



HAL
open science

Évolution de la canalisation génétique dans un modèle quantitatif de réseau de régulation

Estelle Rünneburger

► **To cite this version:**

Estelle Rünneburger. Évolution de la canalisation génétique dans un modèle quantitatif de réseau de régulation. Bio-Informatique, Biologie Systémique [q-bio.QM]. Université Paris-Saclay, 2016. Français. NNT : 2016SACLS547 . tel-01443620

HAL Id: tel-01443620

<https://theses.hal.science/tel-01443620>

Submitted on 23 Jan 2017

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

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

NNT : 2016SACLS547

THÈSE DE DOCTORAT
DE L'UNIVERSITÉ PARIS-SACLAY
PRÉPARÉE À L'UNIVERSITÉ PARIS-SUD

École doctorale n°577
Structure et Dynamique des Systèmes Vivants
Spécialité de doctorat : Sciences de la vie et de la Santé

par

MME ESTELLE RÜNNEBURGER

Évolution de la canalisation génétique dans un modèle quantitatif
de réseau de régulation.

Thèse présentée et soutenue à Gif-sur-Yvette, le 19 décembre 2016.

Composition du Jury :

Mme.	CHRISTINE DILLMANN	Professeur Université Paris-Sud	(Présidente du jury)
M.	FRÉDÉRIC AUSTERLITZ	Directeur de recherche CNRS	(Rapporteur)
M.	SYLVAIN BILLIARD	Maitre de conférence Université Lille 1	(Rapporteur)
M.	OLIVIER LESPINET	Professeur Université Paris-Sud	(Examineur)
M.	ÉTIENNE RAJON	Maitre de conférence UCB Lyon 1	(Examineur)
M.	ARNAUD LE ROUZIC	Chargé de recherche CNRS	(Directeur de thèse)

Remerciements

« Try not. Do... or do not. There is no try. »

Yoda

À Arnaud Le Rouzic, merci d’avoir accueilli une M2 trop timide pour son bien et qui en plus ne savait pas coder et de lui avoir donné sa chance ! Tes larges connaissances, ta pédagogie, ta disponibilité et ta rigueur scientifique m’auront beaucoup apporté tout au long de ces presque quatre années.

À Pierre Capy, Catherine Montchamp-Moreau, Dominique Joly et Myriam Harry, merci de m’avoir ouvert les portes du laboratoire EGCE. Un grand merci également aux équipes administratives ou informatiques, qui ne m’ont toujours pas mise dehors malgré mes fréquentes incursions dans leurs domaines réservés, sourire timide et petit « J’aurais une question... » ou « J’ai un problème... » sur les lèvres.

À Thomas Hansen et Christine Mayer, un grand merci pour m’avoir accueilli à Oslo et pour les longues discussions (parfois trop) pointues.

À tous ceux qui vont vivre le labo, à l’équipe IGGIPOP, à ceux que je croise l’hiver pendant la pause du matin, café pour eux et chocolat pour moi, à ceux qui rendent les repas de midi si instructifs dans tant de domaines si variés, à ceux qui sont dans la même galère doctorale que moi, et aux autres encore que je ne fais que saluer en les croisant, merci tout simplement. C’est grâce à vous si le labo est si agréable.

À l’ensemble de mes co-bureaux, Quentin, Bastien, Andréas, Éric et Thomas, merci pour les discussions sérieuses et pour toutes les autres. Vous avez apporté à cette petite pièce des fous-rires, de la chaleur (dans tous les sens du terme, ce qui est très appréciable en hiver), des menaces, des dessins bizarres sur tableau blanc, des vidéos étranges, des insectes flippants... Bref, je m’y suis sentie comme chez moi.

À une certaine communauté de batraciens, merci pour le soutien moral devant mes craquages répétés, le dépôt de tonnes de chocolat virtuel et le proof-reading de dernière minute.

Aux amis, Arnaud, Rachel, Anne-Lise, Maryem, Jean-Loup et puis les autres, vous savez tous très bien pourquoi je vous remercie, ce n'est pas la peine d'épiloguer.

À mes parents, je n'ai pas assez de mots pour vous remercier de tout ce que vous avez fait et faites pour moi. Je vous aime.

Et enfin, au dernier carré, aux fidèles d'entre les fidèles, Romain, Laure, Cam, Louve, Aramis, ElNat. Pour vous, j'irais au bout du multivers. May the Force be with you. Always.

Table des matières

1	Introduction	1
1.1	Problématique générale de la thèse	3
1.2	La canalisation	4
1.2.1	Waddington : le précurseur	4
1.2.2	Études ultérieures	6
1.2.3	Importance évolutive de la canalisation	7
1.3	Les réseaux de gènes	11
1.3.1	Spécificités des réseaux de gènes	12
1.3.2	Différents types de réseaux	13
1.3.3	Les réseaux de régulation	14
1.4	Contexte de la thèse	21
2	Programme et modèle	23
2.1	Modèle théorique	25
2.1.1	Du génotype au phénotype	25
2.1.2	De l'individu à la population	29
2.1.3	Estimation de la canalisation	31
2.1.4	Autres modèles implémentés	32
2.2	Workflow d'une analyse complète	32
2.2.1	Routines de lancement	32
2.2.2	Programme de simulation "Simevolv"	36
2.3	Simulation développementale sous R	40
3	Résultats : propriétés du modèle	41
3.1	Exploration du processus développemental	43
3.1.1	Dynamique de l'expression génique	44
3.1.2	Instabilité développementale	46
3.1.3	Conclusion	47
3.2	Équilibre mutation - sélection - dérive	48
3.2.1	Évolution temporelle	49

3.2.2	Influence des paramètres du réseau	50
3.2.3	Conclusion	52
3.3	Mode de reproduction	53
3.3.1	Diploïdie versus haploïdie	54
3.3.2	Régime de reproduction	56
3.3.3	Conclusion	58
4	Résultats : pourquoi et comment la canalisation génétique évolue dans des réseaux de régulation	59
5	Discussion et perspectives	77
5.1	Logiciel et modèle	79
5.1.1	Avantages et limites du logiciel de simulation	79
5.1.2	Approximations et paramétrisation du modèle	80
5.1.3	Calcul de la canalisation	82
5.2	Discussion générale et perspectives	83
5.2.1	Apport de la modélisation et généralisation des résultats	83
5.2.2	Importance de la sélection	84
5.2.3	Origine de la canalisation génétique	85
5.3	Conclusion	91
	Liste des symboles	95
	Bibliographie	96

Liste des figures

1.1	Conrad Hal Waddington (8/11/1905 – 26/09/1975)	4
1.2	Exemples de mutants de <i>Drosophila willistoni</i> .	5
1.3	Paysage épigénétique.	6
1.4	Plasticité vs canalisation.	8
1.5	Exemple de réseau neutre.	9
1.6	Effet de l'épistasie sur le phénotype.	13
1.7	Modélisation par équation différentielle d'un réseau de régulation.	14
1.8	Représentation schématique d'un réseau de régulation.	15
2.1	Exemple de matrice d'interaction et réseau de régulation correspondant.	26
2.2	Comportement au cours du développement.	27
2.3	Niveau d'expression du gène i en fonction de la force de la régulation.	27
2.4	Calcul du phénotype de l'individu en fin de développement.	28
2.5	Représentation schématique de la reproduction sexuée et de la recombinaison.	29
2.6	Forces de sélection applicables sur le réseau.	30
2.7	Paysage adaptatif.	30
2.8	Exemple de script de lancement.	34
2.9	Représentation schématique du programme Simevolv.	37
2.10	Temps de simulation pour différents paramètres.	39
3.1	Dynamique de l'expression génique au cours du développement.	44
3.2	Expression génique à la fin du développement.	45
3.3	Évolution de l'instabilité développementale.	46
3.4	Réponse à la sélection.	50
3.5	Réponse à la sélection.	50
3.6	Impact des paramètres sur la réponse à la sélection : valeur sélective.	51
3.7	Impact des paramètres sur la réponse à la sélection : expression génique.	51
3.8	Sélection en modèle haploïde et diploïde.	54
3.9	Diploïdie vs haploïdie.	55

3.10	Influence du taux d'autofécondation.	56
3.11	Influence du taux de recombinaison.	57
5.1	Impact de la sélection contre l'instabilité développementale.	86
5.2	Impact des perturbations environnementales.	88
5.3	Scores de canalisation dans des réseaux aléatoires.	90
5.4	Corrélations entre les scores de canalisation dans des réseaux aléatoires.	90

Liste des tableaux

2.1	Exemple de fichier d'entrée.	33
2.2	Exemple de fichier de sortie.	35
2.3	Jeu de paramètre par défaut pour les simulations du développement. . .	40

Chapitre 1

Introduction

Sommaire

1.1	Problématique générale de la thèse	3
1.2	La canalisation	4
1.2.1	Waddington : le précurseur	4
1.2.2	Études ultérieures	6
1.2.3	Importance évolutive de la canalisation	7
1.2.3.1	Rôle évolutif de la canalisation	7
	Mutations génétiques cryptiques	7
	Plasticité phénotypique	8
	Réseaux neutres	9
1.2.3.2	Apparition de la canalisation	10
	Canalisation environnementale	10
	Canalisation génétique	10
1.3	Les réseaux de gènes	11
1.3.1	Spécificités des réseaux de gènes	12
	Pléiotropie	12
	Épistasie	12
1.3.2	Différents types de réseaux	13
1.3.3	Les réseaux de régulation	14
1.3.3.1	Modélisation des réseaux de régulations	14
	Systèmes d'équations différentielles	14
	Le modèle de Wagner	15
1.3.3.2	Apports du modèle de Wagner et de ses dérivés	16
	Canalisation et assimilation génétique	16
	Évolvabilité	18
	Reproduction	19

Architecture génétique	20
1.4 Contexte de la thèse	21

1.1 Problématique générale de la thèse

L'une des caractéristiques les plus frappantes du développement des organismes est sans doute son extrême stabilité. Malgré les changements d'environnement ou les mutations qui pourraient affecter son déroulement, il reste globalement et remarquablement similaire entre tous les individus d'une même espèce (à l'exception des cas pathologiques).

Cette stabilité est due d'une part au contrôle strict qui s'exerce sur l'expression des gènes tout au long du développement, et d'autre part à une certaine réduction de l'effet des perturbations. Ce tamponnage de l'effet des perturbations, fussent-elles génétiques, micro- ou encore macro-environnementales, est plus connu sous le nom de canalisation.

Cet aspect pourtant fondamental du développement des individus reste encore imparfaitement compris. Certaines hypothèses pressentent une relation étroite entre la manière dont sont organisés les gènes impliqués dans le développement d'un caractère (on parle alors de l'architecture génétique du caractère) et l'état de canalisation dudit caractère.

Les travaux présentés dans les pages suivantes constituent une tentative d'étude de cette hypothèse, afin d'élucider les liens pouvant exister entre l'architecture génétique sous-jacente à un trait phénotypique et certains aspects évolutifs de ces traits, ici leur état de canalisation.

Pour cela, j'ai choisi d'étudier de manière détaillée les réseaux de régulation, dont les propriétés complexes sont particulièrement adaptées à l'étude de l'apparition et de l'évolution de la canalisation. Pour ce faire, j'ai utilisé une méthodologie fondée à la fois sur des représentations issues de la biologie théorique, ainsi que sur des notions empruntées directement à la génétique quantitative. La biologie théorique m'a fourni un modèle de réseaux de régulation couramment utilisé (celui du modèle de Wagner), repris et adapté à notre étude. La biologie quantitative, quant à elle, m'a apporté des index utiles pour la quantification des aspects évolutifs des réseaux et, plus particulièrement, la manière de calculer le score de canalisation.

Dans les sections suivantes, je m'efforcerai de dresser un état des lieux des connaissances actuelles sur la canalisation, ainsi que sur les réseaux de régulation et leurs propriétés.



FIGURE 1.1 – CONRAD HAL WADDINGTON (8/11/1905 ANGLETERRE – 26/09/1975 ÉCOSSE). Anglais touche à tout (biologie du développement, paléontologie, génétique, mais aussi poésie et peinture), père fondateur de la biologie des systèmes, de l'épigénétique et de l'évo-dévo.

1.2 La canalisation

1.2.1 Waddington : le précurseur

En 1942, Conrad Hal Waddington publie un article fondateur intitulé "*Canalization of Development and the Inheritance of Acquired Characters*" (Canalisation du développement et hérédité des caractères acquis, WADDINGTON [1942]). Il y tente une unification entre les partisans de la théorie évolutive (darwinisme) et ceux de la théorie naturaliste (ou lamarckienne, portant sur l'hérédité des caractères acquis), en se basant sur sa connaissance des processus développementaux pour expliquer comment certains traits induits par des effets de l'environnement peuvent s'intégrer au génotype et ainsi se transmettre aux générations suivantes. Waddington reprendra d'ailleurs ce mécanisme et l'étudiera plus en détail dans les années ultérieures, lui donnant le nom d'assimilation génétique (WADDINGTON [1953]).

L'un des arguments de Waddington est de faire remarquer que les processus développementaux sont, dans des organismes soumis à la sélection naturelle, « ajustés de manière à produire un même résultat, peu importe les variations mineures des conditions au cours de son processus ». Ce mécanisme, qu'il nomme canalisation (*canalization*) ou encore tamponnage (*buffering*), est selon lui soutenu par plusieurs observations.

D'un point de vue embryologique et développemental, il remarque qu'un tissu ne peut pas prendre une « voie » médiane entre les différents destins qu'il est susceptible d'atteindre : il se dirigera soit vers l'un, soit vers l'autre, mais pas vers une forme intermédiaire. D'une manière plus concrète, une cellule pourra donner du tissu musculaire ou du tissu nerveux, mais ne pourra pas devenir un mélange des deux. De même, les

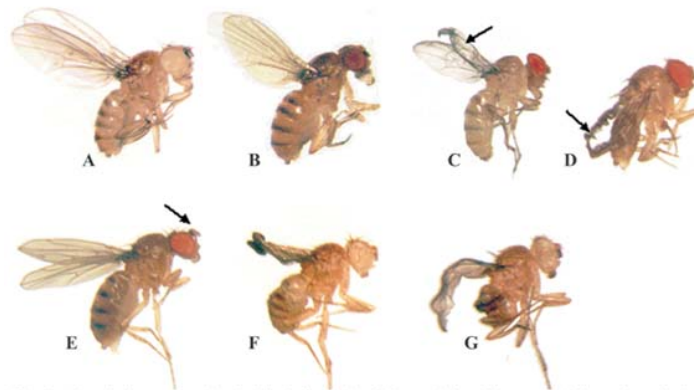


Fig. 2. Altered phenotypes obtained in individuals of *Drosophila willistoni* 17A2 isostrain submitted to temperature stress. Picture shows female individuals with altered morphologies: A. *white* (white eyes); B. *sepia* (brown eyes); C. *blistered* (arrow indicates the presence of blisters on the wings); D. *Curly* (curved wings indicated by the arrow); E. with apparently fused antennae (indicated by the arrow) - see figure 5 for details of the mutant structure; F. Female *white* and *Curly*; G. *white* and *blistered*.

FIGURE 1.2 – EXEMPLES DE MUTANTS DE *Drosophila willistoni*. Les différentes mutations présentées ici conduisent toutes à d'importants changements morphologiques. Image tirée de D'ÁVILA et collab. [2008].

différents tissus coexistants dans un organisme sont très différenciés les uns des autres, et il est impossible de trouver des états intermédiaires suggérant la possibilité d'un passage progressif d'un type à un autre. Waddington note également que ce qui est valable pour un tissu l'est aussi à l'échelle des organes, voire des organismes.

D'un point de vue plus génétique, Waddington parle de « constance de l'état sauvage » : les formes sauvages (non mutées) d'un organisme soumis à la sélection naturelle sont bien moins diverses que les formes mutantes. Par exemple, alors qu'il existe un grand nombre de mutations chez *Drosophila willistoni* conduisant à des phénotypes particuliers et reconnaissables (voir l'exemple figure 1.2), deux drosophiles sauvages de la même espèce sont pour ainsi dire indiscernables.

Waddington déduit de ces deux observations que le phénotype doit être tamponné (ou canalisé) contre des variations mineures provenant de l'environnement, mais aussi contre celles de son propre génome. Cela permet d'assurer la production d'un phénotype optimal en dépit des effets perturbateurs du hasard. Il remarque également que cela n'est pas vrai dans toutes les conditions, ce mécanisme étant rompu de façon notable chez les mutants ou dans des cas pathologiques. Pour Waddington, ceci implique que la canalisation est « une caractéristique d'un système, qui est construite par la sélection naturelle » : en d'autres termes, la canalisation des systèmes développementaux serait bel et bien due à l'action de la sélection naturelle.

Pour illustrer sa théorie de la canalisation développementale, Waddington utilise la désormais célèbre métaphore du paysage épigénétique (voir figure 1.3, WADDINGTON [1957]). Il imagine une vallée descendant d'une colline et se subdivisant en d'autres vallées secondaires. En haut de la vallée, on dépose une bille, qui roulera le long de la

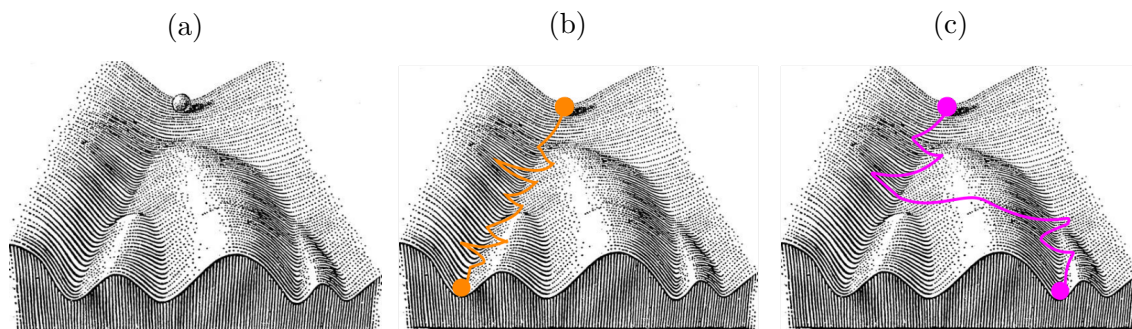


FIGURE 1.3 – PAYSAGE ÉPIGÉNÉTIQUE. (a) Modèle de paysage épigénétique proposé par [Waddington \[1957\]](#). La bille représente une cellule et les différentes vallées représentent les destins cellulaires possibles qui s’offrent à elle. (b) Phénotype canalisé : les perturbations au cours du développement ne changent pas le devenir de la cellule. (c) Phénotype décanalisé : une des perturbations est assez importante pour modifier le devenir de la cellule et l’orienter vers une nouvelle voie.

pente jusqu’à s’arrêter en bas de l’une des vallées. Dans l’interprétation la plus courante, la bille représente une cellule, et les vallées, les voies possibles qu’elle peut emprunter au cours du développement, la menant à des destins cellulaires (types de tissus formés par exemple) différents. Dans cette visualisation, plus les pentes de la vallée sont hautes et abruptes, plus le phénotype final sera canalisé : toutes les modifications du trajet de la bille qui ne seront pas assez importantes pour lui faire passer la ligne de crête n’auront aucun impact sur son devenir.

1.2.2 Études ultérieures

Après les articles fondateurs de Waddington, de nombreux chercheurs se sont emparés, directement ou indirectement, de ce sujet d’étude. Lors des décennies suivantes, la canalisation a pu être étudiée de différentes manières et sous des noms différents qui recouvrent aujourd’hui des notions distinctes, bien que liées (canalisation, stabilité développementale, plasticité phénotypique, asymétrie fluctuante, etc.). Ce flou autour des notions était notamment dû à des définitions vagues, avant que des tentatives d’harmonisations ne soient menées dans les années 2000 (voir [Debat et David \[2001\]](#) pour une revue).

À l’heure actuelle, si la définition réduite de « tamponnage des voies développementales contre les mutations ou les perturbations environnementales » est toujours admise, la canalisation est plus généralement définie comme une réduction, voire une suppression, de la variance de la valeur du phénotype observé entre les différents individus de la population ([Stearns et collab. \[1995\]](#); [Wagner et collab. \[1997\]](#)).

Si les définitions ont évolué au fil des études successives, il en est de même pour le cadre général du terme. Notamment, le terme de canalisation (ou son synonyme

robustesse) a également été repris pour des études qui se sont éloignées du cadre strict du développement pour s'intéresser à tous les niveaux d'organisation d'un individu, de la séquence nucléotidique jusqu'aux réseaux sous-tendant la synthèse d'une enzyme en passant par la conformation des protéines (voir [WAGNER \[2013\]](#) pour une synthèse).

Tout au long de ce manuscrit, j'utiliserai indifféremment les termes de canalisation ou de robustesse en me restreignant uniquement aux processus de tamponnage impliqués lors du développement des individus.

1.2.3 Importance évolutive de la canalisation

Dans cette partie, je propose une revue des résultats marquants sur le rôle de la canalisation dans l'évolution, ainsi que sur les pistes envisagées pour l'apparition d'un tel phénomène.

1.2.3.1 Rôle évolutif de la canalisation

Si ni les origines évolutives de la canalisation ni les mécanismes selon lesquels elle agit ne sont encore totalement élucidés ([SIEGAL et LEU \[2014\]](#)), le rôle que peut jouer ce processus dans l'évolution reste lui aussi en proie à quelques incertitudes.

Mutations génétiques cryptiques

La réflexion la plus évidente serait de dire que la canalisation, en diminuant la variabilité phénotypique au sein des populations, les empêche de pouvoir atteindre un nouvel optimum. Ainsi, la canalisation aurait pour conséquence de ralentir, voire de stopper l'évolution du trait en question, qui se retrouverait bloqué dans un état sous-optimal ([ANCEL et collab. \[2000\]](#)).

Pourtant, des études aussi bien théoriques qu'expérimentales ont montré que des traits canalisés peuvent, sur le long terme, avoir des capacités d'adaptation augmentées et non diminuées. L'une des raisons pouvant expliquer ce phénomène réside dans la notion de variation génétique cryptique ([HERMISSON et WAGNER \[2004\]](#); [LE ROUZIC et CARLBORG \[2008\]](#); [MASEL et TROTTER \[2010\]](#)). Le raisonnement est le suivant : les mutations (ou variations génotypiques) canalisées ne provoquent pas de modifications dans le phénotype. Elles ne sont donc pas éliminées par l'effet de la sélection naturelle et peuvent envahir la population par dérive génétique, augmentant ainsi la quantité de variation génétique disponible dans la population, sans augmenter la variation phénotypique. Si un changement majeur (génétique ou environnemental) se produit, la canalisation peut alors disparaître ou être relâchée. Cela libère ou révèle les mutations jusque là cachées, ce qui a pour conséquence d'augmenter la variance phénotypique et la variance génétique (standing genetic variation) disponibles pour l'adaptation.

Plasticité phénotypique

La canalisation est également souvent opposée à la notion de plasticité phénotypique. Si la canalisation peut être définie comme l'invariance du phénotype face à un changement environnemental, la plasticité, elle, est définie par les variations du phénotype le long d'une norme de réaction en réponse à une perturbation environnementale (voir figure 1.4, [VIA et collab. \[1995\]](#)). Cependant, ces deux mécanismes ne sont peut-être pas aussi antinomiques qu'ils le paraissent de prime abord.

Pour certains auteurs, la plasticité phénotypique relèverait d'une certaine forme de canalisation : s'il est vrai que l'on observe des phénotypes différents selon les conditions, au sein d'une même condition les phénotypes sont peu différenciés. Ainsi, un organisme soumis à une température A ou B présenterait bien deux apparences différentes... mais tous les individus soumis à la même température B auraient une apparence similaire. La différence entre canalisation et plasticité phénotypique serait due à la fréquence des changements environnementaux : un environnement stable conduirait à de la canalisation obtenue par sélection stabilisante, alors que des changements fréquents d'environnement favoriseraient l'apparition des phénomènes de plasticité ([LEVY et SIEGAL \[2012\]](#)).

L'assimilation génétique décrite par [WADDINGTON \[1953\]](#) serait un cas très particulier de plasticité phénotypique. Davantage liée à la notion de décanalisation qu'à celle de canalisation, elle peut être définie comme le processus par lequel un phénotype originellement induit en réponse à des perturbations environnementales devient encodé dans le génome, après plusieurs générations, sous l'effet de la sélection naturelle ou artificielle. Le nouveau phénotype passe ainsi d'induit à constitutif, et apparaît même en absence du déclencheur initial ([SCHARLOO \[1991\]](#); [WADDINGTON \[1957\]](#)).

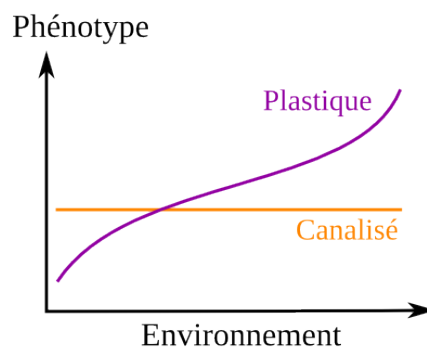


FIGURE 1.4 – PLASTICITÉ VS CANALISATION. Une norme de réaction est une représentation de la relation entre les phénotypes possibles produits par un même génotype d'une part, et la gamme de valeurs environnementales (température, salinité...) auxquelles il est exposé d'autre part. La norme de réaction orange correspond à un génotype canalisé (le phénotype est identique, quelle que soit la valeur de la variable environnementale), la violette, à un génotype plastique (le phénotype varie en fonction de l'environnement).

Cette assimilation peut arriver lorsqu’une mutation ou une perturbation a des effets suffisamment importants pour changer de voie développementale (décanalisation, voir figure 1.3c) et se tourner vers une autre voie. Si le nouveau phénotype obtenu est favorisé, les génotypes les plus sensibles à cette mutation ou perturbation (et donc les plus susceptibles de dévier à nouveau) vont être sélectionnés. Au fil des générations, cette voie de développement secondaire deviendra plus importante (avec davantage de génotypes y conduisant) et sera à son tour canalisée, jusqu’à devenir constitutive. Il est à noter que ce processus semble n’être possible que pour des perturbations arrivant pendant le développement de l’individu, et non à l’état adulte.

Réseaux neutres

Enfin, la canalisation génétique joue également un rôle dans l’établissement et/ou le maintien des réseaux neutres (aussi appelés réseaux génotypiques). Un réseau neutre représente l’ensemble des génotypes possibles codants pour le même phénotype. Ces génotypes pouvant être reliés les uns aux autres par des mutations ponctuelles (WAGNER [2008, 2012]), il est ainsi possible de passer d’un génotype à un autre, même éloigné, en traversant le réseau par une suite de mutations qui ne changeront pas le phénotype des individus (voir figure 1.5). En empêchant les variations phénotypiques dues aux mutations, la canalisation génétique permet d’élargir le réseau neutre : plus le

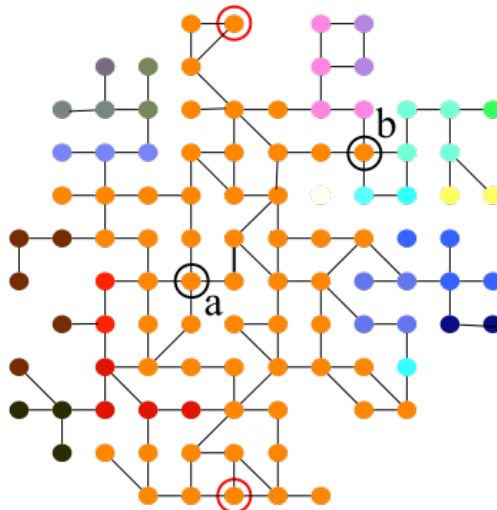


FIGURE 1.5 – EXEMPLE DE RÉSEAU NEUTRE. Chaque point représente un génotype et chaque couleur correspond à un phénotype identique. Les ronds oranges forment un large réseau génotypique et il est possible de passer de l’un des génotypes cerclés de rouge à l’autre en se déplaçant le long de ce réseau par des mutations ponctuelles qui ne changent pas le phénotype. Le génotype cerclé de noir (a) est robuste : une mutation ne changera pas son phénotype. Le génotype (b) n’est pas robuste : une mutation a 3 chances sur 4 de conduire à un phénotype différent.

réseau sera grand, plus les génotypes seront robustes aux mutations, dans un processus d'auto-entretien.

1.2.3.2 Apparition de la canalisation

En 1957, Waddington montre par des expériences de sélection artificielle que la canalisation peut se transmettre de génération en génération. Il s'agit donc d'un mécanisme héritable. Waddington développe en parallèle la notion de sélection canalisante, sous-entendant que la canalisation dispose de bases génétiques (WADDINGTON [1957]). Il échoue cependant à entrevoir la distinction entre canalisation environnementale et canalisation génétique, et considère qu'un seul mécanisme tamponne ces deux types de variations.

Il faudra attendre des études ultérieures (voir par exemple GIBSON [2000]; STEARNS et collab. [1995]; WAGNER et collab. [1997]) pour voir apparaître cette distinction, accompagnée de l'idée sous-jacente que ces deux mécanismes auraient des bases différentes (bien que potentiellement chevauchantes).

Canalisation environnementale

La canalisation environnementale peut se définir par la stabilité d'un phénotype dans une large gamme d'environnements. Le terme d'environnement est ici utilisé pour parler de tout type de perturbations non héritable, qu'elles soient externes (macro-environnement : température, pression...) ou internes (micro-environnement : bruit causé par la variation de concentration d'une molécule, par exemple).

La canalisation environnementale est sans doute apparue de multiples fois : chaque type de perturbation aurait conduit à l'apparition de mécanismes tampons qui lui sont propres. Ainsi, un changement de température peut être régulé par l'activation de protéines de choc thermique (telle que la protéine Hsp90 chez la drosophile, cas d'école des explications biochimiques de la canalisation, voir RUTHERFORD et LINDQUIST [1998]). Il est à noter que ces mécanismes de canalisation, pour la plupart spécifiques, ne possèdent pas de manière systématique la capacité à réagir à une perturbation différente de celle qui a conduit à leur mise en place.

Canalisation génétique

La canalisation génétique, quant à elle, peut se définir par la stabilité d'un phénotype en dépit de perturbations génétiques (mutations). À l'heure actuelle, trois hypothèses principales (non nécessairement exclusives) sont envisagées pour expliquer son apparition et son maintien (voir MASEL et SIEGAL [2009]; DE VISSER et collab. [2003] pour des revues).

La première hypothèse postule que la canalisation génétique serait issue d'un processus adaptatif : elle aurait évolué pour limiter les variations du phénotype face aux perturbations génétiques. Les génotypes mutés, mais canalisés, donnent des phénotypes identiques (et donc proches de l'optimum de la population), et gardent une bonne valeur sélective (ou *fitness*). La canalisation serait ainsi favorisée par la sélection.

La deuxième hypothèse, ou hypothèse congruente, veut que la canalisation génétique ne soit qu'un effet secondaire de la canalisation environnementale (ANCEL et collab. [2000]; WAGNER et collab. [1997]). Certains des mécanismes tampons utilisés pour lutter contre les perturbations venant de l'environnement auraient ainsi été capables de limiter également les effets des mutations.

Enfin, la dernière hypothèse, dite hypothèse intrinsèque, postule que la canalisation génétique ne serait qu'une propriété intrinsèque des systèmes génétiques complexes, due à leur optimisation ou à leur organisation. SIEGAL et BERGMAN [2002] incluent dans cette hypothèse, le scénario selon lequel l'apparition de la canalisation génétique serait un effet secondaire de la nécessaire stabilité du phénotype au cours du développement.

Dans le but de tester ces différentes hypothèses cherchant à expliquer l'apparition de la canalisation, j'ai choisi de travailler sur des architectures génétiques complexes et plus spécifiquement sur les réseaux de régulation.

1.3 Les réseaux de gènes

Il est unanimement admis que la plupart des traits phénotypiques d'un organisme, qu'ils soient morphologiques (taille, forme), métaboliques (taux d'expression d'une enzyme) ou encore comportementaux (mode d'approvisionnement en nourriture), sont contrôlés par un ensemble de gènes (ou de loci au sens large si on y intègre les domaines de régulation, les marques épigénétiques, voire les produits des gènes) interagissant les uns avec les autres de manière plus ou moins complexe.

Représenter ces interactions sous forme de réseaux est devenu une pratique courante, autant d'un point de vue biologique que du point de vue de la modélisation. Cela permet d'une part une visualisation plus rapide de l'intrication des processus, et d'autre part l'utilisation d'outils d'analyse tirés des théories mathématiques sur les réseaux ou les graphes (les notions de complexités ou de connectivité par exemple).

1.3.1 Spécificités des réseaux de gènes

Bien que plusieurs types de réseaux puissent être définis (réseaux de régulation, voies métaboliques, etc.), tous possèdent des caractéristiques communes nées des interactions qu'ils représentent (voir [BARABASI et OLTVAI \[2004\]](#) pour une revue sur les caractéristiques structurelles des réseaux).

Pléiotropie

On parle de pléiotropie lorsque l'effet d'un seul gène influence plusieurs traits phénotypiques : une mutation sur un locus pléiotrope aura ainsi des conséquences sur deux (ou plus) caractères (voir [STEARNS \[2010\]](#) et [PAABY et ROCKMAN \[2013\]](#) pour des revues).

La pléiotropie est particulièrement étudiée en termes de sélection artificielle. En effet, les nombreux effets des mutations pléiotropes peuvent compliquer la sélection d'un trait particulier, ou résulter en le maintien secondaire d'autres traits inintéressants ou contreproductifs.

D'autres études ont également abordé les relations entre pléiotropie et réseaux de gène, tant dans des réseaux biologiques que dans des modèles bio-informatiques. Un des résultats communs est la modularité (opposée à l'universalité) de la pléiotropie au sein des réseaux de gènes : la plupart des mutations pléiotropes auraient ainsi des effets importants à une échelle réduite (au sein de leur module), mais des effets négligeables au niveau de l'organisme entier ([STEARNS \[2010\]](#)).

De même, dans un réseau, il existe quelques loci très pléiotropes (ayant un impact sur de nombreux traits phénotypiques), mais la majeure partie ne l'est pas et n'a d'impact que sur un seul trait ([FEATHERSTONE et BROADIE \[2002\]](#); [LI et collab. \[2006\]](#); [WAGNER et collab. \[2008\]](#)).

Épistasie

L'épistasie peut être définie comme la dépendance entre les effets d'un gène et ceux d'autres gènes qui vont déterminer son « fond génétique ». S'il existe de l'épistasie entre deux gènes, l'effet de la mutation combinée des deux gènes sera plus important (épistasie positive) ou moins important (épistasie négative) que la somme de leurs effets lors d'une mutation simple. On parle d'épistasie de signe lorsqu'une mutation a un effet opposé en la présence ou en l'absence d'une autre mutation (figure 1.6).

Les relations entre épistasie et évolution sont particulièrement complexes (car difficiles à analyser ou à calculer) et loin d'être complètement décrites. Pour autant, l'impact de l'épistasie est sans doute non négligeable ([HANSEN \[2013\]](#); [WOLF et collab. \[2000\]](#)).

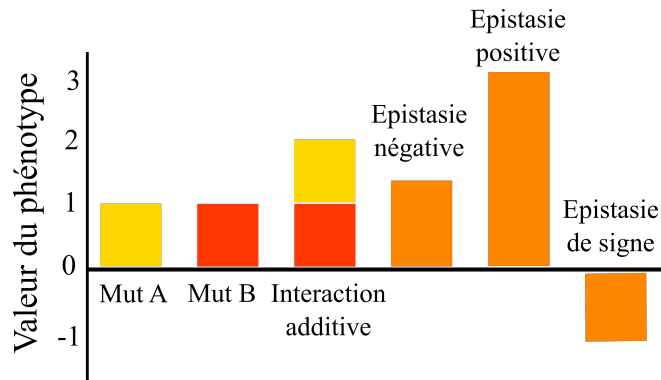


FIGURE 1.6 – EFFET DE L’ÉPISTASIE SUR LE PHÉNOTYPE. Les deux mutations A et B ont un effet de valeur 1 sur le phénotype. En absence d’épistasie, elles s’additionnent pour produire un phénotype de valeur $1 + 1 = 2$. La présence d’épistasie va modifier cette valeur phénotypique résultante, soit en l’augmentant (épistasie positive, valeur de 3), soit en la diminuant (épistasie négative, valeur de 1,5). La présence d’épistasie de signe conduit à une inversion de la valeur phénotypique initiale (-1).

Une des conséquences de l’épistasie sur l’évolution touche à l’aspect du paysage adaptatif (une représentation graphique de la valeur sélective de chacun des génotypes possibles). Un fort taux d’épistasie tendra à lui donner un aspect moins lisse, constitué de plusieurs pics de haute valeur sélective (ou maximums locaux). Ce changement dans le paysage adaptatif peut compliquer l’atteinte du phénotype optimum : en effet, certains génotypes vont rester coincés sur des sous-optimums locaux sans avoir la capacité de dépasser les « vallées » représentées par des génotypes de faible valeur sélective.

1.3.2 Différents types de réseaux

De nombreux types de réseaux génétiques cohabitent au sein d’un organisme, chacun possédant des fonctions et des rôles différents (réseaux de régulation, voie de biosynthèse, cascade de signalisation, etc.). S’ils peuvent être individualisés pour les besoins d’études spécifiques, il est cependant important de noter que tous ces réseaux sont (dans la réalité biologique) fortement intriqués, un réseau de régulation pouvant par exemple réguler l’expression d’une enzyme impliquée dans la synthèse d’une protéine nécessaire au bon fonctionnement d’une voie de signalisation.

Les voies métaboliques sont des ensembles de réactions biochimiques conduisant à la synthèse ou à la dégradation d’une molécule biologique (ou métabolite), catalysées par une série d’enzymes agissant de manière séquentielle (CLARK [1991]; KACSER et BURNS [1979]). Peu connectées et linéaires, les voies métaboliques se montrent très sensibles à l’absence de l’un de leurs composants.

Les réseaux de régulation peuvent être définis comme un ensemble d’entités (gènes ou protéines) en interaction avec elles-mêmes et avec d’autres entités extérieures. Plus précisément, ce terme est généralement employé pour représenter l’effet des différents

facteurs de transcription sur le niveau d'expression de gènes donnés. Leur niveau de complexité élevé et souvent hiérarchisé (BARABASI et OLTVAI [2004]) couplé à la possible redondance entre facteurs de transcription les rendant particulièrement adaptés à l'étude de la canalisation, c'est sur cette famille que j'ai décidé de concentrer mes recherches.

1.3.3 Les réseaux de régulation

1.3.3.1 Modélisation des réseaux de régulations

Pouvoir étudier de manière poussée les réseaux de régulation et les modifications qu'ils subissent au cours du développement en réponse à la sélection est d'une importance majeure en biologie de l'évolution. Cependant, cela peut s'avérer compliqué d'un point de vue expérimental, notamment quand il s'agit de traiter des notions aussi complexes que l'épistasie ou la pléiotropie.

Le développement de modèles théoriques s'avère donc être un atout non négligeable pour étudier les liens entre les caractéristiques des réseaux et leurs propriétés évolutives, et ce à court comme à long terme. Ainsi, de nombreux formalismes mathématiques tentant de s'approcher de la réalité des réseaux de régulations ont été mis en place au fil des années.

Systemes d'équations différentielles

Une part importante des études se fonde sur des systèmes d'équations différentielles (voir par exemple POLYNIKIS et collab. [2009] pour une revue). Cette méthode de modélisation cherche à représenter de manière explicite la variation des taux d'expressions et des concentrations des différents produits des gènes au cours de la vie de la cellule.

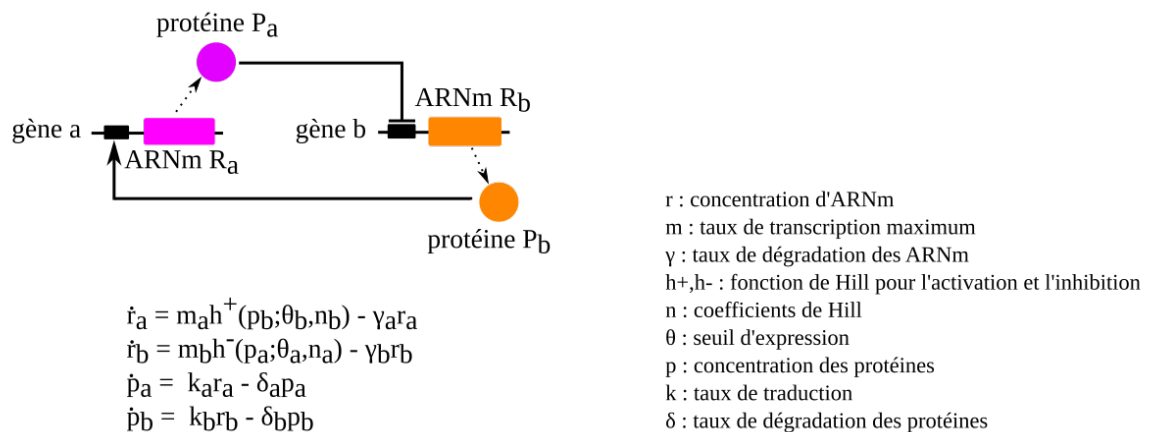


FIGURE 1.7 – MODÉLISATION PAR ÉQUATION DIFFÉRENTIELLE D'UN RÉSEAU DE RÉGULATION. D'après POLYNIKIS et collab. [2009].

Pour cela, elle prend en compte un grand nombre de paramètres biochimiques, tels que les affinités de fixation des facteurs de transcription, les constantes de dissociation des récepteurs et des ligands, ou encore les constantes cinétiques des enzymes.

Un exemple simple de réseau à deux gènes, accompagné de l'un des (nombreux) modèles qui peuvent lui être appliqués, est présenté en figure 1.7.

Outre la complexité de la résolution numérique de telles équations (croissant avec le nombre de gènes impliqués dans le réseau), le désavantage majeur de ce type de modèles tient justement dans leur exhaustivité. Si l'utilisation de très nombreuses variables permet de s'approcher au plus près de la réalité biologique, il est également nécessaire de pouvoir paramétrer ces variables afin de leur donner des gammes de valeurs possibles ou optimales. Beaucoup de ces paramètres ne pouvant être directement inférés de l'observation de systèmes biologiques réels, cette tâche peut s'avérer longue et coûteuse en tests et en essais.

Le modèle de Wagner

Le modèle dit « de Wagner » ([WAGNER \[1996\]](#)) a la particularité de modéliser non seulement le développement d'un individu (passage du génotype au phénotype au cours de plusieurs pas de temps), mais aussi l'évolution de populations entières, à court comme à long terme, ce qui le rend particulièrement adapté à l'étude des propriétés évolutives des réseaux.

Le génotype est représenté par un réseau de régulation (voir figure 1.8). Du point de vue de la modélisation, ces interactions sont codées sous la forme d'une matrice carrée W , qui peut prendre des valeurs positives (activateur), négatives (répresseur) ou égales à zéro (pas d'interaction entre les gènes). La proportion d'interactions différentes de zéro est déterminée par une valeur c , représentant la densité de connexion du réseau.

L'expression des gènes du réseau constitue le phénotype de l'individu. À l'état

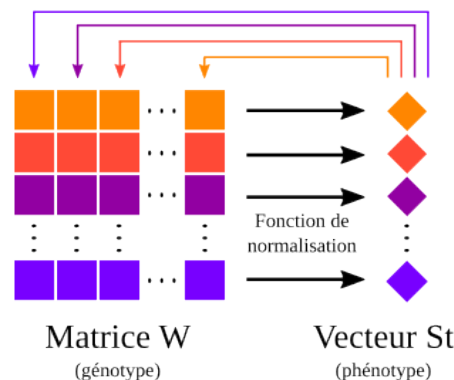


FIGURE 1.8 – REPRÉSENTATION SCHEMATIQUE D'UN RÉSEAU DE RÉGULATION SELON LE MODÈLE DE WAGNER. Chaque gène est régulé par les produits des autres gènes, via des sites de fixation. D'après [SIEGAL et BERGMAN \[2002\]](#).

initial, l'expression des gènes a la valeur $S_0 = 0$. Elle évolue au cours des différents pas de temps du développement selon la formule : $S_{t+1} = \sigma(WS_t)$, où la fonction $\sigma(x)$ est une fonction à seuil qui force les valeurs du phénotype à prendre les valeurs $S_t = -1$ (gène totalement réprimé) ou $S_t = +1$ (gène exprimé à sa valeur maximale). Le développement de l'individu continue de s'incrémenter jusqu'à ce que les gènes du réseau atteignent leur stabilité développementale S^∞ , c'est à dire lorsque $S_{t+1} = S_t$.

La complexité des interactions entre génotype (matrice d'interactions) et phénotype (expression des gènes) au cours du développement se traduit par une relation hautement non linéaire entre génotype et phénotype. Même dans les cas les plus simples (un ou quelques loci), l'expression des gènes à l'équilibre ne peut pas s'exprimer avec des fonctions classiques, ce qui rend tout développement mathématique extrêmement complexe.

La valeur sélective d'un individu est calculée en fonction de la proximité entre le S^∞ qu'il atteint et un S^{opt} optimal fixé. Plus le S^∞ et le S^{opt} sont éloignés, plus la valeur sélective diminue. Si aucun équilibre développemental n'est atteint, l'individu est considéré comme non viable, et sa valeur sélective tombe à zéro. L'absence d'équilibre à la fin du développement est assez courante dans les modèles de Wagner, souvent caractérisés par une forte cyclicité d'un ou plusieurs gènes du réseau, dont l'expression oscille alors entre deux (ou plus) valeurs distinctes.

Les individus se reproduisent de manière sexuée, proportionnellement à leur valeur sélective. Lors de la reproduction, ils sont soumis à la recombinaison et aux mutations (changements au sein de la matrice W).

La population, quant à elle, est soumise à l'action de la dérive génétique et de la sélection stabilisante pour le S^{opt} .

Ce modèle initial a été largement repris et modifié au cours des deux décennies suivantes (voir [LE CUNFF et PAKDAMAN \[2012\]](#) pour une revue), donnant lieu à des avancées significatives dans l'étude des architectures génétiques, de l'évolvabilité, ou encore de la canalisation. La section suivante présente une revue de ces avancées.

1.3.3.2 Apports du modèle de Wagner et de ses dérivés

Canalisation et assimilation génétique

Dès son article fondateur, [WAGNER \[1996\]](#) s'intéresse au rôle de la canalisation dans les réseaux géniques (bien qu'il n'utilise pas ce terme, préférant parler de stabilité épigénétique). Il décrit l'aspect canalisant de la sélection stabilisante, en observant que de longues périodes de ce régime conduisent à une réduction significative du nombre de mutations causant un changement dans l'expression du phénotype, et remarque également que ce résultat dépend de certaines propriétés du réseau, notamment la taille et le nombre de connexions (ce que nous appelons complexité) : plus le réseau est

grand et complexe, plus la canalisation est forte.

Des études ultérieures désavouent cette relation complexité-canalisation. **LECLERC [2008]** montre ainsi qu'en prenant en compte le cout de la complexité des réseaux (et notamment le cout de l'introduction d'une nouvelle interaction), la canalisation est plus importante quand les réseaux sont peu complexes et présentent une faible densité d'interactions.

L'idée de l'influence de la sélection stabilisante sur la canalisation par **WAGNER [1996]** est contredite par des études ultérieures. Notamment, **SIEGAL et BERGMAN [2002]** montrent que l'apparition de la canalisation génétique peut se faire même en l'absence de sélection vers un optimum phénotypique donné. Elle serait due principalement à la nécessité d'avoir une expression des gènes stable tout au long du développement. La canalisation serait donc une propriété émergente de tout développement complexe et ne résulterait pas d'un processus évolutif. Les auteurs se placent ainsi dans le courant de l'hypothèse intrinsèque.

L'hypothèse intrinsèque est également soutenue par l'étude de **HUERTA-SANCHEZ et DURRETT [2007]**, qui examine plusieurs modèles dérivés de celui de Wagner et montre que, tant que la taille de population est suffisante, la canalisation apparaît dans tous les cas : elle ne serait donc pas dépendante des détails du modèle, mais serait un effet secondaire de la sélection pour un meilleur succès reproducteur (ou une meilleure valeur sélective).

L'hypothèse congruente, qui explique l'apparition de la canalisation génétique comme un dérivé de la canalisation contre les perturbations environnementales, est quant à elle soutenue par les travaux de **LE CUNFF et PAKDAMAN [2012]**. Les auteurs démontrent que sélectionner l'expression des gènes vers un optimum donné au cours du développement augmente également la stabilité d'autres phénotypes cryptiques, et que cette forme de canalisation rend le réseau moins sensible face à d'autres perturbations, dont les mutations. Ainsi, l'évolution sélectionnerait des individus robustes à des perturbations qu'ils ont déjà expérimentées, mais aussi face à des perturbations nouvelles.

Les études de **CILIBERTI et collab. [2007a,b]** se placent à l'interface entre hypothèse congruente et hypothèse évolutive. Les auteurs ajoutent au modèle une perturbation de l'état initial (« bruit ») et étudient l'évolution de la canalisation dans des réseaux présentant des topologies différentes. Ils montrent ainsi que la canalisation aux mutations et la canalisation aux perturbations de l'état initial sont corrélées, avec quelques réseaux présentant un taux de canalisation bien plus élevé que les autres.

Ils montrent également que tous les réseaux permettant d'atteindre un phénotype donné sont connectés, formant un grand réseau neutre (qu'ils appellent métagraphe) qui peut être traversé par des changements progressifs de topologie. Ainsi, un réseau très sensible aux mutations ou aux perturbations peut devenir progressivement très

canalisé, prouvant l’aspect évolutif de la canalisation. Ces métagraphes peuvent également faciliter l’apparition de variants génotypiques, selon un mécanisme proche de celui des réseaux neutres ou de la variabilité cryptique.

Waddington avait lié la notion de canalisation à celle d’assimilation génétique. Cette idée a été reprise par **MASEL [2004]**, qui ajoute au modèle de Wagner des perturbations environnementales similaires aux chocs thermiques des expériences de **WADDINGTON [1953]** pendant le développement des individus. Elle montre qu’un processus d’assimilation génétique peut avoir lieu même en l’absence de sélection sur le phénotype (et donc que la sélection sur la stabilité de l’expression génique au cours du développement est suffisante). Tout comme la canalisation, l’assimilation génétique serait donc une propriété émergente des processus développementaux complexes.

Évolvabilité

La capacité des réseaux d’interactions (modélisés selon le modèle de Wagner) à répondre à la sélection est étudiée dès le début des années 2000.

Ainsi, **BERGMAN et SIEGAL [2003]** introduisent la possibilité de supprimer un gène du réseau, et montrent qu’une telle délétion conduit à une perte de la canalisation ainsi qu’à une augmentation de la diversité phénotypique et à une accélération de l’adaptation à un nouvel optimum, ceci étant dû à l’accumulation (pendant la phase canalisée) de variants génétiques cryptiques. Les auteurs valident ces observations par des études d’invalidation génétique (knockout) chez des levures.

Le lien entre canalisation et potentiel évolutif (soit la capacité d’une population à produire de la variabilité phénotypique) est également repris par **ESPINOSA-SOTO et collab. [2011a,b]**. Pour cela, les auteurs ajoutent au modèle des perturbations environnementales internes (comme **CILIBERTI et collab. [2007b]**) et externes (suivant **MASEL [2004]**) et montrent que la canalisation du phénotype promeut la variabilité phénotypique en réponse à des perturbations environnementales, mais pas en réponse à des mutations.

L’évolvabilité est également étudiée pour elle-même, les auteurs cherchant les conditions les plus favorables à sa mise en place. Ainsi, **DRAGHI et WAGNER [2009]** introduisent des environnements changeants (sélection fluctuante sur des phénotypes qui changent de manière périodique) et montrent que la sélection naturelle facilite l’adaptation rapide à de nouveaux environnements (augmentation de l’évolvabilité), et ce indépendamment du mode de reproduction.

Dans une démarche assez similaire, **FIERST [2011a]** ajoute au modèle une population évoluant dans des environnements plastiques : à chaque génération, les individus sont aléatoirement placés dans deux environnements A ou B, qui diffèrent par les valeurs d’initialisation du vecteur S_0 et par les optimums de fitness S^{opt} . Elle démontre ainsi

qu'une population soumise à une sélection vers une plasticité phénotypique accrue s'adapte plus rapidement à un nouvel environnement qu'une population uniquement soumise à un environnement stable. Elle montre également que cela est influencé par la pression de sélection dans l'environnement initial : une sélection faible facilite la plasticité phénotypique et donc la capacité d'adaptation.

PINHO et collab. [2015] étudient quant à eux l'effet du bruit environnemental (perturbation de l'état initial ou perturbation du développement) sur la diversité phénotypique. Ils montrent que ces perturbations tendent à diminuer l'accessibilité à de nouveaux phénotypes lorsque l'expression des gènes est binaire (0 ou 1), mais qu'elles tendent à l'augmenter légèrement si les gènes peuvent prendre davantage de valeurs différentes d'expression. Le bruit environnemental pourrait ainsi augmenter la diversité phénotypique, mais uniquement dans des conditions particulières.

Enfin, **WILDER et STANLEY [2015]** s'intéressent à la différence entre évolvabilité au niveau individuel (la capacité d'un seul génotype à générer de nouveaux phénotypes) et évolvabilité au niveau populationnel (la capacité de l'ensemble des génotypes de la population à générer de nouveaux phénotypes), en testant plusieurs types de sélection. Ils montrent que la sélection divergente (favorisant la diversité phénotypique) et les processus neutres favorisent une évolvabilité populationnelle, alors que les sélections fluctuantes et stabilisantes favorisent une évolvabilité individuelle.

Reproduction

AZEVEDO et collab. [2006] sont les premiers à avoir utilisé le modèle de Wagner pour l'étude des modes de reproduction. Ils démontrent que les populations sexuées sont plus fortement canalisées (dû au rôle de la recombinaison), ce qui tend à favoriser l'émergence d'épistasie négative (sur la valeur sélective). La reproduction sexuée sélectionne ainsi des propriétés génétiques qui favorisent sa propre maintenance. À l'inverse, les populations asexuées sont moins canalisées et favorisent l'apparition d'épistasie positive.

Ces travaux sont repris par **MACCARTHY et BERGMAN [2007]**, qui testent la compétition entre populations sexuées et populations asexuées en présence d'épistasie. Ils montrent ainsi que, si le mode de reproduction et l'épistasie sont laissés libres d'évoluer, la reproduction asexuée surpasse la reproduction sexuée, et mettent également en évidence une corrélation nette entre le taux d'épistasie antagoniste et le temps de fixation du mode de reproduction asexuée.

Les résultats de ces deux travaux sont cependant réfutés par l'étude de **LOHAUS et collab. [2010]**. Les auteurs s'intéressent aux bénéfices à long terme de la reproduction sexuée et observent comme précédemment que les populations sexuées sont plus canalisées que les asexuées. Cependant, eux n'y trouvent pas de preuve du rôle de

l'épistasie négative. Les auteurs notent également que c'est la robustesse à la recombinaison (c'est-à-dire le fait d'être peu sensible aux réarrangements du génome lors de la gamétogénèse) qui diminue le cout du sexe (permettant ainsi aux populations de résister à un envahissement par des individus asexués), et non un plus faible taux de mutation.

Cet effet de la recombinaison a également été étudié du point de vue de la canalisation. Cependant, l'étude de [MARTIN et WAGNER \[2009\]](#) n'a montré qu'un impact réduit de la recombinaison comparé à celui des mutations ponctuelles.

Dans une approche très différente, [FIERST \[2011b\]](#) modifie le modèle de Wagner pour y introduire un trait présentant un dimorphisme sexuel. Elle montre ainsi que les caractères dimorphiques augmentent non seulement la canalisation génétique, mais également l'évolvabilité, ce qui serait dû aux déséquilibres de liaison intra- et inter-sexes.

Architecture génétique

Dès les premières études qui mèneront à la mise au point de son modèle, Wagner s'intéresse à l'importance de l'architecture génétique sur l'évolution ([WAGNER \[1994\]](#)). Il montre qu'une grande partie des gènes d'un réseau sont dupliqués, et propose que l'évolution des réseaux de régulation passe majoritairement par des événements de duplication.

Les études sur l'architecture génétique ne seront reprises que quelques années plus tard. [SIEGAL et collab. \[2007\]](#) étudient les relations entre les propriétés évolutives d'un réseau et sa topologie, notamment sa complexité (ou connectivité) et le nombre d'interactions par gène, mais ne trouvent aucune corrélation fiable. En revanche, d'autres études ont montré une corrélation de cette même topologie avec des réponses à la sélection secondaire comme la canalisation (voire la partie [1.3.3.2](#)).

Si l'importance de l'architecture sur les capacités d'évolution a été étudiée, la question inverse a également été posée : comment évolue le réseau génétique en réponse à la sélection appliquée sur le phénotype qui en découle ? L'étude de [SEVIM et RIKVOLD \[2008\]](#) démontre ainsi que l'effet de l'évolution de la canalisation génétique sur la dynamique du réseau de régulation est mineur et que les réseaux les plus robustes sont également les plus chaotiques.

Enfin, d'autres études se sont intéressées à décrire le comportement du modèle. C'est par exemple le cas de [RHONE et collab. \[2011\]](#), qui s'intéressent aux relations entre la force de la sélection appliquée sur les gènes et la force des régulations à l'intérieur de la matrice. Ils montrent d'une part l'existence d'une corrélation positive entre ces deux aspects, et d'autre part le fait que les gènes répondant le mieux à la sélection sont ceux qui sont le plus faiblement régulés et/ou sont les régulateurs les plus puissants.

PINHO et collab. [2012], eux, ont étudié la dynamique de l'expression des gènes pendant le développement. Ils démontrent que la majeure partie des réseaux d'un modèle de Wagner sont cycliques (une observation déjà bien connue), et que cela est dû aux complexes interactions épistatiques entre les gènes : plus les réseaux sont peu connectés, plus ils ont de chances d'être stables. Les auteurs montrent également que ces conclusions sont robustes aux variations d'autres paramètres du réseau.

Enfin, **ESPINOSA-SOTO et WAGNER [2010]** mettent en évidence la manière dont la modularité (l'organisation du réseau en sous-réseaux bien définis et interagissant peu, voire pas) peut apparaître dans des réseaux comme un effet secondaire d'une spécialisation dans l'activité des gènes : elle aurait pour rôle de diminuer les interférences entre des groupes de gènes ayant des fonctions séparées.

1.4 Contexte de la thèse

L'étude des réseaux de régulation et, plus spécifiquement, des dérivés du modèle de Wagner, a permis d'éclaircir certaines des conditions d'évolution de la canalisation dans une architecture génétique complexe. Cependant, aucune des trois hypothèses expliquant l'origine de ce processus n'a encore été formellement validée ou infirmée. Cela peut être dû au fait que beaucoup d'études sont difficilement comparables, notamment parce qu'elles utilisent des paramètres ou des méthodes de calcul des variables évolutives (dont la canalisation) différents. Par ailleurs, la plupart portent sur des valeurs qualitatives, et non quantitatives, ce qui est susceptible d'influencer les résultats observés.

Dans la suite de ce manuscrit, je présenterai les résultats de mes trois années de thèse, qui visaient à déterminer, d'une part, la capacité des réseaux génétiques complexes à faire évoluer de la canalisation génétique, et d'autre part, les mécanismes impliqués dans sa mise en place. Cette approche *in silico* a été réalisée à partir d'un modèle quantitatif de réseau de régulation, contrastant ainsi avec la majorité des études préalables.

Je m'intéresserai dans un premier temps aux propriétés du processus développemental. L'analyse de réseaux de régulation aléatoires me permettra d'étudier l'état initial des simulations, de déterminer certaines de ses caractéristiques intrinsèques du modèle et de paramétrer les simulations ultérieures.

Je passerai ensuite de l'échelle développementale à l'échelle populationnelle, en explorant la capacité des réseaux de régulation à répondre à la sélection directionnelle et stabilisante sur un grand nombre de générations. Suite à la mise en évidence de

la présence de canalisation génétique dans le modèle, j'étudierai cette question sous l'angle des modes de reproduction et de leur impact éventuel.

Enfin, dans un article paru dans BMC Evolutionary Biology, je présenterai une série de simulations portant plus spécifiquement sur l'évolution de la canalisation génétique dans des réseaux de régulation et sur les mécanismes impliqués dans sa mise en place. Ces études sont effectuées sur des réseaux quantitatifs et qualitatifs, afin de déterminer l'impact de cet aspect du modèle sur les résultats obtenus.

Pour finir, je m'intéresserai brièvement aux hypothèses concurrentes sur l'origine de la canalisation, dans des travaux encore préliminaires.

Chapitre 2

Programme et modèle

Sommaire

2.1	Modèle théorique	25
2.1.1	Du génotype au phénotype	25
2.1.1.1	Le génotype	25
2.1.1.2	Le phénotype	26
2.1.2	De l'individu à la population	29
2.1.2.1	Reproduction	29
2.1.2.2	Sélection	29
2.1.3	Estimation de la canalisation	31
2.1.4	Autres modèles implémentés	32
2.2	Workflow d'une analyse complète	32
2.2.1	Routines de lancement	32
2.2.1.1	Fichier d'entrée	33
2.2.1.2	Fichier de sortie	36
2.2.2	Programme de simulation "Simevolv"	36
2.2.2.1	Cahier des charges	36
2.2.2.2	Organisation interne du programme	37
	GeneticMap	37
	Allele	37
	Haplotype	38
	Genotype	38
	Phenotype	38
	Individual	38
	Population	38
	Fitness	38
	Environment et Canalization	39

Architecture	39
2.2.2.3 Paramétrisation	39
2.3 Simulation développementale sous R	40

Dans le but d'éclaircir les conditions menant à l'évolution de la canalisation dans des architectures génétiques complexes (et plus spécifiquement dans des réseaux de régulation), une grande partie de mon travail de thèse a consisté en la mise au point et en l'implémentation d'un programme de simulation nommé Simevolv. Suivant le modèle initialement décrit par A. Wagner ([WAGNER \[1996\]](#)), ce programme permet d'une part la modélisation du développement d'un individu (passage du génotype au phénotype) et d'autre part la modélisation de l'évolution des populations sur de longues périodes de temps.

Dans ce chapitre, je détaillerai le modèle utilisé, avant de m'intéresser au fonctionnement global du programme Simevolv.

2.1 Modèle théorique

Dérivé du modèle de Wagner, notre modèle en garde les deux échelles de temps (développement et population) et la représentation des régulations entre gènes sous la forme d'une matrice d'interactions, mais s'en éloigne par plusieurs aspects.

2.1.1 Du génotype au phénotype

Chaque individu de la population est décrit par deux composants majeurs : son génotype et son phénotype. Le génotype est représenté par un réseau de régulation, codant pour des facteurs de transcription, dont l'expression à la fin du développement de l'individu constitue le phénotype (voir figure 1.8). Le génotype demeure constant au cours du développement de l'individu, alors que le phénotype évolue de manière dynamique.

2.1.1.1 Le génotype

Le génotype de l'individu est calculé comme la moyenne des deux haplotypes paternel et maternel, eux-mêmes constitués de L loci.

De manière concrète, un génotype (ou un haplotype) est représenté sous la forme d'une matrice carrée W de taille $L \times L$, qui représente les interactions entre les produits des gènes et les facteurs de transcription au sein du réseau (voir figure 2.1). Chaque ligne i de cette matrice représente un allèle, ici décrit comme une série de sites potentiels de fixation pour des facteurs de transcription j . Chacune des cases w_{ij} de la matrice représente l'effet du produit du gène j sur l'expression du gène i . Cet effet peut être

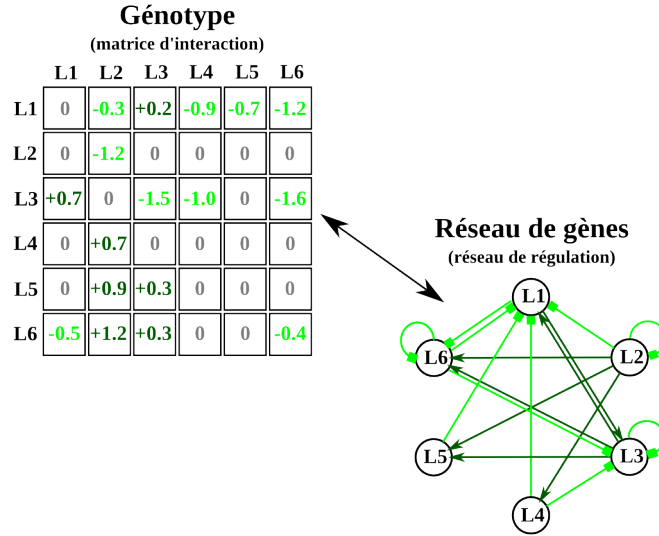


FIGURE 2.1 – EXEMPLE DE MATRICE D’INTERACTION ET RÉSEAU DE RÉGULATION CORRESPONDANT. Réseau de régulation à $L = 6$ loci, complexité de $c_m = 0,5$. Le vert sombre correspond aux activateurs, le vert clair aux répresseurs.

nul ($w_{ij} = 0$), activateur ($w_{ij} > 0$) ou répresseur ($w_{ij} < 0$); plus la valeur absolue est élevée, plus l’effet est fort.

La fréquence des interactions ($w_{ij} \neq 0$) est déterminée par un facteur de complexité c_m ; une valeur $c_m = 0,4$ indique par exemple que 40% des cases sont différentes de zéro. Pour chaque w_{ij} , une valeur est tirée dans une loi uniforme entre 0 et 1. Si cette valeur est supérieure à c_m , la case prend la valeur zéro (pas d’interactions). Si elle est inférieure à c_m , la case représente une interaction dont la force est tirée dans une loi gaussienne (dont la moyenne et la variance peuvent être paramétrées). Les valeurs par défaut sont $\mathcal{N}(0;0,1)$ à l’initialisation du réseau, de telle sorte qu’il y ait autant de répresseurs que d’activateurs et que la force des régulateurs reste faible. Un second paramètre c_d détermine de manière similaire la fréquence des interactions sur la diagonale de la matrice, permettant de gérer la présence ou l’absence d’auto-régulation.

2.1.1.2 Le phénotype

Le phénotype de l’individu est défini comme l’expression de chacun des facteurs de transcription codés par le génotype. Il change à chaque pas de temps du développement d’un individu (voir figure 2.2).

En pratique, le phénotype est représenté sous la forme d’un vecteur S_t de taille L , aux valeurs comprises entre 0 et 1. $S_t = 0$ indique que le gène n’est pas exprimé, $S_t = 1$ qu’il est exprimé à son taux maximal, et une valeur de $S_t = 0,2$ qu’il est exprimé à 20% de sa capacité maximale. En absence de toute régulation, l’expression constitutive des gènes du réseau est fixée à une valeur de $a = 0,2$. Cette valeur paraissait biologiquement

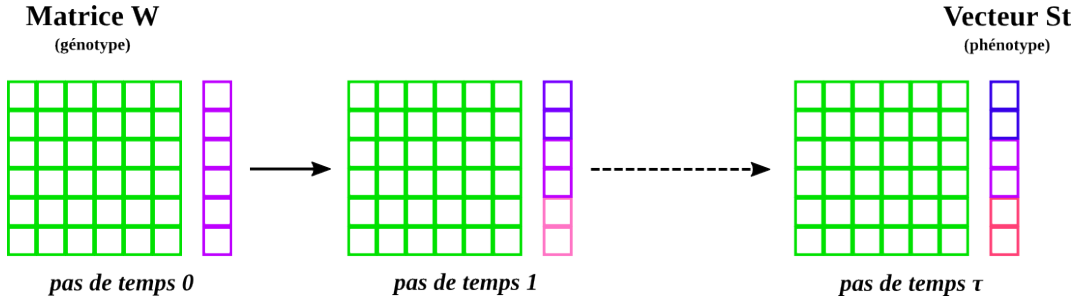


FIGURE 2.2 – COMPORTEMENT AU COURS DU DÉVELOPPEMENT. La matrice W ne change pas au cours des pas de temps du développement, alors que le vecteur S_t évolue.

plus réaliste que celle de $a = 0,5$ fréquemment trouvée dans la littérature, tout en permettant la régulation par répression.

Au début du développement, l'expression des L gènes du réseau est fixé à a . L'expression change ensuite au cours de $\tau = 16$ pas de temps, en fonction du génotype et de l'expression de chaque gène au pas de temps précédent, selon la formule suivante :

$$S_{t+1} = F(W S_t), \quad (2.1)$$

avec

$$F(s) = \begin{pmatrix} f(s_1) \\ f(s_2) \\ \vdots \\ f(s_L) \end{pmatrix} \quad (2.2)$$

où

$$f(x) = \frac{1}{1 + \left(\frac{1}{a} - 1\right) \exp\left(\frac{-x}{a(1-a)}\right)} \quad (2.3)$$

La fonction normalisatrice $f(x)$ est une sigmoïde (voir figure 2.3), agissant sur chacune des valeurs de S_t , et dont le rôle principal est d'assurer la mise à l'échelle

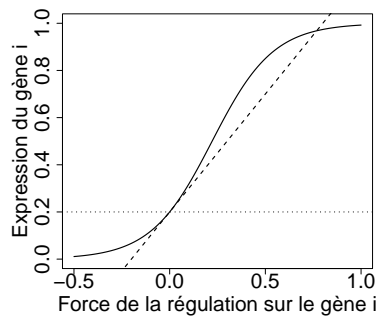


FIGURE 2.3 – NIVEAU D'EXPRESSION DU GÈNE i EN FONCTION DE LA FORCE DE LA RÉGULATION. $S_{i,t+1} = f(\sum_j w_{ij} S_{jt})$. La ligne discontinue représente la pente $df/dx = 1$ au point $f(x) = 0$; la ligne pointillée représente l'expression constitutive des gènes ($a = 0,2$).

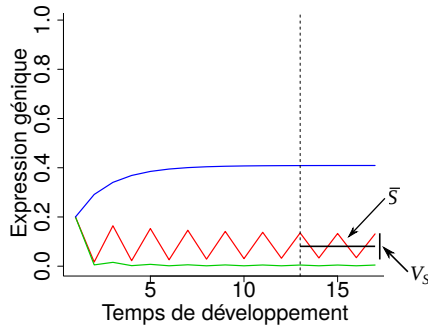


FIGURE 2.4 – CALCUL DU PHÉNOTYPE DE L’INDIVIDU EN FIN DE DÉVELOPPEMENT. Calculé sur les $\tau_c = 4$ derniers pas de temps du développement, le phénotype dépend de deux facteurs : l’expression moyenne des gènes \bar{S} et la variance de l’expression génique V_S (représentant l’instabilité développementale).

des S_t entre 0 et 1 (en effet, si $S_t \in [0, 1]$, $W \in [-\infty, +\infty]$, nécessitant cette remise à l’échelle). Elle permet également d’avoir une pente de $df/dx = 1$ quand $x = 0$, ce qui met à l’échelle la matrice W de telle sorte que la valeur des w_{ij} tende vers des changements observables de l’expression génique (c’est à dire des changements proportionnels aux changements dans la force de régulation).

L’une des caractéristiques des réseaux de gènes est leur cyclicité (voir section 3.1.2) : au cours du développement, certains gènes présentent une expression instable qui peut ou non s’amortir au cours des pas de temps. Cette instabilité développementale peut être quantifiée en calculant la variance V_S de l’expression génique au cours du temps.

L’environnement peut influencer sur le développement de l’individu de différentes manières. Dans ce modèle, cet effet de l’environnement peut être décrit de deux façons différentes : perturbation de l’état initial et perturbation au cours du développement. Les perturbations de l’état initial sont codées comme une modification du vecteur initial S_0 : un modificateur tiré dans une loi Normale $\mathcal{N}(0, \sigma_i)$ est ajouté de manière aléatoire à l’expression de chacun des gènes du vecteur. Le principe est le même pour les perturbations développementales : à chaque pas de développement, chacun des gènes est modifié par une valeur tirée d’une loi Normale $\mathcal{N}(0; \sigma_d)$.

La valeur des S_t est calculée sur les $\tau_c = 4$ derniers pas de temps de développement, permettant d’éviter les variations de l’expression génique apparaissant dans les premiers stades du développement. Le phénotype de l’individu dépend ainsi de deux facteurs : l’expression moyenne des gènes \bar{S} , ainsi que l’instabilité développementale, définie comme la variance de l’expression génique V_S au cours du temps du développement (voir figure 2.4). Ces deux scores sont calculés pour chacun des gènes du réseau, et peuvent ensuite être moyennés à l’échelle de l’individu.

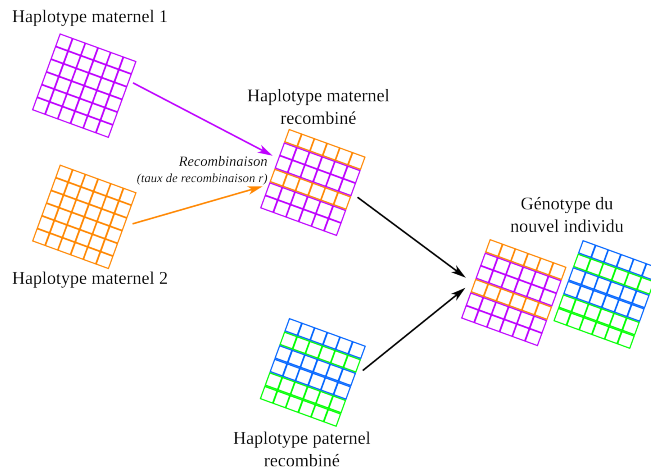


FIGURE 2.5 – REPRÉSENTATION SCHÉMATIQUE DE LA REPRODUCTION SEXUÉE ET DE LA RECOMBINAISON.

2.1.2 De l'individu à la population

2.1.2.1 Reproduction

La population initiale est composée de N individus diploïdes, tous clones, c'est-à-dire qu'ils possèdent tous le même génotype (la même matrice W). Les G générations suivantes, de taille constante, sont produites en simulant de la reproduction sexuée entre individus hermaphrodites.

Les parents de chaque nouvel individu sont tirés au sort dans la population, proportionnellement à leur valeur sélective (voir section 2.1.2.2 ci-dessous). Les gamètes sont formés en tirant aléatoirement chacun des allèles (chacune des lignes de la matrice) parmi les deux haplotypes de chacun des deux parents et en les recombinant selon un taux de recombinaison r (voir figure 2.5).

Le nouveau génotype ainsi formé subit ensuite des mutations selon un taux de mutation par génome haploïde μ . Les mutations affectent les gènes au hasard : un nombre est tiré dans une loi uniforme et si sa valeur est inférieure au taux de mutation μ/L , le gène est considéré comme muté. Un élément non-zéro de la ligne correspondant à ce gène de la matrice W est tiré au hasard puis modifié par l'ajout d'un modificateur tiré dans une loi Normale \mathcal{N} (valeur initiale, σ_m). Ainsi, un individu peut subir plusieurs mutations, mais chaque locus n'est muté qu'une fois.

2.1.2.2 Sélection

La valeur sélective ω d'un individu est une mesure de l'effet de la sélection naturelle, et décrit la capacité d'un individu à se reproduire. Dans notre modèle, elle dépend de deux choses : la distance entre l'expression génique atteinte et une expression génique optimale (ω_d , représentant la réponse à la sélection stabilisante), et l'instabilité déve-

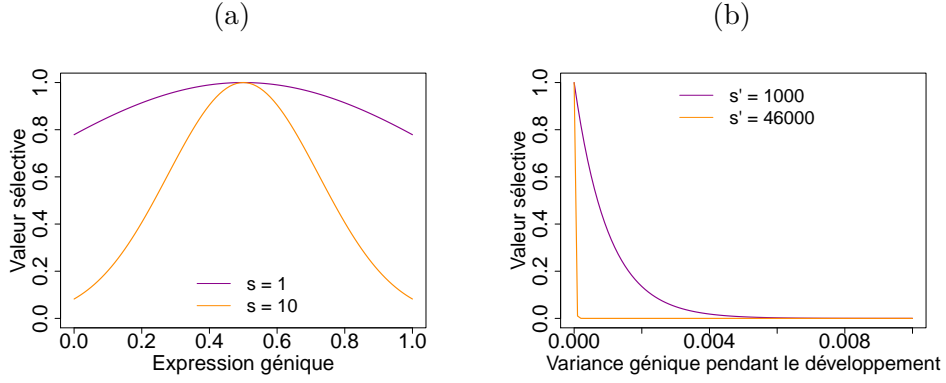


FIGURE 2.6 – FORCES DE SÉLECTION APPLICABLES SUR LE RÉSEAU. Dans les deux cas, la courbe orange présente une sélection plus forte que la courbe violette. (a) Sélection stabilisante sur la distance à l'optimum s . (b) Sélection sur l'instabilité développementale s' .

loppementale (ω_k , réponse à la sélection contre l'instabilité développementale). Elle est calculée selon :

$$\omega = \omega_d \times \omega_k, \quad (2.4)$$

avec

$$\omega_d = \exp\left[s \sum_{i=1}^{\ell} (\bar{S}_i - \theta_i)^2\right] \quad (2.5)$$

et

$$\omega_k = \exp\left[-s' \sum_i^L V_{S_i}\right], \quad (2.6)$$

où \bar{S}_i est l'expression moyenne du gène i , θ_i l'expression optimale du gène i , $\ell \leq L$ le nombre de gènes du réseau sur lesquels la sélection stabilisante a un impact direct, V_{S_i} l'instabilité développementale, s la force de la sélection stabilisante (voir figure 2.6a)

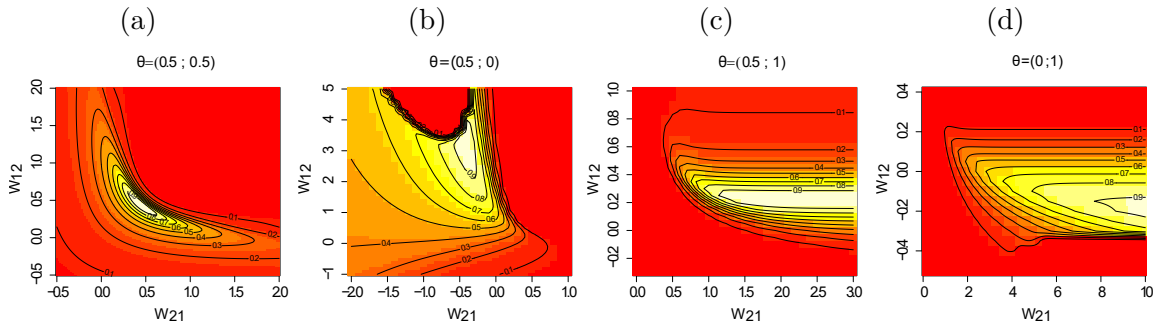


FIGURE 2.7 – PAYSAGE ADAPTATIF. Réseau de deux gènes, pas d'autorégulation (valeurs de la diagonale de la matrice fixées à zéro), $s = 10$, $s' = 46000$ (tous les gènes sont contre-sélectionnés pour l'instabilité développementale). Plus la couleur est proche du blanc, meilleure est la valeur sélective. (a) Optimums des gènes 1 et 2 fixés à $\theta_1 = \theta_2 = 0,5$. (b) $\theta_1 = 0,5$ et $\theta_2 = 0$. (c) $\theta_1 = 0,5$ et $\theta_2 = 1$. (d) $\theta_1 = 0$ et $\theta_2 = 0$.

et s' la force de la sélection contre l'instabilité développementale (voir figure 2.6b).

La complexité des relations épistatiques couplée à ces pressions de sélection se traduit par un paysage adaptatif complexe, même dans des cas simplifiés (figure 2.7 : réseau de deux gènes, absence d'autorégulation). Ainsi, vouloir modifier l'expression de l'un des gènes du réseau implique de modifier également celui des autres.

2.1.3 Estimation de la canalisation

La canalisation peut-être calculée pour les différents types de perturbations : génétique (mutations) et environnementale (perturbations du vecteur d'expression). Tous ces scores de canalisation sont calculés pour chacun des gènes i du réseau et stockés sous la forme d'un vecteur de taille L . La formule générale est la suivante :

$$C_i = -\log \text{Var}(M_i) \quad (2.7)$$

où $\text{Var}(M_i)$ est la variance de l'expression génique du gène i , parmi les mutants qui diffèrent de l'individu initial par une seule perturbation ou mutation. Concrètement, lors du calcul de la canalisation, chaque individu de la population est cloné cent fois. Chacun de ces clones subit une seule mutation ou perturbation selon les modalités décrites ci-dessous. Chacun des clones ainsi modifiés subit un processus développemental complet et l'expression génique est calculée. La variance entre les clones est calculée pour chaque gène, puis transformée logarithmiquement.

Pour la canalisation génétique C_{mut} , la mutation touche la matrice W de la même façon que les mutations réalisées dans la simulation principale. Elle se fait sur un haplotype, un locus et un allèle tiré au hasard, mais dispose de son propre σ_m pour des raisons de reproductibilité (notamment lors des analyses faisant varier le σ_m dans les simulations proprement dites).

Pour les canalisations environnementales C_{init} et C_{dev} , les perturbations se font comme suit. Les perturbations de l'état initial touchent le vecteur S_0 , et consistent en l'ajout (à chaque locus du vecteur) d'un modificateur tiré dans une loi Normale centrée sur la valeur initiale et de déviation standard σ_i . Les perturbations au cours du développement se produisent à chaque pas de temps et touchent tous les gènes (modificateurs tirés dans une loi Normale centrée sur les valeurs de S_t et de déviation standard σ_d). Pour toutes les études présentées dans ce manuscrit, à l'exception des études préliminaires présentées en section 5.2.3.2, les modificateurs environnementaux sont fixés à 0 (pas de perturbations environnementales).

2.1.4 Autres modèles implémentés

Des modèles alternatifs ont été implémentés dans le programme de simulation. Ils n'ont cependant pas été testés ou explorés au cours de la thèse, le sujet de recherche s'étant vite concentré sur une étude plus exhaustive du modèle principal.

Le premier est un modèle additif. Le génotype est représenté comme un vecteur de taille L dont les valeurs initiales sont tirées dans une loi normale. Le phénotype est simplement la somme des valeurs génotypiques, les effets des gènes étant indépendants les uns des autres.

Le second est un modèle multilinéaire qui rajoute de l'épistasie au modèle additif (HANSEN et WAGNER [2001]). Le génotype est toujours codé comme un vecteur de taille L . Le phénotype est la somme des valeurs génotypiques, à laquelle s'ajoutent des composantes épistatiques.

Les suivants sont des implémentations de deux variants du modèle de Wagner provenant de la littérature (SIEGAL et BERGMAN [2002]; WAGNER [1996]). Ils diffèrent du modèle utilisé majoritairement dans cette étude et décrit plus haut par l'allure de la fonction $f(x)$, et par l'intervalle bornant l'expression des gènes S_t (voir l'introduction pour les détails.)

Le dernier est un modèle booléen binaire, développé et utilisé par Christine Mayer, collaboratrice de l'Université d'Oslo. Le génotype est encodé comme un vecteur binaire (0/1) de L gènes, le phénotype est un vecteur binaire d'une taille pouvant aller de 1 à L gènes, et la transition de l'un à l'autre est déterminée par des connecteurs logiques (AND, OR, NOT).

2.2 Workflow d'une analyse complète

Une analyse complète (des fichiers de paramètres jusqu'aux sorties graphiques) fait intervenir plusieurs langages de programmation : les routines de lancement sont scriptées en Bash, le coeur des simulations (le programme Simevolv) est codé en C++, et les analyses statistiques et graphiques sont réalisées sous R (version 3.0.2 Frisbee Sailing).

2.2.1 Routines de lancement

Les routines de lancement sont scriptées en Bash. Elles indiquent notamment le chemin du fichier d'entrée utilisé par Simevolv (le fichier de paramètre, voir section 2.2.1.1 ci-dessous), le nombre de simulations devant être réalisées à partir du même jeu

d'entrée (correspondant à des réplicats différents), le nombre de coeurs du processeur pouvant être impliqué dans l'étude (et donc le nombre de simulations réalisées en parallèle), ainsi que les noms et chemins des fichiers de sorties (sorties du logiciel Simevolv, voir section 2.2.1.2 ci-dessous).

Ces routines permettent également de modifier les fichiers d'entrée, et de pouvoir ainsi lancer plusieurs analyses successives qui diffèrent par les valeurs prises par une variable donnée (tester des tailles de populations différentes, par exemple). Un exemple annoté peut être trouvé en figure 2.8.

2.2.1.1 Fichier d'entrée

Le fichier d'entrée consiste en un fichier de paramètres au format .txt donnant toutes les valeurs permettant l'initialisation du programme. Un exemple de fichier récapitulant les paramètres de base utilisés pour nos analyses peut être trouvé en table 2.1.

	Notation	Paramètre	Valeur
Modèle théorique utilisé		TYPE_ARCHI	m2
Nombre de générations	G	SIMUL_GENER	10 000
Nombre de générations séparant deux sorties	g	SIMUL_OUTPUT	1000
Taille de la population	N	INIT_PSIZE	5000
Taux de ploïdie		GENET_PLOIDY	2
Taux d'autofécondation	f	GENET_SELFING	0,0
Nombre de loci	L	GENET_NBLOC	6
Taux de recombinaison	r	GENET_RECRATES	0,5
Initialisation des allèles	$\mathcal{N}(mean, sd)$	INIT_ALLELES	0,0 0,1
Population initiale clonale ou variable		INIT_CLONAL	clonal
Complexité de la matrice W	c_m	INIT_CONNECT	0,5
Complexité de la diagonale de la matrice W	c_d	INIT_CONDIAG	0
Valeurs de l'état initial S_0		TYPE_SO	basal
Expression constitutive	a	INIT_BASAL	0,2
Nombre de pas de temps	τ	DEV_TIMESTEPS	16
Nombre de pas de temps pour le calcul des \bar{S} et V_S	τ_c	DEV_CALCSTEPS	2
Possibilité de muter les $w_{ij} = 0$		TYPE_ALLELES	zero
Calcul des mutations (par haplotype ou par locus)		GENET_MUTTYPE	individual
Taux de mutation	μ	GENET_MUTRATES	0,01
Effet des mutations	σ_m	GENET_MUTSD	0,1
Effet des perturbations de l'état initial	σ_i	ENVIRO_SDINIT	0
Effet des perturbations au cours du développement	σ_d	ENVIRO_SDDYNAM	0
Type de sélection pour l'optimum phénotypique	ω_d	FITNESS_TYPE	gaussian
Force de la sélection stabilisante	s	FITNESS_STRENGTH	10 10 0 0 0 0
Optimum phénotypique	θ	FITNESS_OPTIMUM	0,5
Type de sélection contre l'instabilité développementale	ω_k	FITNESS_STAB	exponential_stab
Force de la sélection contre l'instabilité développementale	s'	FITNESS_STABSTR	46000
Calcul des scores d'instabilité développementale		OUT_UNSTAB	yes
Calcul des valeurs génotypiques (valeurs des w_{ij})		OUT_GENO	yes
Nombre de tests pour le calcul de la canalisation		OUT_CANAL_TESTS	100
Effet des σ_m pour le calcul de la canalisation	σ_{mc}	OUT_CANAL_MUTSD	0,1
Effet des σ_i pour le calcul de la canalisation	σ_{ic}	OUT_CANAL_SDINIT	0
Effet des σ_d pour le calcul de la canalisation	σ_{dc}	OUT_CANAL_SDDYNAM	0

TABLEAU 2.1 – EXEMPLE DE FICHER D'ENTRÉE. Paramètres par défaut du modèle tel qu'utilisé pour les chapitres 3.3 et 4, soit un réseau de 6 loci, tous contre-sélectionnés pour l'instabilité développementale et dont 2 sont sous pression de sélection stabilisante.

```

#!/bin/bash
# LANCE UNE SIMULATION COMPLETE (standardisation , variables et repetitions)

# Variables
MAIN=~ /Models/Simulations/Constraints
DATE=$(date +%Y-%V)
NAME=CONS_param_basal
var_name=INIT_BASAL # parametre etudie
var=(0.1 0.3 0. 0.5 0.7 0.9) # valeurs prises par le parametre
n_rep=120 # nombre de replicats
n_proc=60 # nombre de processeurs utilisables
param=param.txt # nom du fichier de parametre

# creation du dossier de la simulation
echo -e "\n Debut de la simulation $NAME \n"
mkdir $MAIN/${DATE}_${NAME}

# pour chacune des valeurs du parametre teste :
for ((v=1 ; v<=${#var[@]} ; v++));
do
    echo -e "\n $var_name=${var[v-1]} \n"
    mkdir $MAIN/${DATE}_${NAME}/${var_name=${var[v-1]}}

    # creation et modification du fichier de parametre pour prendre en compte la
    # nouvelle valeur
    cp $MAIN/$param $MAIN/${DATE}_${NAME}/${var_name=${var[v-1]}}/param.txt
    sed -i -e "s/$var_name [[:blank:]].* /$var_name\t${var[v-1]}/" $MAIN/${DATE}
    _${NAME}/${var_name=${var[v-1]}}/param.txt

    # creation du launchfile
    for r in $(seq $n_rep)
    do
        echo ~/Models/Scripts/simevolv/bin/Release/Simul_Prog -p $MAIN/${DATE}
        _${NAME}/${var_name=${var[v-1]}}/param.txt -o $MAIN/${DATE}_${NAME}/
        $var_name=${var[v-1]}/rep_${r}.txt
    done > $MAIN/${DATE}_${NAME}/${var_name=${var[v-1]}}/launchfile.txt

    # lecture du launchfile : lancement des simulations , parallelisees sur les
    # $n_proc processeurs du serveur
    parallel -a $MAIN/${DATE}_${NAME}/${var_name=${var[v-1]}}/launchfile.txt -j
    $n_proc

    # analyse du resultat des repetitions (fichiers recapitulatifs avec moyenne et
    # variance entre les replicats)
    if [ $n_rep = 1 ]
    then
        ~/Models/Scripts/fonctions/analyze-repetitions.R -path $MAIN/${DATE}
        _${NAME}/${var_name=${var[v-1]}} -ptrn rep_ -rem TRUE -outm $MAIN/${
        DATE}_${NAME}/${var_name=${var[v-1]}}/final_mean.csv -outs $MAIN/${
        DATE}_${NAME}/${var_name=${var[v-1]}}/final_sd.csv -outg one
    else
        ~/Models/Scripts/fonctions/analyze-repetitions.R -path $MAIN/${DATE}
        _${NAME}/${var_name=${var[v-1]}} -ptrn rep_ -rem TRUE -outm $MAIN/${
        DATE}_${NAME}/${var_name=${var[v-1]}}/final_mean.csv -outs $MAIN/${
        DATE}_${NAME}/${var_name=${var[v-1]}}/final_sd.csv -outg mean
    fi
done

echo -e "\n Fin de la simulation $NAME \n"

```

FIGURE 2.8 – EXEMPLE DE SCRIPT DE LANCEMENT. Ce script permet de tester différentes valeurs d’expression constitutive.

Gen	MPhen1	MPhen2	MPhen3	VPhen1	VPhen2	VPhen3	MUnstab1	MUnstab2	MUnstab3	VUnstab1	VUnstab2	VUnstab3	MFit	VFit
1	0.27671	0.15089	0.25541	1.5821E-015	6.9389E-016	-1.3789E-015	0	0	-1.3878E-017	0	0	0	0.6074	-1.2157E-014
1000	0.49687	0.11193	0.092508	0.0012321	0.000050248	0.0025146	1.1263E-009	4.2543E-011	2.8764E-009	3.1271E-017	3.5187E-020	7.5713E-015	0.9879	0.00006657
2000	0.50608	0.11176	0.014221	0.0013027	0.0003014	0.00077538	4.3430E-009	1.3322E-010	5.4238E-011	2.7927E-016	2.4391E-019	1.6355E-019	0.98331	0.0014341
3000	0.49912	0.11907	0.014436	0.0016727	0.0004042	0.00016133	3.2622E-008	1.1232E-009	6.1272E-010	7.9749E-014	1.0024E-016	5.1923E-017	0.98412	0.0021606
4000	0.50506	0.11774	0.0031057	0.0010138	0.000031902	0.000054687	6.3802E-009	2.6204E-010	6.4134E-011	1.790E-015	1.0638E-017	3.7756E-019	0.98979	0.00075345
5000	0.49633	0.11743	0.00064769	0.00087843	0.000040534	3.0324E-006	1.2779E-009	5.3916E-011	3.6984E-013	2.2929E-017	3.2146E-020	2.9188E-023	0.99155	0.00072348
6000	0.50472	0.11837	0.00083162	0.00085898	0.000022062	1.4657E-006	2.5954E-009	6.3818E-011	2.1189E-012	4.4931E-017	6.9564E-022	0.99224	0.001289	
7000	0.50602	0.11758	0.000089139	0.00078656	0.000025969	0.000000025	8.248E-010	6.2112E-011	1.06E-014	2.2984E-020	9.2153E-027	0.99262	0.00084775	
8000	0.50201	0.11827	0.000024188	0.00078117	0.000032912	1.2031E-009	9.4105E-010	7.7805E-011	8.6118E-015	8.3751E-018	2.5963E-019	2.5926E-026	0.99231	0.00033247
9000	0.50467	0.11886	0.00005297	0.00095164	0.00011296	1.6135E-008	5.7916E-010	3.9416E-011	2.3314E-015	6.2573E-019	2.4851E-021	2.8826E-028	0.99186	0.0012775
10000	0.48976	0.11835	0.000018513	0.00097362	0.000085763	7.7558E-010	1.306E-009	1.6901E-010	1.0197E-014	2.0857E-017	7.1972E-019	6.258E-026	0.98948	0.00050906

Gen	FitOpt1	FitOpt2	FitOpt3	MGenCan1	MGenCan2	MGenCan3	VGenCan1	VGenCan2	VGenCan3	GenCanFit	MInitCan1	MInitCan2	MInitCan3	VInitCan1
1	0.5	0.5	0.5	-5.7716	-8.4011	-6.4265	1.2967	1.7875	1.8805	-3.1994	-20	-20	-20	0
1000	0.5	0.5	0.5	-5.4435	-8.4294	-8.6528	1.5411	4.0552	2.7413	-4.5492	-20	-20	-20	0
2000	0.5	0.5	0.5	-5.5651	-8.5173	-10.805	1.7777	4.4688	1.9322	-4.8098	-20	-20	-20	0
3000	0.5	0.5	0.5	-5.6844	-8.6144	-11.061	1.8884	4.6781	2.3231	-4.853	-20	-20	-20	0
4000	0.5	0.5	0.5	-5.6477	-8.5526	-14.925	1.9496	5.0727	8.5355	-4.8943	-20	-20	-20	0
5000	0.5	0.5	0.5	-5.5478	-8.4606	-16.968	1.9112	4.6453	7.2345	-4.7599	-20	-20	-20	0
6000	0.5	0.5	0.5	-5.5867	-8.5047	-16.053	1.9372	4.6652	7.0377	-4.8579	-20	-20	-20	0
7000	0.5	0.5	0.5	-5.6117	-8.5673	-18.655	2.09	4.4189	3.1713	-4.901	-20	-20	-20	0
8000	0.5	0.5	0.5	-5.6474	-8.5692	-19.428	1.8701	4.5731	1.2185	-5.0016	-20	-20	-20	0
9000	0.5	0.5	0.5	-5.5089	-8.3802	-19.006	1.9685	5.0531	2.3937	-4.6773	-20	-20	-20	0
10000	0.5	0.5	0.5	-5.6179	-8.5553	-19.521	1.9071	4.3958	1.0046	-4.8743	-20	-20	-20	0

Gen	VInitCan2	VInitCan3	InrCanFit	MDynamCan1	MDynamCan2	MDynamCan3	VDynamCan1	VDynamCan2	VDynamCan3	DynamCanFit	MeanAll1	MeanAll2	MeanAll3	MeanAll4
1	0	0	-20	-20	-20	-20	0	0	0	-20	0.15364	-0.19739	0.18858	0.19339
1000	0	0	-20	-20	-20	-20	0	0	0	-20	0.15407	-0.19827	-0.28545	1.1825
2000	0	0	-20	-20	-20	-20	0	0	0	-20	0.15311	-0.20129	-0.95382	1.2596
3000	0	0	-20	-20	-20	-20	0	0	0	-20	0.15467	-0.20021	-0.99464	1.1761
4000	0	0	-20	-20	-20	-20	0	0	0	-20	0.15335	-0.19974	-1.8421	1.2375
5000	0	0	-20	-20	-20	-20	0	0	0	-20	0.15333	-0.20447	-2.3831	1.2222
6000	0	0	-20	-20	-20	-20	0	0	0	-20	0.14987	-0.19769	-2.1295	1.2592
7000	0	0	-20	-20	-20	-20	0	0	0	-20	0.15109	-0.19969	-2.85	1.2704
8000	0	0	-20	-20	-20	-20	0	0	0	-20	0.15399	-0.19939	-3.2152	1.2364
9000	0	0	-20	-20	-20	-20	0	0	0	-20	0.15267	-0.19662	-2.9886	1.245
10000	0	0	-20	-20	-20	-20	0	0	0	-20	0.13826	-0.20439	-3.3078	1.2472

Gen	MeanAll5	MeanAll6	MeanAll7	MeanAll8	MeanAll9	VarAll10	VarAll11	VarAll12	VarAll13	VarAll14	VarAll15	VarAll16	VarAll17	VarAll18	VarAll19
1	0	0	-0.014167	0	-1.2837E-016	-5.8287E-016	7.0777E-016	-6.3144E-016	0	0	0	0	-1.3607E-017	0	7.8605E-019
1000	0	0	0.030703	0	0.00078109	0.00036503	0.059486	0.021733	0	0.010365	0	0	0.065686	0	0.057873
2000	0	0	0.041636	0	0.0074876	0.00848764	0.058399	0.018493	0	0	0	0	0.16906	0	0.03004
3000	0	0	0.35132	0	0.00052622	0.0013184	0.074408	0.021216	0	0	0	0	0.28072	0	0.13881
4000	0	0	0.10495	0	0.0011235	0.00052874	0.30682	0.011673	0	0	0	0	0.29807	0	0.27572
5000	0	0	-0.23455	0	0.00068436	0.00068082	0.2922	0.012909	0	0	0	0	0.29807	0	0.73267
6000	0	0	0.20093	0	0.00045813	0.00040943	0.25543	0.013509	0	0	0	0	0	0	0.36013
7000	0	0	-0.78913	0	0.00032916	0.00032796	0.26838	0.012753	0	0	0	0	0	0	0.67665
8000	0	0	-0.93216	0	0.00067782	0.00070925	0.21005	0.019408	0	0	0	0	0	0	0.41331
9000	0	0	-0.95736	0	0.0006928	0.00018922	0.17222	0.012995	0	0	0	0	0	0	0.12594
10000	0	0	-1.0228	0	0.00050707	0.00077951	0.064557	0.019272	0	0	0	0	0.063627	0	0.1353

TABLEAU 2.2 – EXEMPLE DE FICHER DE SORTIE. Le fichier correspond à un réseau de trois gènes ; les moyennes et variances sont calculées pour l'ensemble de la population. Gen : générations. MPhen et VPhen : moyenne et variance de l'expression génique \bar{S} de chacun des gènes. MUnstab et VUnstab : moyenne et variance de l'instabilité développementale V_S pour chacun des gènes. MFit et Vfit : moyenne et variance de la valeur sélective ω . FitOpt : optimum θ fixé pour chacun des gènes. MGenCan et VGenCan : moyenne et variance de la canalisation génétique C_{mut} pour chacun des gènes. GenCanFit : valeur sélective sur la canalisation génétique. MInitCan et VInitCan : moyenne et variance de la canalisation des perturbations de l'état initial C_{init} pour chacun des gènes. InrCanFit : valeur sélective sur la canalisation des perturbations de l'état initial. MDynamCan et VDynamCan : moyenne et variance de la canalisation des perturbations développementales C_{dev} pour chacun des gènes. DynamCanFit : valeur sélective sur la canalisation des perturbations de l'état initial. MeanAll et VarAll : moyenne et variances des valeurs de chacun des w_{ij} .

2.2.1.2 Fichier de sortie

Le fichier de sortie du programme Simevolv consiste en un tableau au format .txt. À chaque génération (ou toutes les g générations), différents indicateurs sont calculés à l'échelle de la population (en moyennant les valeurs obtenues pour chacun des N individus de la population). Ces indicateurs sont :

- le génotype (les valeurs des w_{ij})
- l'expression des L gènes du réseau \bar{S}
- l'instabilité de l'expression des gènes du réseau V_S
- la valeur sélective ω
- la canalisation génétique de chacun des gènes du réseau C_{mut}
- la canalisation aux perturbations de l'état initial, pour chacun des gènes du réseau C_{init}
- la canalisation aux perturbations développementales, pour chacun des gènes du réseau C_{dev}

Un exemple de fichier de sortie peut être trouvé en table 2.2.

Les résultats des différentes simulations (les réplicats, correspondant chacun à une population) peuvent également être analysés sous R pour donner deux fichiers récapitulatifs, le premier donnant la moyenne des différents index sur les différents réplicats, et le second leur variance.

2.2.2 Programme de simulation "Simevolv"

Lorsque j'ai commencé ma thèse, le programme Simevolv existait déjà sous une forme adaptée à l'étude du modèle multilinéaire (développé par José Alvarez-Castro et Arnaud Le Rouzic). Ce programme préliminaire a été fortement modifié afin de permettre l'implémentation d'autres modèles et notamment celui de notre modèle de réseau de régulation.

2.2.2.1 Cahier des charges

L'idée principale lors de la mise au point du programme Simevolv était d'avoir une interface pouvant fonctionner de manière indépendante du modèle utilisé, c'est à dire prenant le même type d'entrée et donnant des sorties identiques, quel que soit le type de modèle utilisé. L'architecture globale du programme est toujours similaire : un individu est constitué d'un génotype et d'un phénotype, une population est faite de plusieurs individus qui se reproduisent pour donner la génération suivante, etc. (voir figure 2.9). Le type de modèle théorique choisi (ou architecture génétique, voir paramètre

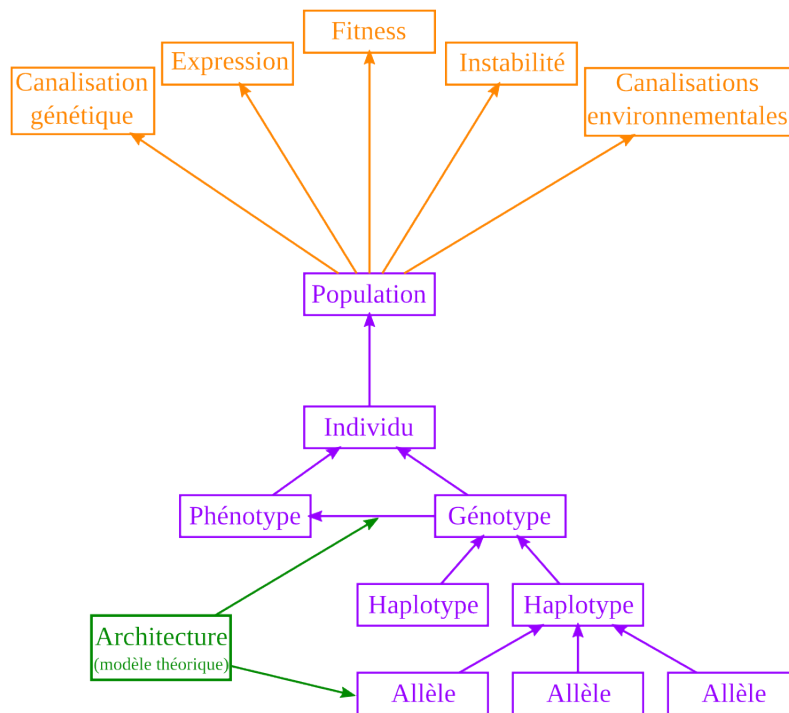


FIGURE 2.9 – REPRÉSENTATION SCHEMATIQUE DU PROGRAMME SIMEVOLV.

TYPE_ARCHI de la table 2.1) apporte des spécificités à cette structure générale : la manière dont sont calculés le génotype et le phénotype (une seule valeur ou plusieurs, par exemple), ainsi que la manière de passer de l’un à l’autre.

La principale limitation du programme est temporelle : chaque simulation doit pouvoir tourner dans un temps raisonnable. Pour cela, nous avons choisi de le coder en C++, un langage compilé rapide et puissant.

2.2.2.2 Organisation interne du programme

Le programme Simevolv est codé en C++. Il est composé de 49 fichiers (24 fichiers C++, 24 headers C++, 1 makefile). Le code est divisé en 63 classes, dont certaines ne sont pas effectives (traces d’anciennes implémentations qui n’ont pas été nettoyées).

Dans la suite de cette section, je détaille les classes principales du programme. D’autres classes existent, notamment pour gérer les aspects aléatoires, statistiques ou mathématiques, ou encore gérer les fichiers de paramètres.

GeneticMap

Cette classe gère la recombinaison entre les allèles.

Allele

La taille des allèles (une ou plusieurs valeurs) est déterminée à cet endroit. C'est également ici que sont réalisés les calculs permettant de passer des allèles au génotype (somme ou moyenne des allèles, par exemple). Enfin, cette classe gère les mutations réalisées lors de la reproduction.

Haplotype

Un haplotype est composé de L allèles. La détermination de la fréquence des mutations et du locus impliqué se fait au sein de cette classe, avant transfert vers "Allele" qui les réalisera.

Genotype

Le génotype est constitué de deux haplotypes, l'un provenant de la mère et l'autre du père (ou les deux du même parent en cas de reproduction asexuée). La production des gamètes et la recombinaison (lors de la reproduction) sont également gérées à cet endroit.

Phenotype

Cette classe récupère le génotype et la valeur du phénotype (calculé dans les classes Architectures). Elle calcule la moyenne et la variance de l'expression génique, ainsi que l'instabilité phénotypique.

Individual

Un individu est composé d'un génotype et d'un phénotype. Cette classe gère la reproduction, et s'occupe du calcul et de la mise à jour de la valeur sélective. C'est également ici que sont créées les collections de mutants à partir desquelles seront calculés les scores de canalisation génétique et environnementale.

Population

La population est une collection d'individus. Le tirage des parents pour la génération suivante est géré par cette classe, puis le mécanisme de la reproduction est transféré vers "Individual". C'est également ici que sont calculés tous les index qui constituent la sortie de la simulation (voir section [2.2.1.2](#)).

Fitness

Les différentes manières de calculer les valeurs sélectives sont implémentées à ce niveau.

Environnement et Canalization

Ces deux classes gèrent les perturbations environnementales et génétiques, et calculent les scores de canalisation associés.

Architecture

C'est à ce niveau qu'est choisie l'architecture génétique utilisée (c'est à dire le modèle théorique) et que sont déterminées la manière dont est codé le génotype et le phénotype, ainsi que la manière de passer de l'un à l'autre. La classe mère est divisée en plusieurs classes filles, chacune gérant les spécificités d'un modèle.

2.2.2.3 Paramétrisation

Tous les paramètres nécessaires à une simulation n'ont pas le même impact sur sa complexité, et donc sur le temps de calcul nécessaire pour la mener à bien. Le temps de simulation étant un critère important, il est nécessaire de trouver le jeu de paramètres qui permet le bon fonctionnement de l'étude (nombre de gènes suffisant, population assez grande pour éviter la dérive, nombre de tests suffisant pour éviter des effets stochastiques, etc.) tout en réduisant autant que possible sa durée, permettant ainsi d'augmenter le nombre de réplicats ou le nombre de simulations.

La figure 2.10 montre l'évolution du temps de simulations en fonction de quelques

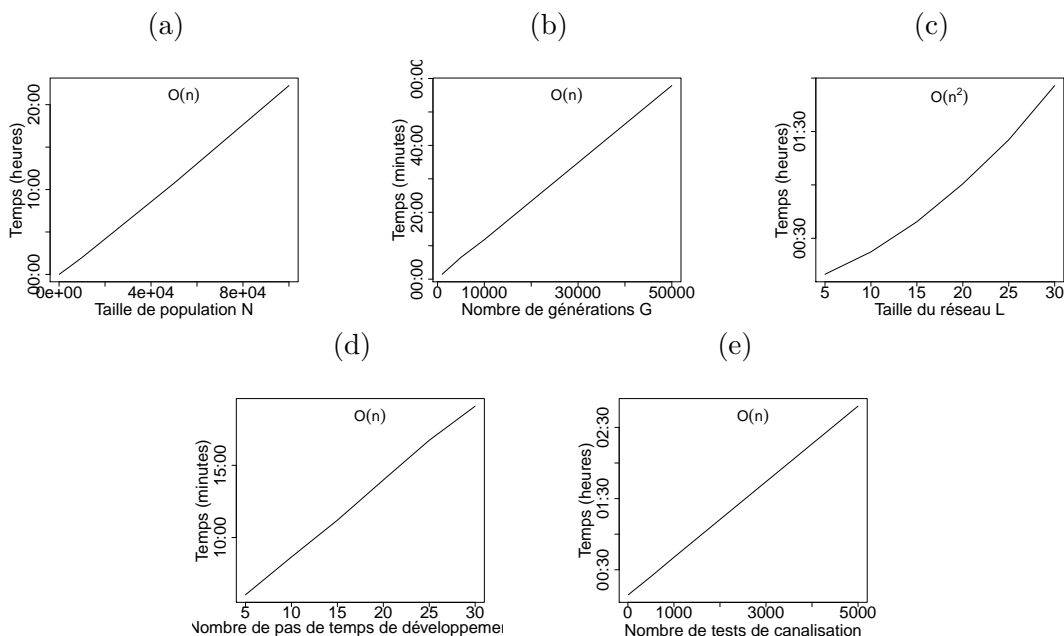


FIGURE 2.10 – TEMPS DE SIMULATION POUR DIFFÉRENTS PARAMÈTRES. Le temps indique la durée réelle nécessaire à l'obtention d'un réplicat. (a) Nombre d'individus dans la population N . (b) Nombre de générations G (statistiques calculées toutes les 1000 générations). (c) Nombre de loci dans le réseau L . (d) Nombre de pas de temps au cours du développement τ . (e) Nombre de tests pour le calcul des canalisations.

paramètres l’influençant. Il est ainsi possible de voir que la durée d’une simulation augmente proportionnellement avec la plupart des paramètres étudiés. Cependant, l’un des paramètres présente une évolution quadratique : doubler sa valeur multiplie le temps de simulation par quatre. Il s’agit du nombre de gènes L . Ceci peut s’expliquer de la façon suivante : cette valeur sert notamment à construire une matrice d’interaction de taille $L \times L$, dont dépend la complexité des calculs ayant lieu au cours du développement d’un individu (voir ci-dessous). Cette valeur ayant une grande importance sur les temps de simulations, nous avons dû nous résoudre à travailler majoritairement sur des réseaux de taille réduite (six gènes).

2.3 Simulation développementale sous R

Un programme permettant de simuler le développement d’un individu (sans notion de population, de sélection ou de reproduction) selon les modalités ci-dessus a également été développé sous R. Il utilise un jeu de paramètres réduit, dont un exemple peut être trouvé en table 2.3.

Ce programme annexe permet d’étudier de manière fine le développement d’un individu à partir d’une matrice W et d’un vecteur initial S_0 , en suivant les changements d’expression génique à chaque pas de temps. S’il serait possible d’implémenter des spécificités similaires dans le logiciel Simevolv, cela complexifierait le logiciel (suivi et stockage de chaque individu au lieu de travailler sur les moyennes au sein des populations, notamment) et augmenterait son temps d’exécution, d’où le choix d’utiliser plutôt un programme secondaire.

	Notation	Valeur
Nombre de loci	L	6
Initialisation des allèles	$\mathcal{N}(mean, sd)$	$\mathcal{N}(0, 0; 0, 1)$
Complexité de la matrice	c_m	0,5
Complexité de la diagonale de la matrice	c_d	1
Expression constitutive	a	0,2
Nombre de pas de temps	τ	16
Nombre de pas de temps pour le calcul des phénotypes	τ_c	4

TABLEAU 2.3 – JEU DE PARAMÈTRE PAR DÉFAUT POUR LES SIMULATIONS DU DÉVELOPPEMENT.

Chapitre 3

Résultats : propriétés du modèle

Sommaire

3.1	Exploration du processus développemental	43
3.1.1	Dynamique de l'expression génique	44
3.1.2	Instabilité développementale	46
3.1.3	Conclusion	47
3.2	Équilibre mutation - sélection - dérive	48
3.2.1	Évolution temporelle	49
3.2.2	Influence des paramètres du réseau	50
3.2.3	Conclusion	52
3.3	Mode de reproduction	53
3.3.1	Diploïdie versus haploïdie	54
3.3.2	Régime de reproduction	56
	Taux d'autofécondation	56
	Taux de recombinaison	57
3.3.3	Conclusion	58

Notre modèle de réseau de gènes, tout comme le modèle originel, peut-être divisé en deux échelles de temps principales : un temps individuel, correspondant au développement des individus, et un temps populationnel, correspondant à l'évolution de la population au fil des générations. Si ces deux processus sont bien évidemment intriqués, chacun présente des caractéristiques propres.

Dans la première section de ce chapitre, je m'intéresse à l'étude du développement des individus (simulation développementale). En me basant sur des matrices W aléatoires, similaires à celles utilisées pour initialiser la population avant évolution, je peux étudier les tendances intrinsèques à ce type de réseaux de régulation quantitatifs.

Dans la seconde section, j'étudie la manière dont le modèle répond (ou non) à la sélection directionnelle puis canalisante sur du long terme (simulation évolutive). De manière plus spécifique, je m'intéresse à l'effet des paramètres du réseau sur cette capacité de réponse.

Enfin, la troisième partie s'intéresse à l'impact des modes de reproduction sur l'évolution des réseaux.

3.1 Exploration du processus développemental

La modélisation du développement des individus est au coeur de notre modèle d'étude. Cependant, ses propriétés propres n'ont été que peu étudiées. En effet, à l'exception de rares travaux (dont PINHO et collab. [2012]), la grande partie des études se concentrent sur la phase populationnelle et évolutive du modèle et les spécificités de la matrice W et du vecteur S_0 ne sont étudiées que sous ce prisme particulier.

Dans cette section, j'ai choisi d'étudier spécifiquement la dynamique développementale en absence de toute pression de sélection, afin de mettre en évidence les propriétés intrinsèques des réseaux de régulation. Je m'intéresse notamment à l'évolution de l'expression génique au cours du développement, sous deux angles distincts. Le premier est la possibilité pour les gènes d'explorer ou non l'intégralité des valeurs possibles (ici, l'intervalle $[0; 1]$) en fonction des paramètres du vecteur d'expression initial S_0 et de la matrice W . Le second est le fort taux d'instabilité développementale (défini comme l'impossibilité pour un gène d'atteindre une expression stable à la fin de son développement) généralement observé dans les modèles de Wagner. Selon PINHO et collab. [2012], cette instabilité serait due en majeure partie à la forte connectivité des réseaux.

L'étude de la dynamique de l'expression génique au cours du développement est réalisée à partir de matrices génotypiques aléatoires. Ces matrices aléatoires sont proches de celles que l'on peut trouver à la génération initiale d'une simulation évolutive, lorsque

la sélection n'a pas encore agi. Elles constituent donc le « point de départ » des études évolutives qui seront présentées dans la suite de ce manuscrit.

Les résultats présentés ici sont issus de simulations de développement (voir section 2.3) basées sur le jeu de paramètres par défaut présenté en table 2.3. Seul le développement des individus est modélisé, il n'y a pas ici de population, de reproduction, de mutation ou encore de sélection. Chaque réplicat des différentes simulations peut être considéré comme un individu : une matrice aléatoire W et le vecteur d'expression S_0 sont créés puis soumis au processus développemental.

3.1.1 Dynamique de l'expression génique

La figure 3.1 présente des cinétiques typiques de l'expression des gènes au cours du développement d'un individu. Chacune des lignes correspond à une simulation différente (soit un jeu de paramètres différent à tester : paramètres standards, force des régulateurs, densité d'interactions, expression constitutive des gènes). Pour chaque si-

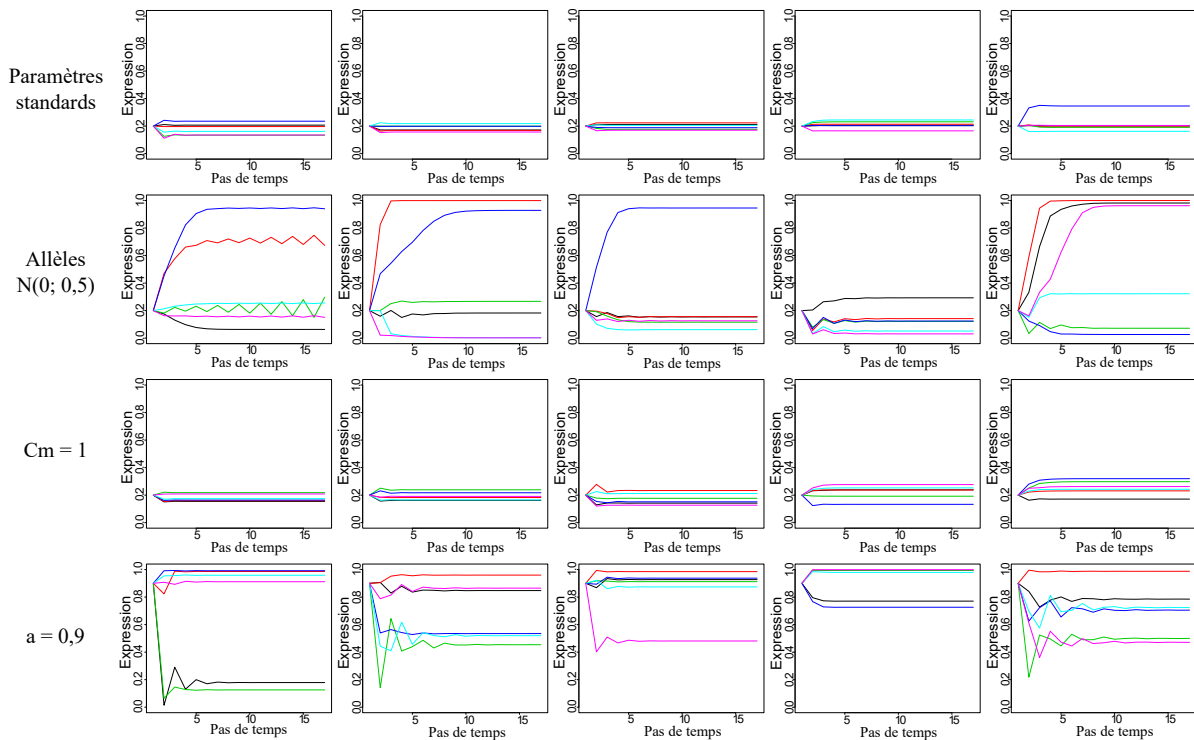


FIGURE 3.1 – DYNAMIQUE DE L'EXPRESSION GÉNIQUE AU COURS DU DÉVELOPPEMENT. Chacune des lignes correspond à une simulation différente (jeu de paramètre différent), et chacun des panneaux correspond à un réplicat (une matrice W indépendante). Les simulations de la première ligne ont des paramètres identiques à ceux présentés en table 2.3. Les trois lignes suivantes diffèrent du jeu de paramètre initial par une valeur seulement. Deuxième ligne : valeurs des w_{ij} plus élevées et plus variables (s.d. = 0,5). Troisième ligne : densité d'interaction maximale ($c_m = 1$). Quatrième ligne : expression constitutive élevée ($a = 0,9$). Au sein de chaque panneau, les six courbes de couleurs correspondent à l'expression des six gènes du réseau.

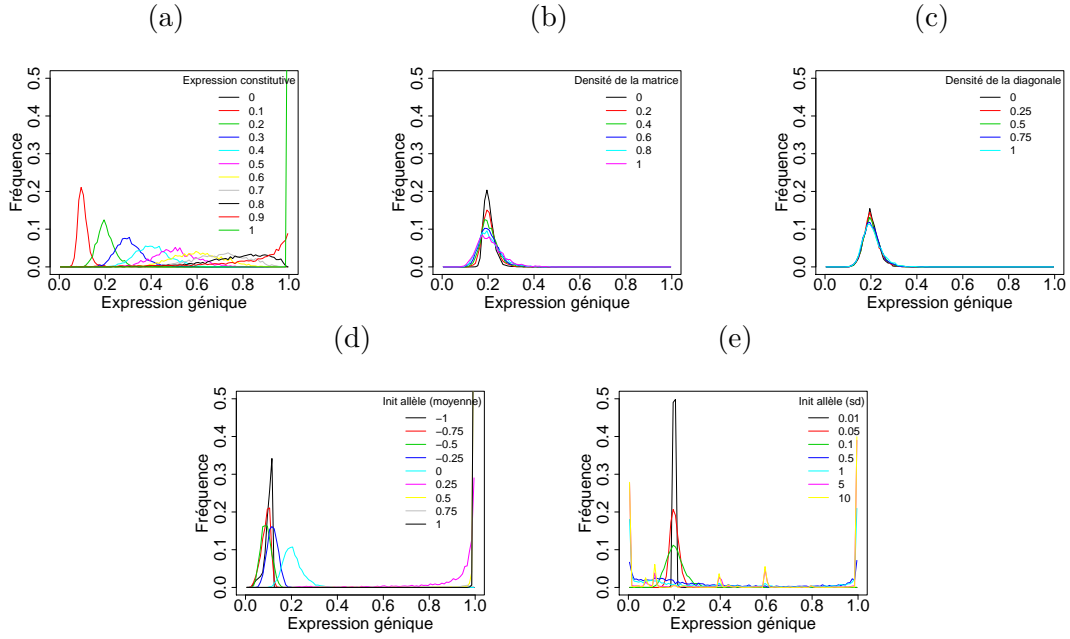


FIGURE 3.2 – EXPRESSION GÉNIQUE À LA FIN DU DÉVELOPPEMENT. Fréquence dans la population des valeurs d’expression génique \bar{S} , calculées sur les 4 derniers pas de temps du développement. Moyenne réalisée sur 500 simulations. (a) Expression constitutive des gènes a . (b) Complexité de la matrice c_m . (c) Densité de la diagonale de la matrice c_d . (d) Initialisation des allèles (valeurs des w_{ij}) : moyenne de la loi gaussienne. (e) Initialisation des allèles (valeurs des w_{ij}) : écart-type de la loi gaussienne.

mulation, cinq réplicats (soit cinq individus) ont été représentés, montrant chacun l’évolution des six gènes du réseau au cours du développement.

Le premier résultat d’intérêt est l’influence des paramètres du réseau sur la gamme de valeurs de l’expression génique. Ainsi, augmenter l’écart-type de la loi normale dans laquelle sont tirés les w_{ij} (et donc augmenter la force des régulateurs, figure 3.1 ligne 1 contre ligne 2) permet d’explorer davantage l’intervalle d’expression $[0; 1]$. Un résultat similaire est obtenu en changeant le niveau d’expression constitutive des gènes (figure 3.1 ligne 1 contre ligne 4). En revanche, la densité des régulations ne montre aucun effet visible (figure 3.1 ligne 1 contre ligne 3).

Ceci est confirmé par la figure 3.2, qui montre la répartition de l’expression des gènes du réseau à la fin du développement. À nouveau, le paramètre le plus influent sur l’exploration de l’espace phénotypique est l’expression constitutive des gènes (figure 3.2a) : plus l’expression constitutive a est élevée, plus la gamme de valeurs atteintes par les gènes est étendue, tout en restant centrée sur la valeur initiale. Le second paramètre d’importance reste l’écart-type utilisé pour l’initialisation des w_{ij} (figure 3.2e).

Le second résultat d’intérêt est l’influence des paramètres sur le comportement cyclique ou instable de l’expression génique au cours du développement. À nouveau, l’écart-type utilisé pour l’initialisation des w_{ij} semble jouer un rôle majeur (figure 3.1 ligne 1 contre ligne 4) : plus il est élevé, plus on observe des comportements cycliques

chez un ou plusieurs gènes des réseaux.

3.1.2 Instabilité développementale

L’instabilité développementale (c’est-à-dire l’incapacité du phénotype à atteindre une valeur stable, calculée comme la variance de l’expression génique V_S sur les derniers pas de temps du développement) est une caractéristique bien connue de ce type de modèle, observable aussi bien chez les variantes basées sur des fonctions normalisatrices à seuil que chez celles utilisant (comme ici) des fonctions sigmoïdes (WAGNER [1996], PINHO et collab. [2012]). Plus que l’aspect de la fonction normalisatrice $f(x)$, ce sont d’autres paramètres qui semblent ainsi impliqués dans la mise en place cette instabilité. Il semble donc intéressant de déterminer plus précisément dans quelles conditions ce type de comportement peut apparaître dans les réseaux.

Pour cela, j’ai testé l’effet des différents paramètres du modèle développemental sur le score d’instabilité (figures 3.3a, 3.3b, 3.3c, 3.3d), à partir du jeu de paramètre par défaut (table 2.3). Les seuls paramètres semblant avoir un effet majeur sur l’instabilité phénotypique sont la moyenne et l’écart-type utilisés pour l’initialisation des valeurs des w_{ij} : plus les régulateurs sont répressifs et plus leur force est importante, plus

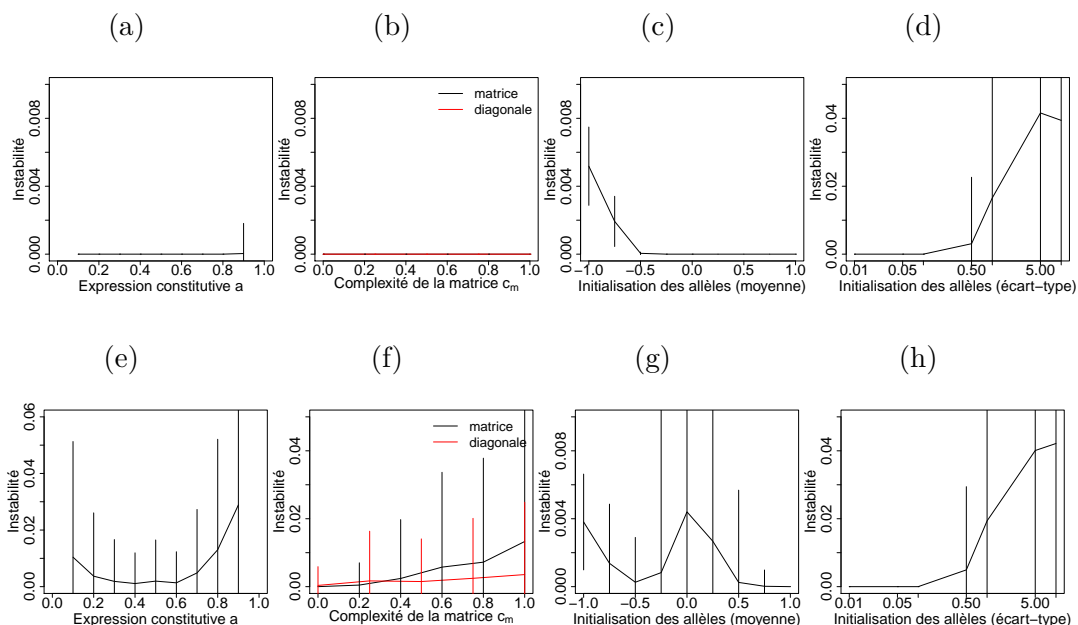


FIGURE 3.3 – ÉVOLUTION DE L’INSTABILITÉ DÉVELOPPEMENTALE. Instabilité développementale calculée comme la variance de l’expression génétique V_S , calculée sur les 4 derniers pas de temps du développement, moyennée sur les 6 gènes du réseau et sur 500 simulations. (a-d) Les paramètres sont identiques à ceux présentés en tableau 2.1. (e-h) L’initialisation des allèles est paramétrée à $\mathcal{N}(0; 0, 5)$. (a)(e) Expression constitutive des gènes a . (b)(f) Complexité de la matrice c_m . (c)(g) Initialisation des allèles (=valeurs des w_{ij}) : moyenne de la loi gaussienne. (d)(h) Initialisation des allèles (=valeurs des w_{ij}) : écart-type de la loi gaussienne.

l'instabilité augmente (figures 3.3c et 3.3d).

Pour pouvoir tester l'influence des autres paramètres, de nouvelles simulations ont été réalisées dans des conditions où l'instabilité développementale peut être observée. Pour cela, les réseaux ont été initialisés grâce à une loi gaussienne $\mathcal{N}(0; \text{s.d.} = 0, 5)$, augmentant ainsi la force des régulateurs (figures 3.3e, 3.3f, 3.3g, 3.3h).

L'expression constitutive (valeurs du vecteur S_0 , figure 3.3e) montre un impact sur l'instabilité : plus la valeur constitutive est extrême (proche de 0 ou 1), plus les gènes ont une tendance à montrer des comportements cycliques. Concernant les paramètres de la matrice W , celui qui a le plus d'influence sur l'instabilité développementale est ici l'écart-type définissant la loi normale dans laquelle seront tirés les w_{ij} . Plus les w_{ij} sont grands, plus l'instabilité développementale augmente (figure 3.3h). De même, plus le réseau est dense et complexe, plus l'instabilité augmente (figure 3.3f), un résultat cohérent avec des études précédentes (PINHO et collab. [2012]). Un réseau totalement connecté présente ainsi une instabilité plus forte qu'un réseau faiblement connecté, et un réseau présentant 100% d'autorégulation est plus instable qu'un réseau en étant dépourvu (pour une complexité de la matrice égale par ailleurs).

3.1.3 Conclusion

J'ai montré durant cette étude que l'expression des gènes est bel et bien sous l'influence des paramètres du réseau. Le paramètre le plus important semble être la force des régulateurs (soit les valeurs des w_{ij} , notamment leur dispersion), qui conditionne à la fois les valeurs d'expression atteignables par le réseau ainsi que l'instabilité de l'expression génique au cours du développement. L'exploration des valeurs de phénotype est également permise par une expression constitutive élevée. Il est cependant possible que cette observation soit due à la forme de la sigmoïde maintenant l'expression génique dans l'intervalle $[0; 1]$.

Si les conditions du réseau permettent l'apparition d'instabilité développementale (autrement dit, si l'écart-type utilisé pour l'initialisation des allèles est élevé), d'autres paramètres secondaires vont avoir une influence sur l'importance de ce phénomène. Notamment, une expression constitutive proche de 0 ou 1 (pas d'expression ou expression maximale) et une forte connectivité au sein du réseau conduisent à une augmentation de l'instabilité.

Ce dernier résultat est en accord avec les observations réalisées par PINHO et collab. [2012], qui observaient une augmentation du score d'instabilité en fonction de la complexité du réseau. Cependant, si eux considéraient ce résultat comme robuste pour tous les autres paramètres du modèle, je montre ici que l'instabilité est principalement sous l'influence de la force des régulateurs.

Dans la plupart des études portant sur les réseaux de régulation, les individus pré-

sentant des gènes instables (ou cycliques) sont considérés comme non viables et leur valeur sélective est grandement pénalisée, les rendant de fait inaptes à la reproduction. Même s'il s'agit là d'un parti-pris biologiquement discutable (une dynamique cyclique serait vue comme non viable par notre définition alors que l'importance des cycles d'expression dans certains processus cellulaires est bien documentée), nous avons fait le choix méthodologique de le reprendre dans la suite de nos études, afin de faciliter la comparaison avec la littérature. Ainsi, les résultats obtenus permettent de paramétrer les simulations évolutives de manière à minimiser autant que possible le nombre d'individus non viables. Il est cependant à noter que notre modèle permet également de lever cette restriction et d'accepter les individus à l'expression génique instable comme étant aptes à la reproduction.

3.2 Équilibre mutation - sélection - dérive

Dans cette section, j'explore la seconde échelle de temps de notre modèle, l'échelle populationnelle. De manière plus spécifique, j'y étudie la capacité de réseaux quantitatifs à répondre à la sélection, directionnelle d'abord, puis stabilisante. Je m'intéresse ici uniquement à la réponse directe à la sélection, c'est-à-dire au changement du phénotype moyen et à l'évolution consécutive de la valeur sélective de la population, et non aux adaptations indirectes, comme la canalisation ou l'évolvabilité.

L'intérêt principal de cette étude est de déterminer dans quelle mesure des réseaux de régulations quantitatifs peuvent répondre à la sélection. À quelle vitesse le phénotype se rapproche-t-il de l'optimum? Est-ce que la population atteint un état d'équilibre, dans lequel les effets des mutations, de la sélection et de la dérive se compensent, ou reste-t-elle à un état instable? Est-ce toujours le cas sur une longue période de temps?

De même, il apparaît intéressant de pouvoir déterminer si cette réponse à la sélection est retrouvée quels que soient les paramètres du réseau, ou si certains réseaux sont plus aptes que d'autres à répondre à la sélection. Si oui, quelles sont leurs caractéristiques? En effet, si des travaux similaires ont déjà été menés, notamment par [SIEGAL et collab. \[2007\]](#), qui ont conclu à une absence de corrélation entre la topologie du réseau et ses propriétés évolutives, de nombreux paramètres n'ont pas été étudiés de manière aussi exhaustive.

Pour tenter de répondre à ces questions, j'utilise dans les simulations suivantes deux jeux de paramètres : "sélection" correspond à un réseau de six gènes, tous sous sélection tant pour un phénotype optimal ($\theta = 0,5$) que contre l'instabilité développementale; "contrôle" correspond à un réseau de six gènes, libre de toute sélection pour

un phénotype optimal, mais très fortement sélectionné contre l'instabilité (il ne s'agit donc pas de dérive génétique pure). Le choix de ne pas prendre la dérive comme témoin a été dicté par la volonté de rester aussi proche que possible de la littérature, qui contre-sélectionne de manière systématique l'instabilité au sein des réseaux.

3.2.1 Évolution temporelle

La figure 3.4 montre l'évolution de la valeur sélective moyenne de la population au long de 10 000 générations, ainsi que celle de ses deux composantes principales : l'expression génique moyenne \bar{S} et la variance de cette expression au sein de la population $Var(\bar{S})$.

À première vue, le modèle semble bien répondre à la sélection : la valeur sélective moyenne augmente très rapidement (en moins de 500 générations, figure 3.4a). Ce résultat est cohérent avec le changement rapide du niveau d'expression des gènes lors de la phase de sélection directionnelle, avant que l'expression ne se stabilise, montrant l'efficacité de la sélection stabilisante (figure 3.4b).

Cependant, même si la valeur sélective atteinte après 10 000 générations est élevée, elle n'en est pas pour autant maximale. Ceci est dû, d'une part (minoritaire) au décalage observable entre l'optimum phénotypique fixé ($\theta = 0,5$) et la valeur moyenne effectivement atteinte par la population ($\bar{S} = 0,48$, figure 3.4b), et d'autre part (majoritaire) à la variance importante de l'expression des gènes entre les individus (figure 3.4c).

Une autre observation importante peut être faite : même après 10 000 générations, la valeur sélective moyenne de la population continue à augmenter légèrement, indiquant que la population n'a toujours pas atteint un état d'équilibre stable entre les effets de la sélection, des mutations et de la dérive. À la place, elle se trouve dans une situation d'équilibre dynamique : l'équilibre mutation - sélection - dérive est bien atteint à chaque point de temps, mais ce point d'équilibre n'est pas fixe et se décale dans le temps. Ce décalage ne vient pas de la composante "moyenne de l'expression génique" de la valeur sélective : en effet, celle-ci est totalement stabilisée (figure 3.4b). En revanche, la variance de l'expression génique au sein de la population, elle, diminue bel et bien au cours du temps (figure 3.4c), vraisemblablement à cause de la diminution de l'effet des mutations du fait de l'augmentation de la canalisation génétique.

Afin de vérifier si cet équilibre dynamique se maintient sur un laps de temps plus long, de nouvelles simulations ont été lancées, selon les mêmes paramètres, mais pour une durée de 50 000 générations (figure 3.5). La valeur sélective se stabilise entre 15 000 et 20 000 générations (figure 3.5a), indiquant que la population a atteint un équilibre mutation - sélection - dérive stable et s'y maintient et la variance de l'expression génique se stabilise dans la même période de temps (figure 3.5c). La durée standard de 10

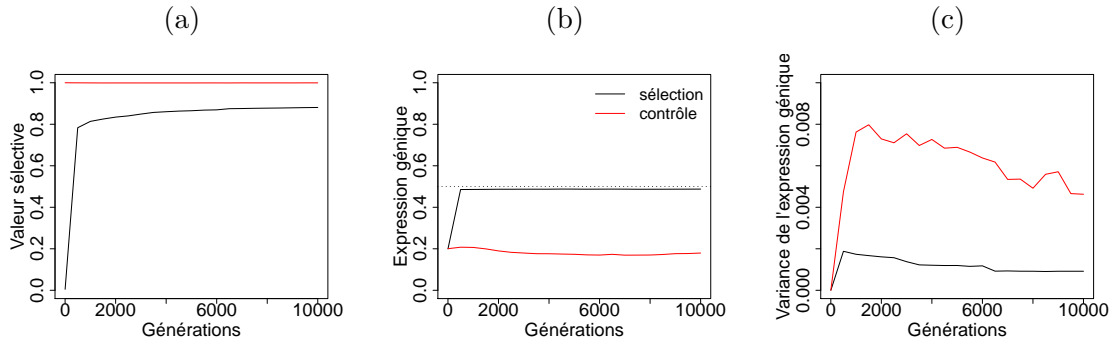


FIGURE 3.4 – RÉPONSE À LA SÉLECTION. 10 000 générations, valeurs calculées toutes les 500; courbes noires : les gènes sont sélectionnés pour un optimum phénotypique de 0,5 et contre l'instabilité phénotypique; courbes rouges : les gènes sont sélectionnés uniquement contre l'instabilité phénotypique. (a) Évolution de la valeur sélective ω . (b) Évolution de l'expression des gènes \bar{S} . (c) Évolution de la variance de l'expression des gènes $Var(\bar{S})$.

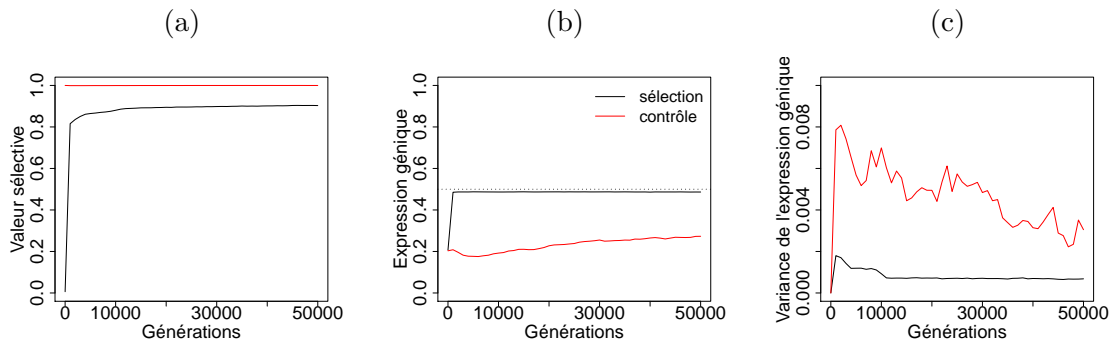


FIGURE 3.5 – RÉPONSE À LA SÉLECTION. 50 000 générations, valeurs calculées toutes les 500; courbes noires : les gènes sont sélectionnés pour un optimum phénotypique de 0,5 et contre l'instabilité phénotypique; courbes rouges : les gènes sont sélectionnés uniquement contre l'instabilité phénotypique. (a) Évolution de la valeur sélective ω . (b) Évolution de l'expression des gènes \bar{S} . (c) Évolution de la variance de l'expression des gènes $Var(\bar{S})$.

000 générations n'est donc pas assez longue pour observer la mise en place d'un état d'équilibre stable au sein de la population.

3.2.2 Influence des paramètres du réseau

La diminution de la valeur sélective moyenne au sein de la population est due à l'impossibilité du modèle à produire une expression génique assez proche de celle de l'optimum fixé, ainsi qu'à la variance phénotypique dans la population. Il est possible que cela soit dû à certaines caractéristiques du modèle, conduisant à des pressions trop importantes sur le réseau pour que celui-ci puisse réguler assez finement l'expression de ses gènes. Pour tester cette hypothèse, j'ai lancé une série de simulations explorant les paramètres du modèle, en suivant l'évolution du phénotype (écart à l'optimum et variance phénotypique, figure 3.7) ainsi que celle de la valeur sélective (figure 3.6).

Concernant les paramètres du réseau, la complexité de la matrice W ne semble

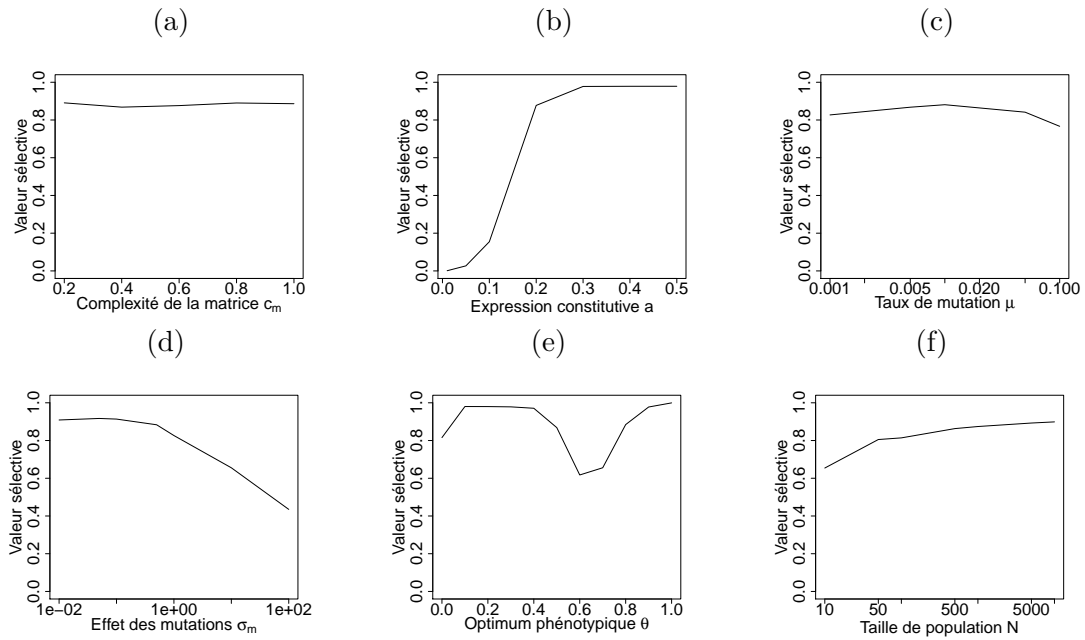


FIGURE 3.6 – IMPACT DES PARAMÈTRES SUR LA RÉPONSE À LA SÉLECTION : VALEUR SÉLECTIVE. La valeur sélective est calculée à la génération $G = 10000$ et moyennée sur 50 réplicats environ. (a) Complexité du réseau c_m . (b) Expression constitutive a . (c) Taux de mutation μ . (d) Effet des mutations σ_m (e) Optimum phénotypique θ . (h) Taille de population N .

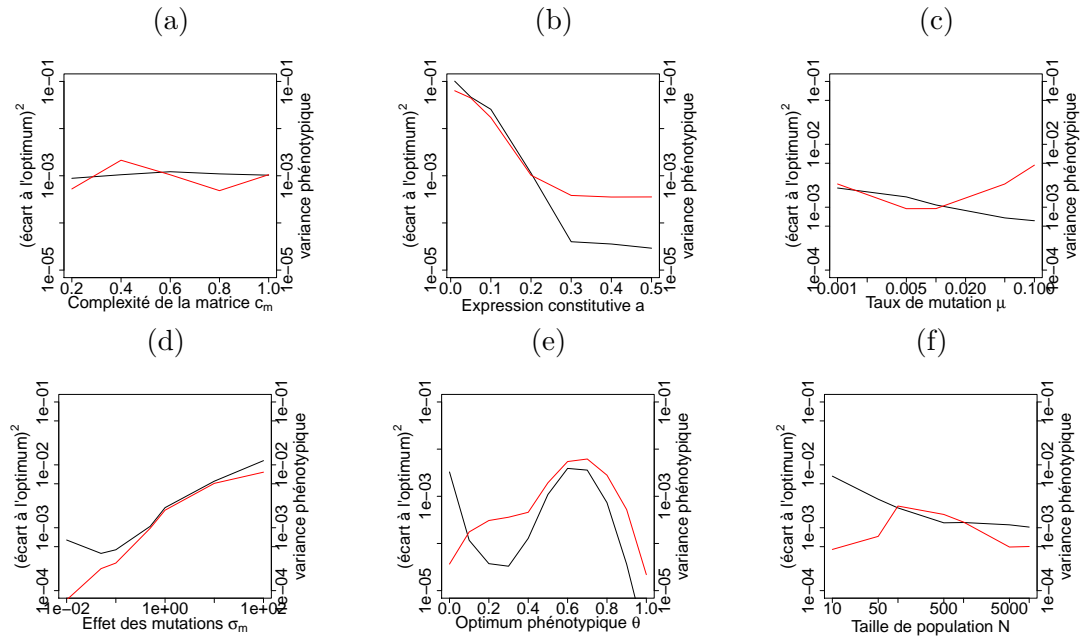


FIGURE 3.7 – IMPACT DES PARAMÈTRES SUR LA RÉPONSE À LA SÉLECTION : EXPRESSION GÉNIQUE. La courbe noire représente le carré de l'écart à l'optimum $(\theta - \bar{S})^2$; la courbe rouge représente la variance de l'expression génique $Var(\bar{S})$. Les valeurs sont calculées à la génération $G = 10000$ et moyennées sur 50 réplicats environ. (a) Complexité du réseau c_m . (b) Expression constitutive a . (c) Taux de mutation μ . (d) Effet des mutations σ_m (e) Optimum phénotypique θ . (h) Taille de population N .

pas avoir d'influence majeure sur la réponse à la sélection (figure 3.7a), même si une faible diminution de la valeur sélective peut-être observée pour les réseaux ayant une complexité supérieure à 50%, diminution qui semble être due à une augmentation de la variabilité phénotypique au sein des populations (figure 3.7a). Ce résultat est cohérent avec l'absence claire d'effet de la complexité du réseau observé par [SIEGAL et collab. \[2007\]](#).

En revanche, l'expression constitutive des gènes montre un effet plus marqué. Ainsi, une expression initiale égale à l'optimum ($\theta = 0,5$) conduit de manière logique à une valeur sélective maximale, qui décroît au fur et à mesure que l'écart entre expression initiale et optimale s'accroît (figure 3.6b), ce qui semble dû tant à un large écart à l'optimum qu'à une forte variabilité phénotypique (figure 3.7b).

L'impact des mutations semble également jouer un rôle non négligeable. Des mutations à fort effet conduisent à une diminution nette de la valeur sélective (figure 3.6d), due aux effets combinés d'une augmentation de l'écart à l'optimum et de la variance phénotypique (figure 3.7d).

L'optimum phénotypique montre lui aussi un impact fort sur la réponse à la sélection. Les valeurs intermédiaires éloignées de l'expression constitutive montrent une augmentation de la distance à l'optimum ainsi que de la variance phénotypique (figure 3.7e), traduit par une forte diminution de la valeur sélective (figure 3.6e). À l'inverse, les valeurs proches de l'expression constitutive des gènes ($a = 0,2$) et les valeurs d'expression maximale montrent une valeur sélective maximale pour la population. Cette difficulté des réseaux à répondre à une sélection vers des valeurs d'expression intermédiaire est consistante avec de précédentes études ([PALMER et FELDMAN \[2009\]](#)).

Enfin, une faible taille de population conduit également à une réponse à la sélection moins efficace (figure 3.6f), ce qui est cohérent avec l'impact accru de la dérive génétique dans de telles conditions.

3.2.3 Conclusion

La première conclusion de cette section est la démonstration que notre modèle de réseau de régulation quantitatif possède bien la capacité de répondre à la sélection directionnelle puis stabilisante. Je montre également que cette réponse est sous la dépendance des paramètres du modèle : ainsi, un réseau présentant une expression constitutive proche de l'expression optimale et soumis à des mutations de petite taille et à des forces de sélection faibles, voit sa capacité de réponse à la sélection maximisée. Ces résultats sont cohérents avec des études précédentes, dont [SIEGAL et collab. \[2007\]](#) ou [PALMER et FELDMAN \[2009\]](#).

La seconde conclusion tirée de cette section, plus marquante, est le fait que même après 10 000 générations, la valeur sélective continue à augmenter, indiquant que la

population n'a toujours pas atteint un état d'équilibre sélection-mutation-dérive stable, mais est toujours dans une phase d'équilibre dynamique. À quoi est dû ce déséquilibre ? Vraisemblablement pas à la dérive : la taille de population (5000 individus) utilisée dans les simulations est suffisante pour que son effet soit négligeable devant celui de la sélection ou des effets des mutations (voir figure 3.6f, qui montre que la valeur sélective atteint un plateau après 1000 individus). Il n'est pas non plus dû à la sélection sur le phénotype : la population y a déjà répondu et l'expression des gènes au cours des générations est très stable.

En revanche, la variance de l'expression génique (variance phénotypique V_P) continue, elle, à diminuer au cours du temps. Celle-ci dépend à la fois des variances environnementale et génotypique, selon la formule $V_P = V_E + V_G$. Dans les simulations présentées tout au long de cette partie, la variance environnementale V_E est fixée à zéro (aucun effet de l'environnement sur l'établissement du phénotype). L'évolution de la variance phénotypique ne peut donc être expliquée que par une diminution de la variance génétique V_G . Plus précisément, elle est due à la diminution de la variance mutationnelle, définie comme la partie de la variance génotypique qui apparaît à chaque génération à cause des mutations. Ces résultats mettent ainsi en évidence la diminution progressive, au fil des générations, de l'impact des mutations, ce qui est la définition même de la canalisation génétique.

En plus de prouver que notre modèle peut répondre de manière directe à la sélection naturelle, cette étude montre qu'il est également capable de favoriser l'évolution de caractéristiques secondaires qui n'ont pas été directement sélectionnées, comme la canalisation qui nous intéresse dans la suite de ce manuscrit.

3.3 Mode de reproduction

Dans cette section, j'explore l'impact de quelques paramètres liés à la reproduction (ploïdie, allo- ou auto-fécondation et recombinaison) sur la capacité des réseaux quantitatifs à répondre à la sélection et à évoluer de la canalisation génétique.

Cette partie a été réalisée alors que les travaux sur l'évolution de la canalisation (voir chapitre 4) étaient déjà bien avancés, ce qui implique des changements dans le jeu de paramètres (détaillé au chapitre 4). À l'inverse des études précédentes, tous les gènes ne sont pas soumis à une pression de sélection directe pour un optimum d'expression. Seuls les deux premiers sont soumis à une telle sélection, les quatre autres restant libres d'évoluer (sous réserve des interactions épistatiques avec les gènes sélectionnés). Tous les gènes restent cependant sélectionnés contre l'instabilité développementale.

De même, je présente ici des résultats portant sur la réponse directe à la sélection et sur l'évolution de la canalisation qui n'ont pas été intégrés à l'article présenté au chapitre 4, mais y sont fortement reliés.

3.3.1 Diploïdie versus haploïdie

Une grande partie des organismes complexes utilisés en tant de modèles génétiques sont diploïdes, qu'il s'agisse d'insectes ou de mammifères. Aussi, c'est assez logiquement que nous avons codé nos individus comme des organismes diploïdes, sans vraiment réfléchir à ce choix. Il aura fallu la remarque d'un des rapporteurs de l'article demandant de justifier ce parti-pris pour que nous nous y intéressions réellement. Il est d'ailleurs à noter que cette question est peu adressée dans la littérature : alors que les modes de reproductions ont été étudiés (notamment les différences entre reproduction sexuée et asexuée ou la recombinaison), la ploïdie ne l'a que peu ou pas été. De même, si la plupart des travaux étudiant les dérivés du modèle de Wagner utilisent un modèle haploïde, cela est rarement indiqué de manière explicite dans les paramètres du modèle.

Pourtant, ce choix n'est pas anodin. D'une part, il a un effet majeur sur les relations entre les gènes. Si l'effet de l'épistasie (relation inter-locus) reste présent chez des individus haploïdes aussi bien que diploïdes, la notion de dominance (relation intra-locus, entre les deux allèles du même locus), elle, n'est présente que chez les organismes diploïdes. D'autre part, la diploïdie peut également jouer un rôle sur la réponse aux mutations : une même mutation peut avoir un effet moins grand dans un organisme diploïde (car elle sera tamponnée par la copie sauvage toujours présente sur l'autre allèle) que dans un organisme haploïde. À l'inverse, les mutations sont plus fréquentes dans un organisme diploïde (dans notre modèle, le taux de mutation est fixé par haplotype et non par génome.)

Aussi, s'il est peu probable que le mode de reproduction influence la capacité du

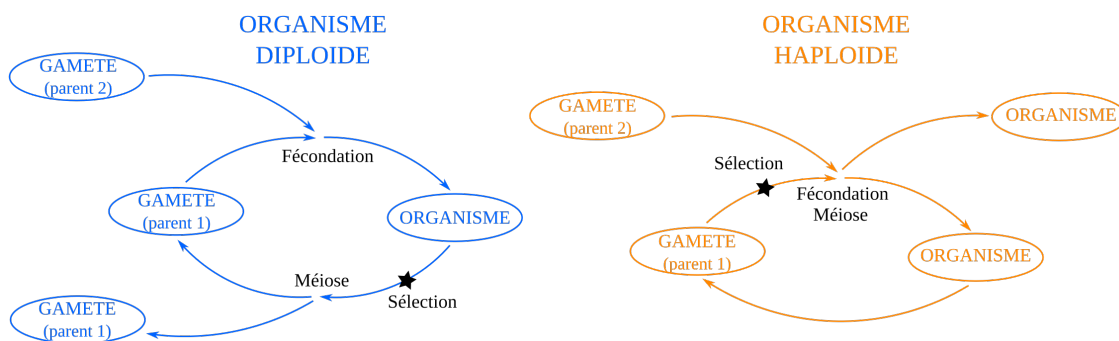


FIGURE 3.8 – SÉLECTION EN MODÈLE HAPLOÏDE ET DIPLOÏDE. Dans le modèle diploïde, la phase de calcul du phénotype et de la sélection se fait après la fécondation. En revanche, elle a lieu juste après la formation des gamètes pour le modèle haploïde.

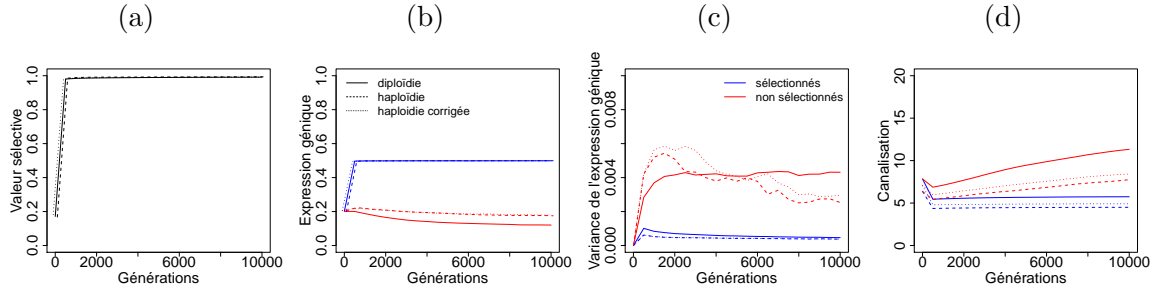


FIGURE 3.9 – DIPLOÏDIE VS HAPLOÏDIE. Réseau de six gènes, deux sous sélection (courbe bleue), quatre non sélectionnés (courbe rouge). Les lignes pleines correspondent au modèle diploïde, les lignes tiretées au modèle haploïde, les lignes pointillées au modèle haploïde couplé à une diminution de l’effet des mutations. (a) Valeur sélective ω . (b) Expression génique \bar{S} . (c) Variance de l’expression génique $Var(\bar{S})$. (d) Canalisation C_{mut} .

réseau à répondre à la sélection, il est possible qu’il ait un impact sur sa capacité à s’adapter aux mutations, et donc modifie l’évolution de la canalisation génétique.

Pour tester ces hypothèses, un modèle haploïde a été implémenté dans le programme Simevolv. Dans ce modèle, le processus de sélection agit directement après la formation des gamètes, et non après la fécondation comme dans le modèle diploïde (voir figure 3.8).

La figure 3.9 montre la réponse à la sélection et l’évolution de la canalisation dans deux populations, l’une diploïde et l’autre haploïde. D’une manière générale, l’évolution des deux populations suit une tendance similaire. Le changement de mode de reproduction n’influe pas sur la valeur sélective moyenne de la population (figure 3.9a), que ce soit sur la distance à l’optimum ou sur la variance phénotypique. Dans les deux cas, les deux gènes sous sélection sont proches de l’optimum et leur variance diminue au cours du temps selon des tendances similaires (figures 3.9b et 3.9c).

Au niveau de canalisation, la tendance est également semblable : la canalisation évolue pour les gènes non sélectionnés, et demeure constante pour les gènes sous sélection directe (figure 3.9d), une observation semblable à celles présentées dans le chapitre 4. Le score de canalisation plus faible des populations haploïdes peut s’expliquer par un effet plus fort des mutations lors du test de canalisation (l’effet n’étant pas moyenné par celui de l’allèle sauvage du second haplotype). Cependant, une tentative de contrebalancer cet effet n’a pas totalement fait disparaître cette différence entre canalisation dans une population diploïde ou haploïde (figure 3.9, lignes pointillées). Une explication alternative pourrait être l’absence de dominance chez les individus haploïdes, mais d’autres mécanismes, comme un changement du nombre d’allèles dans la population, pourraient également intervenir.

3.3.2 Régime de reproduction

Les résultats présentés dans cette partie sont issus du travail de Thomas Riolland, stagiaire de M1 du master SDUEE de l'UPMC. Son stage, réalisé d'avril à juin 2016, a donné lieu à un rapport intitulé "Impact du régime de reproduction sur l'évolution des réseaux de gènes". Il s'y intéresse majoritairement à l'influence du taux d'autofécondation (une approximation pour un régime de reproduction sexuée ou asexuée) et du taux de recombinaison sur l'évolution de la canalisation.

Le régime de reproduction a déjà été étudié à partir des dérivés du modèle de Wagner (voir section 1.3.3.2). Le résultat principal tiré de ces études est que les populations sexuées sont plus fortement canalisées que les asexuées, même si les auteurs ne sont pas en accord sur la raison de cette différence. Ainsi, AZEVEDO et collab. [2006] y voient le rôle de la recombinaison et de l'épistasie, alors que MARTIN et WAGNER [2009] démontrent que la recombinaison ne joue qu'un rôle mineur, et que LOHAUS et collab. [2010] ne trouve aucun rôle de l'épistasie.

Cependant, ces travaux ont été effectués à partir de modèles qualitatifs. Au cours de son stage, Thomas Riolland les a repris dans le but de voir si ces résultats sur l'impact du régime de reproduction sont extensibles à un modèle quantitatif.

Taux d'autofécondation

Le taux d'autofécondation f permet de déterminer la proportion d'individus se reproduisant de manière asexuée. Une valeur de $f = 0$ correspond à une population complètement alloféconde : les deux parents sont tirés proportionnellement à leur valeur sélective, et chacun fournit un des deux haplotypes qui subiront ensuite un évènement de recombinaison. Une valeur de $f = 1$ correspond à une population complètement autoféconde ; dans ce cas, le premier parent sélectionné (toujours en fonction de sa valeur

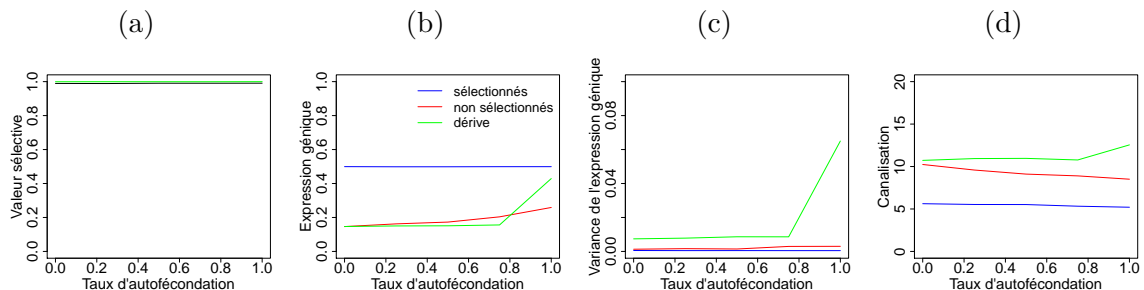


FIGURE 3.10 – INFLUENCE DU TAUX D'AUTOFÉCONDATION. Réseau de six gènes, deux sous sélection (courbe bleue), quatre non sélectionnés (courbe rouge). La courbe verte correspond à un réseau témoin uniquement soumis à l'effet de la dérive génétique. Les valeurs sont calculées à la génération $G = 10000$ et moyennées sur 50 réplicats environ. (a) Valeur sélective ω . (b) Expression génique \bar{S} . (c) Variance de l'expression génique $Var(\bar{S})$. (d) Canalisation C_{mut} .

sélective) fournit les deux haplotypes (qui seront eux aussi recombinés). Les valeurs intermédiaires correspondent à des populations mixtes dans lesquelles les individus peuvent avoir des comportements auto- ou allo- féconds.

La figure 3.10 montre l'effet du taux d'autofécondation sur la réponse à la sélection et sur la canalisation. Ce paramètre n'a pas d'effet sur la valeur sélective de la population (figure 3.10a), ni sur la distance à l'optimum des deux gènes sous sélection (figure 3.10b), ni sur la variance de l'expression génique (figure 3.10c). Concernant les gènes non sélectionnés, augmenter le taux d'autofécondation augmente leur niveau d'expression, ainsi que leur variation phénotypique.

En revanche, le taux d'autofécondation a un impact clair sur le score de canalisation (figure 3.10d). Une population complètement autoféconde est ainsi plus sensible aux mutations qu'une population alloféconde, un résultat cohérent avec des études précédentes (AZEVEDO et collab. [2006]; LOHAUS et collab. [2010]). Cette augmentation de la sensibilité aux mutations dans les populations autofécondes, due principalement aux gènes non sélectionnés, est à rapprocher de l'augmentation parallèle de la variation phénotypique (figure 3.10c).

Une simulation témoin correspondant à un réseau soumis uniquement à la dérive génétique (aucune pression de sélection sur le phénotype ni sur l'instabilité développementale) a également été étudiée. En situation de dérive pure, les populations autofécondes présentent des phénotypes plus variables (figures 3.10c et 3.10b) et sont également plus canalisées que les populations allofécondes (figure 3.10d).

Taux de recombinaison

Le taux de recombinaison r joue sur la probabilité de recombinaison entre les différents loci (c'est à dire entre les différentes lignes de la matrice, voir figure 2.5).

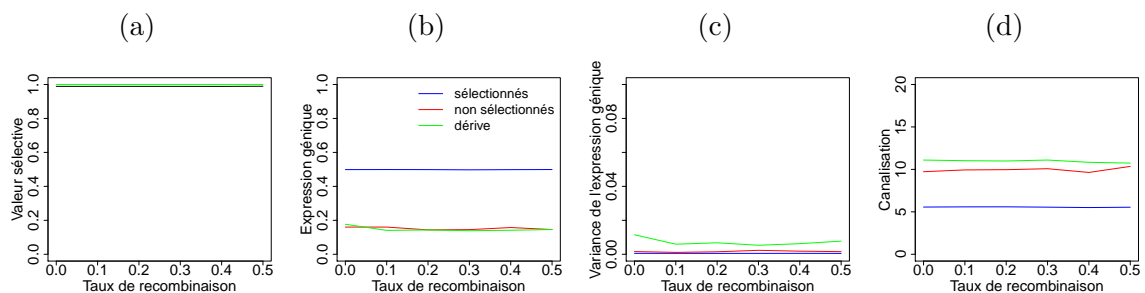


FIGURE 3.11 – INFLUENCE DU TAUX DE RECOMBINAISON. Population alloféconde. Réseau de six gènes, deux sous sélection (courbe bleue), quatre non sélectionnés (courbe rouge). La courbe verte correspond à un réseau témoin uniquement soumis à l'effet de la dérive génétique. Les valeurs sont calculées à la génération $G = 10000$ et moyennées sur 50 réplicats environ. (a) Valeur sélective ω . (b) Expression génique \bar{S} . (c) Variance de l'expression génique $Var(\bar{S})$. (d) Canalisation C_{mut} .

Il ne montre aucune influence sur les capacités d'évolution du système, ni sur la réponse à la sélection, ni sur la capacité à évoluer de la canalisation (figure 3.11), et ce aussi bien dans les réseaux soumis à sélection que dans les réseaux témoins. Ce résultat est cohérent avec une étude précédente, qui avait montré le faible impact de ce paramètre sur la résistance aux mutations (MARTIN et WAGNER [2009]).

3.3.3 Conclusion

Les premiers résultats présentés dans cette section montrent que des populations haploïdes et diploïdes présentent des capacités évolutives similaires : dans les deux cas, la réponse à la sélection et la capacité à faire évoluer de la canalisation suivent des tendances semblables. Cependant, le score de canalisation reste plus faible pour les populations haploïdes, même lorsque l'on prend en compte les biais sur l'effet des mutations. Il est possible que cette différence soit due à la dominance présente dans les populations diploïdes, qui aurait alors une importance moindre que celle envisagée.

Les seconds résultats montrent l'absence d'effet du taux de recombinaison sur les capacités évolutives du modèle, et notamment sur la canalisation. L'évolution de la résistance aux mutations diffère ainsi grandement de l'évolution du taux de mutation, très sensible à la recombinaison (SNIEGOWSKI et collab. [2000]). En effet, contrairement à la relation mutateurs / cibles qui peut être brisée par la recombinaison, les relations épistatiques présentes dans le réseau ont des effets dispersés sur l'ensemble des loci.

Enfin, les résultats obtenus montrent l'impact important du taux d'autofécondation sur l'évolution de la canalisation. Les populations allofécondes s'avèrent ainsi bien moins sensibles aux mutations que les populations autofécondes. Une telle observation peut potentiellement expliquer le développement de la reproduction sexuée (qui peut être rapprochée d'une population complètement alloféconde), ce malgré le coût élevé qu'elle inflige aux organismes, une thèse déjà supportée par AZEVEDO et collab. [2006]; LOHAUS et collab. [2010]. Le cas d'une population à reproduction clonale n'a pas été étudié ici, mais il semble raisonnable de penser qu'elle se conduirait de manière similaire à une population autoféconde.

Chapitre 4

Résultats : pourquoi et comment la canalisation génétique évolue dans des réseaux de régulation

Je présente dans l'article ci-après nos travaux portant sur l'évolution de la canalisation dans des réseaux de régulations quantitatifs.

Le modèle théorique est similaire à celui présenté dans le chapitre 2 et déjà utilisé au chapitre 3. En revanche, au lieu des réseaux de six gènes tous sous sélection des études précédente, les réseaux utilisés dans cette étude ne présentent que deux gènes sous pression de sélection directe, les quatre autres étant sous sélection indirecte du fait de l'épistasie au sein du réseau. J'attire ici l'attention du lecteur sur les différences de notations des paramètres, qui peuvent varier entre celles utilisées tout au long de ce manuscrit et celles utilisées dans l'article.

Nous avons démontré qu'il est possible d'évoluer de la canalisation dans des réseaux de régulation et que cette capacité est principalement due aux gènes non sélectionnés. Nous avons également montré que le score de canalisation atteint par les réseaux est sous la dépendance des paramètres du réseau. Ainsi, des mutations fréquentes et à fort effet, des réseaux dont plusieurs gènes sont libres d'évoluer, et des optimums phénotypiques qualitatifs (pas d'expression ou expression maximale) conduisent à des populations plus fortement canalisées.

Ces résultats nous ont permis de proposer un double mécanisme qui puisse expliquer l'évolution de la canalisation dans les réseaux de régulation. Le premier mécanisme est celui de la réduction de la cible mutationnelle, qui propose que les gènes non utiles à l'établissement du phénotype sous pression de sélection directe puissent être retirés du réseau sans impacter ledit phénotype. Le second est celui de la redondance de la régulation génique, qui postule que du fait de leur présence en plusieurs copies, des séquences régulatrices peuvent être perdues sans conséquence sur le phénotype correspondant. En diminuant l'impact des mutations sur les phénotypes sélectionnés, ces deux mécanismes augmentent la canalisation globale du réseau.

Cet article a été soumis chez BMC Evolutionay Biology en juin 2016, accepté en octobre 2016 et publié en novembre 2016.

RESEARCH ARTICLE

Open Access



Why and how genetic canalization evolves in gene regulatory networks

Estelle Rünneburger and Arnaud Le Rouzic*

Abstract

Background: Genetic canalization reflects the capacity of an organism's phenotype to remain unchanged in spite of mutations. As selection on genetic canalization is weak and indirect, whether or not genetic canalization can reasonably evolve in complex genetic architectures is still an open question. In this paper, we use a quantitative model of gene regulatory network to describe the conditions in which substantial canalization is expected to emerge in a stable environment.

Results: Through an individual-based simulation framework, we confirmed that most parameters associated with the network topology (complexity and size of the network) have less influence than mutational parameters (rate and size of mutations) on the evolution of genetic canalization. We also established that selecting for extreme phenotypic optima (nil or full gene expression) leads to much higher canalization levels than selecting for intermediate expression levels. Overall, constrained networks evolve less canalization than networks in which some genes could evolve freely (i.e. without direct stabilizing selection pressure on gene expression).

Conclusions: Taken together, these results lead us to propose a two-fold mechanism involved in the evolution of genetic canalization in gene regulatory networks: the shrinkage of mutational target (useless genes are virtually removed from the network) and redundancy in gene regulation (so that some regulatory factors can be lost without affecting gene expression).

Keywords: Genetic architecture, Quantitative genetics, Individual-based simulations, Evolution of epistasis

Background

Canalization reflects the capacity of an organism's developmental process to maintain a constant phenotype in spite of perturbations. This concept was first introduced by Waddington [1], who noticed the striking robustness properties of the development in complex organisms. This seminal work inspired further conceptual progress, in particular about the distinction between robustness to environmental versus genetic perturbations [2]. While environmental canalization can be easily explained by natural selection, evolution toward a lower sensitivity to mutations (which will be called indifferently genetic canalization or genetic robustness in this paper) is less straightforward. As genetic canalization tends to decrease the adaptation potential (by preventing the appearance of new phenotypes), it also paradoxically enhances long

term evolvability, by allowing the accumulation of cryptic genetic variation (i.e. mutations sheltered from natural selection as they do not affect the selected phenotypes, and which can be released afterwards and thus contribute to rapid adaptation) [3–5]. Genetic canalization is thus largely overlapping with other evolutionary properties of populations, as exemplified by studies based on population genetic models [6–9]. In particular, counter-intuitively, robustness and evolvability may not systematically be opposite properties of genetic systems [10, 11].

Integrating evolution of genetic variation into the Darwinian framework is thus of outstanding interest in evolutionary biology [12]. However, the complexity of this task is such that it has to rely on formal modelling of realistic genetic architectures involving genetic interactions (epistasis). Simple models of gene regulatory networks appear as good candidates for studying the evolution of epistatic systems. In particular, even small regulatory

*Correspondence: arnaud.le-rouzic@egce.cnrs-gif.fr
Laboratoire Évolution, Génomes, Comportement, Écologie, CNRS-IRD-Univ. Paris-Sud-Université Paris-Saclay, 91198 Gif-sur-Yvette, France

networks display a wide range of different interaction patterns [13]. The gene regulation model proposed by A. Wagner [14, 15] has been widely used to understand the evolution of canalization in regulatory networks [4, 5, 16, 17]. Based on a regulation network encoded as an interaction matrix, it implements a dynamical model of gene expression along the development, and facilitates the theoretical study of long-term evolution under various selection pressures.

Independent implementations of this model, most of the time in individual-based evolutionary simulations, have confirmed consistently that genetic canalization can evolve without selecting directly for robustness. For instance, the sensitivity to mutations can decrease (i.e. genetic canalization can increase) as a by-product of stabilizing selection on gene expression at the end of the development [15, 18]. Using a similar framework, Siegal & Bergman [16] suggested that selection on the stability of development also leads to higher canalization, even in the absence of stabilizing selection for a phenotypic optimum. Other studies based on the same theoretical background have subsequently highlighted the relationship between the evolution of genetic canalization and structural properties of the network (such as density of interactions [15, 16, 19, 20], topology [21–23], size of the network [16]) or between canalization and various evolutionary properties (such as selection [24] or reproduction regime [25–27]).

Most of these studies confirm that robustness is expected to evolve from complex genetic architectures, such as gene regulatory networks. However, (i) there is little empirical evidence that genetic canalization can actually evolve in real populations (in particular in sexual species) [28], (ii) quantitative characters are far from being completely canalized, as evidenced by the prevalence of monogenic diseases or lethal mutations, and by the high level of genetic variance in most natural populations [29], and (iii) little is known about how canalization translates into observable properties from real genetic systems. In this paper, we aim to challenge theoretically the limits and the prediction power of the “canalizing gene networks” theory. First, we defined the conditions in which substantial genetic canalization is expected to appear (or not appear) through indirect selection (“why canalization evolves”). Second, we described the mechanisms involved in the evolution of such genetic canalization in the simulated gene networks (“how canalization evolves”). These results thus bring expectations and conceptual tools for a future empirical validation of the canalization theory.

Methods

In order to address the conditions and mechanisms involved in the evolution of genetic canalization, we used a simulation framework derived from the Wagner model

[14], modified in such a way that gene expression could be treated as quantitative values. This setting makes it possible to use traditional quantitative genetics tools, and to relate the results to the evolution of quantitative measurements of gene expression.

Genotype-to-phenotype map

Each individual is characterized by its genotype (the matrix W , representing the interactions among transcription factors within the regulatory network), from which its phenotype (expression levels for all genes in the network) is calculated.

More specifically, the $L \times L$ interaction matrix W describes the way the L genes interact, each W_{ij} representing the effect of gene i on the expression of gene j [15–17]. This effect can be nil (no interaction, $W_{ij} = 0$), activating ($W_{ij} > 0$) or repressing ($W_{ij} < 0$). Each line of the matrix stands for an allele, i.e. a series of potential fixation sites for regulation factors in the gene promoter. In the default set of simulations, individuals are diploid, and their genotype is obtained by averaging both parental alleles. We also ran simulations with haploid individuals, which showed very similar results (Additional file 1: Figures S1 and S2).

The S_t vector represents the expression level of the L genes of the network at developmental time t ; gene expressions scale between 0 (no expression) and 1 (maximum expression). In absence of regulation, transcription leaks to a constitutive gene expression value a , set by default to $a = 0.2$ (20 % of the max expression). For simplicity, at the beginning of the development, gene expression values are all set to a . Gene expressions then change during the 16 developmental time-steps of the individual, according to:

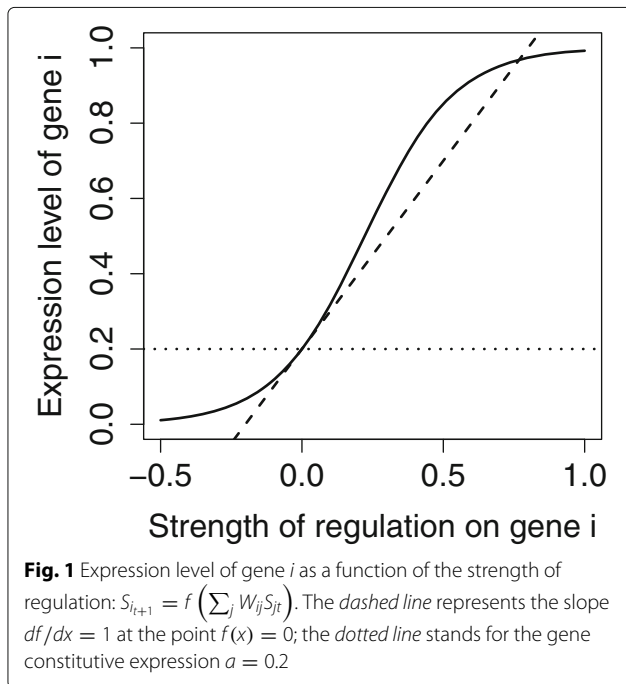
$$S_{t+1} = f(WS_t), \quad (1)$$

where $f(x)$ is a sigmoid function (Fig. 1) such as:

$$f(x) = \frac{1}{1 + \left(\frac{1}{a} - 1\right) \exp\left(\frac{-x}{a(1-a)}\right)}. \quad (2)$$

This function ensures the scaling of the S_t vector in $(0; 1)$, and has been parametrized so that the slope $df/dx = 1$ when $x = 0$. Biologically, this sets the scale of the W matrix so that W_{ij} tends to the induced change in gene expression in absence of other source of regulation.

For each individual, we measured (i) the vector of mean gene expressions (\bar{S}) during the 4 last developmental time steps, and (ii) the vector of the variance of gene expressions (V_S) during the same period. In this context, assuming that the development is long enough to reach a stationary state, V_S can be used as a measurement of developmental stability ($V_S > 0$ indicates a cyclic equilibrium state).



Population genetics model

Individual-based simulations were run for G non-overlapping generations. Successive generations of constant population size N were produced by simulating sexual reproduction among hermaphrodite individuals. In the haploid model, selection was performed after gametogenesis, while selection took place after fertilization in the diploid model. For each offspring, two parents were drawn with probabilities proportional to their fitness (see below). Gametes were generated by picking randomly one allele in each parent, assuming free recombination between the L loci (i.e. between parental alleles, represented by matrix rows). There was no recombination between regulatory sites (within matrix rows).

Mutations occurred at rate μ per haploid genome, during gametogenesis. Mutations can affect one or several loci in an individual, but each locus can be mutated only once (with a probability μ/L). For each mutated locus, a random non-zero element on the matrix row is modified by adding a random Gaussian modifier centred around the former value and of standard deviation σ_m .

Simulations were initialized with N genetically identical individuals (same W matrix). The complexity c of the network specifies the frequency of interactions ($W_{ij} \neq 0$), these non-zero entries were set randomly. The strength of these interactions was initially drawn in a Gaussian distribution $\mathcal{N}(0.0, 0.1)$, indicating a balance between weak enhancers and repressors at the beginning of the simulations. The mean \bar{W}_{ij} at the end of the simulation was used as an indicator of the average direction of regulations, this

measurement being analogous to the network excitation score defined in [30].

The fitness ω_n of an individual n (proportional to the probability of reproduction) depends on the proximity to the target gene expression (d_n) and the gene expression stability during development (k_n), such that $\omega_n = d_n \times k_n$. The fitness component d_n follows a bell-shaped distribution

$$d_n = \exp\left[-s \sum_{i=1}^{\ell} (\bar{S}_{ni} - \theta_i)^2\right] \tag{3}$$

where \bar{S}_{ni} is the mean expression of gene i in individual n , θ_i the phenotypic (expression) optimum for gene i , and s the strength of stabilizing selection. $\ell \leq L$ represents here the number of genes in the network for which the expression has a direct impact on fitness.

In addition, genotypes leading to cyclic gene expression (unstable development) were affected by a fitness penalty. In practice, the fitness component associated with developmental instability in individual n was

$$k_n = \exp\left[-s' \sum_i^L V_{S_{ni}}\right], \tag{4}$$

the selection coefficient being arbitrarily set to $s' = 46,000$, such that individuals with a substantial variance in gene expression at the end of their development $V_S > 10^{-4}$ are unlikely to reproduce ($k_n < 0.01$).

Implementation and parameterization

The core program was written in C++, the simulation pipeline in Bash and the statistical analysis in R. Simulations were run with the default parameters presented in Table 1 unless stated otherwise. Simulations in which at least one gene was disconnected from the network (no regulators) were discarded from the analysis. Results were

Table 1 Default parameters of the model

	Parameter	Value
Number of generations	G	10,000
Population size	N	5,000
Number of loci	L	6
Number of loci under selection	ℓ	2
Initialization of the alleles		$\mathcal{N}(0.0, 0.1)$
Mutation rate per haplotype	μ	0.01
Mutation effect	σ_m	0.5
Matrix complexity	c	0.5
Constitutive expression	a	0.2
Strength of stabilizing selection	s	10
Phenotypic optimum	θ	0.5

averaged over 200 replicates for each parameter set, assuring that the graphical representations are not affected by sampling error.

For each gene i in the network, simulations provided the mean and the variance of mean gene expressions \bar{S}_i in the population every generation. In addition, we estimated genetic canalization in the following way. Each gene i in individual n was featured by a canalization score $C_{n_i} = -\log \text{Var}(M_{n_i})$, where $\text{Var}(M_{n_i})$ was the variance in the expression of gene i among mutants that differ from individual n by a single haploid mutation (drawn in a Gaussian distribution of mean 0 and s.d. 0.5) anywhere in the network. As a consequence, individuals displaying high robustness to mutations (small $\text{Var}(M_{n_i})$) were featured with high canalization scores. In practice, the genotype of each individual n in the population was mutated 100 times, and the average canalization in the population \bar{C}_i was obtained by averaging out all C_{n_i} . An alternative measurement $C'_{n_i} = \log \sqrt{\text{Var}(M_{n_i})}/S_{n_i}$, analogous to a coefficient of variation scaled by the phenotypic expression, gives essentially the same results (Additional file 2: Figure S3).

In addition, the mechanisms underlying the evolution of canalization were investigated through two specific indices, R_s and D_s , that were calculated on every gene of the network. R_{s_i} is the redundancy score of gene i , i.e. the average effect of knocking down another gene of the network on the expression of i . D_{s_i} is the indispensability score for gene i and quantifies the average effect of knocking down this gene (gene expression forced to zero) on the selected genes (average absolute value of gene expression differences). Hence, a gene with a low R_s tends to be regulated in a redundant way (knocking down a single regulator is not enough to affect its expression substantially), while a gene with a low D_s is dispensable (its extinction does not influence the rest of the network).

Results

Evolution of gene expression and genetic canalization

We first studied how the simulated populations responded to selection and how canalization evolved. In the default parameter set (Table 1), two genes (out of six) were under selection pressure toward an intermediate optimal expression of $\theta = 0.5$, while the others could evolve freely. In parallel, we ran equivalent control simulations without any selection on gene expression. In these control cases, gene expression could thus drift freely, but gene networks remained constrained by selection against developmental instability.

Unsurprisingly, genes under stabilizing selection consistently reached a value close to their optimal expression (Fig. 2a and c). Such selection response was fast (less than a few hundred generations) and was associated with a substantial decanalization of the network

(Fig. 2b), which is an expected consequence of directional selection. Two comparisons are meaningful. Comparing selected (blue line) and unselected (red line) genes of a same network highlights patterns only due to the direct effect of stabilizing selection. Comparing unselected gene (red lines) and control simulation (black line) underlines the effects of indirect stabilizing selection on correlated genes. Freely-evolving genes, although epistatically associated with selected genes, displayed a gene expression pattern close to the control case.

However, this similarity between unselected and control case genes vanished when other network properties were considered, as the distribution of regulatory genotypes (W_{ij} values) was different both in mean and variance (Fig. 2d). Control gene networks tended to evolve (on average) negative W ($\bar{W}_{ij} = -0.40 \pm \text{s.e.} = 0.01$), indicating a repressive matrix due to the strong selection against developmental instability. This repressive matrix was also observed for selected genes ($\bar{W}_{ij} = -0.50 \pm 0.03$), but was almost absent for unselected genes ($\bar{W}_{ij} = -0.03 \pm 0.02$). Moreover, both selected and unselected genes showed a peak around 0, corresponding to a strong signal for selected loss of interaction.

Further, the effect of mutations in the network on selected genes remained constant through time, while unselected genes evolved toward canalization (Fig. 2b). After 10,000 generations, selected genes were relatively decanalized, as a canalization score of $C \sim 5$ corresponds to a large phenotypic effect of mutations (s.d. = 0.08 for a trait ranging from 0 to 1). In contrast, unselected genes were substantially canalized ($C \sim 11$ is equivalent to a phenotypic s.d. = 0.004). These results mean that a random mutation anywhere in the network has on average less effect after several thousand generations, this phenomenon being largely due to the canalization of unselected genes.

Conditions for evolving canalization

We then tested the influence of each parameter on the evolution of canalization (Fig. 3). Changing the complexity c of the network (Fig. 3a) had limited effects on the evolution of canalization. Varying the gene constitutive expression a (Fig. 3b) showed that the canalization score is higher when the constitutive expression is close to the phenotypic optimum $\theta = 0.5$ (in other words, evolving canalization is easier close to the constitutive expression). The overlap of non-selected genes and control simulations is not a general phenomenon, as lower a makes evolution of canalization more difficult in selected networks, while larger a makes canalization easier. The largest levels of canalization in selected networks could be found for symmetric sigmoid functions ($a = 0.5$), which is a common (but somehow unrealistic) setting in similar models.

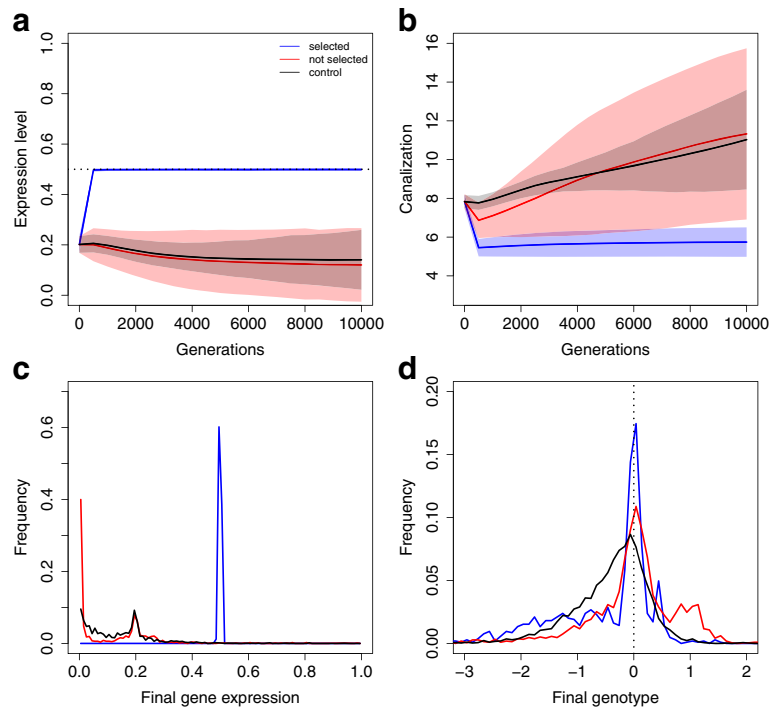


Fig. 2 Evolution of canalization. Genes under direct and indirect selection pressures evolved differently. **a** Evolution of gene expression. *Blue*: selected genes (two out of six genes in the network); *red*: unselected genes; *black*: control simulations (no selected genes). The horizontal dotted line stands for the selected expression level. Filled areas stand for the s.d. across 200 replicates, illustrating the amount of stochasticity in the simulations. Statistics were measured every 500 generations. **b** Evolution of canalization score C_i (negative log of the variance of mutational effects). **c** Distribution of gene expression values S_i at the last generation ($G = 10,000$). **d** Distribution of regulatory effects (W_{ij}) at the last generation (immutable 0s are discarded from the analysis). The vertical dotted line stands for the absence of regulation

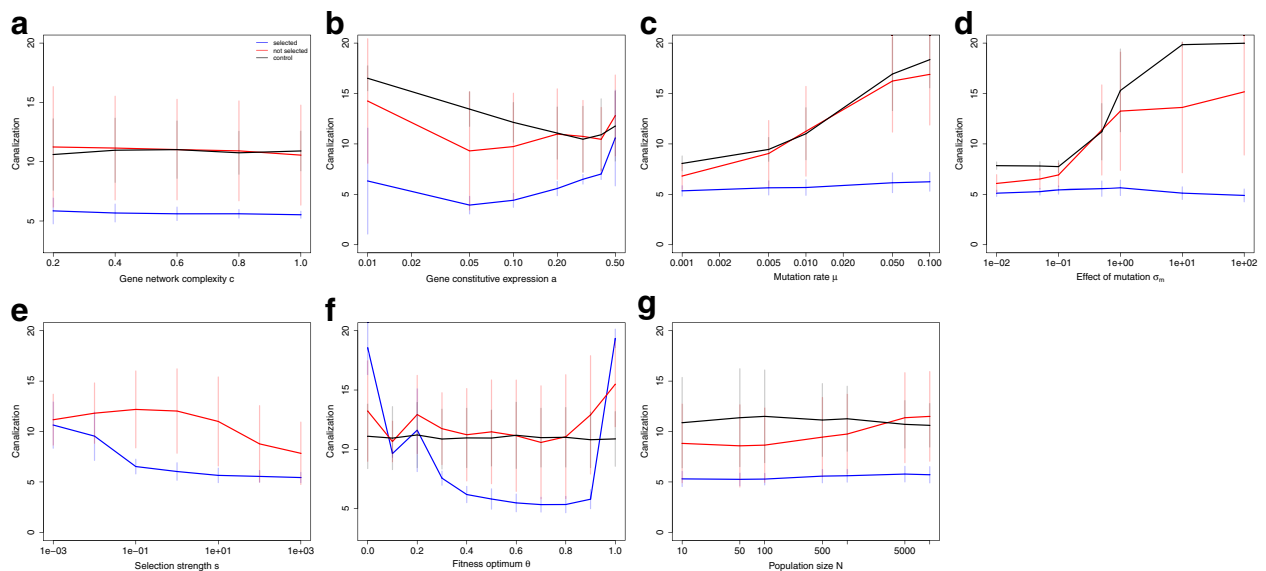


Fig. 3 Why canalization evolves. Average and s.d. canalization scores (negative log of the variance of mutational effects) at $G = 10,000$ generations; two genes out of six are under direct selection pressure. **a** Network complexity c . **b** Constitutive gene expression a . **c** Mutation rate μ . **d** Effect of a mutation σ_m . **e** Strength of stabilizing selection s . **f** Fitness optimum θ . **g** Population size N

In contrast, inflating mutational parameters (mutation rate μ and mutation size σ_m , Fig. 3c and d) increased the network robustness to mutations by orders of magnitude (from $C \sim 7$ to $C \sim 15$ on a logarithmic scale), equivalent to a drop in mutational standard deviation by a factor 60 (from s.d. (M_n) = 0.03 to s.d. (M_n) = 0.0005). This evolution of canalization concerned to a large extent unselected genes only. Overall, the effect of the different parameters on genetic canalization are scaled down for smaller mutation rates, but the main patterns remain very similar (Additional file 3: Figure S4). This is consistent with the theoretical prediction that selection for mutational robustness depends directly on the size and frequency of deleterious mutations. Selection to decrease mutation size remains a weak secondary selection strength, which is illustrated by the fact that it was less efficient in small populations (Fig. 3g).

Selection for developmental stability (control simulations) featured high levels of canalization. This is consistent with the unconstrained evolution of these networks toward a low-expression, developmentally-stable state (Fig. 2c). Paradoxically, the average canalization score decreased notably when adding stabilizing selection toward a phenotypic optimum (Fig. 3e). This indicates that the production of a stable optimal gene expression was possible (i.e. the population lies close to the fitness optimum), but this precludes the evolution of canalization.

Intermediate versus extreme gene expression

Strikingly, the value of the selected optimum had a huge effect on genetic canalization (Fig. 3f). Extreme expression levels (i.e. no expression or maximal expression) could be extremely canalized, while intermediate expression values were unable to become mutationally robust.

To better explore this difference between extreme or intermediate expression level, we ran additional simulations in which one gene out of six was selected toward 0, the second toward 1, and the others four remained free of direct selection pressure, mimicking on/off selection pressures (Fig. 4). In general, the average expression level in such networks was larger, and about 50 % of unselected genes were highly expressed at the end of the simulation (Fig. 4c). Both selected and unselected genes were canalized very efficiently (faster than in control simulations, Fig. 4b). Average regulatory effects were also larger than in control simulations, and were clearly positive in both cases ($\overline{W}_{ij} = 0.27 \pm 0.07$ and $\overline{W}_{ij} = 0.22 \pm 0.03$ for selected and unselected genes respectively, $\overline{W}_{ij} = -0.40 \pm 0.01$ for the control simulation, Fig. 4d).

This phenomenon can be explained by the saturation of regulatory effects in the vicinity of extreme expression levels. Indeed, genes achieved full expression or repression through the unlimited accumulation of multiple

regulatory factors. This was confirmed by the analysis of redundancy scores (Fig. 5a). The figure illustrates how selected genes tend to become more robust to knock-downs of regulatory factors when selected for extreme optima, while genes selected for intermediate expression are unable to change their redundancy level.

Degrees of freedom in the network

In the default simulations, only two genes out of six were submitted to direct selection pressure, allowing four degrees of freedom in the network (i.e. expression levels of these four genes could change due to drift and/or indirect selection without affecting fitness). We explored the effect of global selection constraints on the evolution of canalization, keeping the number of genes unchanged and applying stabilizing selection on a subset of genes ranging from one to all six genes (Fig. 6a). We observed that networks with more degrees of freedom tended to evolve more canalization than completely constrained networks. More specifically, both directly selected and unselected genes were more canalized in networks with only one target gene. Interestingly, from two to six genes under selection, canalization could not evolve for target genes, but the rest of the network could still benefit from additional degrees of freedom.

We also ran simulations in which the number of genes under selection did not change (two selected genes), but the global number of genes in the networks varied. The proportion of selected genes thus changed accordingly, while keeping constant the global selection pressure (Fig. 6b). Adding degrees of freedom in such a way enhanced the canalization score, and which was not only due to an increase in network size (control simulations showed decreasing canalization). Interestingly, robustness stabilized for unselected genes above 10 genes, but kept increasing for genes under selection.

The analysis of the indispensability scores (i.e. the influence of each gene on the expression of selected genes) illustrates how unselected genes lose their influence on the rest of the network during evolution, parallel to the increase in genetic canalization (Fig. 5b).

Discussion

Model approximations

Although our approach remains theoretical and cannot accurately represent all details of molecular and developmental mechanisms, gene network-based models derived from Wagner [15] have been widely used to study the evolutionary properties of complex genetic architectures. In an attempt to get more insights into the evolutionary changes associated with genetic canalization, we slightly modified this traditional framework toward more realistic assumptions, namely: (i) as in recent modeling work [19], gene expressions are quantitative variables ranging

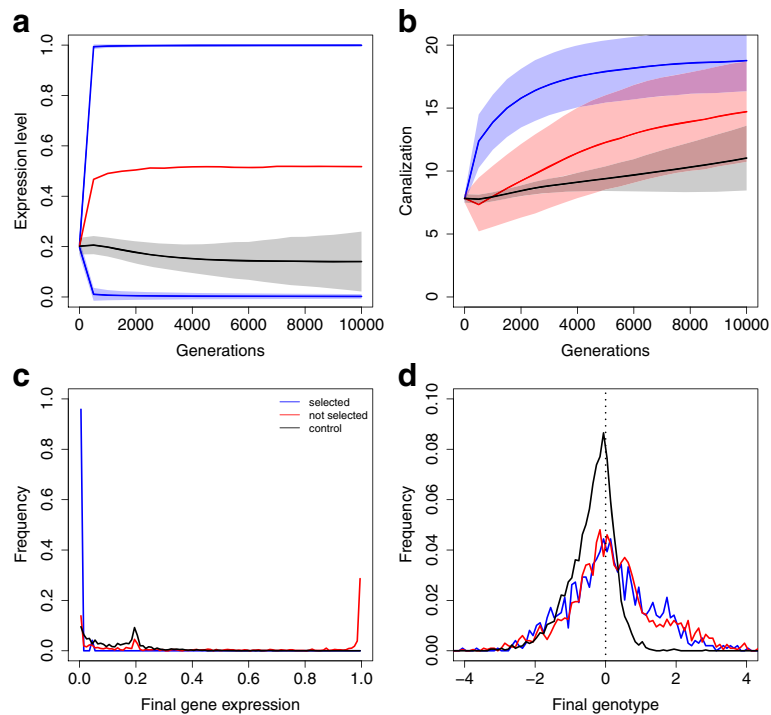


Fig. 4 Extreme phenotypic optima. In a six-gene network, one gene is selected for the minimum expression, one for the maximum expression, the four remaining being free to evolve. **a** Evolution of gene expression. Filled areas stand for the s.d. across 200 replicates (not shown for the non-selected genes as their distribution is bimodal). Statistics were measured every 500 generations. **b** Evolution of canalization score (negative log of the variance of mutational effects). **c** Distribution of gene expression values at the last generation ($G = 10,000$). **d** Distribution of regulatory effects (W_{ij}) at the last generation (immutable 0s are discarded from the analysis). The vertical dotted line stands for the absence of regulation

from 0 (no expression) and 1 (maximum expression); and (ii) we used an asymmetric regulation sigmoid function, featuring a 20 % constitutive gene expression simulating transcription leak. This contrasts with most gene network models, which generally consider that unregulated genes are either completely silent or substantially

expressed, none of these alternatives appearing as satisfactory when focusing on the strength of selection on regulatory sites. The shape of the sigmoid normalizing function is of importance, as its slope conditions the genotype-phenotype mapping, and explains the change in the effect of mutations as a function of gene expression

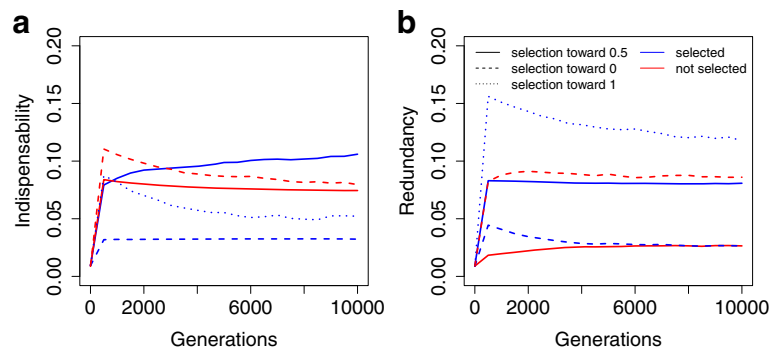


Fig. 5 Redundancy and indispensability *Blue*: selected genes (two out of six genes in the network); *red*: unselected genes; *solid lines*: simulation from Fig. 2, with genes selected toward intermediate optima; *dashed and solid lines*: simulation from Fig. 4, with genes selected for extreme optima, respectively 0 and 1. **a** Redundancy index. Low values indicate a gene regulated in a redundant way. **b** Indispensability index. Low values indicate a dispensable gene

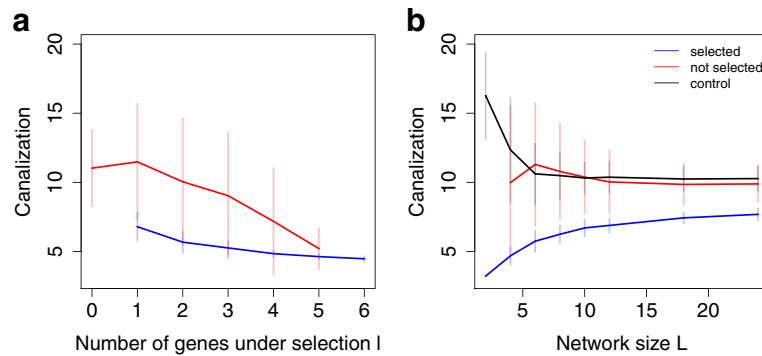


Fig. 6 Degrees of freedom. **a** Number of genes under selection in a six-gene network. **b** Network size (two genes under selection)

(intermediate expressions being more difficult to canalize than than extreme expressions [31]).

In our simulation study, we had to deal with computational constraints, limiting the population size, the number of genes, and the number of simulated generations. However, we could show that population size has only a minor influence on the final results, and we are thus confident that our results obtained with $N = 5,000$ individuals can be extrapolated to more realistic conditions. Consistently with the literature, network size quantitatively affects the evolution of canalization, larger networks being more robust [15, 16]. This effect, however, remains moderate compared to mutational parameters. In order to speed up simulations so that 10,000 generations are enough to get insights into long-term evolutionary properties, we used the common trick of increasing the mutation rate close to the largest existing estimates ($\mu = 0.01$ per gamete and per generation). Simulations with smaller mutation rates (Additional file 3: Figure S4) confirmed the qualitative validity of the large mutation rate simulation strategy.

Why canalization evolves

In this paper, our aim was to identify the conditions in which genetic canalization can evolve in gene networks, as well as the parameters influencing it. In particular, we confirmed that the evolution of canalization was mainly influenced by mutational parameters, canalization being maximal for high mutation rates and high mutation sizes. This result is consistent with previous studies [32, 33]. The impact of mutations on gene network evolution is two-fold: (i) mutations allow the population to spread over the neutral network space and to explore it faster [11], and (ii) a higher frequency of deleterious mutants increases the strength of indirect selection in favor of mutationally-robust genotypes [34].

We also established that the second parameter of great influence on the level of canalization in a network is the gene expression level. Much more genetic robustness

occurs when natural selection favors extreme gene expression values (all or nothing) rather than intermediate ones. Many previous studies generally considered selection toward extreme optima, thus focusing on a particular case that is the most favorable to the evolution of canalization. This new result thus highlights that the theoretical literature probably overestimates the generality of genetic canalization.

Most parameters influencing the gene network structure and topology (number of genes, network complexity) seem to have only moderate effects on canalization (see e.g. [21, 30] for a similar result). Hence, the propensity for a given network to be canalized can hardly be deduced from its basic topological features.

As a consequence, our results suggest that gene networks in various organisms may realistically harbour different canalization levels. The most canalized networks are expected to be found in genetic systems involving large population sizes, high mutation rates, and low selection levels (both in terms of selection strength and number of genes that influence directly the selected phenotypes). Larger gene networks are also expected to evolve (slightly) larger levels of canalization. In addition to these classical population genetic parameters, our results evidence a direct impact of the selected expression level on genetic canalization.

How canalization evolves

It is usually considered that regulatory gene networks are typical instances of complex systems, whose emergent properties cannot be understood by reductionist approaches based on single genes. Genetic or environmental canalization are often described as examples of such emergent properties. Unfortunately, this reasoning tends to consider genetic systems as undecipherable black boxes described by their properties rather than their internal mechanisms. This makes it extremely complicated to deduce evolutionary properties of a system from basic genetic knowledge (function and biochemical properties

of genes and proteins). Here, we identified two mechanisms prone to explain the general canalization features of gene regulatory networks without involving emergent properties *stricto-sensu*: (i) shrinkage of mutational target, and (ii) redundancy of regulatory information (Fig. 7).

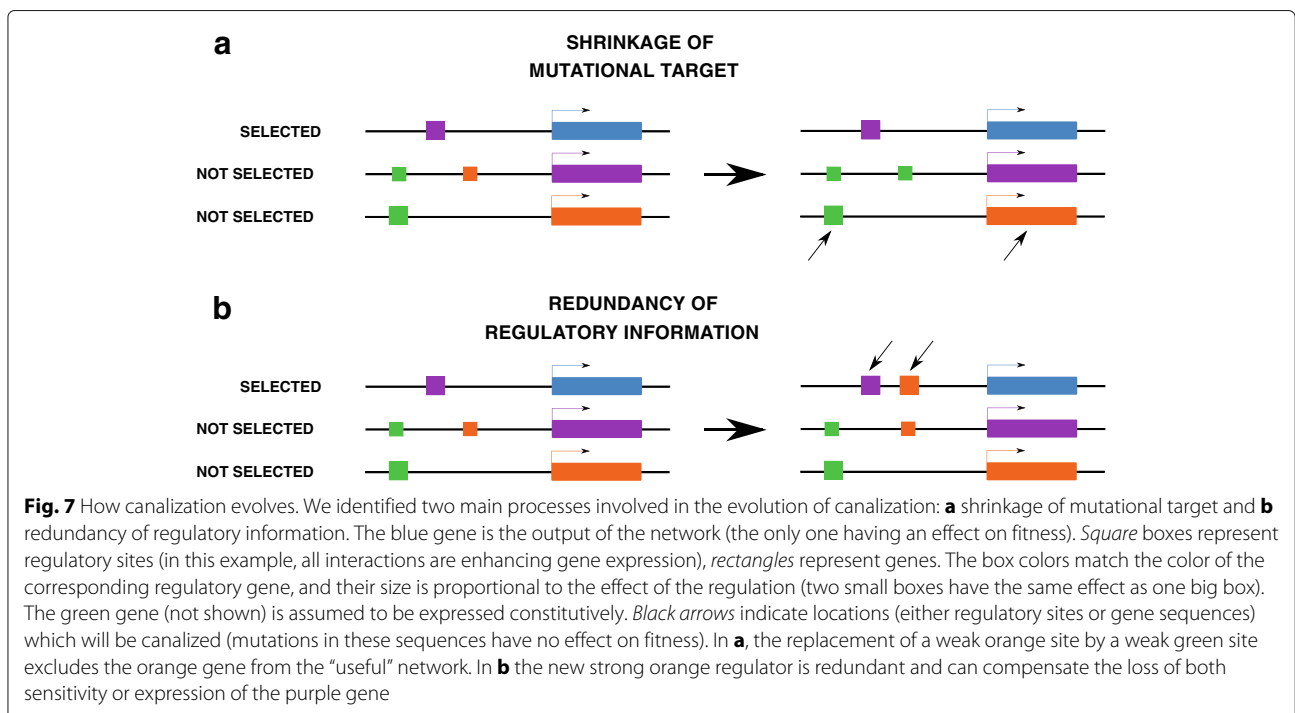
First, when some genes are not under the direct influence of natural selection, evolution in stable environments tends to favour genotypes in which such non-selected genes can be altered without influencing the important output of the network. This phenomenon is equivalent to a reduction of the network size, as some genes become virtually disconnected from the rest of the network. As a consequence, the number of mutations affecting the ‘useful’ genes in the network tends to decrease, and this shrinkage of mutational target leads to a progressive reduction in the mutational variance along generations. Interestingly, the directly-selected genes remain affected by *cis* mutations, but the lesser amount of epistatic interactions decreases the probability of being affected by *trans* mutations.

The second mechanism relies on the accumulation of redundant regulatory information, in such a way that the loss of one of the regulatory factors does not alter the expression of the target genes. In our model, this can happen only for target genes selected for extreme expression levels, due to a saturation effect linked to the shape of the sigmoid regulation-expression function (Eq. 2). This mechanism explains why canalization can evolve so easily for extreme optimum selection levels, which is the usual fitness landscape in Wagner-like models, even those based on sigmoid regulation expression functions

[16, 21, 25, 35]. These two mechanisms are non-exclusive, i.e. mutational target shrinkage is generally associated with redundancy. “Useless” genes become innocuous by evolving low interactions with selected genes and accumulating redundant repressors that canalize them into a developmentally stable, low-expression state. Whether or not these “simple” mechanisms are sufficient to explain the evolution of canalization in gene networks remains an open question. As a matter of fact, in our simulations, we could hardly observe evolution toward canalization in conditions where both of these mechanisms were hampered (direct selection on all genes, which prevents mutational target shrinkage, and selection for intermediate expression values, which prevents redundancy). This observation does not necessarily mean that more subtle emergent properties cannot evolve from complex gene networks. In particular, we have not modeled any complex selection pressure (e.g. favoring phenotypic plasticity [4, 15] or fluctuating selection targets [30, 36]).

Conclusion

The knowledge on how natural selection affects gene regulatory networks is so tenuous that inferring realistic selection conditions for the evolution of canalization in real gene networks has to rely on speculations. On one hand, it seems unlikely that all transcription factors are systematically targeted by direct selection. Most regulatory genes are only known to affect other transcription factors, and most networks seem to have a substantial amount of degrees of freedom (more network



genes than independent output). On the other hand, assuming that natural selection favours systematically extreme gene expression levels (no expression or maximal expression) is clearly more problematic. For instance, most transcriptomic studies focus on gene expression differences between genotypes or environmental conditions, bringing convincing evidence that quantitative (and sometimes modest) expression variations may have major impacts on physiology and development. It would be indeed surprising that selection optima on complex physiological processes are not intermediate. Our results thus suggest that gene networks underlying such subtle regulatory processes may be less prone to evolving strong long-term genetic canalization than previously thought.

Additional files

Additional file 1: Figures S2 and S1. Ploidy level. **Figure S2.** Effect of ploidy on the evolution of expression and canalization. Solid lines stand for diploid populations, dashed lines for haploid populations, and dotted lines for haploid populations with $\sigma_m = 2^{1/2}\sigma_m$ for the canalization tests (to compensate the increase in mutational effect that comes from the second unchanged haplotype in diploids). Subfigure indexes match figure numbers from the main text. (2a)(2b) Evolution of gene expression and canalization in networks in which genes are selected towards an intermediate optimum. (4a)(4b) Evolution of gene expression and canalization in networks in which genes are selected towards extreme expression. **Figure S3.** Effect of simulation parameters on canalization in haploid vs. diploid populations. Average and s.d. canalization scores at $G = 10,000$ generations; two genes out of six are under direct selection pressure. Solid lines stand for diploid populations, dashed lines for haploid populations. Subfigure indexes match figure numbers from the main text. (3a) Network complexity c . (3b) Constitutive gene expression a . (3c) Mutation rate μ . (3d) Effect of a mutation σ_m . (3e) Strength of stabilizing selection s . (3f) Fitness optimum θ . (5b) Population size N . (5b) Network size L . (ZIP 14 kb)

Additional file 2: Figure S3. Scale effect on canalization scores. **Figure S4.** Canalization measured as coefficient of variation. Evolution of canalization when measured relative to the phenotypic expression (i.e. in a similar way as a coefficient of variation: $C_{n_i} = \log \sqrt{\text{Var}(M_{n_i})}/S_{n_i}$). In practice, we computed $\bar{C}_n \simeq \frac{1}{2}\bar{C}_n - \log \bar{S}_n$. Subfigure indexes match figure numbers from the main text, the Y axis represents \bar{C} averaged over different categories of genes (selected, non-selected, and for subfigure (4b), selected for optima of 0 (top blue line) and 1 (bottom blue line)). (PDF 6 kb)

Additional file 3: Effect of mutation rate. **Figure S4.** Canalization still evolves at low mutation rates. Average canalization scores at $G = 10,000$ generations and $\mu = 0.0001$; two genes out of six are under direct selection pressure. Subfigure indexes match figure numbers from the main text. (3a) Network complexity c . (3b) Constitutive gene expression a . (3e) Strength of stabilizing selection s . (3f) Fitness optimum θ . (3g) Population size N . (5b) Network size L . (PDF 6 kb)

Acknowledgements

We thank all the members of the PHC Aurora "Genetic Architectures" project for insightful discussions. We acknowledge Andréas Odorico for useful comments on the manuscript. The English text was reviewed by Malcolm Eden.

Availability of data and materials

The current version of the software and the corresponding simulation results are available from the Zenodo database (doi:10.5281/zenodo.164420).

Authors' contributions

ER and ALR designed the experiments and the simulation framework. ER ran the simulations, analysed the data and designed figures. Both authors contributed equally to the manuscript. Both authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Received: 13 May 2016 Accepted: 14 October 2016

Published online: 08 November 2016

References

- Waddington CH. The Strategy of the Genes. A Discussion of Some Aspects of Theoretical Biology. London: Allen & Unwin; 1957.
- Debat V, David P. Mapping phenotypes: canalization, plasticity and developmental stability. *Trends Ecol Evol.* 2001;16(10):555–561.
- Le Rouzic A, Carlborg Ö. Evolutionary potential of hidden genetic variation. *Trends Ecol Evol.* 2008;23(1):33–37.
- Espinosa-Soto C, Martin OC, Wagner A. Phenotypic robustness can increase phenotypic variability after nongenetic perturbations in gene regulatory circuits. *J Evol Biol.* 2011;24(6):1284–1297.
- Espinosa-Soto C, Martin OC, Wagner A. Phenotypic plasticity can facilitate adaptive evolution in gene regulatory circuits. *BMC Evol Biol.* 2011;11(1):1.
- Rice SH. The evolution of canalization and the breaking of Von Baer's laws: modeling the evolution of development with epistasis. *Evolution.* 1998;52(3):647–656.
- Hansen TF, Wagner GP. Modeling genetic architecture: a multilinear theory of gene interaction. *Theor Popul Biol.* 2001;59(1):61–86.
- de Visser J, Hermisson J, Wagner GP, Meyers LA, Bagheri-Chaichian H, Blanchard JL, Chao L, Cheverud JM, Elena SF, Fontana W, et al. Perspective: evolution and detection of genetic robustness. *Evolution.* 2003;57(9):1959–1972.
- Hansen TF. Why epistasis is important for selection and adaptation. *Evolution.* 2013;67(12):3501–3511.
- Hermisson J, Wagner GP. The population genetic theory of hidden variation and genetic robustness. *Genetics.* 2004;168(4):2271–2284.
- Wagner A. Robustness and evolvability: a paradox resolved. *Proc R Soc Lond Ser B Biol Sci.* 2008;275(1630):91–100.
- Pigliucci M, Müller GB. *Evolution, the Extended Synthesis.* Cambridge: MIT Press; 2010.
- Gjuvsland AB, Hayes BJ, Omholt SW, Carlborg Ö. Statistical epistasis is a generic feature of gene regulatory networks. *Genetics.* 2007;175(1):411–420.
- Wagner A. Evolution of gene networks by gene duplications: a mathematical model and its implications on genome organization. *Proc Natl Acad Sci.* 1994;91(10):4387–4391.
- Wagner A. Does evolutionary plasticity evolve? *Evolution.* 1996;50(3):1008–1023.
- Siegal ML, Bergman A. Waddington's canalization revisited: developmental stability and evolution. *Proc Natl Acad Sci.* 2002;99(16):10528–10532.
- Masel J. Genetic assimilation can occur in the absence of selection for the assimilating phenotype, suggesting a role for the canalization heuristic. *J Evol Biol.* 2004;17(5):1106–1110.
- Le Cunff Y, Pakdaman K. Phenotype–genotype relation in Wagner's canalization model. *J Theor Biol.* 2012;314:69–83.
- Siegal ML, Promislow DE, Bergman A. Functional and evolutionary inference in gene networks: does topology matter? *Genetica.* 2007;129(1):83–103.
- Leclerc RD. Survival of the sparsest: robust gene networks are parsimonious. *Mol Syst Biol.* 2008;4(1):213.
- Ciliberti S, Martin OC, Wagner A. Robustness can evolve gradually in complex regulatory gene networks with varying topology. *PLoS Comput Biol.* 2007;3(2):15.
- Ciliberti S, Martin OC, Wagner A. Innovation and robustness in complex regulatory gene networks. *Proc Natl Acad Sci.* 2007;104(34):13591–13596.
- Espinosa-Soto C, Wagner A. Specialization can drive the evolution of modularity. *PLoS Comput Biol.* 2010;6(3):1000719.
- Wilder B, Stanley K. Reconciling explanations for the evolution of evolvability. *Adapt Behav.* 2015;23(3):171–179.

25. Azevedo RB, Lohaus R, Srinivasan S, Dang KK, Burch CL. Sexual reproduction selects for robustness and negative epistasis in artificial gene networks. *Nature*. 2006;440(7080):87–90.
26. Martin OC, Wagner A. Effects of recombination on complex regulatory circuits. *Genetics*. 2009;183(2):673–684.
27. Lohaus R, Burch CL, Azevedo RB. Genetic architecture and the evolution of sex. *J Hered*. 2010;101(suppl 1):142–157.
28. Ogbunugafor C, McBride R, Turner P. Predicting virus evolution: the relationship between genetic robustness and evolvability of thermotolerance. In: *Cold Spring Harbor Symposia on Quantitative Biology*. New York: Cold Spring Harbor Laboratory Press; 2009. p. 109–118.
29. Hansen TF, Pélabon C, Houle D. Heritability is not evolvability. *Evol Biol*. 2011;38(3):258–277.
30. Draghi J, Wagner G. The evolutionary dynamics of evolvability in a gene network model. *J Evol Biol*. 2009;22(3):599–611.
31. Palmer ME, Feldman MW. Dynamics of hybrid incompatibility in gene networks in a constant environment. *Evolution*. 2009;63(2):418–431.
32. Wagner GP, Booth G, Bagheri-Chaichian H. A population genetic theory of canalization. *Evolution*. 1997;51(2):329–347.
33. Elena SF, Sanjuán R. The effect of genetic robustness on evolvability in digital organisms. *BMC Evol Biol*. 2008;8(1):284.
34. Proulx SR, Phillips PC. The opportunity for canalization and the evolution of genetic networks. *Am Nat*. 2005;165(2):147–161.
35. Carneiro MO, Taubes CH, Hartl DL. Model transcriptional networks with continuously varying expression levels. *BMC Evol Biol*. 2011;11(1):1.
36. Watson RA, Wagner GP, Pavlicev M, Weinreich DM, Mills R. The evolution of phenotypic correlations and “developmental memory”. *Evolution*. 2014;68(4):1124–1138.

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

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

Submit your manuscript at
www.biomedcentral.com/submit



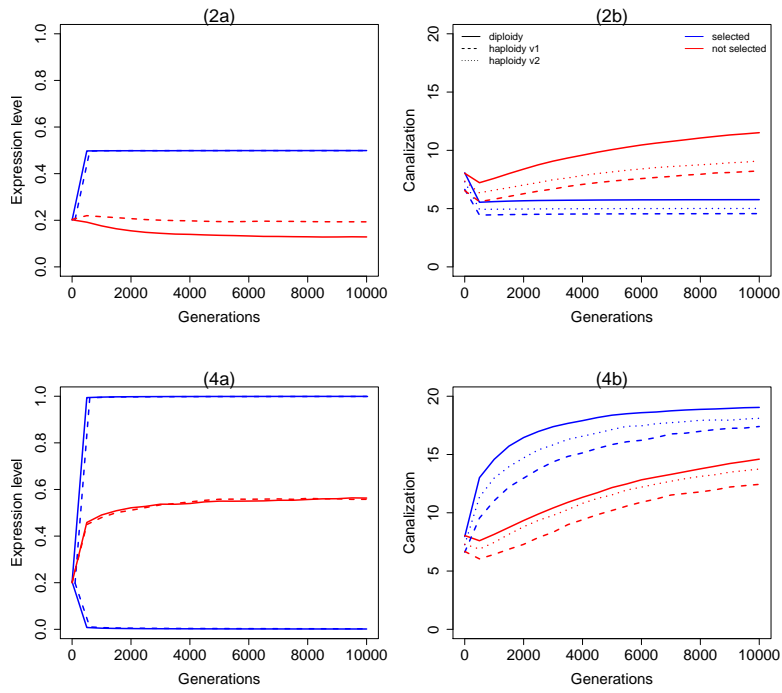


FIGURE 4.1 – Additional file 1-1 - Ploidy level

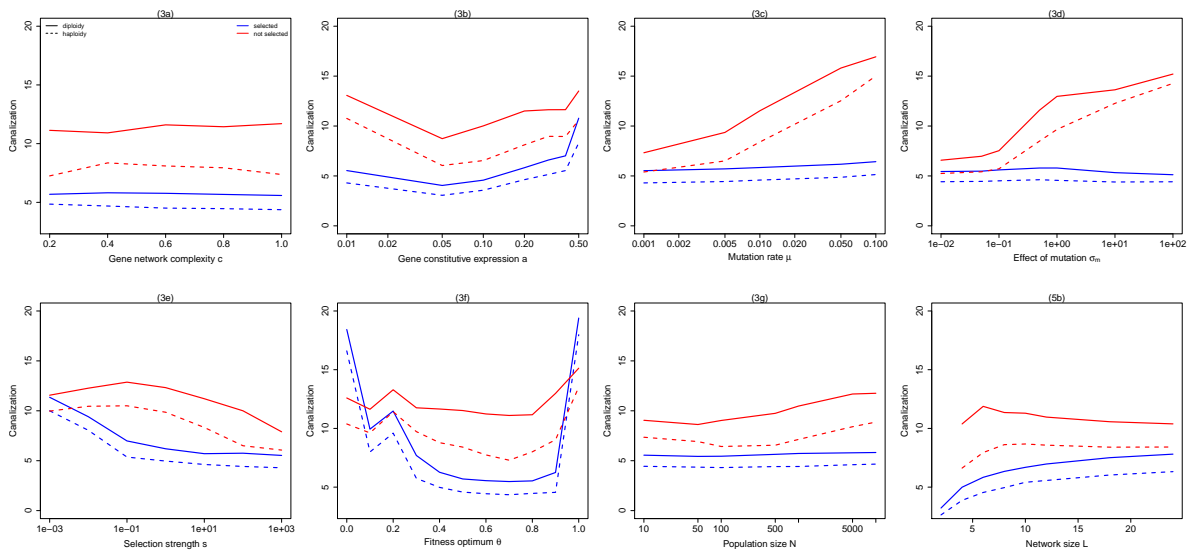


FIGURE 4.2 – Additional file 1-2 - Ploidy level

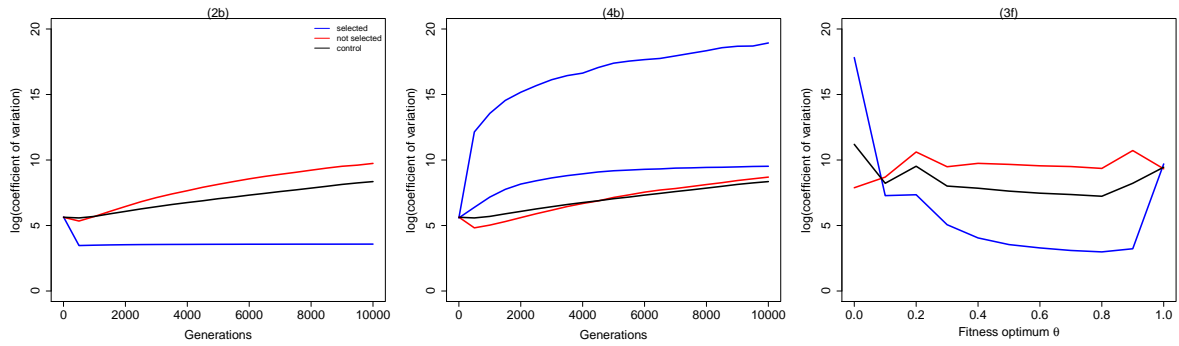


FIGURE 4.3 – Additional file 1 - Scale on canalisation score

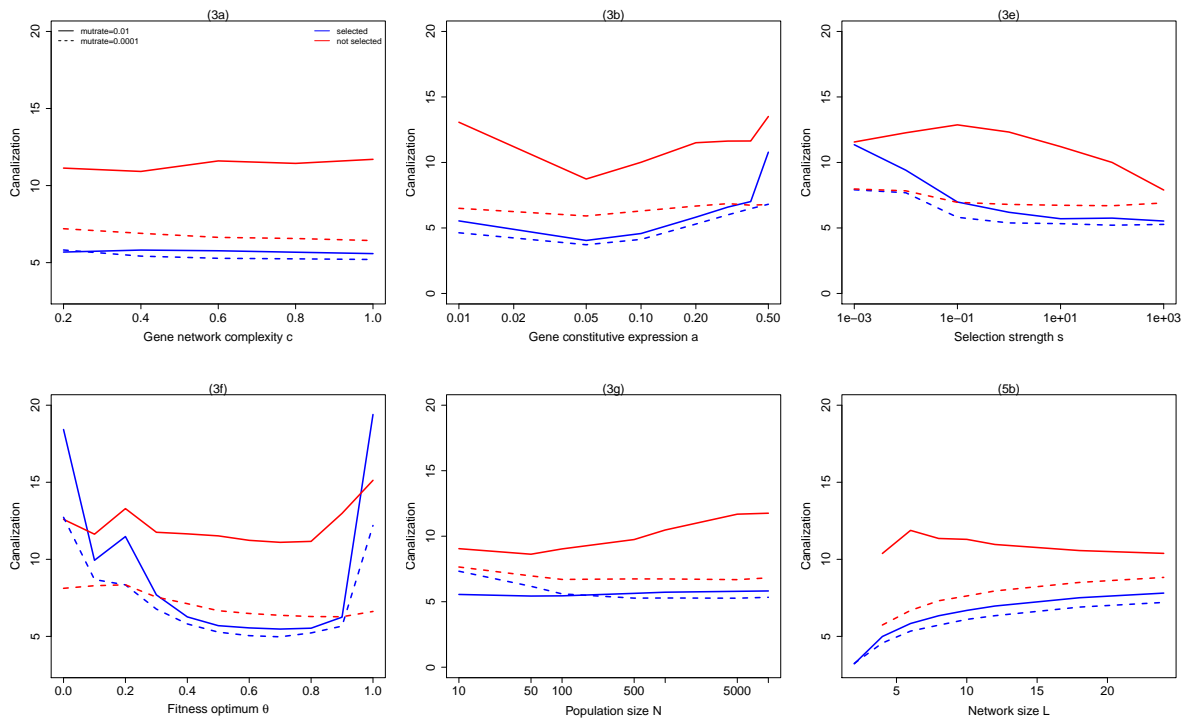


FIGURE 4.4 – Additional file 3 - Effect of mutation rate

Chapitre 5

Discussion et perspectives

Sommaire

5.1	Logiciel et modèle	79
5.1.1	Avantages et limites du logiciel de simulation	79
5.1.2	Approximations et paramétrisation du modèle	80
	Aspect quantitatif de l'expression génique	80
	Aspect quantitatif de l'optimum phénotypique	80
	Expression constitutive	81
	Fonction sigmoïde	81
	Mutation	81
	Dérive	81
	Taille du réseau	82
	Topologie des réseaux aléatoires	82
5.1.3	Calcul de la canalisation	82
5.2	Discussion générale et perspectives	83
5.2.1	Apport de la modélisation et généralisation des résultats	83
5.2.2	Importance de la sélection	84
5.2.3	Origine de la canalisation génétique	85
	5.2.3.1 Impact de la stabilité développementale	85
	5.2.3.2 Différents types de canalisation	88
5.3	Conclusion	91

5.1 Logiciel et modèle

Au sein de cette section, je m'intéresse uniquement au programme Simevolv vu comme un outil. Je tente d'adresser ici ses avantages et ses limites, ainsi que ceux du modèle en réseau de gène quantitatif. De même, j'explicite les choix de paramétrisation faits au cours des différentes analyses, et détaille le raisonnement qui nous y a conduits. Enfin, je m'intéresse plus spécifiquement à la mesure de la canalisation et aux changements qui pourraient lui être apportés.

5.1.1 Avantages et limites du logiciel de simulation

Le programme Simevolv a initialement été pensé comme une plateforme d'analyses multi-modèles. L'idée était de fournir une base commune (voir figure 2.9) permettant des analyses standardisées, sur laquelle les spécificités de chacun des modèles étudiés pouvaient être greffées (via l'utilisation de classes Architectures particulières, voir 2.2.2.2). En effet, si le génotype n'est pas codé de la même manière entre notre modèle et un modèle additif, par exemple, ou que le calcul du phénotype à partir du génotype varie, un individu reste représenté comme étant la combinaison d'un génotype et d'un phénotype, et une population reste un ensemble d'individus sur lesquels il est possible d'appliquer des forces de sélection et des analyses statistiques. Quelques modèles alternatifs d'architectures génétiques ont ainsi été implémentés dans le programme (voir 2.1.4).

Malgré notre décision de nous concentrer uniquement sur notre modèle principal, cette possibilité d'étudier d'autres types d'architecture reste présente. Si les modifications apportées tout au long de la thèse ont été nombreuses, l'aspect modulaire du programme (dû aux particularités du C++) a permis de les intégrer de manière souple, sans modifier de manière trop importante le reste du programme. Le programme a également été simplifié et de nouvelles fonctionnalités globales ont été ajoutées, comme la possibilité d'utiliser successivement plusieurs fichiers de paramètres (permettant par exemple une gestion plus simple de la sélection fluctuante.)

À l'heure actuelle, notre équipe possède une vision relativement détaillée du logiciel de simulation, de ce qui est possible ou pas de faire, des modules déjà implémentés, de ceux qu'il serait intéressant d'y rajouter et de la manière dont cela peut-être fait. Le logiciel a démontré qu'il était assez flexible pour supporter les modifications, même importantes. Malgré tout, des reliquats d'anciennes fonctions, désormais obsolètes, persistent et certaines fonctionnalités ne sont pas codées de façon optimale (dont la gestion des phénotypes). Aussi, une phase de nettoyage du logiciel et d'optimisation de certains processus pourrait s'avérer nécessaire avant de nouvelles études.

5.1.2 Approximations et paramétrisation du modèle

Aspect quantitatif de l'expression génique

Pouvoir utiliser la puissance des outils apportés par la génétique quantitative était un des prérequis du projet. Aussi, exprimer le phénotype (l'expression des gènes) de manière quantitative était une évidence. Pour cela, il a été décidé que les valeurs du vecteur S_t pourraient prendre toutes les valeurs possibles comprises dans l'intervalle $[0, 1]$.

Cela s'écarte de la majorité des études précédentes par deux points. Le premier est le nombre important d'études (proche de 50%) dans lesquelles le vecteur S_t est traité de manière qualitative (pas d'expression ou expression maximale), avec une fonction de normalisation à seuil (comme le modèle original, voir [WAGNER \[1996\]](#)) et non une sigmoïde, introduite pour la première fois par [SIEGAL et BERGMAN \[2002\]](#).

Le second est que la grande partie des études précédentes considèrent les valeurs de S_t comme pouvant être comprises dans l'intervalle $[-1, +1]$, avec -1 correspondant à une absence d'expression, +1 à une expression maximale, et 0 à une expression intermédiaire. Suivant [MASEL \[2004\]](#) et quelques autres études, j'ai choisi de privilégier un intervalle d'expression $[0, 1]$, qui semble plus réaliste d'un point de vue biologique. Si un gène est régulé de telle sorte qu'il ne s'exprime pas, il paraît plus naturel de considérer que son expression est nulle et non négative, ce qui pourrait laisser penser à une inversion de son rôle. De plus, ce choix d'échelle peut également être rapproché des valeurs obtenues en protéomique (qui suit de manière quantitative l'expression des protéines au cours du temps), permettant d'envisager l'injection de données réelles dans notre modèle.

Aspect quantitatif de l'optimum phénotypique

Cet aspect quantitatif de l'expression génique est également retrouvé au niveau de l'optimum phénotypique devant être atteint par les gènes sous sélection. Alors que la quasi-totalité des travaux dérivés du modèle de Wagner utilisent un optimum qualitatif (ou binaire : pas d'expression ou expression maximale), j'ai fait le choix d'étudier des optimums quantitatifs (toutes les valeurs possibles dans l'intervalle $[0, 1]$) et d'étudier l'importance de ce choix sur l'évolution des réseaux. Cette approche se rapproche des travaux de [PINHO et collab. \[2015\]](#), une des rares études dans laquelle les auteurs ont fait le choix d'explorer plusieurs valeurs d'optimum comprises dans l'intervalle $[0, 1]$.

D'un point de vue biologique, ce choix s'avère également judicieux et interprétable : il est en effet connu que pour de nombreux gènes, la régulation n'est pas affaire de tout ou rien, mais plutôt du maintien d'un niveau d'expression précis qui peut (potentiellement) évoluer au cours du temps.

Expression constitutive

Dans la description originelle du modèle, l'expression constitutive des gènes (c'est-à-dire leur expression en absence de toute régulation) est fixée à $a = 0$, donc à 50% de l'expression maximale (pour rappel, l'intervalle d'expression est $[-1, +1]$ chez [WAGNER \[1996\]](#)). Ce choix de considérer l'expression constitutive comme étant à 50% a été repris dans la majorité des études suivantes.

Au long de cette étude, je me suis écartée de ce parti-pris en fixant l'expression constitutive à $a = 0,2$ (soit 20% de l'expression maximale). Même si ce choix en particulier est assez arbitraire (0,1 ou 0,15 ou 0,25 auraient tout aussi bien pu être choisis), il me semble malgré tout plus cohérent biologiquement parlant. En effet, l'expression constitutive réelle des gènes peut s'avérer très variable et est généralement assez basse, nécessitant la présence de régulateurs pour d'exprimer à un niveau plus élevé.

Fonction sigmoïde

La fonction sigmoïde permet de borner le phénotype dans l'intervalle $[0,1]$ et assure également la mise à l'échelle vis-à-vis de la valeur des w_{ij} , permettant la comparaison entre des simulations aux w_{ij} différents. Sans cette mise à l'échelle (assurée par la présence de l'expression constitutive a dans la formule), il faudrait adapter à chaque simulation l'effet des mutations qui touchent la matrice W .

Par manque de temps, l'impact de la forme de la fonction normalisatrice sur les résultats obtenus n'a pas pu être analysé.

Mutation

L'un des problèmes majeurs conduisant au choix des paramètres par défaut du modèle est un problème temporel. Afin de permettre de garder un nombre de générations assez limité pour que le modèle tourne dans un temps correct tout en ayant une vision assez claire des propriétés évolutives à long terme, nous avons choisi d'utiliser un taux de mutation assez élevé ($\mu = 0,01$ par gamète et par génération). Des simulations ont été menées avec des taux de mutation inférieurs, menant à des résultats qualitativement similaires.

Dérive

Pour limiter l'influence de la dérive, nous avons travaillé avec des populations de taille conséquente (sans pour autant pénaliser le temps de simulation), soit $N = 5000$ individus. Au cours des analyses précédentes, certains résultats ont d'ailleurs confirmé que la taille de population, passé un seuil sous lequel la dérive prédomine ($N = 1000$ individus), n'avait qu'une influence mineure sur la canalisation et la réponse à la sélection.

Taille du réseau

Comme expliqué en chapitre 2.2.2.3, la taille réduite du réseau ($L = 6$) est principalement due à des contraintes de temps. S'il est vrai que les réseaux biologiques sont généralement d'une taille supérieure, ce choix nous a malgré tout permis d'observer les propriétés générales de ce type de modèle. De plus, les résultats obtenus semblent robustes même pour des réseaux de taille supérieure.

Topologie des réseaux aléatoires

La topologie des réseaux n'est ni codée (hormis le taux d'interaction) ni vérifiée dans le modèle, à cause de la complexité à produire des matrices aléatoires respectant un cahier des charges (topologie, densité des connexions, etc.) précis. Une étape de vérification au moment des analyses statistiques permet simplement d'éliminer les répliquats montrant des gènes déconnectés du reste du réseau (aucune variation de leur expression au cours des générations).

De même, aucune analyse statistique des propriétés topologiques des matrices de régulation n'a encore été implémentée (sens et force des régulateurs, nombre de connexions par gènes, etc.). De tels outils d'analyses existent pourtant (notamment en théorie des graphes) et pourraient servir à des études complémentaires.

5.1.3 Calcul de la canalisation

La canalisation est généralement décrite comme une diminution de la variabilité phénotypique au sein de la population. Nous avons repris cette définition pour le calcul du score de canalisation, défini comme $C = -\log \text{Var}(M)$. $\text{Var}(M)$ étant la variance de l'expression génique parmi une population de mutants qui diffèrent de l'individu initial par une seule perturbation ou mutation, nous mesurons ici la dispersion des phénotypes dans la population de mutants.

En revanche, l'écart entre les phénotypes atteints par les mutants et le phénotype de l'individu initial n'est pas pris en compte. Pourtant, intégrer cette dimension au score de canalisation serait intéressant. En effet, il est théoriquement envisageable (bien qu'improbable) d'avoir des individus clones à l'expression très proche les uns des autres (donc fortement canalisées d'après notre mode de calcul), mais ayant malgré tout un phénotype moyen éloigné de celui de l'individu initial, signe que les perturbations ont malgré tout modifié le comportement du réseau (ce qui indiquerait une faible canalisation).

Dans l'article présenté au chapitre 4, nous introduisons deux manières supplémentaires d'étudier la canalisation : le score de redondance et le score d'indispensabilité. Le score de redondance est l'effet moyen (sur un gène donné) de l'extinction d'un autre gène du réseau. Si ce score est faible, alors ce gène est régulé de manière redondante

et éteindre un seul régulateur n'est pas suffisant pour modifier substantiellement son expression. Le score d'indispensabilité quantifie l'effet de l'extinction d'un gène donné sur l'expression des gènes sous pression de sélection. Si ce score est faible, le gène éteint est dispensable et son extinction n'influence pas le reste du réseau. Des scores faibles de redondance et d'indispensabilité peuvent ainsi favoriser l'apparition de la canalisation dans les réseaux de régulations.

Cependant, ces scores ont été mis au point très tardivement (lors de la révision de l'article en vue d'une resoumission), aussi n'avons-nous encore que peu de recul sur leur intérêt ou sur la manière la plus adéquate de calculer ce genre d'indices.

5.2 Discussion générale et perspectives

Je mène dans cette section une discussion générale sur les résultats obtenus au cours de ce travail de thèse. En effet, l'organisation même du présent manuscrit a conduit à discuter des résultats particuliers au sein des chapitres correspondants.

5.2.1 Apport de la modélisation et généralisation des résultats

Le travail présenté tout au long de ce manuscrit a été réalisé à partir d'un modèle théorique, ce qui implique nécessairement des approximations et des simplifications. Il ne s'agit pas dans ce type d'approche de représenter toute la complexité du développement d'un organisme biologique, mais d'étudier un système assez simplifié pour permettre la mise en évidence de faits généraux, tout en restant assez détaillé pour approcher une certaine réalité biologique.

Ainsi, notre modèle permet de prendre en compte plusieurs paramètres d'intérêt biologique, que ce soit dans le type de reproduction ou dans les interactions géniques. Des notions complexes comme l'épistasie (via la complexité du réseau) ou la dominance peuvent ainsi être étudiées, ce qui s'avère généralement compliqué sur des modèles réels.

Cependant, il serait particulièrement intéressant de confronter les données obtenues à des données réelles. Les progrès substantiels de la transcriptomique permettent désormais la reconstruction fine de réseaux de régulation réels (sens et force des interactions, notamment), tout en suivant de manière dynamique l'expression des gènes du réseau au cours du temps. Notre modèle de simulation permet l'implémentation de telles données réelles : au lieu de partir de matrices W et de vecteurs d'expression initiale S_0 tirés de manière aléatoire, il est ainsi possible d'initialiser les simulations avec une matrice et un vecteur d'expression préremplis pour se rapprocher au mieux des

données expérimentales. En confrontant les résultats obtenus avec ceux des données réelles, cette approche permettrait de valider (ou d’invalidier, le cas échéant) la portée prédictive de telles études théoriques.

Il serait ainsi envisageable d’implémenter des données issues d’expérience de sélection artificielle sur des organismes unicellulaires, notamment celles impliquant un suivi complet des expressions géniques au cours du développement des individus, mais également au cours des générations successives. Cela permettrait d’étudier l’échelle développementale (est-ce que l’évolution des expressions géniques donnée par le logiciel est cohérente avec les données réelles?), mais aussi, et surtout, l’échelle populationnelle. Par exemple, il serait possible de suivre l’évolution de la topologie du réseau ou de la force des régulations, avant, pendant et après une phase de sélection directionnelle ou stabilisante, et de comparer les données obtenues aux données réelles.

5.2.2 Importance de la sélection

Au cours de ces travaux de thèse, j’ai validé l’effet décanalisant de la sélection directionnelle, ainsi que celui canalisant de la sélection stabilisante, du moins dans des réseaux de régulation (quantitatifs aussi bien que qualitatifs). Cependant, je ne me suis pas intéressée à l’impact d’autres types de sélection sur l’apparition et le maintien de la canalisation. Peu d’études ont d’ailleurs été réalisées sur le sujet, du moins avec ce type de modèles. Si **FIERST [2011a]** ou **WILDER et STANLEY [2015]** se sont intéressés à la question, en testant l’importance des environnements variables ou en implémentant différents modes de sélections, leurs résultats portaient sur l’évolvabilité et la capacité d’adaptation des individus et des populations, pas sur la canalisation.

Pourtant, il est peu probable que des phénomènes de canalisation puissent apparaître dans une population soumise à une sélection autre que stabilisante. Notamment, la sélection fluctuante serait un objet d’étude particulièrement intéressant. En effet, si la sélection stabilisante vise au maintien d’un phénotype optimal sur de longues périodes de temps, nécessitant de ce fait le tamponnage des perturbations qui pourraient éloigner les individus ou la population de cet optimum, il n’en va pas de même pour la sélection fluctuante. Cette dernière est caractérisée par le changement plus ou moins rapide de l’optimum phénotypique auquel est soumise la population, et nécessite donc une certaine variabilité phénotypique, ce qui semble contre intuitif avec l’évolution de la canalisation. Pour autant, des études antérieures ont montré que de la sélection fluctuante pouvait bel et bien évoluer de la canalisation génétique (**KAWECKI [2000]**; **LE ROUZIC et collab. [2013]**). Il serait donc intéressant de pouvoir déterminer si ce résultat peut être reproduit avec notre modèle, et dans quelles conditions (notamment la fréquence des changements) la canalisation peut être favorisée.

5.2.3 Origine de la canalisation génétique

Grâce à notre modèle de réseaux de gènes quantitatifs, j'ai montré que des réseaux de régulation quantitatifs peuvent faire évoluer de la canalisation génétique dans certaines conditions. La canalisation est ainsi maximisée dans des réseaux avec peu de gènes sous pression de sélection directe et soumis à une forte activité mutationnelle. Ces observations semblent aller dans le sens de l'hypothèse évolutive, qui postule que la canalisation serait adaptative (WADDINGTON [1942]; WAGNER [1996]).

En revanche, certains résultats semblent aller dans le sens de l'hypothèse intrinsèque (notamment soutenue par HUERTA-SANCHEZ et DURRETT [2007]; SIEGAL et BERGMAN [2002] et qui veut que la canalisation ne soit qu'une conséquence de l'architecture génétique des réseaux de régulations). Notamment, j'ai montré que des réseaux dépourvus de toute contrainte évolutive autre que la sélection contre l'instabilité au cours du développement peuvent également faire évoluer de la canalisation. Des études plus poussées sur les relations entre ces deux mécanismes semblent donc nécessaires afin d'avoir une vision plus complète du processus d'évolution de la canalisation dans des réseaux de régulation génétiques.

Il reste une hypothèse cherchant à expliquer l'apparition de la canalisation qui n'a pas du tout été approchée au cours des travaux précédents. Il s'agit de l'hypothèse congruente, qui veut que la canalisation génétique soit une conséquence de l'évolution de la canalisation contre les perturbations environnementales (ANCEL et collab. [2000]; LE CUNFF et PAKDAMAN [2012]). Il serait ainsi intéressant de pouvoir étudier l'effet des perturbations environnementales (modification de l'état initial ou perturbations au cours du développement) sur la canalisation génétique. De même, étudier les interactions entre la canalisation génétique et les canalisations environnementales serait une piste de travail intéressante.

Dans la suite de cette section, je présente quelques résultats préliminaires. Les premiers s'attachent à la relation entre la canalisation génétique et l'effet de la sélection contre l'instabilité développementale. Les seconds étudient l'hypothèse congruente et l'impact des perturbations environnementales. Entamées en fin de thèse, ces quelques études ne sont pas assez poussées pour être exploitables en l'état, mais peuvent malgré tout donner des pistes de réflexion pour des travaux futurs.

5.2.3.1 Impact de la stabilité développementale

J'étudie ici la relation entre canalisation et sélection contre l'instabilité développementale (rappelons que selon WAGNER [1996] la première serait un effet secondaire de la seconde). Dans certaines des études présentées plus tôt dans cette thèse, j'ai montré que des réseaux soumis uniquement à de la sélection contre l'instabilité développemen-

tale pouvaient faire évoluer de la canalisation, et ce même en l'absence de sélection stabilisante vers un optimum donné. Le but de cette partie est d'étudier l'aspect canalisant (ou non) de la stabilité développementale et de le mettre en relation avec la canalisation due à la sélection stabilisante.

Pour cela, je compare quatre simulations distinctes (figure 5.1). La première, "dérive", présente un réseau non soumis à la sélection, ni vers un optimum d'expression, ni contre l'instabilité développementale : il s'agit donc de dérive génétique pure. La seconde, "optimum", présente un réseau de six gènes, dont deux sont soumis à la sélection stabilisante vers un optimum phénotypique donné ($\theta = 0,5$ sur les 4 derniers pas de temps du développement), mais l'instabilité développementale est autorisée. Dans la troisième, "stabilité", tous les gènes sont contre sélectionnés contre l'instabilité développementale, mais ne sont soumis à aucune force de sélection stabilisante. Enfin, la dernière simulation, "optimum+stabilité" est identique à celle présentée au chapitre 4 : tous les gènes sont contre-sélectionnés contre l'instabilité développementale, et deux d'entre eux sont sous pression de sélection stabilisante pour un optimum donné ($\theta = 0,5$).

Le premier résultat marquant est le comportement du réseau soumis uniquement à l'effet de la dérive génétique (courbe noire). D'une manière générale, si certains gènes tendent à rester à leur expression initiale, la plupart tendent progressivement vers une expression maximale de 1, expliquant l'augmentation de l'expression des gènes (en moyenne et en variance, figures 5.1a et 5.1b). Ce comportement (en absence de toute

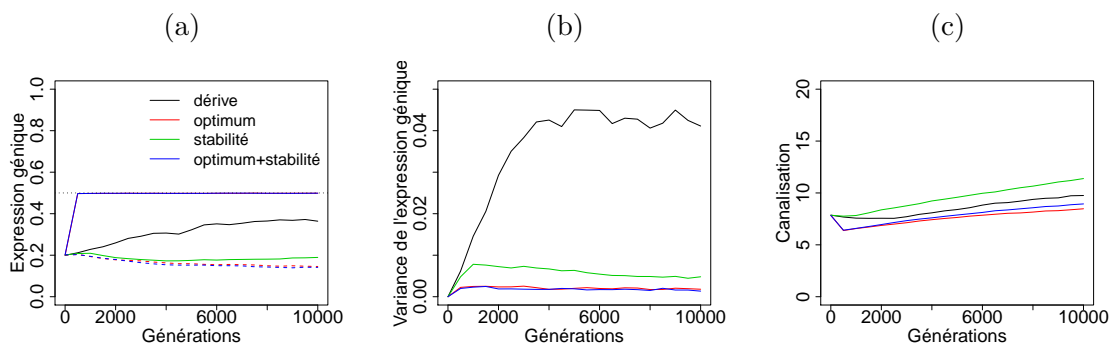


FIGURE 5.1 – IMPACT DE LA SÉLECTION CONTRE L'INSTABILITÉ DÉVELOPPEMENTALE. La courbe noire "dérive" représente un réseau soumis uniquement à la dérive (pas de sélection ni sur le phénotype ni sur l'instabilité). La courbe rouge "optimum" représente un réseau dont deux gènes (sur six) sont soumis à la sélection stabilisante. La courbe verte "stabilité" représente un réseau non soumis à la sélection stabilisante, mais dont les gènes sont sélectionnés contre l'instabilité développementale. La courbe bleue "optimum+stabilité" correspond à un réseau soumis aux deux forces de sélection : tous les gènes sont sélectionnés contre l'instabilité développementale, et deux gènes (sur six) sont soumis à la sélection stabilisante. (a) Expression génique \bar{S} . Dans le cas des réseaux soumis à la sélection stabilisante, les deux gènes sous sélection sont représentés par une ligne pleine, les quatre autres par une ligne pointillée. (b) Variance de l'expression génique $Var(\bar{S})$. (c) Canalisation C_{mut} .

pression de sélection) est dû à la forme particulière de la fonction normalisatrice $f(x)$, qui tend à pousser l'expression des gènes vers les extrêmes. L'expression des gènes à leur valeur maximale ou minimale a également pour conséquence une augmentation mécanique du score de canalisation, vraisemblablement par la mise en place de la redondance des régulations (figure 5.1c).

Le comportement d'un réseau soumis uniquement à la sélection contre l'instabilité phénotypique (courbe verte) montre un comportement très différent d'un réseau soumis à la dérive pure. L'évolution de l'expression des gènes (tant en moyenne qu'en variance) semble contrainte et se rapproche de celle observée dans les réseaux soumis à la sélection stabilisante (figures 5.1a et 5.1b).

Comme montré précédemment, les réseaux uniquement soumis à de la sélection contre l'instabilité développementale peuvent non seulement faire évoluer de la canalisation, mais la faire évoluer à des niveaux supérieurs à ceux obtenus en la présence de sélection stabilisante (figure 5.1c). La sélection contre l'instabilité développementale est donc bien intrinsèquement canalisante, comme envisagé par **WAGNER [1996]** et **SIEGAL et BERGMAN [2002]**. Cette importance de l'hypothèse intrinsèque est également soutenue par l'aspect canalisant de la dérive génétique.

En revanche, les deux simulations soumises à l'effet de la sélection stabilisante (courbes rouge et bleu) montrent des niveaux de canalisation moindres, indiquant un certain effet décanalisant de la sélection stabilisante. Ceci est d'ailleurs à rapprocher des observations de l'article présenté au chapitre 4, qui montrent que la canalisation des réseaux de régulation est due aux gènes qui ne sont pas directement sous pression de sélection stabilisante.

Il est également intéressant de noter que ces deux simulations suivent des trajectoires similaires, voire superposables. Ainsi, ajouter ou retirer l'effet de la sélection contre l'instabilité développementale n'influe pas sur la capacité de réponse du réseau à la sélection, ni sur le score de canalisation. Il est probable que cela soit dû à une pression plus importante exercée sur le réseau par la sélection directionnelle que par la sélection contre l'instabilité développementale.

Si ces résultats semblent aller dans le sens de l'hypothèse intrinsèque développée par **SIEGAL et BERGMAN [2002]**, cette étude préliminaire n'est cependant pas suffisante pour trancher. Il serait notamment intéressant de répéter cette étude avec des jeux de paramètres augmentant la probabilité d'observer de l'instabilité développementale au début des simulations. En effet, la section 3.1.2 a permis de démontrer que le jeu de paramètres utilisé ici conduit à un taux d'instabilité des matrices initiales très faible, ce qui pourrait être évité en augmentant l'écart-type servant à initialiser les allèles.

De même, seuls deux cas de sélection contre l'instabilité développementale ont été étudiés ici : pas de sélection ou contre-sélection totale. Il est possible que des analyses

réalisées avec des forces de sélection contre l'instabilité développementale intermédiaires donnent des résultats différents. Enfin, il est probable que la canalisation du réseau dépende d'un équilibre entre les effets des deux forces de sélection. Des études plus poussées faisant intervenir une large gamme de force de sélection (tant vers un optimum fixé que contre la variation de l'expression génique pendant le développement) semblent ainsi nécessaires.

5.2.3.2 Différents types de canalisation

Je m'intéresse dans cette partie à l'effet de la présence de perturbations environnementales sur l'évolution de la canalisation génétique, afin de déterminer si soumettre le réseau à des perturbations autres que génétique conduit à un changement dans le score de canalisation génétique, comme le préconise l'hypothèse congruente.

Pour cela, j'ai lancé une nouvelle série de quatre simulations (figure 5.2). Dans la première simulation, "mutation", le réseau n'est soumis qu'à l'effet des mutations (et est donc équivalent aux simulations présentées jusqu'ici). Dans la deuxième, "init", des perturbations de l'état initial ont été ajoutées au réseau selon les modalités décrites en section 2.1.3 (avec des perturbations d'effet $\sigma_i = 0, 1$). Dans la troisième, "dynam", le réseau est soumis à des perturbations se produisant tout au long du développement (d'effet $C_{init} = 0,01$). Enfin, les deux types de perturbations environnementales ont été ajoutés au réseau dans la dernière simulation, "init+dynam".

Perturber l'état initial (courbe rouge, à comparer à la noire) ne semble pas perturber les capacités d'évolution du réseau, que ce soit au niveau de la valeur sélective (figure 5.2a) ou au niveau de la canalisation génétique (figure 5.2d).

Ajouter des perturbations développementales (courbe verte et bleue) diminue à la fois la valeur sélective moyenne de la population et le score de canalisation génétique

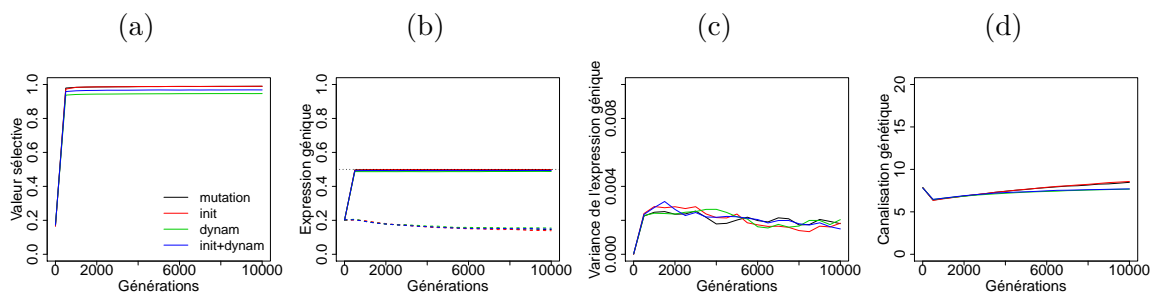


FIGURE 5.2 – IMPACT DES PERTURBATIONS ENVIRONNEMENTALES. La courbe rouge "mutation" représente un réseau uniquement soumis aux mutations. D'autres types de perturbations sont également ajoutées au réseau pour les autres simulations : perturbation de l'état initial (courbe rouge "init"), perturbations au cours du développement (courbe verte "dynam"), ou les deux (courbe bleue "init+dynam"). (a) Valeur sélective ω . (b) Expression génique \bar{S} . (c) Variance de l'expression génique $Var(\bar{S})$. (d) Canalisation génétique C_{mut} .

obtenu. Cela est dû à la manière d'implémenter ces perturbations dans le modèle : en effet, le phénotype est modifié directement, ce qui rajoute de la variation d'expression des gènes qui ne peut pas être contrebalancée par le réseau.

Il semble ainsi que soumettre le réseau à d'autres types de perturbations (ici, perturbation de l'état initial ou perturbations de l'expression génique à chaque pas de temps) ne conduise pas à une augmentation de la canalisation génétique, comme le voudrait l'hypothèse congruente. Cependant, cette étude reste trop préliminaire pour conclure de manière définitive sur ce point.

L'étude de la canalisation environnementale a également été approchée dans des réseaux aléatoires (correspondant aux réseaux que l'on peut trouver avant évolution), dans le but de déterminer s'il existait ou non des corrélations observables entre les différents types de canalisation (canalisation génétique, canalisation contre l'instabilité au cours du développement, canalisation contre les perturbations de l'état initial, canalisation contre les perturbations au cours du développement).

Pour cela, des études ont été réalisées à l'aide du programme en R présenté en section 2.3. Le score d'instabilité développementale, "instab", est à nouveau défini comme la variance de l'expression génique V_S sur les 4 derniers pas de temps. En parallèle, trois nouveaux scores de canalisation, "gen", "init" et "dev", ont été créés selon les modalités présentées en section 2.1.3 : les mutations sont codées comme des modifications de la matrice W , les perturbations initiales modifient le vecteur S_0 et les perturbations développementales touchent les vecteurs S_t tout au long du développement. Chaque individu ainsi modifié subit un processus développemental complet, puis la variance de l'expression génique entre les individus ($\text{Var}(M)$) est calculée. Cette variance est transformée logarithmiquement pour donner le score de canalisation, selon la formule $C = -\log \text{Var}(M)$.

La figure 5.3 présente l'évolution de ces scores de canalisation en fonction des paramètres du réseau. Comme déjà démontré au chapitre 3 pour la seule canalisation génétique, les deux paramètres ayant le plus d'influence sur les scores de canalisation sont la moyenne et l'écart type utilisés pour l'initialisation des w_{ij} (c'est-à-dire l'effet et la force des régulateurs), et ce pour les quatre scores (figures 5.3d et 5.3e).

La canalisation des perturbations de l'état initial (courbe verte) semble également modifiée par la taille et la complexité du réseau (figures 5.3b et 5.3c) : plus le réseau est grand et dense, plus le score de canalisation diminue. L'inverse peut être observé pour les perturbations développementales (courbe bleu).

Les corrélations entre ces quatre scores ont également été calculées (figure 5.4). Le tracé incomplet de la figure 5.4e peut-être expliqué par l'absence de variance (canalisation parfaite) pour les scores "init" et "instab" de la figure 5.3, empêchant le calcul

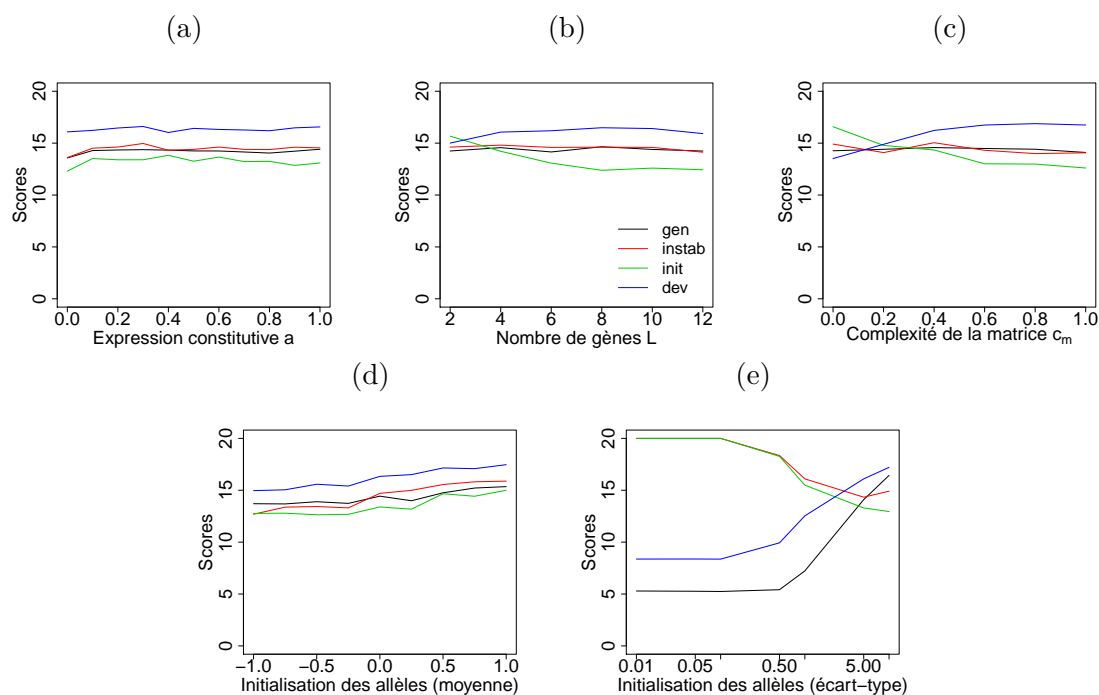


FIGURE 5.3 – SCORES DE CANALISATION DANS DES RÉSEAUX ALÉATOIRES. (a) Valeur de l’expression constitutive a . (b) Nombre de gènes L . (c) Complexité de la matrice c_m . (d) Initialisation des allèles : moyenne de la loi gaussienne. (e) Initialisation des allèles : écart-type de la loi gaussienne.

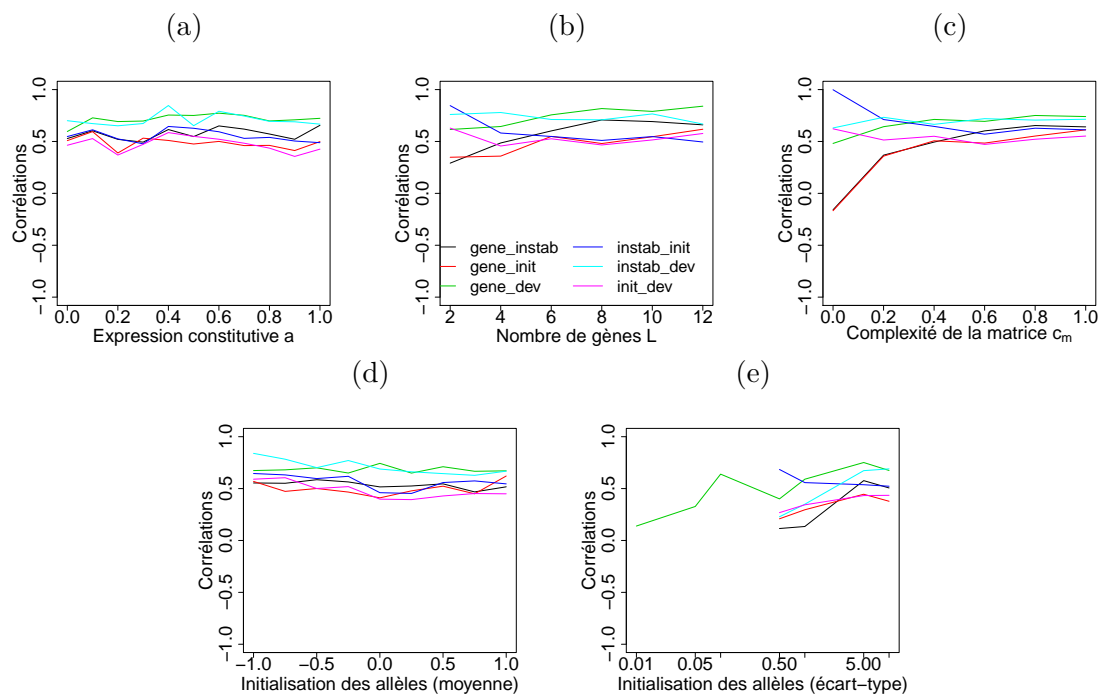


FIGURE 5.4 – CORRÉLATIONS ENTRE LES SCORES DE CANALISATION DANS DES RÉSEAUX ALÉATOIRES. (a) Valeur de l’expression constitutive a . (b) Nombre de gènes L . (c) Complexité de la matrice c_m . (d) Initialisation des allèles : moyenne de la loi gaussienne. (e) Initialisation des allèles : écart-type de la loi gaussienne.

du score de corrélation.

Le résultat majeur de cette figure est l'existence de corrélations positives et significatives entre tous ces différents scores. Cependant, certaines corrélations semblent être sous la dépendance de la force des régulateurs (écart-type des w_{ij} , figure 5.4e). De même, il est possible d'observer que les corrélations les plus fortes, indépendamment des paramètres du réseau, sont celles entre la canalisation génétique et la canalisation des perturbations durant le développement (courbe verte), et celles entre instabilité développementale et canalisation des perturbations durant le développement (courbe bleu clair).

En l'état, ces résultats ne permettent pas de trancher de manière certaine sur les possibles relations entre ces différentes formes de canalisation et d'autres analyses sont encore nécessaires. De plus, il ne s'agit ici que d'analyses réalisées sur des réseaux aléatoires, en l'absence de toute force évolutive. Il serait particulièrement intéressant de tester ces relations dans des réseaux évolués et de voir si ces liens persistent ou si de nouveaux peuvent apparaître.

5.3 Conclusion

Au cours de ma thèse, j'ai montré que les réseaux de régulation quantitatifs sont aptes à répondre à la sélection directionnelle et stabilisante. J'ai également démontré que le régime de reproduction n'a que peu d'impact sur cette capacité et que des populations haploïdes ou diploïdes présentent un comportement similaire.

Un autre résultat marquant est l'absence d'équilibre mutation-dérive-sélection après des milliers de générations, majoritairement due à une diminution progressive de la variabilité phénotypique et, plus précisément, de son composant mutationnel. Cela nous a permis de mettre en évidence la présence de canalisation génétique dans notre modèle de réseau.

En étudiant plus précisément les paramètres permettant l'évolution de la canalisation dans les réseaux de régulation quantitatifs, j'ai rappelé l'effet majeur des paramètres mutationnels (fréquence et effet), mais également l'importance de l'optimum pour lequel le phénotype est sélectionné. En effet, lorsque le phénotype est sélectionné pour des valeurs extrêmes (ou qualitatives), les réseaux sont plus aptes à faire évoluer de la canalisation que lorsque le phénotype optimal est à un niveau d'expression intermédiaire. Cela est certainement dû à la difficulté d'avoir une régulation fine, alors que des valeurs extrêmes peuvent être atteintes et maintenues par le biais de la redondance dans la régulation.

Cette observation de l'impact de la valeur phénotypique optimum est particulièrement intéressante, car elle apporte un nouvel éclairage aux études antérieures sur l'évolution de la canalisation, généralement réalisées avec des modèles qualitatifs. Elle permet également de mieux comprendre les mécanismes qui peuvent prendre place dans des organismes réels, bien plus proches d'un modèle quantitatif que d'un modèle qualitatif. Nos travaux suggèrent ainsi que les organismes réels sont moins aptes à produire des niveaux importants de canalisation génétique que ce qui avait été préalablement envisagé.

Pour autant, ces résultats restent sujets à caution, car potentiellement modifiables par l'ajout au modèle des perturbations environnementales, qui n'ont pas encore été étudiées en profondeur, mais touchent de manière omniprésente les organismes biologiques.

Liste des symboles

a Expression constitutive des gènes.

C_{dev} Canalisation des perturbations au cours du développement.

C_{init} Canalisation des perturbations de l'état initial.

C_{mut} Canalisation génétique.

c_m Complexité de la matrice W , représentant la proportion de $w_{ij} \neq 0$.

c_d Complexité de la diagonale de matrice W .

f Taux d'autofécondation.

G Nombre de générations dans une simulation.

g Nombre de générations entre deux calculs des valeurs de sortie.

L Nombre de loci dans le réseau.

ℓ Nombre de loci sous sélection dans le réseau.

μ Taux de mutation par individu.

σ_m Effet des mutations.

N Taille de la population.

r Taux de recombinaison entre loci.

s Force de la sélection stabilisante.

s' Force de la sélection contre l'instabilité développementale.

σ_d Effet des perturbation des vecteurs S_t au cours du développement.

σ_i Effet des perturbation du vecteur initial S_0 .

\bar{S} Expression génique ou phénotype (moyenne de l'expression sur les τ_c derniers pas de temps du développement).

V_S Variance de l'expression génique ou instabilité (variance de l'expression sur les τ_c derniers pas de temps du développement).

S_t Vecteur de taille L représentant l'expression des gènes du réseau (phénotype de l'individu).

S_0 Vecteur de taille L représentant l'expression initiale du réseau.

θ Valeur de l'expression optimale des gènes sous sélection.

τ Nombre de pas de temps de développement d'un individu.

τ_c Nombre de pas de temps de développement sur lesquels sont calculés la moyenne et la variance de l'expression génique.

ω Valeur sélective d'un individu.

ω_d Composant "distance" de la valeur sélective d'un individu.

ω_k Composant "instabilité" de la valeur sélective d'un individu.

W Matrice de taille $L \times L$ représentant le génotype d'un individu.

w_{ij} Case de la matrice W , représentant l'effet du gène i sur l'expression du gène j .

Bibliographie

- ANCEL, L. W., W. FONTANA et collab.. 2000, «Plasticity, evolvability, and modularity in rna», *Journal of Experimental Zoology*, vol. 288, n° 3, p. 242–283.
- AZEVEDO, R. B., R. LOHAUS, S. SRINIVASAN, K. K. DANG et C. L. BURCH. 2006, «Sexual reproduction selects for robustness and negative epistasis in artificial gene networks», *Nature*, vol. 440, n° 7080, p. 87–90.
- BARABASI, A.-L. et Z. N. OLTVAI. 2004, «Network biology : understanding the cell's functional organization», *Nature reviews genetics*, vol. 5, n° 2, p. 101–113.
- BERGMAN, A. et M. L. SIEGAL. 2003, «Evolutionary capacitance as a general feature of complex gene networks», *Nature*, vol. 424, n° 6948, p. 549–552.
- CILIBERTI, S., O. C. MARTIN et A. WAGNER. 2007a, «Innovation and robustness in complex regulatory gene networks», *Proceedings of the National Academy of Sciences*, vol. 104, n° 34, p. 13 591–13 596.
- CILIBERTI, S., O. C. MARTIN et A. WAGNER. 2007b, «Robustness can evolve gradually in complex regulatory gene networks with varying topology», *PLoS Comput Biol*, vol. 3, n° 2, p. e15.
- CLARK, A. 1991, «Mutation-selection balance and metabolic control theory.», *Genetics*, vol. 129, n° 3, p. 909–923.
- D'ÁVILA, M. F., R. N. GARCIA, E. L. LORETO et V. L. D. S. VALENTE. 2008, «Analysis of phenotypes altered by temperature stress and hipermutability in *Drosophila willistoni*», *Iheringia. Série Zoologia*, vol. 98, n° 3, p. 345–354.
- DEBAT, V. et P. DAVID. 2001, «Mapping phenotypes : canalization, plasticity and developmental stability», *Trends in Ecology & Evolution*, vol. 16, n° 10, p. 555–561.
- DRAGHI, J. et G. WAGNER. 2009, «The evolutionary dynamics of evolvability in a gene network model», *Journal of evolutionary biology*, vol. 22, n° 3, p. 599–611.

- ESPINOSA-SOTO, C., O. C. MARTIN et A. WAGNER. 2011a, «Phenotypic plasticity can facilitate adaptive evolution in gene regulatory circuits», *BMC Evolutionary Biology*, vol. 11, n° 1, p. 1.
- ESPINOSA-SOTO, C., O. C. MARTIN et A. WAGNER. 2011b, «Phenotypic robustness can increase phenotypic variability after nongenetic perturbations in gene regulatory circuits», *Journal of Evolutionary Biology*, vol. 24, n° 6, p. 1284–1297.
- ESPINOSA-SOTO, C. et A. WAGNER. 2010, «Specialization can drive the evolution of modularity», *PLoS Comput Biol*, vol. 6, n° 3, p. e1000719.
- FEATHERSTONE, D. E. et K. BROADIE. 2002, «Wrestling with pleiotropy : genomic and topological analysis of the yeast gene expression network», *Bioessays*, vol. 24, n° 3, p. 267–274.
- FIERST, J. 2011a, «A history of phenotypic plasticity accelerates adaptation to a new environment», *Journal of evolutionary biology*, vol. 24, n° 9, p. 1992–2001.
- FIERST, J. L. 2011b, «Sexual dimorphism increases evolvability in a genetic regulatory network», *Evolutionary Biology*, vol. 38, n° 1, p. 52–67.
- GIBSON, G. 2000, «Canalization in evolutionary genetics : a stabilizing theory?», *Bioessays*, vol. 22, n° 4, p. 372–380.
- HANSEN, T. F. 2013, «Why epistasis is important for selection and adaptation», *Evolution*, vol. 67, n° 12, p. 3501–3511.
- HANSEN, T. F. et G. P. WAGNER. 2001, «Modeling genetic architecture : a multilinear theory of gene interaction», *Theoretical population biology*, vol. 59, n° 1, p. 61–86.
- HERMISSON, J. et G. P. WAGNER. 2004, «The population genetic theory of hidden variation and genetic robustness», *Genetics*, vol. 168, n° 4, p. 2271–2284.
- HUERTA-SANCHEZ, E. et R. DURRETT. 2007, «Wagner’s canalization model», *Theoretical Population Biology*, vol. 71, n° 2, p. 121–130.
- KACSER, H. et J. BURNS. 1979, «Molecular democracy : who shares the controls?», *Biochemical Society Transactions*, vol. 7, n° 5, p. 1149–1160.
- KAWECKI, T. J. 2000, «The evolution of genetic canalization under fluctuating selection», *Evolution*, vol. 54, n° 1, p. 1–12.
- LE CUNFF, Y. et K. PAKDAMAN. 2012, «Phenotype–genotype relation in Wagner’s canalization model», *Journal of Theoretical Biology*, vol. 314, p. 69–83.

- LE ROUZIC, A., J. M. ÁLVAREZ-CASTRO et T. F. HANSEN. 2013, «The evolution of canalization and evolvability in stable and fluctuating environments», *Evolutionary Biology*, vol. 40, n° 3, p. 317–340.
- LE ROUZIC, A. et Ö. CARLBORG. 2008, «Evolutionary potential of hidden genetic variation», *Trends in ecology & evolution*, vol. 23, n° 1, p. 33–37.
- LECLERC, R. D. 2008, «Survival of the sparsest : robust gene networks are parsimonious», *Molecular systems biology*, vol. 4, n° 1, p. 213.
- LEVY, S. F. et M. L. SIEGAL. 2012, «The robustness continuum», dans *Evolutionary Systems Biology*, Springer, p. 431–452.
- LI, R., S.-W. TSAIH, K. SHOCKLEY, I. M. STYLIANOU, J. WERGEDAL, B. PAIGEN et G. A. CHURCHILL. 2006, «Structural model analysis of multiple quantitative traits», *PLoS Genet*, vol. 2, n° 7, p. e114.
- LOHAUS, R., C. L. BURCH et R. B. AZEVEDO. 2010, «Genetic architecture and the evolution of sex», *Journal of heredity*, vol. 101, n° suppl 1, p. S142–S157.
- MACCARTHY, T. et A. BERGMAN. 2007, «Coevolution of robustness, epistasis, and recombination favors asexual reproduction», *Proceedings of the National Academy of Sciences*, vol. 104, n° 31, p. 12 801–12 806.
- MARTIN, O. C. et A. WAGNER. 2009, «Effects of recombination on complex regulatory circuits», *Genetics*, vol. 183, n° 2, p. 673–684.
- MASEL, J. 2004, «Genetic assimilation can occur in the absence of selection for the assimilating phenotype, suggesting a role for the canalization heuristic», *Journal of evolutionary biology*, vol. 17, n° 5, p. 1106–1110.
- MASEL, J. et M. L. SIEGAL. 2009, «Robustness : mechanisms and consequences», *Trends in genetics*, vol. 25, n° 9, p. 395–403.
- MASEL, J. et M. V. TROTTER. 2010, «Robustness and evolvability», *Trends in Genetics*, vol. 26, n° 9, p. 406–414.
- PAABY, A. B. et M. V. ROCKMAN. 2013, «The many faces of pleiotropy», *Trends in Genetics*, vol. 29, n° 2, p. 66–73.
- PALMER, M. E. et M. W. FELDMAN. 2009, «Dynamics of hybrid incompatibility in gene networks in a constant environment», *Evolution*, vol. 63, n° 2, p. 418–431.
- PINHO, R., E. BORENSTEIN et M. W. FELDMAN. 2012, «Most networks in wagner’s model are cycling», *PloS one*, vol. 7, n° 4, p. e34 285.

- PINHO, R., V. GARCIA et M. W. FELDMAN. 2015, «Phenotype accessibility and noise in random threshold gene regulatory networks», *PloS one*, vol. 10, n° 4, p. e0119972.
- POLYNIKIS, A., S. HOGAN et M. DI BERNARDO. 2009, «Comparing different ode modelling approaches for gene regulatory networks», *Journal of theoretical biology*, vol. 261, n° 4, p. 511–530.
- RHONE, B., J.-T. BRANDENBURG et F. AUSTERLITZ. 2011, «Impact of selection on genes involved in regulatory network : a modelling study», *Journal of evolutionary biology*, vol. 24, n° 10, p. 2087–2098.
- RUTHERFORD, S. L. et S. LINDQUIST. 1998, «Hsp90 as a capacitor for morphological evolution», *Nature*, vol. 396, n° 6709, p. 336–342.
- SCHARLOO, W. 1991, «Canalization : genetic and developmental aspects», *Annual Review of Ecology and Systematics*, vol. 22, p. 65–93.
- SEVIM, V. et P. A. RIKVOLD. 2008, «Chaotic gene regulatory networks can be robust against mutations and noise», *Journal of theoretical biology*, vol. 253, n° 2, p. 323–332.
- SIEGAL, M. L. et A. BERGMAN. 2002, «Waddington’s canalization revisited : developmental stability and evolution», *Proceedings of the National Academy of Sciences*, vol. 99, n° 16, p. 10 528–10 532.
- SIEGAL, M. L. et J.-Y. LEU. 2014, «On the nature and evolutionary impact of phenotypic robustness mechanisms.», *Annual review of ecology, evolution, and systematics*, vol. 45, p. 496–517.
- SIEGAL, M. L., D. E. PROMISLOW et A. BERGMAN. 2007, «Functional and evolutionary inference in gene networks : does topology matter ?», *Genetica*, vol. 129, n° 1, p. 83–103.
- SNIEGOWSKI, P. D., P. J. GERRISH, T. JOHNSON, A. SHAVER et collab.. 2000, «The evolution of mutation rates : separating causes from consequences», *Bioessays*, vol. 22, n° 12, p. 1057–1066.
- STEARNS, F. W. 2010, «One hundred years of pleiotropy : a retrospective», *Genetics*, vol. 186, n° 3, p. 767–773.
- STEARNS, S. C., M. KAISER et T. J. KAWECKI. 1995, «The differential genetic and environmental canalization of fitness components in drosophila melanogaster», *Journal of Evolutionary Biology*, vol. 8, n° 5, p. 539–557.

- VIA, S., R. GOMULKIEWICZ, G. DE JONG, S. M. SCHEINER, C. D. SCHLICHTING et P. H. VAN TIENDEREN. 1995, «Adaptive phenotypic plasticity : consensus and controversy», *Trends in Ecology & Evolution*, vol. 10, n° 5, p. 212–217.
- DE VISSER, J., J. HERMISSON, G. P. WAGNER, L. A. MEYERS, H. BAGHERI-CHAICHIAN, J. L. BLANCHARD, L. CHAO, J. M. CHEVERUD, S. F. ELENA, W. FONTANA et collab.. 2003, «Perspective : evolution and detection of genetic robustness», *Evolution*, vol. 57, n° 9, p. 1959–1972.
- WADDINGTON, C. H. 1942, «Canalization of development and the inheritance of acquired characters», *Nature*, vol. 150, n° 3811, p. 563–565.
- WADDINGTON, C. H. 1953, «Genetic assimilation of an acquired character», *Evolution*, p. 118–126.
- WADDINGTON, C. H. 1957, *The strategy of the genes. A discussion of some aspects of theoretical biology*, Allen & Unwin, London.
- WAGNER, A. 1994, «Evolution of gene networks by gene duplications : a mathematical model and its implications on genome organization», *Proceedings of the National Academy of Sciences*, vol. 91, n° 10, p. 4387–4391.
- WAGNER, A. 1996, «Does evolutionary plasticity evolve?», *Evolution*, p. 1008–1023.
- WAGNER, A. 2008, «Robustness and evolvability : a paradox resolved», *Proceedings of the Royal Society of London B : Biological Sciences*, vol. 275, n° 1630, p. 91–100.
- WAGNER, A. 2012, «The role of robustness in phenotypic adaptation and innovation», *Proceedings of the Royal Society of London B : Biological Sciences*, vol. 279, n° 1732, p. 1249–1258.
- WAGNER, A. 2013, *Robustness and evolvability in living systems*, Princeton University Press.
- WAGNER, G. P., G. BOOTH et H. BAGHERI-CHAICHIAN. 1997, «A population genetic theory of canalization», *Evolution*, p. 329–347.
- WAGNER, G. P., J. P. KENNEY-HUNT, M. PAVLICEV, J. R. PECK, D. WAXMAN et J. M. CHEVERUD. 2008, «Pleiotropic scaling of gene effects and the ‘cost of complexity’», *Nature*, vol. 452, n° 7186, p. 470–472.
- WILDER, B. et K. STANLEY. 2015, «Reconciling explanations for the evolution of evolvability», *Adaptive Behavior*, vol. 23, n° 3, p. 171–179.

WOLF, J. B., E. D. BRODIE et M. J. WADE. 2000, *Epistasis and the evolutionary process*, Oxford University Press, USA.

Titre : Évolution de la canalisation génétique dans un modèle quantitatif de réseau de régulation.

Mots clés : architecture génétique ; génétique quantitative ; modélisation

Résumé : La canalisation génétique est définie comme la capacité d'un organisme à avoir un développement constant en dépit des mutations qui l'affectent. À l'heure actuelle, trois hypothèses majoritaires cherchent à expliquer l'apparition de ce processus : évolutive, congruente et intrinsèque. Pour tester ces hypothèses, j'ai choisi d'étudier les réseaux de régulation. Pour cela, j'ai réutilisé un modèle théorique pour simuler *in silico* l'évolution des architectures génétiques, et les analyser par les outils de la génétique quantitative.

J'ai d'abord étudié les comportements évolutifs de notre modèle et sa capacité de réponse à la sélection stabilisante. Outre l'analyse de l'impact des paramètres du modèle, j'ai mis en évidence l'absence d'équilibre mutation - sélection - dérive après des milliers de générations du fait de l'augmentation

progressive de la canalisation.

J'ai ensuite montré que les réseaux soumis à des mutations fréquentes et fortes, sélectionnés vers des optimums phénotypiques extrêmes, et dans lesquels certains gènes sont laissés libres d'évoluer sont plus aptes à faire évoluer de la canalisation génétique. Ces résultats nous ont amenés à proposer un double mécanisme impliqué dans l'évolution de la canalisation dans les réseaux de régulation : la réduction de la cible mutationnelle et la redondance de la régulation génique.

Je termine ce manuscrit en présentant quelques pistes d'études complémentaires, portant notamment sur l'étude de la canalisation contre les perturbations environnementales et l'utilisation de modèles alternatifs.

Title : Evolution of genetic architecture in quantitative model of gene regulatory network.

Keywords : genetic architecture ; quantitative genetics ; modelling

Abstract : Genetic canalization is defined as the capacity of an organism to undergo a normal development even when the genome is altered by mutations. Currently, three main hypotheses are prone to explain the apparition of such a process: evolutionary, congruent and intrinsic.

To test these hypotheses, I chose to study gene regulatory networks. To this end, I used a theoretical model, ran *in silico* simulations, and analyzed the genetic architecture by using quantitative genetics tools.

I first studied the evolutionary behavior of the model, and its capacity to respond to stabilizing selection. In addition to the sensitivity analysis to model parameters, I evidenced the absence of mutation-

selection-drift equilibrium after several thousand generations, which reveals the evolution of canalization.

I also showed that networks submitted to frequent and large mutations, and/or selected toward extreme phenotypic optima are more prone to evolve genetic canalization. This result leads us to propose a two-fold mechanism able to explain the evolution of canalization in gene regulatory networks: shrinkage of mutational targets and redundancy in genetic regulation.

At the end of this manuscript, I propose some possible future studies, such as the study of canalization towards environmental perturbations, and use of alternative models.

