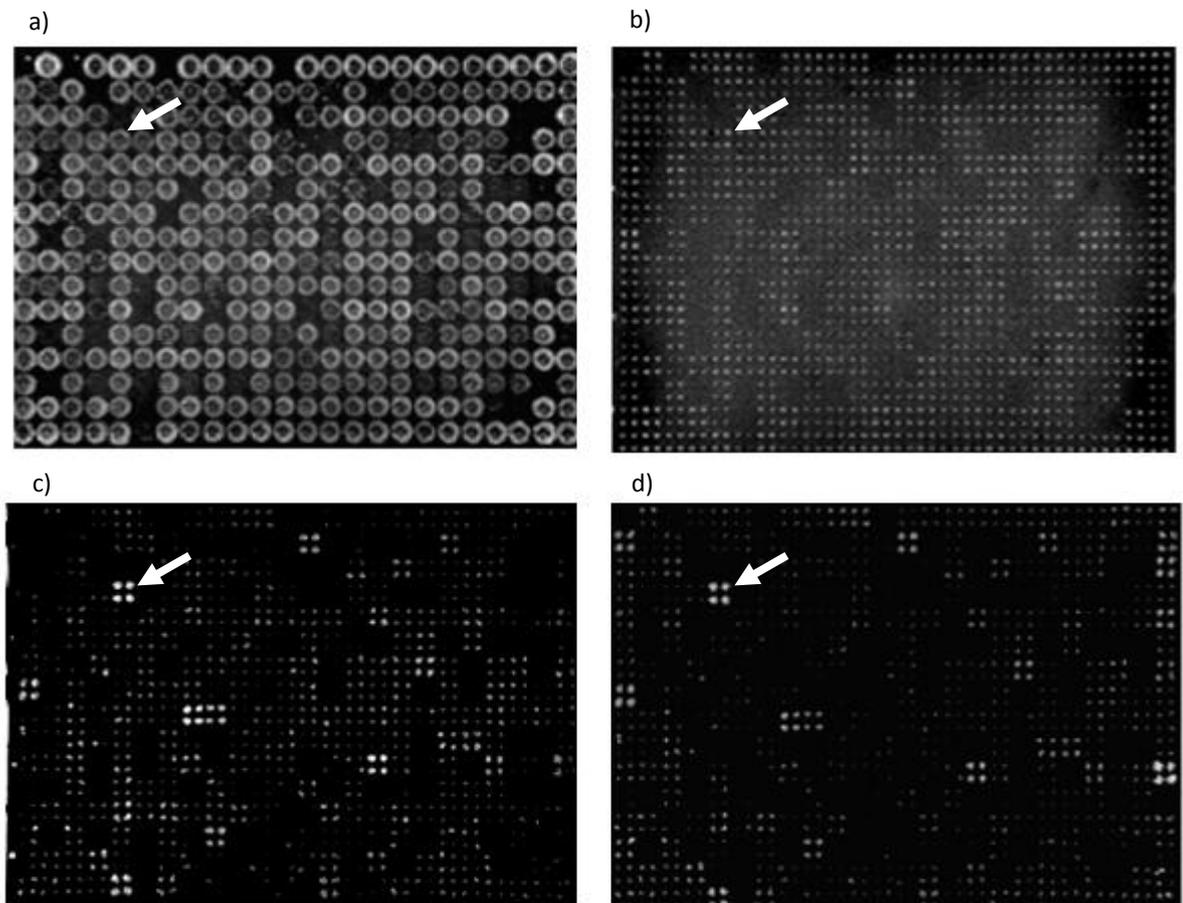


Figure 59. Exemple de crible simple hybride obtenu avec la séquence LAE. (a) une souche modifiée portant le gène rapporteur *his3* sous le contrôle de la séquence cible à tester est transformée individuellement par 384 plasmides exprimant des FTs de drosophile fusionnés au TAD de Gal4, puis déposée selon un dépôt unique sur des boîtes contenant un milieu His- Ura- Trp-. Après 3 à 4 jours de culture à 30°C, à l'aide d'un robot, chaque colonie est repiquée en quatre exemplaires sous forme de carrés sur milieu His- Ura- Trp- sans ou avec du 3-AT (respectivement, 20 mM (c) et 40 mM (d) dans cet exemple). La croissance sur des boîtes contenant du 3-AT indique une liaison potentielle d'un facteur de transcription d'intérêt à la séquence d'ADN cible testée.

faible nombre de copies, j'ai réalisé des tests d'auto-activation (Figure 58). Ces tests consistent à examiner la capacité des souches modifiées à croître sur un milieu His-, et ceci en présence de concentrations croissantes de 3-AT, un compétiteur de croissance. Ainsi les souches qui ont les plus faibles niveaux d'expression basale seront inhibées par les concentrations les plus faibles de 3-AT. Ces dernières sont alors repiquées et amplifiées pour réaliser le crible.

1.4. CRIBLE SIMPLE HYBRIDE

Les souches stables (*LAE-his3* et *S5-his3*) sélectionnées lors du test d'auto-activation, sont par la suite transfectées individuellement par chacun des plasmides exprimant les différentes protéines chimères BDGal4-TF de la librairie issue de la drosophile. Cette étape est effectuée en grande partie par un robot (Tecon Evo), doté d'une tête de pipetage à 384 pointes, d'incubateurs et d'une centrifugeuse, permettant de réaliser de manière autonome l'ensemble des étapes de transfection ainsi que le repiquage sur des boîtes de culture. Ainsi chaque souche rendue compétente est d'abord transfectée par chacun des 718 plasmides pAD-STF de la librairie répartis dans deux plaques à 384 puits, puis repiquée sur deux boîtes de culture His- et Trp- (Figure 59). L'absence de tryptophane dans le milieu permet de maintenir le vecteur d'expression pAD-STF, qui complémente une mutation impliquée dans la synthèse du tryptophane. Par la suite les boîtes contenant les différentes transfections sont incubées à 30°C pendant trois à quatre jours, jusqu'à ce qu'elles se développent convenablement.

Enfin pour évaluer la validité des interactions positives entre les facteurs de transcription de la librairie et les séquences *cis*-régulatrices à tester, les souches transfectées sont repiquées sur des boîtes de culture contenant un milieu minimum His- Trp-, ainsi que des concentrations croissantes de 3-AT (10, 20, 40 et 80 mM). Chaque souche transfectée est repiquée, en utilisant le robot 'Singer Rotor', en quatre exemplaires sous forme de carrés, ce qui permet de passer de 384 à 1536 dépôts (Figure 59). Cette disposition facilite la lecture de la plaque et l'identification des faux positifs. Théoriquement si le facteur de transcription testé ne reconnaît pas la séquence *cis*-régulatrice d'intérêt la croissance de la souche va être inhibée même avec des concentrations faibles de 3-AT. En revanche si le facteur de transcription interagit avec la séquence testée la croissance va être induite (due à l'expression de *his3*).

Après 5 à 6 jours d'incubation à 30°C, les boîtes sont examinées et celles qui sont jugées lisibles (en moyenne, 30 spots d'activation sont facilement distinguables) sont photographiées et analysées par le programme d'analyse d'image TIDY ('transcription factor-DNA interaction detection in yeast') qui détermine selon les critères fixés (par exemple taille moyenne des 4 dépôts comparée à la taille moyenne de tous les dépôts) les interactions significatives.

1.5. RESULTATS DU CRIBLE SIMPLE HYBRIDE

Afin de limiter le nombre de candidats et de s'affranchir des aléas expérimentaux chaque souche de levure sélectionnée a été testée en duplicat et uniquement les candidats considérés comme positifs dans les deux expériences réalisées séparément ont été retenus. Ainsi à l'issue de ce crible, j'ai identifié 8 régulateurs potentiels des gènes *bab*. Parmi lesquels 4 facteurs qui s'expriment dans les disques imaginaux de pattes et d'antenne : C15/Cll qui s'exprime au centre des disques; Snail qui s'exprime dans des groupe de cellules dispersés dans les disques; Pan/dTCF qui se trouve en aval de la voie Wg et est impliqué dans la formation de l'axe P/D; Stat92E qui est impliqué dans la formation de l'axe D/V et l'expression correcte de *wg* et *dpp*. Nous avons identifié également des facteurs potentiels qui n'ont pas à ce jour été décrits comme ayant une fonction dans le développement des pattes ou de l'antenne : CG34376, CrebB-17A, HLH3B et Abd-B. Il est à noter qu'Abd-B est exprimé dans le disque génital, et pourrait contribuer à réguler les gènes *bab* dans ce tissu. D'autres interactions moins fortes que celles décrites précédemment ont pu être notées, comme par exemple Sp1 et d'autres CGs qui sont étudiés par d'autres membres de l'équipe. Enfin, il est à noter également que certains facteurs comme Ss, Rn et Bab1/2 n'étaient pas disponibles dans la collection et n'ont donc pas pu être testés.

Snail	2(LAEc1 run1et2)
Pan/dTCF	2(LAEc1 run1et2)
Stat92E	2(LAEc1 run1et2)
C15	2(BPS5c1 run1et2)
Abd-B	2(BPS5c1 run1et2)
CrebB-17A	2(BPS5c1 run1et2)
HLH3B	2(BPS5c1 run1et2)
CG34376	2(BPS5c1 run1et2)

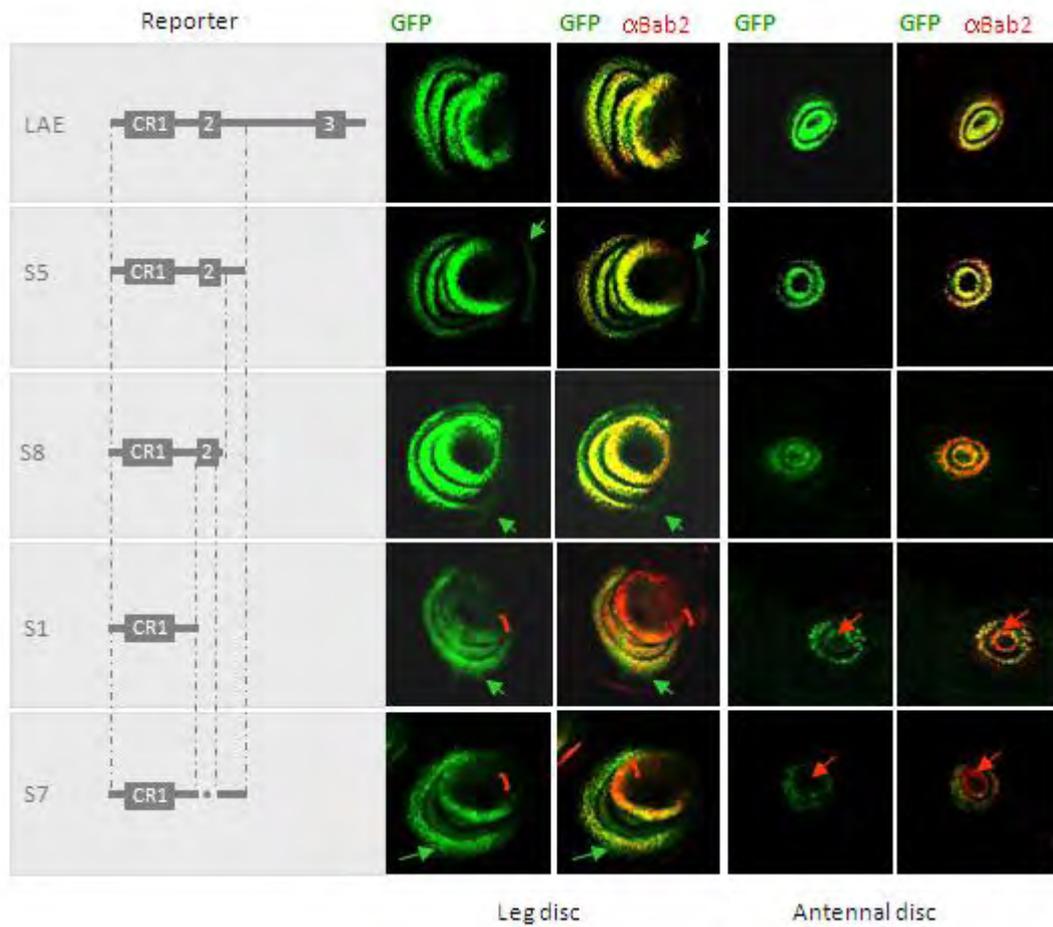


Figure 60. La boîte CR2 est nécessaire à l'activation de l'anneau distal, dans les tissus de pattes et d'antenne (crochets et flèches rouges, respectivement). A noter pour l'ensemble des constructions délétées, une faible dé-répression proximale dans les disques de pattes (flèches vertes).

2. CARACTERISATION FONCTIONNELLE D'AUTRES SEQUENCES CIS-REGULATRICES SPECIFIQUES DE L'ACTIVATION DISTALE

2.1. LA BOITE CR2 DU LAE EST REQUISE POUR L'ACTIVATION DE LA REGION DISTALE.

L'analyse fonctionnelle du LAE a montré que le fragment S1 (137 pb) représente la séquence minimale nécessaire pour avoir un signal dans tout le domaine d'expression de *bab* (Figure 3 de la publication). Toutefois, avec la construction S1 l'intensité relative du signal dans les anneaux proximaux et distaux s'est inversée. En fait, le signal devient plus faible au niveau de l'anneau distal patte et antenne alors que normalement l'inverse est observé, aussi bien pour le gène *bab2* endogène que pour le transgène *LAE-GFP*. Ceci suggère, que des sites impliqués principalement dans l'activation dans le domaine distal ont été délétés. Etant donné que cette diminution distale n'est pas visible avec la construction S5 les sites impliqués dans l'activation distale doivent donc se situer dans les 92 pb qui le différencient de S1 et qui incluent la boîte CR2. Pour évaluer le rôle de la partie non conservée qui se trouve du côté 3' de la boîte CR2, j'ai alors analysé l'effet de sa délétion dans le contexte S5 (fragment S8). Le fait de garder la boîte CR2 permet d'avoir un niveau d'expression dans la partie distale semblable à S5 (Figure 60), ce qui suggère que la boîte CR2 a un rôle activateur et que ce rôle serait plutôt dans les anneaux distaux. Pour le confirmer, j'ai délété précisément la boîte CR2 dans le contexte du fragment S5, pour générer la construction S7. Un défaut d'activation distale a alors été observé, ce qui démontre que la boîte CR2 est requise pour l'activation distale, aussi bien dans les tissus de pattes que d'antennes.

2.2. LA REGION INTENSITE 3' CONTIENT DES REGIONS ACTIVATRICES DISTALES

L'analyse fonctionnelle du LAE a révélé des séquences activatrices distribuées sur l'ensemble de la région 3', sans effets notoires sur la spécificité. Pour évaluer le rôle activateur de la boîte CR3, j'ai fusionné le fragment S1 avec les 177 pb de l'extrémité 3' du LAE, (construction S2 ; Figure 61). J'ai alors observé une forte expression de tous les anneaux et une restauration de l'activation distale normale dans les disques de pattes et d'antenne alors que cette construction ne contient pas la boîte CR2. Ce résultat montre que la région 3' rajoutée

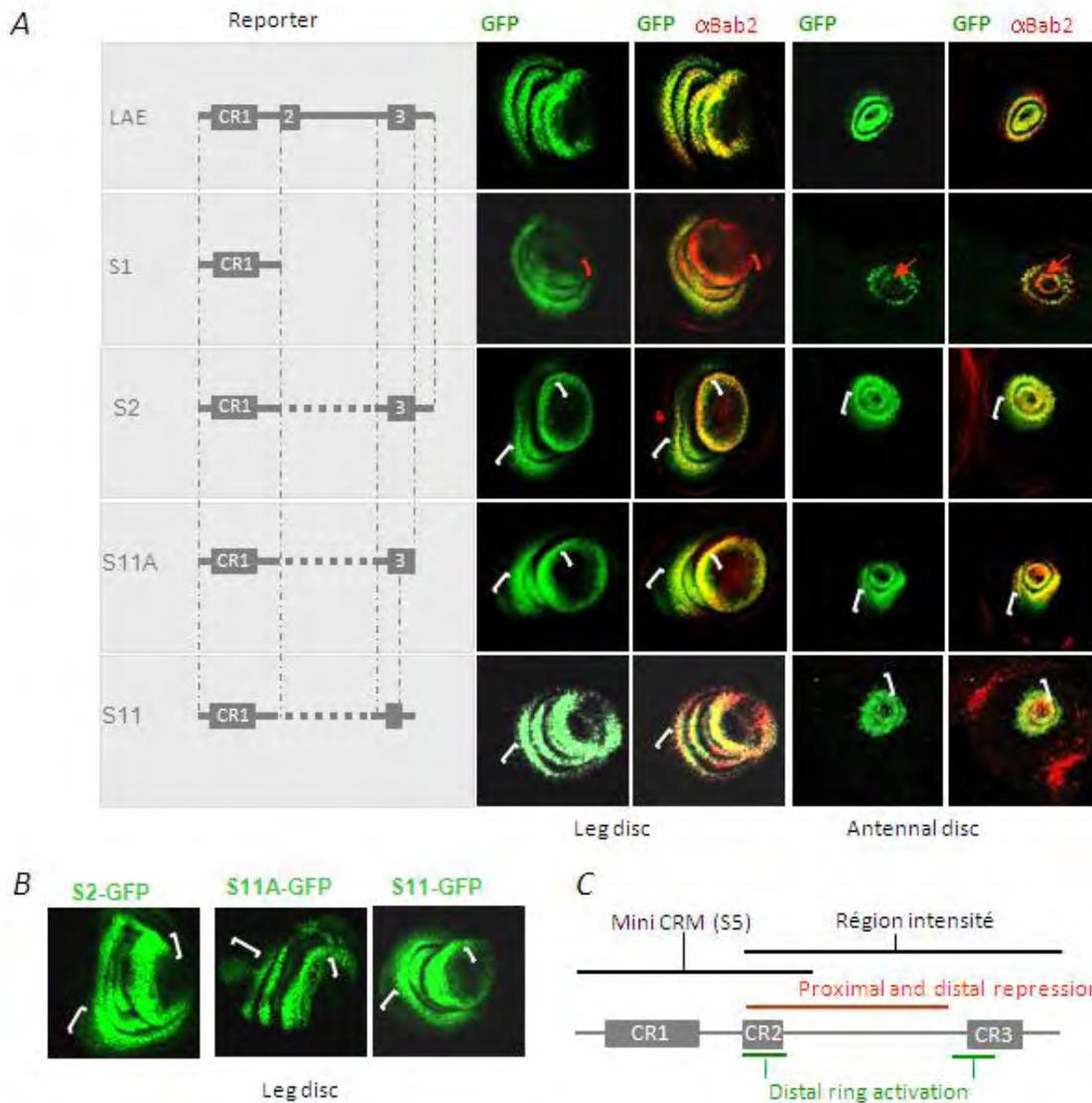


Figure 61. La région intensité 3' contient des régions activatrice et répressive. En A sont montrés les patrons d'expression des différents rapporteurs GFP examinés. La région 3' couvrant la moitié 5' de la boîte CR3 est suffisante pour restaurer le défaut d'activation distale de la construction S1, suggérant un rôle redondant avec la boîte CR2. A noter une forte dé-répression distale et proximale (crochets blancs) dans les disques de pattes, et une faible dé-répression proximale pour le disque d'antenne, suggérant la présence de site(s) répresseur(s) dans la région centrale délétée. En B, sont présentées des images focales illustrant les dé-répressions proximale et distale des constructions restaurées en activation distale. L'ensemble des éléments *cis*-régulateurs de la région intensité est résumé dans la partie C.

contient des séquences impliquées dans l'activation distale ayant une redondance fonctionnelle avec la boîte CR2. Afin de localiser précisément ces, séquences j'ai modifié la construction *S2* en enlevant l'extrémité 3' non conservée (construction *S11A*) ainsi que la moitié 3' de la boîte CR3 (construction *S11* ; Figure 61). Ces deux constructions testées permettent toujours de corriger le problème d'activation dans la partie distale, à nouveau dans les tissus de pattes et d'antenne. Ces données nous permettent de conclure que des séquences impliquées dans l'activation distale se trouvent au niveau des 97 pb couvrant la moitié 5' de boîte CR3 et que ces séquences sont fonctionnellement redondantes avec la boîte CR2.

3. EVALUATION DU ROLE DE BOWL DANS LA REPRESSION DES GENES *BAB*

Au cours de l'analyse fonctionnelle de la région 3', pour les constructions *S2*, *S11A* et *S11*, des dé-répressions ont été observées dans les régions les plus distales et les plus proximales du domaine d'expression de *bab2*, particulièrement dans les disques de pattes (Figure 61), révélant que la région centrale délétée (251 pb) contient des séquences impliquées dans la répression distale et proximale. Nous avons alors cherché à identifier la nature du ou des répresseur(s) impliqué(s). Bowl est le seul facteur de transcription liant l'ADN de manière spécifique, qui a été décrit comme répresseur des gènes *bab* dans les parties proximale et distale des pattes (de Celis Ibeas and Bray, 2003; Greenberg and Hatini, 2009; Hatini et al., 2005) et dans la partie proximale du disque d'antenne (Greenberg and Hatini, 2009).

Une analyse *in silico* de la distribution des sites potentiels Bowl sur l'ensemble du LAE a révélé la présence de deux sites consensus, tous les deux situés dans la région de 251 bp impliquée dans la répression. Le premier est situé dans la boîte CR2 et le second 120 pb en 3' du premier (Figure 62). De manière significative, le site de la boîte CR2 est jouté par un site non consensus, en orientation inverse. Ces données suggèrent que Bowl puisse se lier à la boîte CR2 sous forme d'un dimère. L'alignement du LAE de 22 espèces de drosophiles a révélé que le site consensus aval (non présent dans le mini-CRM BPS5) est également sous forte pression sélective. Comme la conservation est restreinte au site Bowl putatif et n'était pas apparente dans notre analyse *in-silico* sur les 12 espèces initialement séquencées, elle n'a donc pas été considérée dans la publication.

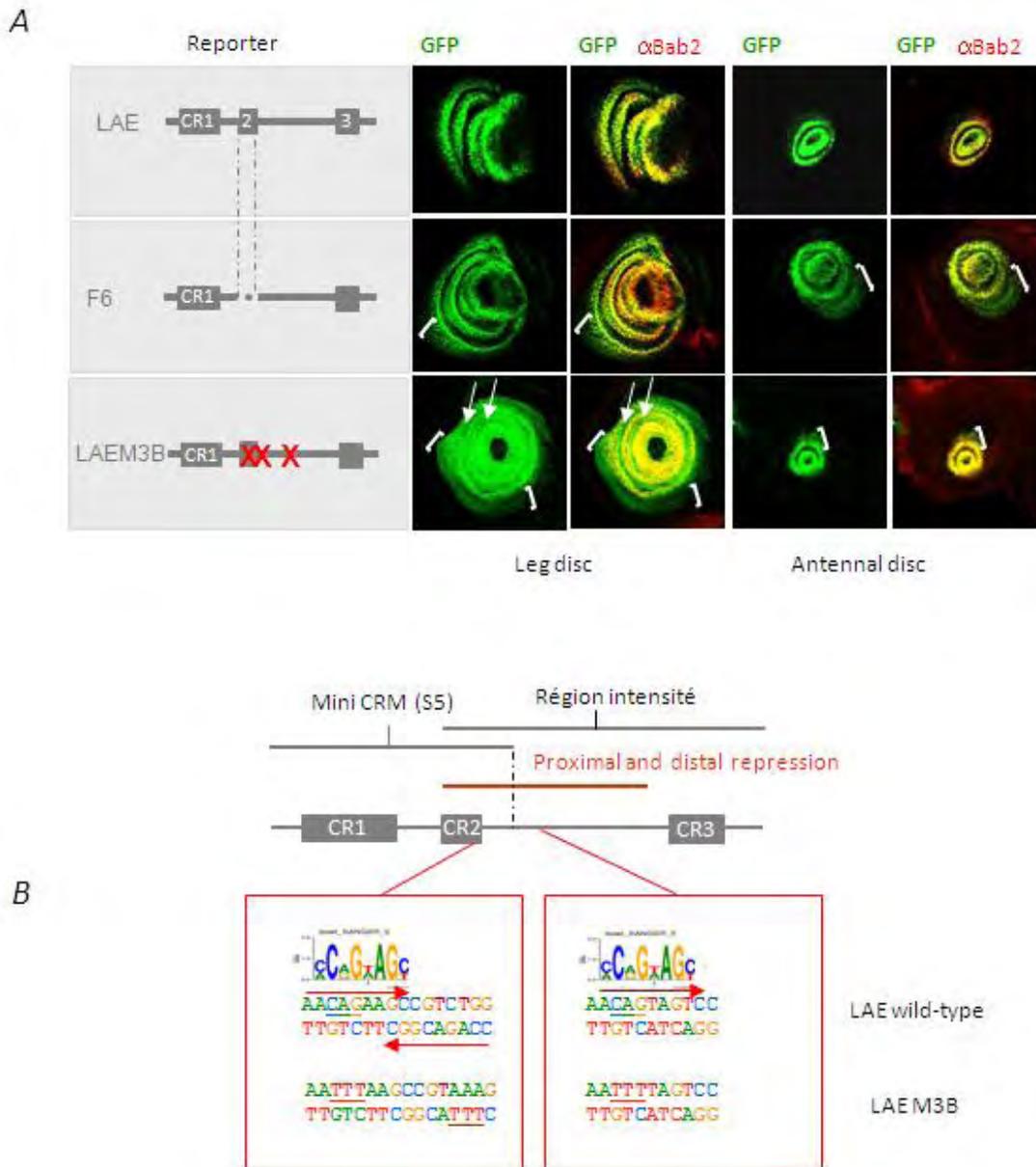


Figure 62. Les sites potentiels de liaison de Bowl sont requis pour réprimer l'activité du LAE en dehors du domaine d'expression normal de *bab2*, proximale pour les disques de pattes et d'antenne, ainsi que distalement et dans les régions inter-anneaux pour ceux de pattes. En A, les patrons d'expression des constructions *F6* et *LAEM3B* sont montrés. Noter les dé-répressions proximales. Dans *F6* la dé-répression distale est certainement masquée par le défaut d'activation distale du à la délétion de l'ensemble de la boîte CR2. En B, les sites potentiels de liaison de Bowl sont alignés avec la séquence consensus. La construction *LAEM3B* est mutée au niveau des trois sites potentiels (CAG en TTT).

Afin d'évaluer le rôle répresseur de la boîte CR2 contenant les deux sites Bowl en configuration inversée, nous avons analysé l'effet de sa délétion précise dans le contexte du LAE (construction *F6*). J'ai observé alors une dé-répression aux niveaux des régions proximale et distale des tissus de patte, et une dé-répression proximale dans les tissus d'antenne.

Ces nouvelles données suggèrent une action directe potentielle de Bowl dans la répression des gènes *bab*, dans les tissus de pattes et d'antenne. Pour évaluer plus précisément le rôle régulateur putatif de Bowl, j'ai muté les trois sites potentiels de liaison de Bowl, dans le contexte du LAE entier (construction LAEM3B). J'ai alors observé une forte dé-répression proximale en t1, ainsi que distale au niveau du segment t5, et même inter-anneaux t2-t3 et t3-t4, au niveau des tissus de pattes. La faible dé-répression distale de la construction *F6* s'explique par la délétion à la fois des sites Bowl et des séquences activatrices distales de la boîte CR2 (cf. ci-dessus). Pour l'antenne, une faible dé-répression a été observée, ce qui pourrait en partie s'expliquer par une redondance fonctionnelle avec d'autres répresseurs proximaux, comme *Dac* et *Hth*. En conclusion, ces résultats suggèrent que Bowl est probablement un répresseur direct des gènes *bab* proximement, distalement et dans les régions inter-anneaux des disques imaginaux de pattes ainsi que dans la partie proximale de son patron d'expression dans les disques d'antennes.

4. CLL/C15 UN INHIBITEUR POTENTIEL DES GENES *BAB* DANS LA REGION DISTALE

L'analyse simple hybride a permis de montrer une interaction entre le fragment S5 et le facteur à homéodomaine *Clawless* (*Cll/C15*). Comme *Cll* s'exprime dans la partie centrale des disques de pattes et d'antenne où l'expression des gènes *bab* est inhibée et qu'un mutant *C15* semble entraîner une dé-répression des gènes *bab* au centre des disques de pattes (Campbell, 2005 données non montrées), nous avons voulu tester la possibilité que *Cll* puisse réprimer les gènes *bab* directement. Etant donné que nous n'avons pas pu mettre en évidence de site responsable de la dérégulation distale dans le LAE et que *Cll* est un facteur de transcription à homéodomaine comme *Dll* on peut proposer que *Cll* agisse comme répresseur de *bab* dans la partie distale en entrant en compétition avec *Dll* au niveau de ses sites de fixation (Noyes et al., 2008).

Comme pour Dll, la séquence LAE contient plusieurs sites potentiels de fixation de Cll. En outre, nous avons constaté que le site composite de liaison de Dll au niveau de la boîte CR1 ressemble au site de fixation de l'hétérodimère Cll/Al (Miyazono et al., 2010), ce qui suggère que Cll seule ou en association avec Al puisse réprimer les gènes *bab* en empêchant Dll de se fixer au niveau de la boîte CR1. Afin de tester cette hypothèse, j'ai examiné par des expériences de gels retards la fixation d'un mélange de protéines Al et Cll, produites *in-vitro* à partir d'un lysat de réticulocyte de lapin, sur le site composite de liaison de Dll. A l'issue de cette expérience, je n'ai pas pu détecter de synergie entre Al et Cll pour fixer la région d'intérêt. Comme les protéines Al et Cll sont mal traduites *in-vitro*, il faudrait refaire cette expérience avec des protéines recombinantes purifiées, produites chez *E. coli*, et utiliser comme contrôle positif la sonde décrite par (Miyazono et al., 2010).

5. RECHERCHE DE FACTEURS CONFERANT L'INTENSITE DU SIGNAL

Lors de notre analyse fonctionnelle du LAE nous avons montré que ce dernier est formé par deux régions : une région intensité et une région spécificité pour laquelle on a pu identifier deux acteurs importants à savoir Dll et Rn. Afin de déterminer le ou les facteurs impliqué(s) dans la force du signal nous avons étudié : (i) l'importance d'autres sites Dll ou de manière générale de type homéodomaine se trouvant hors de la région spécificité (ii) la régulation de *bab* par lui-même et enfin (iii) le rôle potentiel de Sp1/Btd dans l'expression de *bab*

5.1. DLL, UN DETERMINANT DE LA FORCE DU SIGNAL ?

Comme Dll est un acteur essentiel dans l'activation de *bab* et que plusieurs sites consensus de fixation capables de lier Dll (gel retard) *in-vitro* se situent dans la région activatrice 3', nous avons voulu tester l'importance des sites potentiels de fixation de Dll et des protéines à homéodomaine de manière générale sur l'activité du LAE. Dans le contexte du LAE entier, j'ai muté deux sites potentiels de fixation de Dll. Le premier se situe dans la région la plus 3' non conservée du LAE et le second dans la boîte CR2 (voir article). En parallèle, j'ai testé l'effet de la mutation du site potentiel de fixation Dll dans la boîte CR2 dans le contexte du mini-CRM *S5*. Pour les deux constructions, je n'ai pas noté de changement net du profil d'expression de la GFP (données non présentées). Ces résultats indiquent que ces sites de fixation pour des facteurs de type homéodomaine ne sont pas

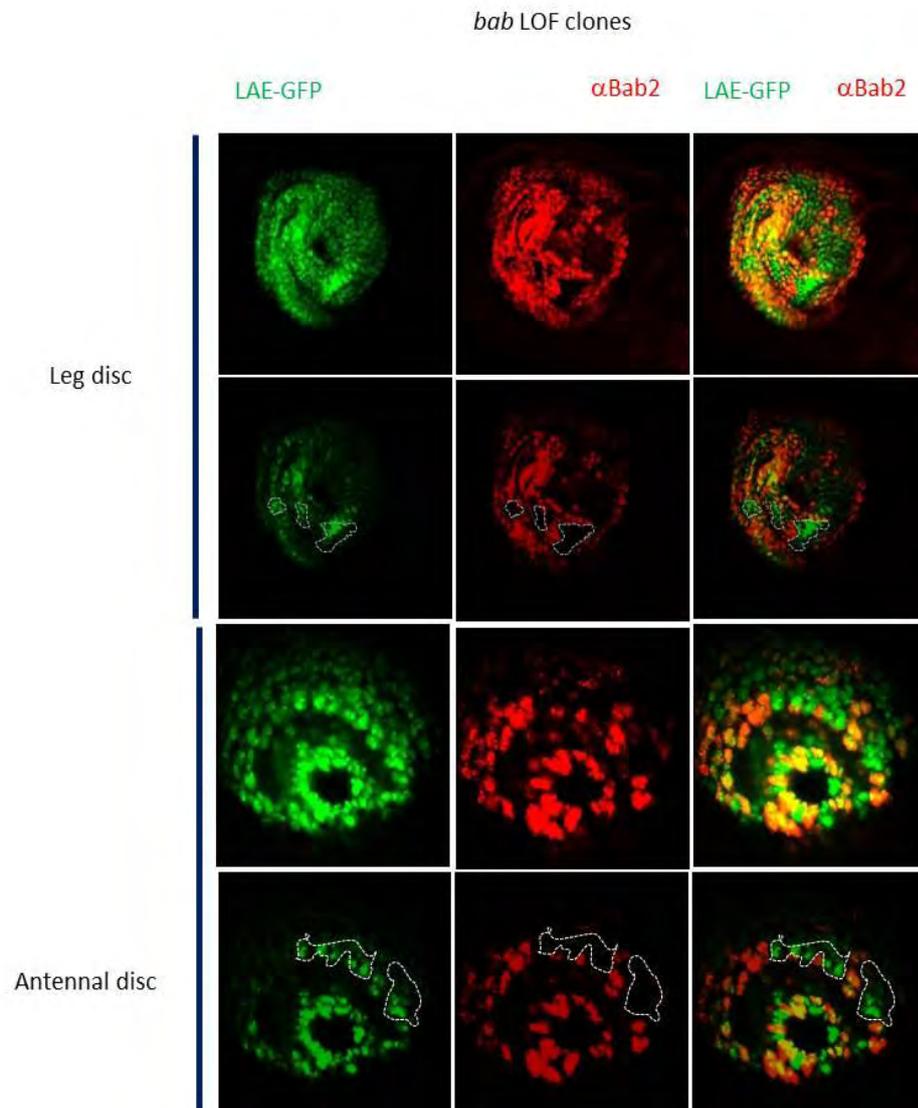


Figure 63. Autorégulation négative des gènes *bab*. Les clones mitotiques mutants pour l'allèle nul *bab*^{AR06} sont révélés par l'absence d'immuno-marquage avec l'anticorps anti-Bab2 (rouge). On remarque une augmentation globale de l'expression du rapporteur *LAE-GFP* dans des clones mutants *bab* révélant une autorégulation avec autonomie cellulaire. Les images ont été traitées de manière à mieux visualiser les cellules mutantes exprimant fortement la GFP.

strictement nécessaires ni pour l'intensité ni pour la spécificité dans le contexte du LAE et du mini-CRM.

5.2. LES GENES *BAB* S'AUTOREGULENT ?

Un mécanisme communément utilisé au cours du développement pour renforcer et ou maintenir l'expression d'un gène est une autorégulation positive. Ainsi dans le cadre de la recherche de facteurs *trans* impliqués dans l'intensité du signal, j'ai étudié la possibilité d'une régulation de *bab* par lui-même. Il est à rappeler qu'il a été montré que la protéine Bab1 peut se fixer sur la région chromosomique (immunolocalisation sur chromosomes polytènes) contenant les gènes *bab* (Lours et al., 2003)

5.2.1. EXPRESSION DU LAE-GFP EN CONTEXTE MUTANT POUR *BAB*

Pour tester cette hypothèse, j'ai examiné dans un premier temps l'expression du gène rapporteur *LAE-GFP* dans des disques *bab*⁻. Cette expérience préliminaire n'a pas révélé d'effet particulier sur l'activité régulatrice du LAE (non montré). Dans un deuxième temps, j'ai analysé l'expression du gène rapporteur *LAE-GFP* dans des clones mitotiques *bab*⁻ (Figure 63) grâce à la génération de lignées *LAE-GFP; bab^{AR07} FRT80B/TM6B* et *Dll^{EM212}Gal4/CyO; arm-LacZ FRT80B/TM6B*. Il ressort de cette analyse que la perte de la fonction *bab* entraîne une surexpression du gène rapporteur *LAE-GFP* et ceci dans la majorité des clones *bab*⁻. Ces données montrent qu'il existe bien une autorégulation (directe ou indirecte) de *bab* mais cette dernière est négative. Bab2 n'est donc pas le facteur permettant d'induire de forts niveaux d'expression du gène *bab2* dans les pattes et l'antenne.

5.2.2. RECHERCHE DE SITES DE FIXATION POTENTIELS DES PROTEINES *BAB*

Etant donné que le LAE contient au niveau de la boîte CR3 une région riche en AT évoquant les sites de liaison décrits pour les protéines Bab (Lours et al., 2003), nous avons pensé que *bab* pourrait s'autoréguler à travers la région CR3. Nous avons ainsi testé la capacité de la protéine Bab2 à fixer la région riche en bases AT de la boîte CR3, afin d'évaluer la possibilité d'une régulation directe. Pour ce faire, j'ai produit la protéine Bab2 *in-vitro* en utilisant du lysat de réticulocyte de lapin, et j'ai testé par la technique de migration retardée en gel natif des complexes ADN/protéine (EMSA), sa capacité à lier efficacement la région AT. Les premières tentatives n'ont pas révélé de fixation de Bab2 à la sonde testée. Il

faudrait refaire cette expérience en utilisant la protéine de fusion GST-BabCD2 correspondant au domaine de fixation à l'ADN de la protéine Bab2 préalablement mis en évidence par Lours et col 2003, ainsi qu'un témoin positif correspondant à la séquence utilisée pour montrer l'interaction entre Bab2 et les régions riches en AT (Lours et al., 2003). Cependant, il est probable que le résultat négatif soit réel. En effet, l'intensité de la GFP pour S11 et S11A est similaire (résultats non montrés) alors qu'on pouvait s'attendre à une diminution d'expression avec S11A qui se différencie précisément de S11 par la région riche en AT testée dans nos expériences *in-vitro*.

5.3.SP1 REGULATEUR SPECIFIQUE PATTE DES GENES *BAB* ?

Sp1 et *Btd* sont deux gènes paralogues codant pour des facteurs de transcription de type doigt à zinc. Ils sont exprimés selon le même patron dès le stade embryonnaire où ils imposent l'identité des appendices de type ventraux (antennes et pattes) et sont requis pour l'expression de *Dll* dans les disques imaginaux en formation (Estella et al., 2003 ; Estella et Mann, 2010). L'expression de ces gènes semble se maintenir au cours du développement car ils sont exprimés dans les disques d'antenne et de pattes au stade L3 tardif. *Sp1/Btd* et *Bab2* ne chevauchent pas dans les disques d'antennes. En revanche dans les pattes, ils englobent tout le domaine d'expression de *bab2*.

L'analyse de la séquence du LAE a montré l'existence de nombreux sites consensus de liaison pour *Sp1* dans la région intensité. De plus, il a été montré par un autre doctorant dans notre équipe (Luis Humberto Mojica) qu'un gain de fonction *Sp1* par 'Flip Out' augmente le niveau d'expression du transgène *LAE-GFP* dans l'antenne mais pas dans les pattes, sans pour autant modifier le niveau d'expression de *Dll*, ce qui suggère que *Sp1* pourrait être un régulateur positif des gènes *bab* dans les pattes.

Discussion

1. UN CRM COMMUN POUR LES TISSUS DE PATTES ET D'ANTENNE

L'expression dynamique des gènes *bab* dans les pattes et l'antenne et le lien évolutif entre ces deux structures, suscite plusieurs questions quant à la nature des séquences *cis*-régulatrices impliquées dans l'expression des gènes *bab* dans les pattes et l'antenne. Pour répondre à ces questions nous avons découpé dans un premier temps les deux fragments BP42/BP47 (11Kb) qui reproduisent l'expression des gènes *bab* dans les disques de pattes (williams et al., 2008) et d'antenne. Nous avons pu ainsi montrer qu'un seul élément *cis*-régulateur hautement conservé de 0,56 Kb, appelé LAE (pour Leg Antennal Enhancer), dirige l'expression de *bab2* (*bab1*) dans l'espace et dans le temps aussi bien dans les disques de pattes que d'antennes, montrant ainsi que les éléments *cis*-régulateurs des gènes *bab* pour les pattes et les antennes sont confondus. Cette constatation est en accord avec le fait que la patte et l'antenne sont deux structures homologues qui auraient évolué à partir d'un appendice ancestral commun (Casares et Mann, 1998; Cummins et al., 2003; Dong et al., 2000; Duncan et al., 1998; Pai et al., 1998; Struhl, 1981). Ainsi on peut conclure que la différence d'expression de *bab2* (*bab1*) entre l'antenne et la patte ne provient pas d'une différence au niveau des régions *cis*-régulatrices mais plutôt d'une expression différentielle des facteurs de transcription *trans*-régulateurs.

Le fait que les CRMs pattes et antenne soient confondus n'empêche pas pour autant la possibilité que des séquences au sein du LAE soient spécifiques de l'antenne ou de la patte. Ceci pourrait expliquer en partie pourquoi l'activité du LAE est plus sensible à la perte des sites Rn (construction S10) dans l'antenne que dans les pattes. Par ailleurs, il est probable que les nombreux sites potentiels Sp1 qui se trouvent principalement dans la partie centrale du LAE soient des sites activateurs pattes spécifiques. En effet, le domaine d'expression de Sp1 englobe celui des gènes *bab* dans les disques de pattes alors qu'il ne chevauche pas du tout avec *bab* dans le disque d'antenne au stade L3 tardif. Cette hypothèse s'appuie aussi sur le fait que l'expression ectopique de Sp1 dans des groupes de cellules antennaires qui expriment normalement les gènes *bab*, augmente le niveau d'expression de ce dernier alors que l'expression ectopique de Sp1 dans les disques de pattes est sans effet sur l'expression de *bab* (résultats non publiés du laboratoire).

Quoique plusieurs CRMs reproduisant l'expression de *Dll* et *dac* dans les pattes aient été identifiés, notre étude est la première en son genre à montrer qu'un seul élément

cis-régulateur est à la fois nécessaire et suffisant pour diriger l'expression d'un même gène dans deux structures homologues. Il serait donc intéressant d'analyser l'expression des CRMs de *Dll* et *dac* dans le disque d'antenne afin de voir s'ils sont spécifiques de la patte où s'ils fonctionnent de manière similaire au LAE. Une telle étude pourrait aider à mieux comprendre l'évolution des pattes et de l'antenne.

2. LE LAE : UN MODELE POUR L'ETUDE DU FONCTIONNEMENT D'UN CRM

L'expression des gènes *bab* dans les disques de pattes et d'antenne est très dynamique, passant d'une expression homogène sur tout le domaine Dll au début du stade L3, à une expression sous forme de quatre anneaux concentriques dans les disques de pattes et deux dans le disque d'antenne à la fin du stade L3. Plusieurs études se sont intéressées à la régulation du gène *bab2* dans les disques de pattes et d'antenne. Ainsi il a été montré que *Dll* est strictement requis pour l'expression de *bab2* et que *ss* est un activateur au niveau des anneaux proximaux (Chu et al., 2002; de Celis et Bray, 2003; Galindo et al., 2002). Il a été montré aussi que la restriction proximale de *bab2* est médiée par *dac* et *bowl* dans les disques de pattes (de Celis Ibeas et Bray, 2003; Greenberg et Hatini, 2009; Hatini et al, 2005), auxquelles se rajoute *sal* et *hth* dans celui d'antenne, alors que la restriction distale au niveau du segment tarsal t1 se fait par *bowl*. Il a été noter qu'en plus de son rôle dans la répression proximale, *Dac* module également le niveau d'expression de *bab2* inter-anneaux (Chu et al., 2002). Enfin, il a été suggéré que la voie EGFR médie la répression distale (Campbell, 2002). Cependant, la régulation en *cis* n'a jamais été abordée pour les gènes *bab* et aucune de ces études n'a pu montrer une implication directe des facteurs précédemment cités dans la régulation des gènes *bab*.

En menant une analyse fonctionnelle du LAE, j'ai pu montrer que la régulation spécifique dépend principalement d'une région de 220 pb (construction S5, ou «mini-CRM»), se trouvant du côté 5' du LAE. J'ai montré aussi que cette région intègre les apports positifs des deux activateurs identifiés Dll et Rn ainsi que négatifs du répresseur Bowl. De manière intéressante, on a montré que l'activation proximale et distale des anneaux *bab* dépend de régions *cis*-régulatrices différentes. Pour la régulation proximale, on a pu identifier deux régions dont une contenant un site de fixation du facteur Rn, requis *in-vivo* pour l'activation directe de *bab2* par Rn. Cependant le(s) facteur(s) (X) se liant à la région LS8

et induisant l'expression de *bab2* au niveau de l'anneau le plus proximal reste(nt) à déterminer. Pour la régulation distale, il est clair que la boîte CR2 contient des sites de fixation d'activateurs distaux qui seraient redondants avec d'autres sites se trouvant du côté 3' de la boîte CR3 du LAE. Le(s) facteur(s) se liant à ces régions restent encore à déterminer.

Dans cette étude, nous avons également identifié une courte séquence nécessaire pour la répression inter-anneau à l'extrémité 3' de CR1 (motif LS1). La nature du répresseur qui lie LS1 reste à identifier, en particulier son lien fonctionnel avec le FT Dac, qui ne lie pas l'ADN et qui fonctionne au sein d'un complexe multiprotéique (Chen et al., 1997). Nous avons montré aussi que Bowl pourrait être un répresseur direct de *bab2* dans les parties distale et proximale et qu'il aurait un rôle dans la répression inter-anneaux. Enfin, nous suggérons que Cll soit directement impliqué dans la répression de *bab2* dans la future région prétarse et l'arista. En fait aucun des sous-fragments ou des formes mutées du LAE n'affichent de dérégulations dans région distale qui correspond au future prétarse, suggérant (i) soit la présence d'éléments d'ADN répressifs fonctionnellement redondants ou (ii) soit un mécanisme de répression passant par une compétition entre les facteurs répresseurs et activateurs sur les mêmes sites de liaison. Or, le site de fixation de Cll est très similaire à celui de Dll et la région LS2/LS3 fixant Dll et qui est critique au fonctionnement du LAE ressemble considérablement au site optimal de fixation de l'hétérodimère Cll/A1. On sait aussi que l'expression ectopique de Cll réprime *bab2* et le *LAE-GFP* de manière cellulaire autonome (Luis Mojica, Résultats non publiés). A ceci se rajoute le fait que la dérégulation de la forme mutée LAEM3B s'étend sur t5 mais pas sur toute la région distale ce qui montre que la répression au niveau du prétarse est due à un facteur qui est exprimé spécifiquement dans cette région. Enfin Cll a été identifié par l'analyse simple hybride aussi bien pour les fragments LAE que S5 ce qui renforce encore d'avantage l'hypothèse d'un rôle de Cll dans la répression distale.

3. L'HOMÉOPROTEINE DLL EST UN ACTIVATEUR DIRECT DE *BAB2* DANS LES PATTES ET L'ANTENNE

Dans cette étude, nous avons fourni des preuves de l'implication de Dll comme activateur direct de *bab2*, se liant aux motifs conservés TAAT qui se trouvent dans la partie 3' de CR1. Ce modèle est d'autant plus renforcé par deux autres constatations: (i) le fait que la

liaison directe de Dll au niveau du LAE *in vivo* à été déjà établie par des expériences d'immunoprécipitation de la chromatine (ChIP) (ModENCODE projet), dans des extraits embryonnaires et (2) le fait qu'une séquence A/T riche, similaire au site de fixation de Dll dans le LAE (AATTTAATGGTAAA; les nucléotides conservés sont soulignés), se trouve présente au sein d'un enhancer antennaire de *ss* qui est une autre cible directe de Dll (Duncan et al., 2010). Ceci suggère une grammaire de CRM conservée pour les gènes régulés par Dll.

4. UN ROLE SPECIFIQUE DE LA PROTEINE A DOIGT DE ZINC ROTUND DANS LA REGULATION PROXIMALE DE *BAB2*

Dans ce travail, nous montrons pour la première fois un rôle de Rotund dans la régulation de l'expression de *bab2* à la différence de ce qui été rapporté par Saint-Pierre et col (2002), qui suggèraient que *bab* ne soit pas régulé par Rn. En fait, dans ce travail les auteurs ont analysé l'effet de la perte de fonction de *rn* sur l'expression de *bab2* dans des disques de patte L3 précoces et ont trouvé que l'expression globale de *bab2* était normale (un seul anneau). Toutefois, ils n'ont pas analysé ou montré le profil d'expression de *bab2* dans des disques tardifs où on observe normalement quatre anneaux concentriques (Saint-Pierre et al., 2002). En effet, d'après les résultats (non publiés) obtenus au laboratoire dans les disques tardifs mutant pour *rn*, *bab* s'exprime en un anneau voire deux au maximum avec une expression très faible de l'anneau le plus proximal.

Etant donné que l'activation de *rn* dans les tissus t1-t3 du tarse dépend de l'activité de *ss* (Natori et al., 2012), nos données nous permettent de proposer que *Ss* (Chu et al., 2002) régule indirectement l'expression de *bab* dans les anneaux proximaux, via l'activation de *rn* dont le produit active directement l'expression de *bab2*. Enfin, nous avons montré que l'expression ectopique de Rn est capable d'activer *bab2* dans tout le domaine *Dll*, en particulier dans la partie la plus distale des disques de pattes et d'antenne. La dérégulation de *bab2* dans la région la plus distale (prétarse particulièrement) peut être expliquée en plus de l'effet activateur direct de Rn par *bab* par la répression des gènes cibles de la voie EGFR par Rn. En effet, un gain de fonction de Rn dans le domaine *Dll* fait perdre l'expression de *C15/cll* et *dLim1* (observations non publiées).

5. *BAB1* ET *BAB2*, DEUX GENES SOUS LE CONTROLE D'UN MEME ENHANCER ?

En prenant en compte du fait que : (i) *bab1* et *bab2* présentent des patrons d'expressions similaires dans les pattes et l'antenne ; (ii) le LAE est l'unique CRM identifié au niveau du locus *bab* (Williams et al 2008) capable de diriger l'expression d'un gène rapporteur selon le même patron que le gène *bab2* dans les pattes et l'antenne ; (iii) le LAE est absolument requis pour l'expression de *bab2* dans les disques de pattes et d'antenne comme on a pu le montrer en utilisant le BAC 26B15 (qui contient le promoteur et une partie du premier exon de *bab1*, la région inter-génique *bab1/bab2* et la totalité du gène *bab2*) et la technique du recombineering, on peut conclure qu'il est fort probable que le LAE régule aussi l'expression du gène *bab1*.

Afin de confirmer ou d'infirmer cette hypothèse. J'ai tout d'abord tenté par hybridation *in situ* de détecter le transcrite correspondant au premier exon de *bab1* mais l'absence de signal observé semble indiquer que l'ARN messager tronqué supposé être produit à partir du BAC 26B15 est probablement instable). Par la suite, j'ai tenté d'intégrer le gène rapporteur GFP au sein du BAC 26B15 en aval du promoteur *bab1* et de tester l'effet de la délétion du LAE dans le contexte du BAC modifié sur l'activité du promoteur *bab1* mais malheureusement, cette expérience n'a pas pu aboutir. Une autre façon de tester cette hypothèse serait de faire des expériences de 3C sur le locus *bab* à partir de disques de pattes et d'antenne. Ces expériences permettront de déterminer s'il existe un rapprochement physique *in vivo* entre le LAE et le promoteur *bab1* dans les disques de pattes et d'antenne ce qui constituerait une preuve indirecte de la régulation conjointe de *bab1* et *bab2* par le LAE.

D'après les travaux de Williams et col (2008) tous les CRMs identifiés au niveau du locus *bab* se trouveraient en un seul exemplaire. Cette constatation suggère que les CRM qui activent *bab1* activent aussi *bab2*. Comme *bab1* et *bab2* sont deux gènes paralogues, il serait logique que les promoteurs proximaux ancestraux de *bab1* et *bab2* fonctionnent de manière assez semblable et qu'ils répondent de manière plus ou moins similaire vis à vis des séquences *cis*-régulatrices préexistantes. Il est même possible que cette propriété constitue une contrainte évolutive qui permette d'avoir une expression forte des gènes *bab*.

6. CONSERVATION DU LAE CHEZ LES DROSOPHILIDAE ET AU-DELA

Le LAE est situé dans la région intergénique de *bab1* et *bab2* non seulement dans toutes les espèces Drosophilidae examinées, mais aussi dans la mouche tsétsé, *Glossina morsitans* (observations non publiées). Cela signifie que le LAE est soumis à une forte pression de sélection. En outre, nos expériences de sauvetage *in-vivo* par le BAC, ainsi que l'analyse systématique de l'expression des constructions rapporteurs couvrant la totalité du locus *bab* (Williams et al., 2008), indiquent l'absence de « shadow enhancers » (Hong et al., 2008) susceptibles de remplacer le LAE. Cette hypothèse est consolidée par le fait que le LAE de *D. virilis* génère un profil d'expression très similaire à celui de *D. melanogaster*, montrant ainsi une conservation fonctionnelle du LAE même chez une espèce de drosophile éloignée. Il est à noter que les boîtes CR1 et CR2 sont en partie conservées chez *Glossina morsitans*, soutenant l'idée qu'ils contiennent des sites de liaison pour les principaux activateurs évolutivement conservés (par exemple, Dll) ainsi que des répresseurs.

7. ORGANISATION TRIDIMENSIONNELLE DU LOCUS *BAB*

Deux allèles mutants du gène *bab1*, *bab^P* et *babA128*, obtenue suite à l'insertion de transposons à proximité du promoteur de *bab1* (Figure 44), n'expriment pas le gène *bab1* mais n'entraînent pas pour autant de phénotypes dans les pattes. En revanche, les mutations *bab^{PRDS}* et *bab^{PR72}* obtenues suite à l'excision de l'élément P contenu dans l'allèle muté *bab^P* (Couderc et al., 2002) entraînent une baisse du niveau d'expression de *bab2* et une fusion des segments tarsaux les plus distaux. Ceci indique que les régions délétées suite à l'excision des éléments transposables sont impliquées dans la régulation de *bab2*. J'ai essayé de localiser grossièrement par PCR la délétion de *bab^{PR72}* et j'ai pu montrer que l'excision ne touche pas le LAE suggérant que cette délétion n'affecte pas le LAE ni le gène *bab2*. D'après les travaux de Jean-Louis Couderc (résultats non publiés) le fragment Fe1D de 4,5Kb englobant le promoteur proximal de *bab2*, le premier exon et intron ainsi qu'une partie du deuxième exon, reproduit de manière imparfaite l'expression de *bab* dans les pattes avec une expression très faible (surtout dans les anneaux les plus distaux) et des anneaux incomplets. Cette expérience montre que le promoteur de *bab2* contient une partie de l'information permettant l'expression

du gène *bab2*. Comme *bab1* et *bab2* sont deux gènes paralogues, on peut imaginer que le phénotype obtenu avec les hypomorphes *bab^{PRDS}* et *bab^{PR72}* est dû à la délétion de la région équivalente à Fe1D au niveau du promoteur de *bab1*. Cette hypothèse suggérerait un rôle du promoteur de *bab1* dans la régulation de *bab2*, éventuellement en améliorant l'interaction entre le promoteur *bab2* et le LAE.

Au cours de mes tests de sauvetage par le BAC 26B15 (qui contient les promoteurs de *bab1* et de *bab2*) et le *LAE-bab2cDNA* (qui contient 120pb du promoteur proximal de *bab2* en plus du promoteur *hsp70*), j'ai noté dans le premier cas 100% de restauration de la taille et des jonctions des targes, mais aussi l'existence de peignes sexuels supplémentaires incomplets dans les targes plus distaux. En revanche, dans le cas du *LAE-bab2* j'ai observé des problèmes de jonction entre t4 et t5 dans 30% des cas ainsi qu'une apparition de peignes sexuels ectopiques plus prononcée qu'avec le BAC. Cette différence suggère l'intervention d'autres régions *cis*-régulatrices secondaires dans l'expression fine de *bab2*. Il serait donc intéressant de tester par la technique de 3C le rapprochement des promoteurs *bab1* et *bab2* ainsi que leur interaction avec le LAE. Si le rapprochement entre les promoteurs *bab1* et *bab2* est avéré ceci pourrait suggérer un effet synergique entre les deux promoteurs et le LAE, permettant de conférer une robustesse au système (même si le LAE reste un élément incontournable). Ce modèle de multiples interactions en *cis* rappelle celui de l'« archipel de régulation » du locus *HoxD* de la souris et pourrait expliquer en partie pourquoi les gènes *bab* ont été conservés dupliqués chez plusieurs espèces au cours de l'évolution.

8. ROLE DU LAE DANS LE DEVELOPPEMENT DES PEIGNES SEXUELS

Les peignes sexuels sont des structures spécifiques des mâles qui se développent au niveau des targes des pattes prothoraciques. Le nombre de peignes et leur taille varie largement d'une espèce de drosophile à une autre. Ainsi par exemple chez *D. melanogaster* on a un seul peigne au niveau du tarse t1, chez *D. pseudoobscura* on en a deux et il n'y en a aucun chez *D. virilis*. Il a été rapporté par Randsholt et Santamaria (2008) que le développement des peignes sexuels nécessite l'expression de *Scr* et que ce dernier est réprimé par *bab*, ainsi la diminution d'expression de *bab2* chez *D. melanogaster* entraîne l'expression ectopique de *Scr* et par la suite l'apparition de peignes sexuels ectopiques au niveau des targes

les plus distaux. Ceci montre que la régulation de l'expression des gènes *bab* est très importante pour la formation des peignes sexuels et il est même probable que la variation du niveau d'expression des gènes *bab* soit au moins en partie à l'origine de la diversité de ce caractère sexuel chez les différentes espèces de drosophile.

Nous avons bien montré que l'activation de *bab2* dans les pattes dépend de Rn proximale et d'un facteur Y inconnu jusqu'à présent distalement. Ainsi on peut très bien imaginer que l'activité du LAE dans les différentes espèces contribue à la diversité des peignes sexuels. Toutefois ceci n'empêche pas l'existence d'autres mécanismes comme par exemple l'évolution de la régulation du gène *Scr*. Il est à noter que chez *D. virilis*, l'espèce la plus éloignée de *D. melanogaster*, qui n'exprime pas *Scr* dans les tarsi et qui n'a pas de caractères sexuels, son LAE contient une insertion d'approximativement 200 pb au centre qui pourrait être synonyme d'une région « intensité » plus longue conférant une activité plus importante. Pour tester cette hypothèse, l'on pourrait tester le sauvetage du phénotype *bab* avec le transgène LAE^{D.virilis}*bab2cDNA* et comparer les phénotypes ainsi que l'expression de *Scr* au sauvetage effectué avec le LAE^{D.melanogaster}*bab2cDNA*. Ainsi si les peignes sexuels disparaissent ou sont réduits, ceci pourrait suggérer que le LAE est à l'origine de la diversité des peignes sexuels chez les Drosophilidae.

9. RN ACTIVATEUR DIRECT DE *BAB2* ET REPRESSEUR INDIRECT

Comme décrit précédemment dans l'article, (Figure 7) l'expression ectopique de Rn dans le domaine Dll (*Dll-Gal4 ; UAS-Rn*) induit l'expression ectopique de *bab2* ainsi que du gène rapporteur *S5-GFP* dans tout le domaine *Dll*. Cependant, on n'a aucune expression du gène rapporteur *S10-GFP* avec ce gain d'expression de Rn ce qui appuie fortement notre hypothèse que Rn induit directement *bab2* à travers le LAE. Il est à noter cependant que dans les conditions normales *S10* exprime la GFP dans les disques de pattes, avec un niveau faible au niveau des anneaux les plus proximaux, ainsi que dans le disque d'antenne au niveau de l'anneau le plus distal. Comme dans un gain de fonction Rn, on perd toute l'expression du gène rapporteur *S10-GFP* ceci montre que Rn est aussi un répresseur indirect de *bab2* (voir article Figure 7). Nous n'avons pour l'instant aucune idée de la nature du facteur activateur dont l'expression pourrait être réprimée par Rn.

Proximal → Distal

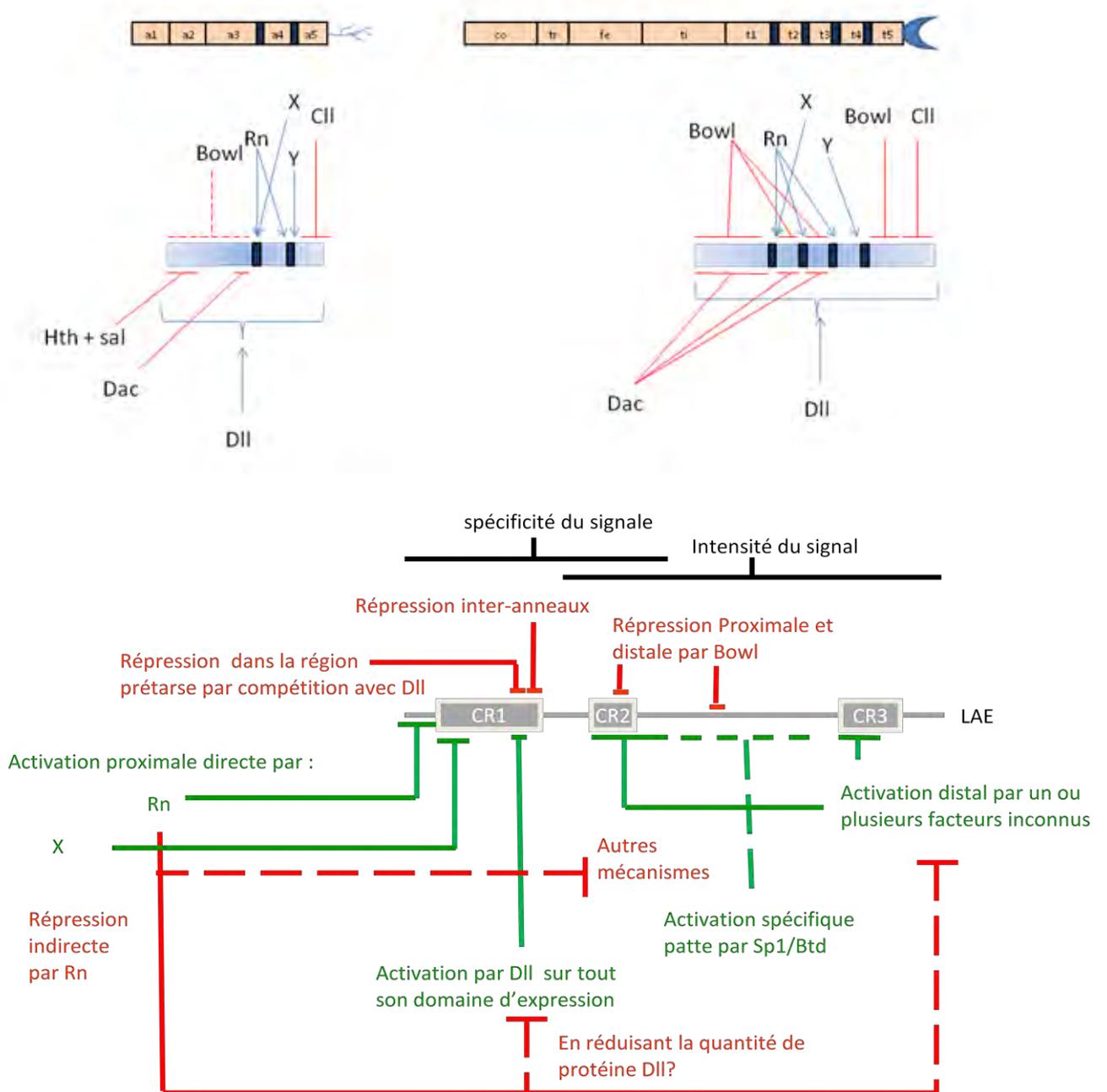


Figure 64. Model du fonctionnement de l'enhancer LAE dirigeant l'expression du gène *bab2* dans les tissus de pattes et l'antenne.

10. MODELE

Ainsi en se basant sur les études précédentes et les nouveaux éléments issus de notre étude nous proposons un nouveau modèle de la régulation des gènes *bab* dans les tissus de pattes et de l'antenne (Figure 64). Nous montrons dans ce modèle l'implication de Rn et d'un facteur X dans l'activation proximale et l'existence d'un facteur Y impliqué dans l'activation distale. Nous montrons aussi que la répression distale dans les pattes est due d'une part à Bowl au niveau du t1 et Cll (avec éventuellement d'autres cibles de la voie EGFR) au niveau du future prétarse. Cependant, l'effet de Bowl dans l'antenne est moins net surtout distalement. En effet, d'après Greenberg et Hatini (2009) Bowl ne s'exprime pas dans la partie la plus distale dans la région équivalente à t1 dans la patte. Dans la partie proximale, l'effet de Bowl est très limité et on a une très faible dérpression, qui ne donne en aucun cas un anneau comme on a pu observer dans les pattes suggérant l'intervention d'autres répresseurs en sus de Bowl. Nous rapportons pour la première fois un modèle où Cll inhiberait *bab* en entrant en compétition avec l'activateur de base Dll et suggérons que ce site de fixation de Dll au niveau de LS2/LS3 constituerait un site optimal de fixation de l'hétérodimère Al/Cll.

Références

- Abu-Shaar, M., and Mann, R.S. (1998). Generation of multiple antagonistic domains along the proximodistal axis during *Drosophila* leg development. *Development* *125*, 3821-3830.
- Agalioti, T., Lomvardas, S., Parekh, B., Yie, J., Maniatis, T., and Thanos, D. (2000). Ordered recruitment of chromatin modifying and general transcription factors to the IFN-beta promoter. *Cell* *103*, 667-678.
- ALLFREY, V.G., FAULKNER, R., and MIRSKY, A.E. (1964). Acetylation and Methylation of Histones and their Possible Role in the Regulation of RNA Synthesis. *Proc Natl Acad Sci U S A* *51*, 786-794.
- Allison, L.A., Wong, J.K., Fitzpatrick, V.D., Moyle, M., and Ingles, C.J. (1988). The C-terminal domain of the largest subunit of RNA polymerase II of *Saccharomyces cerevisiae*, *Drosophila melanogaster*, and mammals: a conserved structure with an essential function. *Mol Cell Biol* *8*, 321-329.
- Angelini, D.R., and Kaufman, T.C. (2005). Insect appendages and comparative ontogenetics. *Dev Biol* *286*, 57-77.
- Arnosti, D.N., Gray, S., Barolo, S., Zhou, J., and Levine, M. (1996). The gap protein knirps mediates both quenching and direct repression in the *Drosophila* embryo. *EMBO J* *15*, 3659-3666.
- Artavanis-Tsakonas, S., Rand, M.D., and Lake, R.J. (1999). Notch signaling: cell fate control and signal integration in development. *Science* *284*, 770-776.
- Austin, R.J., and Biggin, M.D. (1995). A domain of the even-skipped protein represses transcription by preventing TFIID binding to a promoter: repression by cooperative blocking. *Mol Cell Biol* *15*, 4683-4693.
- Ayer, S., and Benyajati, C. (1990). Conserved enhancer and silencer elements responsible for differential *Adh* transcription in *Drosophila* cell lines. *Mol Cell Biol* *10*, 3512-3523.
- Baek, H.J., Kang, Y.K., and Roeder, R.G. (2006). Human Mediator enhances basal transcription by facilitating recruitment of transcription factor IIB during preinitiation complex assembly. *J Biol Chem* *281*, 15172-15181.
- Baena-López, L.A., Baonza, A., and García-Bellido, A. (2005). The orientation of cell divisions determines the shape of *Drosophila* organs. *Curr Biol* *15*, 1640-1644.
- Bailey, A.M., and Posakony, J.W. (1995). Suppressor of hairless directly activates transcription of enhancer of split complex genes in response to Notch receptor activity. *Genes Dev* *9*, 2609-2622.
- Balamotis, M.A., Pennella, M.A., Stevens, J.L., Wasylyk, B., Belmont, A.S., and Berk, A.J. (2009). Complexity in transcription control at the activation domain-mediator interface. *Sci Signal* *2*, ra20.
- Bannister, A.J., and Kouzarides, T. (2011). Regulation of chromatin by histone modifications. *Cell Res* *21*, 381-395.
- Barmina, O., and Kopp, A. (2007). Sex-specific expression of a HOX gene associated with rapid morphological evolution. *Dev Biol* *311*, 277-286.
- Barolo, S. (2012). Shadow enhancers: frequently asked questions about distributed *cis*-regulatory information and enhancer redundancy. *Bioessays* *34*, 135-141.
- Barolo, S., and Levine, M. (1997). hairy mediates dominant repression in the *Drosophila* embryo. *EMBO J* *16*, 2883-2891.

- Becam, I., Fiuza, U.M., Arias, A.M., and Milán, M. (2010). A role of receptor Notch in ligand cis-inhibition in *Drosophila*. *Curr Biol* 20, 554-560.
- Bedford, M.T., and Richard, S. (2005). Arginine methylation an emerging regulator of protein function. *Mol Cell* 18, 263-272.
- Berger, S.L. (2007). The complex language of chromatin regulation during transcription. *Nature* 447, 407-412.
- Bernstein, B.E., Mikkelsen, T.S., Xie, X., Kamal, M., Huebert, D.J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K., *et al.* (2006). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* 125, 315-326.
- Bickel, R.D., Schackwitz, W.S., Pennacchio, L.A., Nuzhdin, S.V., and Kopp, A. (2009). Contrasting patterns of sequence evolution at the functionally redundant bric à brac paralogs in *Drosophila melanogaster*. *J Mol Evol* 69, 194-202.
- Biddie, S.C., John, S., Sabo, P.J., Thurman, R.E., Johnson, T.A., Schiltz, R.L., Miranda, T.B., Sung, M.H., Trump, S., Lightman, S.L., *et al.* (2011). Transcription factor API potentiates chromatin accessibility and glucocorticoid receptor binding. *Mol Cell* 43, 145-155.
- Bilodeau, S., Kagey, M.H., Frampton, G.M., Rahl, P.B., and Young, R.A. (2009). SetDB1 contributes to repression of genes encoding developmental regulators and maintenance of ES cell state. *Genes Dev* 23, 2484-2489.
- Bishop, S.A., Klein, T., Arias, A.M., and Couso, J.P. (1999). Composite signalling from Serrate and Delta establishes leg segments in *Drosophila* through Notch. *Development* 126, 2993-3003.
- Blau, J., Xiao, H., McCracken, S., O'Hare, P., Greenblatt, J., and Bentley, D. (1996). Three functional classes of transcriptional activation domain. *Mol Cell Biol* 16, 2044-2055.
- Blochlinger, K., Jan, L.Y., and Jan, Y.N. (1993). Postembryonic patterns of expression of cut, a locus regulating sensory organ identity in *Drosophila*. *Development* 117, 441-450.
- Boehm, A.K., Saunders, A., Werner, J., and Lis, J.T. (2003). Transcription factor and polymerase recruitment, modification, and movement on dhsp70 in vivo in the minutes following heat shock. *Mol Cell Biol* 23, 7628-7637.
- Boeing, S., Rigault, C., Heidemann, M., Eick, D., and Meisterernst, M. (2010). RNA polymerase II C-terminal heptarepeat domain Ser-7 phosphorylation is established in a mediator-dependent fashion. *J Biol Chem* 285, 188-196.
- Bonasio, R., Tu, S., and Reinberg, D. (2010). Molecular signals of epigenetic states. *Science* 330, 612-616.
- Bouazoune, K., and Brehm, A. (2006). ATP-dependent chromatin remodeling complexes in *Drosophila*. *Chromosome Res* 14, 433-449.
- Boube, M., Joulia, L., Cribbs, D.L., and Bourbon, H.M. (2002). Evidence for a Mediator of RNA Polymerase II Transcriptional Regulation Conserved from Yeast to Man. *Cell* 110, 143-151.
- Bourbon, H.M. (2008). Comparative genomics supports a deep evolutionary origin for the large, four-module transcriptional mediator complex. *Nucleic Acids Res* 36, 3993-4008.
- Bray, S.J. (2006). Notch signalling: a simple pathway becomes complex. *Nat Rev Mol Cell Biol* 7, 678-689.

- Brook, W.J., and Cohen, S.M. (1996). Antagonistic interactions between wingless and decapentaplegic responsible for dorsal-ventral pattern in the *Drosophila* Leg. *Science* *273*, 1373-1377.
- Bryant, P.J. (1970). Cell lineage relationships in the imaginal wing disc of *Drosophila melanogaster*. *Dev Biol* *22*, 389-411.
- Bryant, P.J. (1978). Pattern formation in imaginal discs. In *The Genetics and Biology of Drosophila*, pp. 229-235.
- Bulger, M., and Groudine, M. (2011). Functional and mechanistic diversity of distal transcription enhancers. *Cell* *144*, 327-339.
- Buratowski, S. (2009). Progression through the RNA polymerase II CTD cycle. *Mol Cell* *36*, 541-546.
- Burke, L.J., and Baniahmad, A. (2000). Co-repressors 2000. *FASEB J* *14*, 1876-1888.
- Buryanov, Y.I., and Shevchuk, T.V. (2005). DNA methyltransferases and structural-functional specificity of eukaryotic DNA modification. *Biochemistry (Mosc)* *70*, 730-742.
- Bushey, A.M., Ramos, E., and Corces, V.G. (2009). Three subclasses of a *Drosophila* insulator show distinct and cell type-specific genomic distributions. *Genes Dev* *23*, 1338-1350.
- Busturia, A., and Bienz, M. (1993). Silencers in abdominal-B, a homeotic *Drosophila* gene. *EMBO J* *12*, 1415-1425.
- Butler, J.E., and Kadonaga, J.T. (2001). Enhancer-promoter specificity mediated by DPE or TATA core promoter motifs. *Genes Dev* *15*, 2515-2519.
- Calhoun, V.C., and Levine, M. (2003). Long-range enhancer-promoter interactions in the *Scr*-*Antp* interval of the *Drosophila* Antennapedia complex. *Proc Natl Acad Sci U S A* *100*, 9878-9883.
- Calhoun, V.C., Stathopoulos, A., and Levine, M. (2002). Promoter-proximal tethering elements regulate enhancer-promoter specificity in the *Drosophila* Antennapedia complex. *Proc Natl Acad Sci U S A* *99*, 9243-9247.
- Campbell, G. (2002). Distalization of the *Drosophila* leg by graded EGF-receptor activity. *Nature* *418*, 781-785.
- Campbell, G. (2005). Regulation of gene expression in the distal region of the *Drosophila* leg by the Hox11 homolog, C15. *Dev Biol* *278*, 607-618.
- Campbell, G., and Tomlinson, A. (1998). The roles of the homeobox genes *aristaless* and *Distal-less* in patterning the legs and wings of *Drosophila*. *Development* *125*, 4483-4493.
- Campbell, G., Weaver, T., and Tomlinson, A. (1993). Axis specification in the developing *Drosophila* appendage: the role of wingless, decapentaplegic, and the homeobox gene *aristaless*. *Cell* *74*, 1113-1123.
- Camporeale, G., Shubert, E.E., Sarath, G., Cerny, R., and Zemleni, J. (2004). K8 and K12 are biotinylated in human histone H4. *Eur J Biochem* *271*, 2257-2263.
- Campos, E.I., and Reinberg, D. (2009). Histones: annotating chromatin. *Annu Rev Genet* *43*, 559-599.
- Cao, R., and Zhang, Y. (2004). The functions of E(Z)/EZH2-mediated methylation of lysine 27 in histone H3. *Curr Opin Genet Dev* *14*, 155-164.

- Carroll, S.B., Grenier, J.K., and Weatherbee, S.D. (2001). *From DNA to Diversity: Molecular Genetics and the Evolution of Animal Design* (Malden, MA: Blackwell Science).
- Carr, A., and Biggin, M.D. (1999). A comparison of in vivo and in vitro DNA-binding specificities suggests a new model for homeoprotein DNA binding in *Drosophila* embryos. *EMBO J* 18, 1598-1608.
- Carroll, S.B. (2005). Evolution at two levels: on genes and form. *PLoS Biol* 3, e245.
- Casares, F., and Mann, R.S. (2001). The ground state of the ventral appendage in *Drosophila*. *Science* 293, 1477-1480.
- Casares, F., Sánchez, L., Guerrero, I., and Sánchez-Herrero, E. (1997). The genital disc of *Drosophila melanogaster*. I. Segmental and compartmental organization. *Dev Genes Evol* 207, 216-228.
- Cavener, D.R., Ottelson, D.C., and Kaufman, T.C. (1986). A rehabilitation of the genetic map of the 84B-D region in *Drosophila melanogaster*. *Genetics* 114, 111-123.
- Cerutti, H., and Casas-Mollano, J.A. (2009). Histone H3 phosphorylation: universal code or lineage specific dialects? *Epigenetics* 4, 71-75.
- Chaib, H., Prébet, T., Vey, N., and Collette, Y. (2011). [Histone methyltransferases: a new class of therapeutic targets in cancer treatment?]. *Med Sci (Paris)* 27, 725-732.
- Chang, B., Chen, Y., Zhao, Y., and Bruick, R. (2007). JMJD6 is a histone arginine demethylase. *Science* 318, 444-447.
- Chau, C.M., Zhang, X.Y., McMahon, S.B., and Lieberman, P.M. (2006). Regulation of Epstein-Barr virus latency type by the chromatin boundary factor CTCF. *J Virol* 80, 5723-5732.
- Chen, W., Zollman, S., Couderc, J.L., and Laski, F.A. (1995). The BTB domain of bric à brac mediates dimerization in vitro. *Mol Cell Biol* 15, 3424-3429.
- Chen, Y., and Struhl, G. (1998). In vivo evidence that Patched and Smoothed constitute distinct binding and transducing components of a Hedgehog receptor complex. *Development* 125, 4943-4948.
- Chiang, A., O'Connor, M.B., Paro, R., Simon, J., and Bender, W. (1995). Discrete Polycomb-binding sites in each parasegmental domain of the bithorax complex. *Development* 121, 1681-1689.
- Chinnadurai, G. (2003). CtBP family proteins: more than transcriptional corepressors. *Bioessays* 25, 9-12.
- Chopra, V.S., Hong, J.W., and Levine, M. (2009). Regulation of Hox gene activity by transcriptional elongation in *Drosophila*. *Curr Biol* 19, 688-693.
- Chu, J., Dong, P.D., and Panganiban, G. (2002). Limb type-specific regulation of bric a brac contributes to morphological diversity. *Development* 129, 695-704.
- Clapier, C.R., and Cairns, B.R. (2009). The biology of chromatin remodeling complexes. *Annu Rev Biochem* 78, 273-304.
- Cohen, S.M. (1990). Specification of limb development in the *Drosophila* embryo by positional cues from segmentation genes. *Nature* 343, 173-177.
- Cohen, S.M. (1993). Imaginal Disc Development. In *The development of Drosophila melanogaster*, B. Michael, and A.M. Arias, eds. (Cold Spring Harbor), p. 747.

- Cohen, S.M., Bronner, G., Kuttner, F., Jurgens, G., and Jackle, H. (1989). Distal-less encodes a homoeodomain protein required for limb development in *Drosophila*. *Nature* 338, 432-434.
- Corden, J.L., Cadena, D.L., Ahearn, J.M., and Dahmus, M.E. (1985). A unique structure at the carboxyl terminus of the largest subunit of eukaryotic RNA polymerase II. *Proc Natl Acad Sci U S A* 82, 7934-7938.
- Corona, D.F., and Tamkun, J.W. (2004). Multiple roles for ISWI in transcription, chromosome organization and DNA replication. *Biochim Biophys Acta* 1677, 113-119.
- Cosma, M.P., Tanaka, T., and Nasmyth, K. (1999). Ordered recruitment of transcription and chromatin remodeling factors to a cell cycle- and developmentally regulated promoter. *Cell* 97, 299-311.
- Couderc, J.L., Godt, D., Zollman, S., Chen, J., Li, M., Tiong, S., Cramton, S.E., Sahut-Barnola, I., and Laski, F.A. (2002). The bric à brac locus consists of two paralogous genes encoding BTB/POZ domain proteins and acts as a homeotic and morphogenetic regulator of imaginal development in *Drosophila*. *Development* 129, 2419-2433.
- Coulter, D.E., Swaykus, E.A., Beran-Koehn, M.A., Goldberg, D., Wieschaus, E., and Schedl, P. (1990). Molecular analysis of odd-skipped, a zinc finger encoding segmentation gene with a novel pair-rule expression pattern. *EMBO J* 9, 3795-3804.
- Courey, A.J., and Tjian, R. (1988). Analysis of Sp1 in vivo reveals multiple transcriptional domains, including a novel glutamine-rich activation motif. *Cell* 55, 887-898.
- Cramer, P. (2004). RNA polymerase II structure: from core to functional complexes. *Curr Opin Genet Dev* 14, 218-226.
- Cuddapah, S., Jothi, R., Schones, D.E., Roh, T.Y., Cui, K., and Zhao, K. (2009). Global analysis of the insulator binding protein CTCF in chromatin barrier regions reveals demarcation of active and repressive domains. *Genome Res* 19, 24-32.
- Cujec, T.P., Cho, H., Maldonado, E., Meyer, J., Reinberg, D., and Peterlin, B.M. (1997). The human immunodeficiency virus transactivator Tat interacts with the RNA polymerase II holoenzyme. *Mol Cell Biol* 17, 1817-1823.
- Davidson, E.H. (2001). *Genomic Regulatory Systems. Development and Evolution* (San Diego, CA: Academic Press).
- de Celis Ibeas, J.M., and Bray, S.J. (2003). Bowl is required downstream of Notch for elaboration of distal limb patterning. *Development* 130, 5943-5952.
- de Celis, J.F., Tyler, D.M., de Celis, J., and Bray, S.J. (1998). Notch signalling mediates segmentation of the *Drosophila* leg. *Development* 125, 4617-4626.
- Dear, T.N., and Rabbitts, T.H. (1994). A *Drosophila melanogaster* homologue of the T-cell oncogene HOX11 localises to a cluster of homeobox genes. *Gene* 141, 225-229.
- Diaz-Benjumea, F.J., Cohen, B., and Cohen, S.M. (1994). Cell interaction between compartments establishes the proximal-distal axis of *Drosophila* legs. *Nature* 372, 175-179.
- Dieci, G., and Sentenac, A. (2003). Detours and shortcuts to transcription reinitiation. *Trends Biochem Sci* 28, 202-209.
- Ding, N., Tomomori-Sato, C., Sato, S., Conaway, R.C., Conaway, J.W., and Boyer, T.G. (2009). MED19 and MED26 Are Synergistic Functional Targets of the RE1 Silencing Transcription Factor in Epigenetic Silencing of Neuronal Gene Expression. *J Biol Chem* 284, 2648-2656. Epub 2008 Dec 2642.

- Dong, P.D., Chu, J., and Panganiban, G. (2000). Coexpression of the homeobox genes *Distal-less* and *homothorax* determines *Drosophila* antennal identity. *Development* *127*, 209-216.
- Dong, P.D., Chu, J., and Panganiban, G. (2001). Proximodistal domain specification and interactions in developing *Drosophila* appendages. *Development* *128*, 2365-2372.
- Dong, P.D., Dicks, J.S., and Panganiban, G. (2002). *Distal-less* and *homothorax* regulate multiple targets to pattern the *Drosophila* antenna. *Development* *129*, 1967-1974.
- Duncan, D., Kiefel, P., and Duncan, I. (2010). Control of the spineless antennal enhancer: direct repression of antennal target genes by *Antennapedia*. *Dev Biol* *347*, 82-91.
- Duncan, D.M., Burgess, E.A., and Duncan, I. (1998). Control of distal antennal identity and tarsal development in *Drosophila* by *spineless-aristapedia*, a homolog of the mammalian dioxin receptor. *Genes Dev* *12*, 1290-1303.
- Déjardin, J., and Cavalli, G. (2005). Epigenetic inheritance of chromatin states mediated by Polycomb and trithorax group proteins in *Drosophila*. *Prog Mol Subcell Biol* *38*, 31-63.
- Eberl, D.F. (1999). Feeling the vibes: chordotonal mechanisms in insect hearing. *Curr Opin Neurobiol* *9*, 389-393.
- Emerald, B.S., Curtiss, J., Mlodzik, M., and Cohen, S.M. (2003). Distal antenna and distal antenna related encode nuclear proteins containing pipsqueak motifs involved in antenna development in *Drosophila*. *Development* *130*, 1171-1180.
- Emmons, R.B., Duncan, D., and Duncan, I. (2007). Regulation of the *Drosophila* distal antennal determinant *spineless*. *Dev Biol* *302*, 412-426.
- Estella, C., and Mann, R.S. (2008). Logic of *Wg* and *Dpp* induction of distal and medial fates in the *Drosophila* leg. *Development* *135*, 627-636.
- Estella, C., and Mann, R.S. (2010). Non-redundant selector and growth-promoting functions of two sister genes, *buttonhead* and *Sp1*, in *Drosophila* leg development. *PLoS Genet* *6*, e1001001.
- Estella, C., McKay, D.J., and Mann, R.S. (2008). Molecular integration of *wingless*, *decapentaplegic*, and autoregulatory inputs into *Distalless* during *Drosophila* leg development. *Dev Cell* *14*, 86-96.
- Estella, C., Rieckhof, G., Calleja, M. and Morata, G. (2003). The role of *buttonhead* and *Sp1* in the development of the ventral imaginal discs of *Drosophila*. *Development* *130*, 5929-5941.
- Estella, C., Voutev, R., and Mann, R.S. (2012). A dynamic network of morphogens and transcription factors patterns the fly leg. *Curr Top Dev Biol* *98*, 173-198.
- Estrada, B., and Sánchez-Herrero, E. (2001). The Hox gene *Abdominal-B* antagonizes appendage development in the genital disc of *Drosophila*. *Development* *128*, 331-339.
- F, C.R. (1998). *The Insects. Structure and Function* (Cambridge University Press.).
- Falvo, J.V., Thanos, D., and Maniatis, T. (1995). Reversal of intrinsic DNA bends in the IFN beta gene enhancer by transcription factors and the architectural protein HMG I(Y). *Cell* *83*, 1101-1111.
- Fileiko, N.A., Kolar, C., West, J.T., Smith, S.A., Hassan, Y.I., Borgstahl, G.E., Zempleni, J., and Lyubchenko, Y.L. (2011). The role of histone H4 biotinylation in the structure of nucleosomes. *PLoS One* *6*, e16299.

- Fisher, C.L., and Fisher, A.G. (2011). Chromatin states in pluripotent, differentiated, and reprogrammed cells. *Curr Opin Genet Dev* 21, 140-146.
- Flemming, W., *Zellsubstanz*, K., FCW, Z., and Vogel, L. (1882).
- Fortini, M.E. (2009). Notch signaling: the core pathway and its posttranslational regulation. *Dev Cell* 16, 633-647.
- Fristrom, D., and Fristrom, J.W. (1993). The metamorphic development of the adult epidermis. In *The Development of Drosophila Melanogaster*, M. Bate, and A. Martinez-Arias, eds. (New York.: Cold Spring Harbor), pp. 843–897.
- Fuda, N.J., Ardehali, M.B., and Lis, J.T. (2009). Defining mechanisms that regulate RNA polymerase II transcription in vivo. *Nature* 461, 186-192.
- Fuse, N., Hirose, S., and Hayashi, S. (1996). Determination of wing cell fate by the escargot and snail genes in *Drosophila*. *Development* 122, 1059-1067.
- Galindo, M.I., Bishop, S.A., Greig, S., and Couso, J.P. (2002). Leg patterning driven by proximal-distal interactions and EGFR signaling. *Science* 297, 256-259.
- Galindo, M.I., and Couso, J.P. (2000). Intercalation of cell fates during tarsal development in *Drosophila*. *Bioessays* 22, 777-780.
- Galindo, M.I., Pueyo, J.I., Fouix, S., Bishop, S.A., and Couso, J.P. (2007). Peptides encoded by short ORFs control development and define a new eukaryotic gene family. *PLoS Biol* 5, e106.
- Gamble, M.J., and Kraus, W.L. (2010). Multiple facets of the unique histone variant macroH2A: from genomics to cell biology. *Cell Cycle* 9, 2568-2574.
- Gasperowicz, M., and Otto, F. (2005). Mammalian Groucho homologs: redundancy or specificity? *J Cell Biochem* 95, 670-687.
- Gaszner, M., and Felsenfeld, G. (2006). Insulators: exploiting transcriptional and epigenetic mechanisms. *Nat Rev Genet* 7, 703-713.
- Gerber, H.P., Seipel, K., Georgiev, O., Höfferer, M., Hug, M., Rusconi, S., and Schaffner, W. (1994). Transcriptional activation modulated by homopolymeric glutamine and proline stretches. *Science* 263, 808-811.
- Ghisletti, S., Barozzi, I., Mietton, F., Polletti, S., De Santa, F., Venturini, E., Gregory, L., Lonie, L., Chew, A., Wei, C.L., *et al.* (2010). Identification and characterization of enhancers controlling the inflammatory gene expression program in macrophages. *Immunity* 32, 317-328.
- Gindhart, J.G., King, A.N., and Kaufman, T.C. (1995). Characterization of the cis-regulatory region of the *Drosophila* homeotic gene *Sex combs reduced*. *Genetics* 139, 781-795.
- Giorgianni, M.W., and Mann, R.S. (2011). Establishment of medial fates along the proximodistal axis of the *Drosophila* leg through direct activation of *dachshund* by *Distalless*. *Dev Cell* 20, 455-468.
- Godt, D., Couderc, J.L., Cramton, S.E., and Laski, F.A. (1993). Pattern formation in the limbs of *Drosophila*: *bric a brac* is expressed in both a gradient and a wave-like pattern and is required for specification and proper segmentation of the tarsus. *Development* 119, 799-812.

- Godt, D., and Laski, F.A. (1995). Mechanisms of cell rearrangement and cell recruitment in *Drosophila* ovary morphogenesis and the requirement of *bric à brac*. *Development* *121*, 173-187.
- Golembo, M., Raz, E., and Shilo, B.Z. (1996). The *Drosophila* embryonic midline is the site of Spitz processing, and induces activation of the EGF receptor in the ventral ectoderm. *Development* *122*, 3363-3370.
- Golembo, M., Yarnitzky, T., Volk, T., and Shilo, B.Z. (1999). Vein expression is induced by the EGF receptor pathway to provide a positive feedback loop in patterning the *Drosophila* embryonic ventral ectoderm. *Genes Dev* *13*, 158-162.
- Goll, M.G., and Bestor, T.H. (2005). Eukaryotic cytosine methyltransferases. *Annu Rev Biochem* *74*, 481-514.
- Gompel, N., Prud'homme, B., Wittkopp, P.J., Kassner, V.A., and Carroll, S.B. (2005). Chance caught on the wing: cis-regulatory evolution and the origin of pigment patterns in *Drosophila*. *Nature* *433*, 481-487.
- Gonzalez, F., Duboule, D., and Spitz, F. (2007). Transgenic analysis of *Hoxd* gene regulation during digit development. *Dev Biol* *306*, 847-859.
- Gonzalez-Crespo, S., Abu-Shaar, M., Torres, M., Martinez, A.C., Mann, R.S., and Morata, G. (1998). Antagonism between extradenticle function and Hedgehog signalling in the developing limb. *Nature* *394*, 196-200.
- González-Crespo, S., and Morata, G. (1995). Control of *Drosophila* adult pattern by extradenticle. *Development* *121*, 2117-2125.
- Gorfinkiel, N., Morata, G., and Guerrero, I. (1997). The homeobox gene *Distal-less* induces ventral appendage development in *Drosophila*. *Genes Dev* *11*, 2259-2271.
- Goto, S., and Hayashi, S. (1997). Specification of the embryonic limb primordium by graded activity of *Decapentaplegic*. *Development* *124*, 125-132.
- Gowher, H., Leismann, O., and Jeltsch, A. (2000). DNA of *Drosophila melanogaster* contains 5-methylcytosine. *EMBO J* *19*, 6918-6923.
- Green, R.B., Hatini, V., Johansen, K.A., Liu, X.J., and Lengyel, J.A. (2002). *Drumstick* is a zinc finger protein that antagonizes *Lines* to control patterning and morphogenesis of the *Drosophila* hindgut. *Development* *129*, 3645-3656.
- Greenberg, L., and Hatini, V. (2009). Essential roles for *lines* in mediating leg and antennal proximodistal patterning and generating a stable Notch signaling interface at segment borders. *Dev Biol* *330*, 93-104.
- Gromöller, A., and Lehming, N. (2000). *Srb7p* is a physical and physiological target of *Tup1p*. *EMBO J* *19*, 6845-6852.
- Guenther, M.G., Levine, S.S., Boyer, L.A., Jaenisch, R., and Young, R.A. (2007). A chromatin landmark and transcription initiation at most promoters in human cells. *Cell* *130*, 77-88.
- Guglielmi, B., van Berkum, N.L., Klapholz, B., Bijma, T., Boube, M., Boschiero, C., Bourbon, H.M., Holstege, F.C., and Werner, M. (2004). A high resolution protein interaction map of the yeast Mediator complex. *Nucleic Acids Res* *32*, 5379-5391.

- Handoko, L., Xu, H., Li, G., Ngan, C.Y., Chew, E., Schnapp, M., Lee, C.W., Ye, C., Ping, J.L., Mulawadi, F., *et al.* (2011). CTCF-mediated functional chromatin interactome in pluripotent cells. *Nat Genet* 43, 630-638.
- Hao, I., Green, R.B., Dunaevsky, O., Lengyel, J.A., and Rauskolb, C. (2003). The odd-skipped family of zinc finger genes promotes *Drosophila* leg segmentation. *Dev Biol* 263, 282-295.
- Hart, M.C., Wang, L., and Coulter, D.E. (1996). Comparison of the structure and expression of odd-skipped and two related genes that encode a new family of zinc finger proteins in *Drosophila*. *Genetics* 144, 171-182.
- Hatini, V., Green, R.B., Lengyel, J.A., Bray, S.J., and Dinardo, S. (2005). The Drumstick/Lines/Bowl regulatory pathway links antagonistic Hedgehog and Wingless signaling inputs to epidermal cell differentiation. *Genes Dev* 19, 709-718.
- Haynie, J.L., and Bryant, P.J. (1986). Development of the eye-antenna imaginal disc and morphogenesis of the adult head in *Drosophila melanogaster*. *J Exp Zool* 237, 293-308.
- Heinzel, T., Lavinsky, R.M., Mullen, T.M., Söderstrom, M., Laherty, C.D., Torchia, J., Yang, W.M., Brard, G., Ngo, S.D., Davie, J.R., *et al.* (1997). A complex containing N-CoR, mSin3 and histone deacetylase mediates transcriptional repression. *Nature* 387, 43-48.
- Heitzler, P., Bourouis, M., Ruel, L., Carteret, C., and Simpson, P. (1996). Genes of the Enhancer of split and achaete-scute complexes are required for a regulatory loop between Notch and Delta during lateral signalling in *Drosophila*. *Development* 122, 161-171.
- Hens, K., Feuz, J.D., Isakova, A., Iagovitina, A., Massouras, A., Bryois, J., Callaerts, P., Celniker, S.E., and Deplancke, B. (2011). Automated protein-DNA interaction screening of *Drosophila* regulatory elements. *Nat Methods* 8, 1065-1070.
- Higashijima, S., Kojima, T., Michiue, T., Ishimaru, S., Emori, Y., and Saigo, K. (1992). Dual Bar homeo box genes of *Drosophila* required in two photoreceptor cells, R1 and R6, and primary pigment cells for normal eye development. *Genes Dev* 6, 50-60.
- Hirose, Y., and Ohkuma, Y. (2007). Phosphorylation of the C-terminal domain of RNA polymerase II plays central roles in the integrated events of eucaryotic gene expression. *J Biochem* 141, 601-608.
- Hong, J.W., Hendrix, D.A., and Levine, M.S. (2008). Shadow enhancers as a source of evolutionary novelty. *Science* 321, 1314.
- Huang, Y., Li, W., Yao, X., Lin, Q.J., Yin, J.W., Liang, Y., Heiner, M., Tian, B., Hui, J., and Wang, G. (2012). Mediator complex regulates alternative mRNA processing via the MED23 subunit. *Mol Cell* 45, 459-469. Epub 2012 Jan 2019.
- Hung, M.S., Karthikeyan, N., Huang, B., Koo, H.C., Kiger, J., and Shen, C.J. (1999). *Drosophila* proteins related to vertebrate DNA (5-cytosine) methyltransferases. *Proc Natl Acad Sci U S A* 96, 11940-11945.
- Huppert, S.S., Jacobsen, T.L., and Muskavitch, M.A. (1997). Feedback regulation is central to Delta-Notch signalling required for *Drosophila* wing vein morphogenesis. *Development* 124, 3283-3291.
- Irvine, K.D., and Wieschaus, E. (1994). fringe, a Boundary-specific signaling molecule, mediates interactions between dorsal and ventral cells during *Drosophila* wing development. *Cell* 79, 595-606.

- Iwasaki, O., Tanaka, A., Tanizawa, H., Grewal, S.I., and Noma, K. (2010). Centromeric localization of dispersed Pol III genes in fission yeast. *Mol Biol Cell* 21, 254-265.
- Jarman, A.P., Sun, Y., Jan, L.Y., and Jan, Y.N. (1995). Role of the proneural gene, atonal, in formation of *Drosophila* chordotonal organs and photoreceptors. *Development* 121, 2019-2030.
- Jiang, J., and Struhl, G. (1996). Complementary and mutually exclusive activities of decapentaplegic and wingless organize axial patterning during *Drosophila* leg development. *Cell* 86, 401-409.
- Jishage, M., Malik, S., Wagner, U., Uberheide, B., Ishihama, Y., Hu, X., Chait, B.T., Gnatt, A., Ren, B., and Roeder, R.G. (2012). Transcriptional regulation by Pol II(G) involving mediator and competitive interactions of Gdown1 and TFIIF with Pol II. *Mol Cell* 45, 51-63.
- Johnston, L.A., and Schubiger, G. (1996). Ectopic expression of wingless in imaginal discs interferes with decapentaplegic expression and alters cell determination. *Development* 122, 3519-3529.
- Joulià, L., Deutsch, J., Bourbon, H.M., and Cribbs, D.L. (2006). The specification of a highly derived arthropod appendage, the *Drosophila* labial palps, requires the joint action of selectors and signaling pathways. *Dev Genes Evol* 216, 431-442.
- Junion, G., Bataillé, L., Jagla, T., Da Ponte, J.P., Tapin, R., and Jagla, K. (2007). Genome-wide view of cell fate specification: ladybird acts at multiple levels during diversification of muscle and heart precursors. *Genes Dev* 21, 3163-3180.
- Junion, G., Spivakov, M., Girardot, C., Braun, M., Gustafson, E.H., Birney, E., and Furlong, E.E. (2012). A transcription factor collective defines cardiac cell fate and reflects lineage history. *Cell* 148, 473-486.
- Juven-Gershon, T., Hsu, J.Y., and Kadonaga, J.T. (2008a). Caudal, a key developmental regulator, is a DPE-specific transcriptional factor. *Genes Dev* 22, 2823-2830.
- Juven-Gershon, T., Hsu, J.Y., Theisen, J.W., and Kadonaga, J.T. (2008b). The RNA polymerase II core promoter - the gateway to transcription. *Curr Opin Cell Biol* 20, 253-259.
- Kaufman, T.C., Seeger, M.A., and Olsen, G. (1990). Molecular and genetic organization of the antennapedia gene complex of *Drosophila melanogaster*. *Adv Genet* 27, 309-362.
- Kenyon, K.L., Ranade, S.S., Curtiss, J., Mlodzik, M., and Pignoni, F. (2003). Coordinating proliferation and tissue specification to promote regional identity in the *Drosophila* head. *Dev Cell* 5, 403-414.
- Keogh, M.C., Podolny, V., and Buratowski, S. (2003). Bur1 kinase is required for efficient transcription elongation by RNA polymerase II. *Mol Cell Biol* 23, 7005-7018.
- Kerber, B., Monge, I., Mueller, M., Mitchell, P.J., and Cohen, S.M. (2001). The AP-2 transcription factor is required for joint formation and cell survival in *Drosophila* leg development. *Development* 128, 1231-1238.
- Kim, J., Sebring, A., Esch, J.J., Kraus, M.E., Vorwerk, K., Magee, J., and Carroll, S.B. (1996). Integration of positional signals and regulation of wing formation and identity by *Drosophila* vestigial gene. *Nature* 382, 133-138.
- Kim, J.H., Cho, E.J., Kim, S.T., and Youn, H.D. (2005). CtBP represses p300-mediated transcriptional activation by direct association with its bromodomain. *Nat Struct Mol Biol* 12, 423-428.

- Kim, T.K., Ebright, R.H., and Reinberg, D. (2000). Mechanism of ATP-dependent promoter melting by transcription factor IIIH. *Science* 288, 1418-1422.
- Kim, Y.J., Bjorklund, S., Li, Y., Sayre, M.H., and Kornberg, R.D. (1994). A multiprotein mediator of transcriptional activation and its interaction with the C-terminal repeat domain of RNA polymerase II. *Cell* 77, 599-608.
- Klymenko, T., and Müller, J. (2004). The histone methyltransferases Trithorax and Ash1 prevent transcriptional silencing by Polycomb group proteins. *EMBO Rep* 5, 373-377.
- Kojima, T. (2004). The mechanism of *Drosophila* leg development along the proximodistal axis. *Dev Growth Differ* 46, 115-129.
- Kojima, T., Sato, M., and Saigo, K. (2000). Formation and specification of distal leg segments in *Drosophila* by dual Bar homeobox genes, BarH1 and BarH2. *Development* 127, 769-778.
- Kojima, T., Tsuji, T., and Saigo, K. (2005). A concerted action of a paired-type homeobox gene, *aristales*, and a homolog of *Hox11/tlx* homeobox gene, *clawless*, is essential for the distal tip development of the *Drosophila* leg. *Dev Biol* 279, 434-445.
- Kondo, T., Hashimoto, Y., Kato, K., Inagaki, S., Hayashi, S., and Kageyama, Y. (2007). Small peptide regulators of actin-based cell morphogenesis encoded by a polycistronic mRNA. *Nat Cell Biol* 9, 660-665.
- Kondo, T., Plaza, S., Zanet, J., Benrabah, E., Valenti, P., Hashimoto, Y., Kobayashi, S., Payre, F., and Kageyama, Y. (2010). Small peptides switch the transcriptional activity of *Shavenbaby* during *Drosophila* embryogenesis. *Science* 329, 336-339.
- Kopp, A., Duncan, I., Godt, D., and Carroll, S.B. (2000). Genetic control and evolution of sexually dimorphic characters in *Drosophila*. *Nature* 408, 553-559.
- Kornberg, R.D. (2005). Mediator and the mechanism of transcriptional activation. *Trends Biochem Sci* 30, 235-239.
- Kothapalli, N., Camporeale, G., Kueh, A., Chew, Y.C., Oommen, A.M., Griffin, J.B., and Zempleni, J. (2005). Biological functions of biotinylated histones. *J Nutr Biochem* 16, 446-448.
- Kouzarides, T. (2002). Histone methylation in transcriptional control. *Curr Opin Genet Dev* 12, 198-209.
- Kouzarides, T. (2007). Chromatin modifications and their function. *Cell* 128, 693-705.
- Kubota, K., Goto, S., Eto, K., and Hayashi, S. (2000). EGF receptor attenuates Dpp signaling and helps to distinguish the wing and leg cell fates in *Drosophila*. *Development* 127, 3769-3776.
- Kubota, K., Goto, S., and Hayashi, S. (2003). The role of Wg signaling in the patterning of embryonic leg primordium in *Drosophila*. *Dev Biol* 257, 117-126.
- Kuehner, J.N., Pearson, E.L., and Moore, C. (2011). Unravelling the means to an end: RNA polymerase II transcription termination. *Nat Rev Mol Cell Biol* 12, 283-294.
- Kulkarni, M.M., and Arnosti, D.N. (2003). Information display by transcriptional enhancers. *Development* 130, 6569-6575.
- Kurant, E., Pai, C.Y., Sharf, R., Halachmi, N., Sun, Y.H., and Salzberg, A. (1998). *Dorsotonal/homothorax*, the *Drosophila* homologue of *meis1*, interacts with *extradenticle* in patterning of the embryonic PNS. *Development* 125, 1037-1048.

- Lau, P.N., and Cheung, P. (2011). Histone code pathway involving H3 S28 phosphorylation and K27 acetylation activates transcription and antagonizes polycomb silencing. *Proc Natl Acad Sci U S A* *108*, 2801-2806.
- Lecourtois, M., and Schweisguth, F. (1995). The neurogenic suppressor of hairless DNA-binding protein mediates the transcriptional activation of the enhancer of split complex genes triggered by Notch signaling. *Genes Dev* *9*, 2598-2608.
- Lecuit, T., and Cohen, S.M. (1997). Proximal-distal axis formation in the *Drosophila* leg. *Nature* *388*, 139-145.
- Lee, H., Kraus, K.W., Wolfner, M.F., and Lis, J.T. (1992). DNA sequence requirements for generating paused polymerase at the start of hsp70. *Genes Dev* *6*, 284-295.
- Lee, T.I., and Young, R.A. (2000). Transcription of Eukaryotic Protein-Coding Genes. *Annu Rev Genet* *34*, 77-137.
- Lehmann, M., Siegmund, T., Lintermann, K.G., and Korge, G. (1998). The pipsqueak protein of *Drosophila melanogaster* binds to GAGA sequences through a novel DNA-binding domain. *J Biol Chem* *273*, 28504-28509.
- Lenhard, B., Sandelin, A., and Carninci, P. (2012). Metazoan promoters: emerging characteristics and insights into transcriptional regulation. *Nat Rev Genet* *13*, 233-245.
- Levine, M. (2010). Transcriptional enhancers in animal development and evolution. *Curr Biol* *20*, R754-763.
- Lilly, B., O'Keefe, D.D., Thomas, J.B., and Botas, J. (1999). The LIM homeodomain protein dLim1 defines a subclass of neurons within the embryonic ventral nerve cord of *Drosophila*. *Mech Dev* *88*, 195-205.
- Lin, Y.S., Carey, M., Ptashne, M., and Green, M.R. (1990). How different eukaryotic transcriptional activators can cooperate promiscuously. *Nature* *345*, 359-361.
- Litt, M., Qiu, Y., and Huang, S. (2009). Histone arginine methylations: their roles in chromatin dynamics and transcriptional regulation. *Biosci Rep* *29*, 131-141.
- Liu, Z., and Karmarkar, V. (2008). Groucho/Tup1 family co-repressors in plant development. *Trends Plant Sci* *13*, 137-144.
- Lours, C., Bardot, O., Godt, D., Laski, F.A., and Couderc, J.L. (2003). The *Drosophila melanogaster* BTB proteins bric à brac bind DNA through a composite DNA binding domain containing a pipsqueak and an AT-Hook motif. *Nucleic Acids Res* *31*, 5389-5398.
- Luger, K., Mäder, A.W., Richmond, R.K., Sargent, D.F., and Richmond, T.J. (1997). Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* *389*, 251-260.
- Lyko, F., Ramsahoye, B.H., and Jaenisch, R. (2000). DNA methylation in *Drosophila melanogaster*. *Nature* *408*, 538-540.
- Malik, S., and Roeder, R.G. (2005). Dynamic regulation of pol II transcription by the mammalian Mediator complex. *Trends Biochem Sci* *30*, 256-263.
- Malik, S., and Roeder, R.G. (2010). The metazoan Mediator co-activator complex as an integrative hub for transcriptional regulation. *Nat Rev Genet* *11*, 761-772. Epub 2010 Oct 2013.

- Malone, C.S., Miner, M.D., Doerr, J.R., Jackson, J.P., Jacobsen, S.E., Wall, R., and Teitell, M. (2001). CmC(A/T)GG DNA methylation in mature B cell lymphoma gene silencing. *Proc Natl Acad Sci U S A* 98, 10404-10409.
- Mardon, G., Solomon, N.M., and Rubin, G.M. (1994). dachshund encodes a nuclear protein required for normal eye and leg development in *Drosophila*. *Development* 120, 3473-3486.
- Marfella, C.G., and Imbalzano, A.N. (2007). The Chd family of chromatin remodelers. *Mutat Res* 618, 30-40.
- Marshall, N.F., and Price, D.H. (1995). Purification of P-TEFb, a transcription factor required for the transition into productive elongation. *J Biol Chem* 270, 12335-12338.
- Martin, C., and Zhang, Y. (2005). The diverse functions of histone lysine methylation. *Nat Rev Mol Cell Biol* 6, 838-849.
- McKay, D.J., Estella, C., and Mann, R.S. (2009). The origins of the *Drosophila* leg revealed by the cis-regulatory architecture of the *Distalless* gene. *Development* 136, 61-71.
- Meinhardt, H. (1983). Cell determination boundaries as organizing regions for secondary embryonic fields. *Dev Biol* 96, 375-385.
- Merika, M., and Thanos, D. (2001). Enhanceosomes. *Curr Opin Genet Dev* 11, 205-208.
- Merika, M., Williams, A.J., Chen, G., Collins, T., and Thanos, D. (1998). Recruitment of CBP/p300 by the IFN beta enhanceosome is required for synergistic activation of transcription. *Mol Cell* 1, 277-287.
- Metzger, E., Yin, N., Wissmann, M., Kunowska, N., Fischer, K., Friedrichs, N., Patnaik, D., Higgins, J.M., Potier, N., Scheidtmann, K.H., *et al.* (2008). Phosphorylation of histone H3 at threonine 11 establishes a novel chromatin mark for transcriptional regulation. *Nat Cell Biol* 10, 53-60.
- Miller, A.C., Lyons, E.L., and Herman, T.G. (2009). cis-Inhibition of Notch by endogenous Delta biases the outcome of lateral inhibition. *Curr Biol* 19, 1378-1383.
- Mirth, C., and Akam, M. (2002). Joint development in the *Drosophila* leg: cell movements and cell populations. *Dev Biol* 246, 391-406.
- Miyazono, K., Zhi, Y., Takamura, Y., Nagata, K., Saigo, K., Kojima, T., and Tanokura, M. (2010). Cooperative DNA-binding and sequence-recognition mechanism of *aristaless* and *clawless*. *EMBO J* 29, 1613-1623.
- Mohrmann, L., and Verrijzer, C.P. (2005). Composition and functional specificity of SWI2/SNF2 class chromatin remodeling complexes. *Biochim Biophys Acta* 1681, 59-73.
- Monge, I., Krishnamurthy, R., Sims, D., Hirth, F., Spengler, M., Kammermeier, L., Reichert, H., and Mitchell, P.J. (2001). *Drosophila* transcription factor AP-2 in proboscis, leg and brain central complex development. *Development* 128, 1239-1252.
- Montavon, T., Soshnikova, N., Mascrez, B., Joye, E., Thevenet, L., Splinter, E., de Laat, W., Spitz, F., and Duboule, D. (2011). A regulatory archipelago controls Hox genes transcription in digits. *Cell* 147, 1132-1145.
- Morata, G. (2001). How *Drosophila* appendages develop. *Nat Rev Mol Cell Biol* 2, 89-97.
- Morata, G., and Lawrence, P.A. (1979). Development of the eye-antenna imaginal disc of *Drosophila*. *Dev Biol* 70, 355-371.

- Morcillo, P., Rosen, C., Baylies, M.K., and Dorsett, D. (1997). Chip, a widely expressed chromosomal protein required for segmentation and activity of a remote wing margin enhancer in *Drosophila*. *Genes Dev* *11*, 2729-2740.
- Morrison, A.J., and Shen, X. (2009). Chromatin remodelling beyond transcription: the INO80 and SWR1 complexes. *Nat Rev Mol Cell Biol* *10*, 373-384.
- Muse, G.W., Gilchrist, D.A., Nechaev, S., Shah, R., Parker, J.S., Grissom, S.F., Zeitlinger, J., and Adelman, K. (2007). RNA polymerase is poised for activation across the genome. *Nat Genet* *39*, 1507-1511. Epub 2007 Nov 1511.
- Myers, L.C., and Kornberg, R.D. (2000). Mediator of transcriptional regulation. *Annu Rev Biochem* *69*, 729-749.
- Nagy, L., Kao, H.Y., Chakravarti, D., Lin, R.J., Hassig, C.A., Ayer, D.E., Schreiber, S.L., and Evans, R.M. (1997). Nuclear receptor repression mediated by a complex containing SMRT, mSin3A, and histone deacetylase. *Cell* *89*, 373-380.
- Nakanishi, S., Sanderson, B.W., Delventhal, K.M., Bradford, W.D., Staehling-Hampton, K., and Shilatifard, A. (2008). A comprehensive library of histone mutants identifies nucleosomal residues required for H3K4 methylation. *Nat Struct Mol Biol* *15*, 881-888.
- Nambu, J.R., Lewis, J.O., Wharton, K.A., and Crews, S.T. (1991). The *Drosophila* single-minded gene encodes a helix-loop-helix protein that acts as a master regulator of CNS midline development. *Cell* *67*, 1157-1167.
- Natori, K., Tajiri, R., Furukawa, S., and Kojima, T. (2012). Progressive tarsal patterning in the *Drosophila* by temporally dynamic regulation of transcription factor genes. *Dev Biol* *361*, 450-462.
- Nelson, C.E., Hersh, B.M., and Carroll, S.B. (2004). The regulatory content of intergenic DNA shapes genome architecture. *Genome Biol* *5*, R25.
- Ng, C.S., and Kopp, A. (2008). Sex combs are important for male mating success in *Drosophila melanogaster*. *Behav Genet* *38*, 195-201.
- Noma, K., Cam, H.P., Marais, R.J., and Grewal, S.I. (2006). A role for TFIIC transcription factor complex in genome organization. *Cell* *125*, 859-872.
- Noyes, M.B., Christensen, R.G., Wakabayashi, A., Stormo, G.D., Brodsky, M.H., and Wolfe, S.A. (2008). Analysis of homeodomain specificities allows the family-wide prediction of preferred recognition sites. *Cell* *133*, 1277-1289.
- Oehler, S., Alberti, S., and Müller-Hill, B. (2006). Induction of the lac promoter in the absence of DNA loops and the stoichiometry of induction. *Nucleic Acids Res* *34*, 606-612.
- Ohtsuki, S., Levine, M., and Cai, H.N. (1998). Different core promoters possess distinct regulatory activities in the *Drosophila* embryo. *Genes Dev* *12*, 547-556.
- Okada, M., and Hirose, S. (1998). Chromatin remodeling mediated by *Drosophila* GAGA factor and ISWI activates fushi tarazu gene transcription in vitro. *Mol Cell Biol* *18*, 2455-2461.
- Orsi, G.A., Couble, P., and Loppin, B. (2009). Epigenetic and replacement roles of histone variant H3.3 in reproduction and development. *Int J Dev Biol* *53*, 231-243.
- Pai, C.Y., Kuo, T.S., Jaw, T.J., Kurant, E., Chen, C.T., Bessarab, D.A., Salzberg, A., and Sun, Y.H. (1998). The Homothorax homeoprotein activates the nuclear localization of another

- homeoprotein, extradenticle, and suppresses eye development in *Drosophila*. *Genes Dev* *12*, 435-446.
- Panganiban, G., Irvine, S.M., Lowe, C., Roehl, H., Corley, L.S., Sherbon, B., Grenier, J.K., Fallon, J.F., Kimble, J., Walker, M., *et al.* (1997). The origin and evolution of animal appendages. *Proc Natl Acad Sci U S A* *94*, 5162-5166.
- Panganiban, G., and Rubenstein, J.L. (2002). Developmental functions of the Distal-less/Dlx homeobox genes. *Development* *129*, 4371-4386.
- Panin, V.M., Papayannopoulos, V., Wilson, R., and Irvine, K.D. (1997). Fringe modulates Notch-ligand interactions. *Nature* *387*, 908-912.
- Panne, D., Maniatis, T., and Harrison, S.C. (2007). An atomic model of the interferon-beta enhanceosome. *Cell* *129*, 1111-1123.
- Parelho, V., Hadjur, S., Spivakov, M., Leleu, M., Sauer, S., Gregson, H.C., Jarmuz, A., Canzonetta, C., Webster, Z., Nesterova, T., *et al.* (2008). Cohesins functionally associate with CTCF on mammalian chromosome arms. *Cell* *132*, 422-433.
- Peng, J., Marshall, N.F., and Price, D.H. (1998). Identification of a cyclin subunit required for the function of *Drosophila* P-TEFb. *J Biol Chem* *273*, 13855-13860.
- Penton, A., and Hoffmann, F.M. (1996). Decapentaplegic restricts the domain of wingless during *Drosophila* limb patterning. *Nature* *382*, 162-164.
- Perales, R., and Bentley, D. (2009). "Cotranscriptionality": the transcription elongation complex as a nexus for nuclear transactions. *Mol Cell* *36*, 178-191.
- Perissi, V., Jepsen, K., Glass, C.K., and Rosenfeld, M.G. (2010). Deconstructing repression: evolving models of co-repressor action. *Nat Rev Genet* *11*, 109-123.
- Perry, M.W., Boettiger, A.N., Bothma, J.P., and Levine, M. (2010). Shadow enhancers foster robustness of *Drosophila* gastrulation. *Curr Biol* *20*, 1562-1567.
- Perry, M.W., Boettiger, A.N., and Levine, M. (2011). Multiple enhancers ensure precision of gap gene-expression patterns in the *Drosophila* embryo. *Proc Natl Acad Sci U S A* *108*, 13570-13575.
- Petrykowska, H.M., Vockley, C.M., and Elnitski, L. (2008). Detection and characterization of silencers and enhancer-blockers in the greater CFTR locus. *Genome Res* *18*, 1238-1246.
- Phillips, J.E., and Corces, V.G. (2009). CTCF: master weaver of the genome. *Cell* *137*, 1194-1211.
- Postlethwait, J.H. (1978). *Clonal analysis of drosophila cuticular pattern* (London: Academic press).
- Postlethwait, J.H., and Schneiderman, H.A. (1971). Pattern formation and determination in the antenna of the homoeotic mutant *Antennapedia* of *Drosophila melanogaster*. *Dev Biol* *25*, 606-640.
- Prazak, L., Fujioka, M., and Gergen, J.P. (2010). Non-additive interactions involving two distinct elements mediate sloppy-paired regulation by pair-rule transcription factors. *Dev Biol* *344*, 1048-1059.
- Prud'homme, B., Gompel, N., Rokas, A., Kassner, V.A., Williams, T.M., Yeh, S.D., True, J.R., and Carroll, S.B. (2006). Repeated morphological evolution through cis-regulatory changes in a pleiotropic gene. *Nature* *440*, 1050-1053.

- Pueyo, J.I., and Couso, J.P. (2004). Chip-mediated partnerships of the homeodomain proteins Bar and Aristaless with the LIM-HOM proteins Apterous and Lim1 regulate distal leg development. *Development* *131*, 3107-3120.
- Pueyo, J.I., and Couso, J.P. (2008). The 11-aminoacid long Tarsal-less peptides trigger a cell signal in *Drosophila* leg development. *Dev Biol* *324*, 192-201.
- Pueyo, J.I., and Couso, J.P. (2011). Tarsal-less peptides control Notch signalling through the Shavenbaby transcription factor. *Dev Biol* *355*, 183-193.
- Pueyo, J.I., Galindo, M.I., Bishop, S.A., and Couso, J.P. (2000). Proximal-distal leg development in *Drosophila* requires the apterous gene and the Lim1 homologue dlim1. *Development* *127*, 5391-5402.
- Rahl, P.B., Lin, C.Y., Seila, A.C., Flynn, R.A., McCuine, S., Burge, C.B., Sharp, P.A., and Young, R.A. (2010). c-Myc regulates transcriptional pause release. *Cell* *141*, 432-445.
- Ramsahoye, B.H., Biniszkiwicz, D., Lyko, F., Clark, V., Bird, A.P., and Jaenisch, R. (2000). Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a. *Proc Natl Acad Sci U S A* *97*, 5237-5242.
- Randsholt, N.B., and Santamaria, P. (2008). How *Drosophila* change their combs: the Hox gene *Sex* combs reduced and sex comb variation among *Sophophora* species. *Evol Dev* *10*, 121-133.
- Rauskolb, C., and Irvine, K.D. (1999). Notch-mediated segmentation and growth control of the *Drosophila* leg. *Dev Biol* *210*, 339-350.
- Reeves, R., and Nissen, M.S. (1990). The A.T-DNA-binding domain of mammalian high mobility group I chromosomal proteins. A novel peptide motif for recognizing DNA structure. *J Biol Chem* *265*, 8573-8582.
- Renfranz, P.J., and Benzer, S. (1989). Monoclonal antibody probes discriminate early and late mutant defects in development of the *Drosophila* retina. *Dev Biol* *136*, 411-429.
- Repiso, A., Bergantiños, C., Corominas, M., and Serras, F. (2011). Tissue repair and regeneration in *Drosophila* imaginal discs. *Dev Growth Differ* *53*, 177-185.
- Rieckhof, G.E., Casares, F., Ryoo, H.D., Abu-Shaar, M., and Mann, R.S. (1997). Nuclear translocation of extradenticle requires homothorax, which encodes an extradenticle-related homeodomain protein. *Cell* *91*, 171-183.
- Rubio, E.D., Reiss, D.J., Welch, P.L., Distche, C.M., Filippova, G.N., Baliga, N.S., Aebersold, R., Ranish, J.A., and Krumm, A. (2008). CTCF physically links cohesin to chromatin. *Proc Natl Acad Sci U S A* *105*, 8309-8314.
- Sanicola, M., Sekelsky, J., Elson, S., and Gelbart, W.M. (1995). Drawing a stripe in *Drosophila* imaginal disks: negative regulation of decapentaplegic and patched expression by engrailed. *Genetics* *139*, 745-756.
- Sarma, K., and Reinberg, D. (2005). Histone variants meet their match. *Nat Rev Mol Cell Biol* *6*, 139-149.
- Sato, S., Tomomori-Sato, C., Parmely, T.J., Florens, L., Zybaylov, B., Swanson, S.K., Banks, C.A., Jin, J., Cai, Y., Washburn, M.P., *et al.* (2004). A set of consensus mammalian mediator subunits identified by multidimensional protein identification technology. *Mol Cell* *14*, 685-691.

- Sayeed, O., and Benzer, S. (1996). Behavioral genetics of thermosensation and hygrosensation in *Drosophila*. *Proc Natl Acad Sci U S A* *93*, 6079-6084.
- Schmidt, J.V., and Bradfield, C.A. (1996). Ah receptor signaling pathways. *Annu Rev Cell Dev Biol* *12*, 55-89.
- Schnepf, B., Grumblin, G., Donaldson, T., and Simcox, A. (1996). Vein is a novel component in the *Drosophila* epidermal growth factor receptor pathway with similarity to the neuregulins. *Genes Dev* *10*, 2302-2313.
- Schwabish, M.A., and Struhl, K. (2007). The Swi/Snf complex is important for histone eviction during transcriptional activation and RNA polymerase II elongation in vivo. *Mol Cell Biol* *27*, 6987-6995.
- Schweitzer, R., and Shilo, B.Z. (1997). A thousand and one roles for the *Drosophila* EGF receptor. *Trends Genet* *13*, 191-196.
- Senger, K., Armstrong, G.W., Rowell, W.J., Kwan, J.M., Markstein, M., and Levine, M. (2004). Immunity regulatory DNAs share common organizational features in *Drosophila*. *Mol Cell* *13*, 19-32.
- Shi, Y., Sawada, J., Sui, G., Affar, e.B., Whetstone, J.R., Lan, F., Ogawa, H., Luke, M.P., and Nakatani, Y. (2003). Coordinated histone modifications mediated by a CtBP co-repressor complex. *Nature* *422*, 735-738.
- Shogren-Knaak, M., Ishii, H., Sun, J.M., Pazin, M.J., Davie, J.R., and Peterson, C.L. (2006). Histone H4-K16 acetylation controls chromatin structure and protein interactions. *Science* *311*, 844-847.
- Shroff, S., Joshi, M., and Orenic, T.V. (2007). Differential Delta expression underlies the diversity of sensory organ patterns among the legs of the *Drosophila* adult. *Mech Dev* *124*, 43-58.
- Shuey, D.J., and Parker, C.S. (1986). Bending of promoter DNA on binding of heat shock transcription factor. *Nature* *323*, 459-461.
- Siggers, T., Duyzend, M.H., Reddy, J., Khan, S., and Bulyk, M.L. (2011). Non-DNA-binding cofactors enhance DNA-binding specificity of a transcriptional regulatory complex. *Mol Syst Biol* *7*, 555.
- Sikorski, T.W., and Buratowski, S. (2009). The basal initiation machinery: beyond the general transcription factors. *Curr Opin Cell Biol* *21*, 344-351.
- Silverstein, R.A., and Ekwall, K. (2005). Sin3: a flexible regulator of global gene expression and genome stability. *Curr Genet* *47*, 1-17.
- Simboeck, E., Sawicka, A., Zupkovitz, G., Senese, S., Winter, S., Dequiedt, F., Ogris, E., Di Croce, L., Chiocca, S., and Seiser, C. (2010). A phosphorylation switch regulates the transcriptional activation of cell cycle regulator p21 by histone deacetylase inhibitors. *J Biol Chem* *285*, 41062-41073.
- Simon, J., Chiang, A., Bender, W., Shimell, M.J., and O'Connor, M. (1993). Elements of the *Drosophila* bithorax complex that mediate repression by Polycomb group products. *Dev Biol* *158*, 131-144.
- Slattery, M., Riley, T., Liu, P., Abe, N., Gomez-Alcala, P., Dror, I., Zhou, T., Rohs, R., Honig, B., Bussemaker, H.J., *et al.* (2011). Cofactor binding evokes latent differences in DNA binding specificity between Hox proteins. *Cell* *147*, 1270-1282.

- Snodgrass, R.E. (1928). Morphology of the insect head and its appendages. In Smithsonian Miscellaneous Collections, pp. 1-158.
- Snodgrass, R.E. (1935). Principles of Insect Morphology (McGraw-Hill, New York.).
- Spitz, F., Gonzalez, F., and Duboule, D. (2003). A global control region defines a chromosomal regulatory landscape containing the HoxD cluster. *Cell* *113*, 405-417.
- St Pierre, S.E., Galindo, M.I., Couso, J.P., and Thor, S. (2002). Control of *Drosophila* imaginal disc development by rotund and roughened eye: differentially expressed transcripts of the same gene encoding functionally distinct zinc finger proteins. *Development* *129*, 1273-1281.
- Struhl, G. (1981). A homoeotic mutation transforming leg to antenna in *Drosophila*. *Nature* *292*, 635-638.
- Swanson, C.I., Evans, N.C., and Barolo, S. (2010). Structural rules and complex regulatory circuitry constrain expression of a Notch- and EGFR-regulated eye enhancer. *Dev Cell* *18*, 359-370.
- Tabata, T., Schwartz, C., Gustavson, E., Ali, Z., and Kornberg, T.B. (1995). Creating a *Drosophila* wing de novo, the role of engrailed, and the compartment border hypothesis. *Development* *121*, 3359-3369.
- Tajiri, R., Misaki, K., Yonemura, S., and Hayashi, S. (2010). Dynamic shape changes of ECM-producing cells drive morphogenesis of ball-and-socket joints in the fly leg. *Development* *137*, 2055-2063.
- Takahashi, H., Parmely, T.J., Sato, S., Tomomori-Sato, C., Banks, C.A., Kong, S.E., Szutorisz, H., Swanson, S.K., Martin-Brown, S., Washburn, M.P., *et al.* (2011). Human mediator subunit MED26 functions as a docking site for transcription elongation factors. *Cell* *146*, 92-104.
- Taverna, S.D., Ueberheide, B.M., Liu, Y., Tackett, A.J., Diaz, R.L., Shabanowitz, J., Chait, B.T., Hunt, D.F., and Allis, C.D. (2007). Long-distance combinatorial linkage between methylation and acetylation on histone H3 N termini. *Proc Natl Acad Sci U S A* *104*, 2086-2091.
- Tempera, I., Klichinsky, M., and Lieberman, P.M. (2011). EBV latency types adopt alternative chromatin conformations. *PLoS Pathog* *7*, e1002180.
- Tempera, I., Wiedmer, A., Dheekollu, J., and Lieberman, P.M. (2010). CTCF prevents the epigenetic drift of EBV latency promoter Qp. *PLoS Pathog* *6*, e1001048.
- Thanos, D., and Maniatis, T. (1995). Virus induction of human IFN beta gene expression requires the assembly of an enhanceosome. *Cell* *83*, 1091-1100.
- Theisen, H., Haerry, T.E., O'Connor, M.B., and Marsh, J.L. (1996). Developmental territories created by mutual antagonism between Wingless and Decapentaplegic. *Development* *122*, 3939-3948.
- Thomas, M.C., and Chiang, C.M. (2006). The general transcription machinery and general cofactors. *Crit Rev Biochem Mol Biol* *41*, 105-178.
- Tsuji, T., Sato, A., Hiratani, I., Taira, M., Saigo, K., and Kojima, T. (2000). Requirements of Lim1, a *Drosophila* LIM-homeobox gene, for normal leg and antennal development. *Development* *127*, 4315-4323.

- Tsukiyama, T., and Wu, C. (1995). Purification and properties of an ATP-dependent nucleosome remodeling factor. *Cell* 83, 1011-1020.
- Tweedie, S., Ng, H.H., Barlow, A.L., Turner, B.M., Hendrich, B., and Bird, A. (1999). Vestiges of a DNA methylation system in *Drosophila melanogaster*? *Nat Genet* 23, 389-390.
- Valenzuela, L., Dhillon, N., and Kamakaka, R.T. (2009). Transcription independent insulation at TFIIC-dependent insulators. *Genetics* 183, 131-148.
- van Berkum, N.L., and Dekker, J. (2009). Determining spatial chromatin organization of large genomic regions using 5C technology. *Methods Mol Biol* 567, 189-213.
- Vaquerizas, J.M., Kummerfeld, S.K., Teichmann, S.A., and Luscombe, N.M. (2009). A census of human transcription factors: function, expression and evolution. *Nat Rev Genet* 10, 252-263.
- Vokes, S.A., Ji, H., Wong, W.H., and McMahon, A.P. (2008). A genome-scale analysis of the cis-regulatory circuitry underlying sonic hedgehog-mediated patterning of the mammalian limb. *Genes Dev* 22, 2651-2663.
- Voss, T.C., Schiltz, R.L., Sung, M.H., Yen, P.M., Stamatoyannopoulos, J.A., Biddie, S.C., Johnson, T.A., Miranda, T.B., John, S., and Hager, G.L. (2011). Dynamic exchange at regulatory elements during chromatin remodeling underlies assisted loading mechanism. *Cell* 146, 544-554.
- Wang, G., Balamotis, M.A., Stevens, J.L., Yamaguchi, Y., Handa, H., and Berk, A.J. (2005). Mediator requirement for both recruitment and postrecruitment steps in transcription initiation. *Mol Cell* 17, 683-694.
- Wang, L., Brown, J.L., Cao, R., Zhang, Y., Kassis, J.A., and Jones, R.S. (2004). Hierarchical recruitment of polycomb group silencing complexes. *Mol Cell* 14, 637-646.
- Wang, L., and Coulter, D.E. (1996). bowel, an odd-skipped homolog, functions in the terminal pathway during *Drosophila* embryogenesis. *EMBO J* 15, 3182-3196.
- Wendt, K.S., Yoshida, K., Itoh, T., Bando, M., Koch, B., Schirghuber, E., Tsutsumi, S., Nagae, G., Ishihara, K., Mishiro, T., *et al.* (2008). Cohesin mediates transcriptional insulation by CCCTC-binding factor. *Nature* 451, 796-801.
- Wheeler, S.R., Stagg, S.B., and Crews, S.T. (2009). MidExDB: a database of *Drosophila* CNS midline cell gene expression. *BMC Dev Biol* 9, 56.
- Wilkins, R.C., and Lis, J.T. (1997). Dynamics of potentiation and activation: GAGA factor and its role in heat shock gene regulation. *Nucleic Acids Res* 25, 3963-3968.
- Williams, J.A., Bell, J.B., and Carroll, S.B. (1991). Control of *Drosophila* wing and haltere development by the nuclear vestigial gene product. *Genes Dev* 5, 2481-2495.
- Williams, T.M., Selegue, J.E., Werner, T., Gompel, N., Kopp, A., and Carroll, S.B. (2008). The regulation and evolution of a genetic switch controlling sexually dimorphic traits in *Drosophila*. *Cell* 134, 610-623.
- Wittkopp, P.J. (2006). Evolution of cis-regulatory sequence and function in Diptera. *Heredity (Edinb)* 97, 139-147.
- Wittkopp, P.J., Carroll, S.B., and Kopp, A. (2003). Evolution in black and white: genetic control of pigment patterns in *Drosophila*. *Trends Genet* 19, 495-504.

- Wittkopp, P.J., True, J.R., and Carroll, S.B. (2002). Reciprocal functions of the *Drosophila* yellow and ebony proteins in the development and evolution of pigment patterns. *Development* *129*, 1849-1858.
- Wood, A.M., Van Bortle, K., Ramos, E., Takenaka, N., Rohrbaugh, M., Jones, B.C., Jones, K.C., and Corces, V.G. (2011). Regulation of chromatin organization and inducible gene expression by a *Drosophila* insulator. *Mol Cell* *44*, 29-38.
- Wray, G.A., Hahn, M.W., Abouheif, E., Balhoff, J.P., Pizer, M., Rockman, M.V., and Romano, L.A. (2003). The evolution of transcriptional regulation in eukaryotes. *Mol Biol Evol* *20*, 1377-1419.
- Wu, J., and Cohen, S.M. (1999). Proximodistal axis formation in the *Drosophila* leg: subdivision into proximal and distal domains by Homothorax and Distal-less. *Development* *126*, 109-117.
- Wu, J., and Cohen, S.M. (2000). Proximal distal axis formation in the *Drosophila* leg: distinct functions of teashirt and homothorax in the proximal leg. *Mech Dev* *94*, 47-56.
- Wysocka, J., Swigut, T., Xiao, H., Milne, T.A., Kwon, S.Y., Landry, J., Kauer, M., Tackett, A.J., Chait, B.T., Badenhorst, P., *et al.* (2006). A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling. *Nature* *442*, 86-90.
- Yang, J., and Corces, V.G. (2012). Insulators, long-range interactions, and genome function. *Curr Opin Genet Dev* *22*, 86-92.
- Yao, L.C., Phin, S., Cho, J., Rushlow, C., Arora, K., and Warrior, R. (2008). Multiple modular promoter elements drive graded brinker expression in response to the Dpp morphogen gradient. *Development* *135*, 2183-2192.
- Zecca, M., Basler, K., and Struhl, G. (1995). Sequential organizing activities of engrailed, hedgehog and decapentaplegic in the *Drosophila* wing. *Development* *121*, 2265-2278.
- Zeitlinger, J., Zinzen, R.P., Stark, A., Kellis, M., Zhang, H., Young, R.A., and Levine, M. (2007). Whole-genome ChIP-chip analysis of Dorsal, Twist, and Snail suggests integration of diverse patterning processes in the *Drosophila* embryo. *Genes Dev* *21*, 385-390.
- Zhu, H., Joliot, V., and Prywes, R. (1994). Role of transcription factor TFIIF in serum response factor-activated transcription. *J Biol Chem* *269*, 3489-3497.
- Zinzen, R.P., Girardot, C., Gagneur, J., Braun, M., and Furlong, E.E. (2009). Combinatorial binding predicts spatio-temporal cis-regulatory activity. *Nature* *462*, 65-70.
- Zollman, S., Godt, D., Privé, G.G., Couderc, J.L., and Laski, F.A. (1994). The BTB domain, found primarily in zinc finger proteins, defines an evolutionarily conserved family that includes several developmentally regulated genes in *Drosophila*. *Proc Natl Acad Sci U S A* *91*, 10717-10721.

AUTHOR: Aissette Baanannou

TITLE: Study of the transcriptional regulation of *bab* genes during *Drosophila* appendages development

PhD SUPERVISORS: Henri-Marc Bourbon and Muriel Boube

PLACE AND DATE OF DEFENSE: IBCG, 118 route de Narbonne, 31062 Toulouse cedex 09. 17 December 2012 at 10 AM

ABSTRACT: A proper spatio-temporal expression of *bric-a-brab2* is required for the proximo-distal (PD) patterning of the *Drosophila* leg and antenna. Here, we deciphered the molecular mechanisms controlling *bab2* expression in developing distal leg (4 rings) and antennal (2 rings) tissues. Using site-targeted GFP reporter assay and BAC recombineering, we show that restricted *bab2* expression relies on a single 0.56 kb *cis*-regulatory module (CRM), which is necessary and sufficient to drive normal Bab2 function in both developing leg and antenna. Through deletion and site-directed mutagenesis approaches, we identified discrete DNA motifs acting either positively or negatively. Finally, loss- and gain-of-function genetic analyses in combination with *in-vitro* electrophoretic mobility shift assays established that the Distal-less homeodomain protein directly activates *bab2* in all rings, whereas the Rotund C2H2 zinc-finger protein is specifically required for direct activation in the most proximal rings. This work paves the way to decipher at the molecular level how to make concentric rings from an epithelial layer in order to organize the PD limb patterning.

KEY WORDS: *Drosophila*, Development, CRM, *bab*.

ADMINISTRATIVE DISCIPLINE: Molecular Genetics

Centre de Biologie du Développement
UMR 5547, Bâtiment 4R3
118, Route de Narbonne
31062 TOULOUSE CEDEX 09