



**HAL**  
open science

# Étude de la variabilité phénotypique, sous influence environnementale, chez *S. mansoni*, parasite de l'homme : implication de l'épigénome par une approche globale

J. M. J. Lepasant

## ► To cite this version:

J. M. J. Lepasant. Étude de la variabilité phénotypique, sous influence environnementale, chez *S. mansoni*, parasite de l'homme : implication de l'épigénome par une approche globale. Sciences de l'environnement. Université de Perpignan, 2012. Français. NNT: . tel-00941903

**HAL Id: tel-00941903**

**<https://theses.hal.science/tel-00941903>**

Submitted on 4 Feb 2014

**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.

UNIVERSITÉ PERPIGNAN VIA DOMITIA

Spécialité : Biologie

Ecole doctorale : Énergie et Environnement

**Étude de la variabilité phénotypique, sous  
influence environnementale, chez  
*Schistosoma mansoni*, parasite de l'homme :  
implication de l'épigénome par une approche  
globale**

Thèse de doctorat présentée par

Julie M.J. LEPESANT

En vue de l'obtention du grade de

DOCTEUR DE L'UNIVERSITÉ DE PERPIGNAN VIA DOMITIA

Sous la direction de

M. Christoph Grunau, Maître de conférence

Membres du Jury

Mme Cristina VIEIRA, Professeur (rapporteur)

M. Christoph GREVELDING, Professeur (rapporteur)

Mme Ana RIVERO, Chargée de recherche (examineur)

M. Guillaume MITTA, Professeur (examineur)

M. Christoph GRUNAU, Maître de conférence (directeur de thèse)

12 Novembre 2012



Petit à petit, l'oiseau fait son nid.





## Résumé

La bilharziose est une maladie parasitaire humaine qui se trouve au second rang mondial, après la malaria, en termes de morbidité et de mortalité induites par le parasite. Plus de 200 millions de personnes sont infestées dans 74 pays et plus de 200 000 personnes décèdent par an (Chitsulo *et al.* 2004; Gryseels *et al.* 2006). *Schistosoma mansoni*, ici étudié, est responsable de la bilharziose intestinale.

Ce parasite possède un cycle de vie complexe, passant obligatoirement à travers deux hôtes : un hôte intermédiaire mollusque d'eau douce et un hôte définitif humain ou murin. Lors du cycle de développement, les différents stades sont exposés à divers environnements. Par exemple, les stades larvaires peuvent être confrontés à plusieurs hôtes intermédiaires (souche de mollusque) ou définitifs (humain ou murin). Mes travaux de thèse concernent l'influence de l'environnement sur l'épigénome de *Schistosoma mansoni*. L'information épigénétique est une source supplémentaire pour la création de variants phénotypiques, nous proposons que certains de ces variants puissent être hérités et sélectionnés, et ainsi conférer un avantage au parasite.

Pour investir cette hypothèse nous avons décidé de modifier l'environnement du parasite de façon biotique et abiotique, lorsqu'il se développe au sein de l'hôte définitif ou lorsqu'il est au contact de l'hôte intermédiaire. En effet, *Schistosoma mansoni* est un endoparasite donc chaque hôte, dans lequel il va se maintenir, représente son environnement direct. Pour cela nous avons utilisé une drogue anti-helminthe et un hôte intermédiaire allopatrique. La réponse du parasite aux différents changements d'environnements a été analysée de la population (traits d'histoire de vie) à la molécule (génome, épigénome, transcriptome). Les nouvelles techniques de biologie moléculaire et de séquençage massif nous ont permis de réaliser des études globales.

Lors de ces travaux de thèse, les différentes évolutions expérimentales ont mené à deux énigmes : (1) « Adaptation rapide de *Schistosoma mansoni* à une drogue anti-helminthes » et (2) « Adaptation rapide de *Schistosoma mansoni* à l'Amérique du Sud ». Les résultats obtenus pour cette dernière ont amené une troisième énigme : (3) « Différenciation sexuelle en l'absence de gène W-spécifique dans un système ZW ».

## Abstract

The schistosomiasis is a parasitic human disease which ranks second only after the malaria, in term of morbidity and mortality. More than 200 million people are infested in around 74 countries and more than 200.000 death per year (Chitsulo *et al.* 2004; Gryseels *et al.* 2006). The study model, *Schistosoma mansoni*, is responsible for the intestinal schistosomiasis.

The life cycle of this parasite is complex with the passage through two obligatory host, a fresh water snail as intermediate host, and human or rodents as definitive hosts. During the life cycle, all the stages could be exposed to various environments. For example, the larval stage could be exposed to different intermediate or definitive hosts. The aim of my research is to study the influence of the environment on the epigenome of *Schistosoma mansoni*. The epigenetic information is a supplementary resource to generate phenotypic variants and we propose that some variants could confer advantage to the parasite, could be selected and heritable through generations.

To investigate this assumption, we decide to modify the environment, essentially the host environment, using biotic and abiotic modifications. Because *Schistosoma mansoni* is an endoparasite, each host in which he will develop, represents his direct environment. In this study we have modified the definitive host environment using antihelminth drugs, and the intermediate host environment using an allopatric strain. The adaptive response of the parasite to various environment is analysed from population level (life history traits) to the molecular level (genome, epigenome, and transcriptome). New techniques of molecular biology and massive sequencing allowed us to do whole genome analyses.

During this PhD, different experimental evolution lead to three different enigma: (1) « Rapid adaptation of *Schistosoma mansoni* to antihelminth drug », (2) « Rapid adaptation of *Schistosoma mansoni* to South America » and (3) « Sexual differentiation without W-specific genes, in a ZW system ».



*À ma sœur et mes parents ...*

## Remerciements

*La fin de la thèse approchant à grands pas, j'ai tendance à devenir nostalgique et j'avoue ne pas savoir par où débiter...*

*Je vais donc commencer par remercier les personnes sans qui cette expérience n'aurait jamais commencé : André, Guillaume, Céline et Christoph. Un grand merci à vous quatre de m'avoir permis de venir travailler dans le laboratoire. Merci pour votre aide, gentillesse et complicité. Merci André de m'avoir soutenue au concours, pour ces nombreux moments en salle du vivant, vos blagues et vos échantillons gentiment offerts. Merci Guillaume pour tout ce que tu m'as donné pendant ces trois superbes années. Merci Christoph d'avoir été un chef hors normes, pour tout ce temps que tu m'as offert, toute ta science qui m'a faite grandir, tous ces pots ou réunions qui ont eu lieu grâce à notre travail acharné, pour ce sujet de thèse extraordinaire... Merci Céline pour ton aide, les potins et les fous rires. Un grand merci à vous quatre, je pense que j'oublie certainement des choses mais le cœur y est.*

*Je tiens aussi à remercier les membres de mon jury, qui ont accepté d'évaluer mon travail durant ces trois années à Perpignan.*

*Je remercie toutes les personnes qui ont interagies de près ou de loin : Anne M., Anne R., Rémi, Juliette, Nolwenn, Bernard, Nath, Sébastien, Gabby, Hélène et Eve. Merci Irma pour tes arrangements « mission », je n'oublierai jamais le fait que tu m'aies mariée à un certain TOTEM... Merci Diane pour ces nombreuses pauses.*

*Merci au trio infernal : Jeff, Richard et David D. pour tous les instants passés dans la salle de biomol, vos innombrables blagues...*

*Un très grand merci à Benjamin pour tous les bons moments et ces fous rires en Roumanie ! :) et merci à Jérôme pour ces heures de manips, ton aide pour les articles (je t'ai surexploité...).*

*Je tiens ensuite à remercier tous mes compagnons de « galère », les CDDs ! Un merci tout particulier à Cécile, Deborah, Valentina, Clémence et Guilhem qui ont été et sont toujours mes épaules, ami(e)s, confident(e)s, précieux tant pour les manips que pour ma santé mentale. J'aurais tellement de choses à vous dire ... Merci pour les discussions interminables, les fous rires, les randos, la piscine, la plage, les vacances, le sport, les restaurants, les apéros qui s'éternisent, votre gentillesse, votre humour et votre amitié. Merci !*

*Je remercie bien sûr Julien, Pauline, Rodrigue, Nico, Sophie et Yves qui m'ont accueillie à mon arrivée et m'ont permis de m'intégrer rapidement. Merci à Marion et Julie, les*

*« croquettes » pour leur joie de vivre et tous ces moments que nous avons partagés! Merci à Sara d'avoir le courage de continuer notre travail.*

*Un merci tout spécial à Jérémie qui m'a permis d'avancer, de décompresser quand il fallait, ces fous rires, ces « parties de tennis de bureau », ces pauses (parfois sans fin), les potins, les apéros, ta gentillesse, toute ton expérience, ton soutien et ton amitié...*

*Je ne pourrais bien sûr pas oublier David, mon super-stagiaire, qui m'a vraiment beaucoup aidée. Merci pour ta gentillesse, tes bons gâteaux et surtout ton travail!*

*Je ne pouvais pas finir ces remerciements sans remercier les « extérieurs » qui m'ont permis d'avancer, prendre du temps pour faire autre chose que travailler, et parfois même m'aider dans mon travail. Marion (Sticot) merci de ton aide pour ces calculs de dilutions infernaux ?! et surtout pour ton soutien. J'aurais tant de choses à te dire à propos de toutes ces choses que nous avons faites ensemble pendant ces trois années dont les barbecues, la plage (qui finit parfois aux urgences...), la semaine d'intégration, les apéros, les concerts, les magasins, les restos, le co-voiturage et autres petites choses qui m'ont permis de ne pas devenir folle (pas plus que d'habitude en tout cas).*

*Un grand merci aux Julies et Soso pour votre amitié.*

*Merci Mel pour le dessin de ma couverture et d'être toujours là. Merci à Christelle, tu es mon roc !*

*Merci aussi à Betty, Elsa, Leïla, Olivier et Gaël (Mac Gyver !) pour votre gentillesse et votre joie de vivre.*

*Ces remerciements sont interminables et je vais donc les finaliser. Je tiens à remercier Viviane Boulo qui m'a permis de découvrir la biomol et la PCRq, merci Viviane de m'avoir donné l'opportunité de travailler ensemble et de faire mes premiers pas dans la recherche. Merci pour ton aide, ton soutien, ta bonne humeur...*

*Merci à toute l'équipe de Banyuls sur Mer qui m'a permis de faire ce stage de M2 qui m'a lancée sur la voie de la thèse. Merci à Chub, Maria et Sandrine pour votre agréable compagnie.*

*Merci à Gilles Bœuf pour avoir été là pendant mon Master à Paris, pour m'avoir soutenue tout au long de ces années, pour m'avoir aidée dans le choix de ma thèse, pour toutes ces discussions autour d'un café, pour ta vision des choses et pour ta gentillesse.*

*Je finirai (enfin !) en remerciant du fond du cœur toute ma famille : mes parents qui m'ont toujours soutenue et ont été mes psychologues ; ma sœur et son mari qui m'ont permis de me*

*rendre compte que j'aimais ce que je faisais (même dans les moments difficiles) et de décompresser lors de vacances chez eux; mes grands-mères, arrière grand mère et tantes qui m'ont soutenue. Je tiens à préciser que les ¾ de ma famille sont persuadés que je travaille encore sur les huitres ou le poisson (famille de goinfres !), mais ils ont toujours été là ! Un grand merci à ma super maman qui est maintenant une spécialiste de l'épigénétique de schisto suite à de nombreuses relectures du manuscrit.*

*Merci à Camille qui m'a supportée (...), aidée, soutenue... On a dit qu'on ne s'étalerait pas, tu sais très bien tout ce que je pense de toi donc je vais simplement te dire « Merci pour tout ».*

*Pour terminer, une petite pensée à tous les petits et gros animaux qui m'ont permis de travailler et de faire une grande découverte hautement scientifique : l'addiction des hamsters aux BN au chocolat.*

*MERCI à vous tous de m'avoir « adoptée », soutenue, aidée à avancer, tous ces moments (les bons comme les mauvais), les nombreux repas (noëls, anniversaires, papiers acceptés, galettes des rois, barbecues, déjeuners, restaurants...) et merci surtout pour toute votre aide et joie de vivre.*

*Je vais maintenant m'arrêter là et vous souhaiter à tous le meilleur pour la suite!*

# TABLE DES MATIÈRES

## PRÉAMBULE

CHAPITRE 1 : Introduction générale.....	p 24
1 Contexte scientifique.....	p 24
1.1 Relation entre l'adaptation et l'épigénétique.....	p 24
1.1.1 Historique.....	p 24
1.1.2 Adaptation vs Acclimatation.....	p 27
1.1.2.1 Le corail comme exemple d'acclimatation.....	p 28
1.1.2.2 Les pinsons des Galápagos comme exemple d'adaptation.....	p 29
1.1.2.3 Les tardigrades comme exemple d'adaptation.....	p 30
1.1.3 La relation entre génotype, épigénotype, environnement et phénotype.....	p 31
1.1.4 Les supports de l'information épigénétique.....	p 34
1.1.5 L'information épigénétique chez <i>S. mansoni</i> : État de l'art.....	p 36
1.1.6 Vers une caractérisation approfondie de l'épigénotype de <i>S. mansoni</i> .....	p 38
1.2 Modèle d'étude.....	p 40
1.2.1 Zones de répartition.....	p 40
1.2.2 Bilharziose.....	p 41
1.2.3 Cycle de vie et reproduction.....	p 41
1.2.4 Environnement-hôte.....	p 44
1.2.5 <i>S. mansoni</i> comme modèle en adaptation évolutive.....	p 47
1.2.6 Traitements.....	p 48
1.2.6.1 Traitements à succès.....	p 48
1.2.6.2 Traitements en échec.....	p 49
1.2.7 Phénomène de compatibilité.....	p 52
1.3 Évolutions expérimentales sous pression de sélection : drogues anti-helminthes et hôte intermédiaire allopatrique.....	p 55
2 Objectifs de la recherche doctorale.....	p 57

## CHAPITRE 2 : Matériel et Méthodes

1 Matériel biologique.....	p 60
1.1 Énigme 1 « Adaptation rapide de <i>S. mansoni</i> à une drogue anti-helminthe ».....	p 60
1.2 Énigme 2 « Adaptation rapide de <i>S. mansoni</i> à l'Amérique du Sud ».....	p 62
1.3 Énigme 3 « Différenciation sexuelle en l'absence de gène W-spécifique dans un système ZW ».....	p 64
2 Analyses statistiques et de génétique des populations.....	p 64
3 Traits d'histoire de vie.....	p 64
4 Techniques de biologie moléculaire et cellulaire.....	p 65
4.1 Sexage et génotypage des stades larvaires.....	p 65
4.2 ImmunoPrécipitation de la Chromatine (ChIP).....	p 66
4.3 Extractions des ARNs totaux.....	p 69
4.3.1 Pour les analyses de transcriptomique par RNA-seq.....	p 69
4.3.2 Pour les quantifications relatives ciblées par PCRq.....	p 70
4.4 FISH ( <i>Fluorescent In Situ Hybridization</i> ).....	p 70
5 Techniques d'analyses bioinformatiques.....	p 70
5.1 Epigénome global : ChIP-seq.....	p 70
5.2 Transcriptome global : RNA-seq.....	p 75
5.3 Génome global : gDNA-seq.....	p 78

## CHAPITRE 3 : Résultats

1 Énigme 1 « Adaptation rapide de <i>S. mansoni</i> à une drogue anti-helminthe ».....	p 80
2 Énigme 2 « Adaptation rapide de <i>S. mansoni</i> à l'Amérique du Sud ».....	p 88
2.1 Étude des candidats génétiques et épigénétiques différents entre deux souches de parasites, issues d'isolats géographiques distincts.....	p 88
2.2 Analyse des séquences promotrices des gènes impliqués dans le polymorphisme de compatibilité, les <i>SmPoMucs</i> .....	p 93
2.3 Évolution expérimentale de l'adaptation d'une souche de parasite <i>S. mansoni</i> à différentes souches d'hôte intermédiaire.....	p 96
2.3.1 Traits d'histoire de vie : mesures phénotypiques.....	p 96

2.3.2 Analyse transcriptomique (RNA-seq) et épigénétique (ChIP-seq).....	p 102
3 Énigme 3 « Différenciation sexuelle en l'absence de gène W-spécifique dans un système ZW ».....	p 115
3.1 Analyse des séquences répétitives et <i>de novo</i> assemblage de ces séquences.....	p 116
3.2 Analyse des changements chromatinien autour des séquences répétitives satellites suggérant un possible mécanisme pour l'émergence des chromosomes sexuels.....	p 122
 CHAPITRE 4 : Discussion.....	 p 129
 CHAPITRE 5 : Conclusions et perspectives.....	 p 159
 CHAPITRE 6: Bibliographie.....	 p 167
 CHAPITRE 7 : Annexes.....	 p 178
1 Publication: <b>Lepesant J.M.J.</b> , Boissier J., Climent D., Cosseau C. & Grunau C. (in prep.). Experimental observations of life traits during the life cycle of <i>Schistosoma mansoni</i> infecting sympatric or allopatric intermediate host.....	p 179
2 Publication: <b>Lepesant J.M.J.</b> , Cosseau C., Boissier J., Freitag M., Portela J., Climent D., Perrin C., Zerlotini A. & Grunau C. (2012a). Chromatin structure changes around satellite repeats on the <i>Schistosoma mansoni</i> female sex chromosome suggest a possible mechanism for sex chromosome emergence. <i>Genome Biology</i> , 13, R14. (IF 9,036).....	p 199
3 Publication: <b>Lepesant J.M.J.</b> , Roquis D., Emans R., Cosseau C., Arancibia N., Mitta G. & Grunau C. (2012b). Combination of de-novo assembly of massive sequencing reads with classical repeat prediction improves identification of repetitive sequences in <i>Schistosoma mansoni</i> . <i>Experimental Parasitology</i> , 130, 470-474. (IF 2,122).....	p 215
4 Publication : <b>Lepesant J.M.J.</b> , Grunau C. & Cosseau C. (2011). Towards an understanding of the epigenetics of schistosomes: a comparative epigenomic study. <i>Memorias do Instituto Oswaldo Cruz</i> , 106, 823-830. (IF 2,058).....	p 221
5 Publication : Perrin C., <b>Lepesant J.M.J.</b> , Roger E., Duval D., Thuillier V., Alliene J.F., Grunau C., Mitta G. & Cosseau C. (under revision). Promoter regions of duplicated	

*Schistosoma mansoni* gene crucial for infections success undergo rapid epigenetic modifications in varying environments while remaining genetically stable.....p 230

6 Curriculum vitae.....p 278

## Préambule

Cette thèse s'intègre dans le cadre de l'obtention du titre de Docteur en Sciences au sein de l'Ecole Doctorale Energie et Environnement de l'Université de Perpignan *via* Domitia. Ce travail a été réalisé au laboratoire Écologie et Évolution des Interactions au sein de l'équipe Épigenétique environnementale et évolutive des interactions.

Cette thèse porte sur l'étude de l'influence environnementale sur l'épigénome d'un parasite humain, *Schistosoma mansoni*. Elle est divisée dans ce manuscrit en cinq parties. La première, consiste en une introduction générale sur le modèle d'étude, sur les connaissances acquises et sur le sujet de thèse. La deuxième partie, présente les différentes techniques utilisées lors de nos expérimentations. La troisième partie, fait état des résultats obtenus pour chaque question scientifique posée. La quatrième partie, propose une discussion générale des résultats obtenus. Enfin la dernière partie, regroupe toutes les références bibliographiques utilisées dans ce manuscrit ainsi que dans les articles publiés lors de ces travaux.

Les résultats issus de cette recherche ont été valorisés par trois articles (Lepessant *et al.* 2011; Lepessant *et al.* 2012a; Lepessant *et al.* 2012b). Deux articles supplémentaires sont en cours de publication (Lepessant *et al.* in prep.; Perrin *et al.* en révision). Tous ces articles, publiés et en soumission, sont présentés tout au long de ce manuscrit et sont disponibles en annexe (1 à 5).

Je tiens à remercier tout particulièrement Mme Cristina Vieira-Heddi et M. Christoph Grevelding d'avoir accepté d'évaluer ce travail, ainsi que Mme Ana Rivero, M. Guillaume Mitta et M. Christoph Grunau de faire partie de mon jury.





## Liste des figures et tableaux

### Figures

Figure 1: Évolution adaptative lamarckienne ou néo-darwinienne.....	p 25
Figure 2: Illustration de la loi de Shelford.....	p 26
Figure 3: Modèle hypothétique de la variabilité phénotypique.....	p 27
Figure 4: Observations microscopiques de coupes transversales de corail.....	p 29
Figure 5: Différents phénotypes de bec chez les géospizes des Galápagos.....	p 30
Figure 6: Observation au microscope électronique à transmission d'un tardigrade.....	p 30
Figure 7: Hérité épigénétique des soins maternels aux jeunes chez les rongeurs.....	p 33
Figure 8: Schéma de la structure de la chromatine.....	p 34
Figure 9: Schéma de la structure de l'euchromatine et de l'hétérochromatine.....	p 35
Figure 10: Effet inhibiteur de la trichostatine A sur les histones déacétylases.....	p 38
Figure 11: Modèle théorique de réponse adaptative selon un changement environnemental.....	p 39
Figure 12: Répartition de la schistosomiase en 2011.....	p 40
Figure 13: Cycle de vie et dimorphisme sexuel de <i>S. mansoni</i> .....	p 42
Figure 14: Photos en microscopie des différents stades de développement de <i>S. mansoni</i> . .....	p 43
Figure 15: Différents environnements auxquels <i>S. mansoni</i> est confronté.....	p 45
Figure 16: Traits d'histoire de vie mesurables tout au long du cycle de développement de <i>S. mansoni</i> .....	p 47
Figure 17: Proposition d'un mécanisme d'action pour l'hycanthone et le lucanthone.....	p 51
Figure 18: Schéma du polymorphisme de compatibilité <i>S. mansoni</i> / <i>B. glabrata</i> .....	p 52
Figure 19: Phénomène de compatibilité pour une souche de parasite <i>SmBre</i> ( <i>S. mansoni</i> Brésil).....	p 53
Figure 20: Synthèse des analyses réalisées au cours de cette thèse.....	p 58
Figure 21: Protocole expérimental utilisé pour les traitements à l'hycanthone.....	p 61
Figure 22: Protocole expérimental de changements d'hôte intermédiaire.....	p 63
Figure 23: Électrophérogrammes pour un génotype mâle et un génotype femelle.....	p 66

Figure 24: Principe d'immunoprécipitation de la chromatine.....	p 67
Figure 25: Quantité d'ADN immunoprécipité en fonction d'une quantité croissante d'anticorps anti H3K4me3.....	p 68
Figure 26: Exemple de boîte à moustaches des scores de qualité pour l'anticorps H3K27me3.....	p 72
Figure 27: Exemple de graphique de distribution des nucléotides par position des séquences pour l'anticorps H3K27me3.....	p 73
Figure 28: Représentation schématique du protocole expérimental utilisé pour les analyses RNA-seq.....	p 78
Figure 29: Traits d'histoire de vie : réponse immunopathologique de l'hôte définitif.....	p 81
Figure 30: Nombre de vers adultes moyens développés suite au traitement hycanthone...	p 82
Figure 31: Synthèse des résultats de l'expérience de traitement à l'hycanthone.....	p 83
Figure 32: Nombre de vers adultes moyens développés suite au traitement lucanthone...	p 84
Figure 33: Résultats de l'expérience de traitement au lucanthone.....	p 85
Figure 34: Synthèse des résultats de l'expérience de traitement au lucanthone : génération 1.....	p 86
Figure 35: Synthèse des résultats de l'expérience de traitement au lucanthone : génération 2.....	p 87
Figure 36: Profil d'enrichissement pour les souches Guadeloupe et Brésil.....	p 92
Figure 37: Enrichissement, obtenu en ChIP-PCRq, pour les marques H3K9ac et H3K9me3 des régions promotrices des gènes <i>SmPoMucs</i> (1, 3.1(r1-r2)) dans la souche Guadeloupe et la souche Brésil.....	p 96
Figure 38: Protocole expérimental de changement d'hôte intermédiaire (rappel).....	p 97
Figure 39: Profil d'émission cercarienne selon le développement du parasite au sein d'un hôte intermédiaire sympatrique ou allopatrique.....	p 99
Figure 40: Prévalences de mollusques infestés dans la condition sympatrique et la condition allopatrique.....	p 99
Figure 41: Sexes-ratios des différents stades de développement dans la condition sympatrique et la condition allopatrique.....	p 100
Figure 42: Diagramme de Venn des transcrits différentiellement exprimés entre la condition sympatrique et allopatrique.....	p 103
Figure 43: Répartition des transcrits différentiellement exprimés dans la condition allopatrique par rapport à la condition sympatrique.....	p 104
Figure 44: Diagramme de répartitions des profils d'expression observés dans la condition allopatrique pour 223 transcrits différentiellement exprimés dans les trois générations....	p 106

Figure 45: Exemple de transcrit présentant un différentiel d'expression associé à des modifications de la structure chromatinienne dans la condition allopatrique.....	p 107
Figure 46: Domaines conservés d'un gène différentiellement exprimé dans la condition allopatrique, associé à des modifications de la structure chromatinienne.....	p 108
Figure 47: Grandes fonctions moléculaires et biologiques dans lesquelles sont impliqués les transcrits différentiellement exprimés dans la condition allopatrique.....	p 109
Figure 48: Domaines conservés d'un gène différentiellement exprimé dans la condition allopatrique.....	p 110
Figure 49: Profils types des marques d'histones (H3K9ac, H3K4me3, H3K9me et H3K27me3), dans la région promotrice, en fonction du niveau d'expression à l'échelle du génome.....	p 112
Figure 50: Schéma de la répartition des différentes marques d'histones (H3K9ac, H3K4me3, H3K9me3 et H3K27me3) autour du <i>Transcription Start Site (TSS)</i> .....	p 114
Figure 51: Expression relative de 7 séquences répétées aux stades miracidia et adultes...	p 119
Figure 52: Représentation schématique du pipeline utilisé pour l'identification et l'annotation des séquences répétées dans le génome de <i>S. mansoni</i> .....	p 121
Figure 53: Localisation, par FISH ( <i>Fluorescent in situ Hybridization</i> ), sur le chromosome W des 8 séquences répétées les plus abondantes.....	p 124
Figure 54: Représentation des taux de transcription des séquences répétées W4, W5 et SMA $\alpha$ fem-1 aux différents stades de développement de <i>S. mansoni</i> .....	p 125
Figure 55: Structure chromatinienne, évaluée par ChIP-seq et ChIP-PCRq, autour des séquences répétées femelles spécifiques W3, W4 et W5 aux différents stades de développement.....	p 126
Figure 56: Hypothèses des expériences de mutation directe ou aléatoire de Luria & Delbrück.....	p 132
Figure 57: Protocole expérimental de Lederberg & Lederberg.....	p 133
Figure 58: Design expérimental de Luria & Delbrück adapté à <i>S. mansoni</i> .....	p 135
Figure 59: Schéma de synthèse des résultats obtenus sur le trait d'histoire de vie sexe-ratio dans les conditions sympatriques et allopatriques.....	p 148
Figure 60: Modèle hypothétique du rôle des séquences répétées femelles spécifiques dans l'hétérochromatisation du chromosome W.....	p 155

## **Tableaux**

Tableau 1: Modifications de la chromatine sur les histones.....	p 36
Tableau 2: Historique des composés présentant des propriétés antischistosomiasés.....	p 49
Tableau 3: Anticorps testés pour les analyses d'immunoprécipitation de la chromatine...	p 69
Tableau 4: Nombre de séquences utilisées à chaque étape de l'analyse ChIP-seq.....	p 74
Tableau 5: Nombre de séquences utilisées à chaque étape de l'analyse RNA-seq.....	p 77
Tableau 6: Régions contenant des différences dans le nombre de séquences alignées entre la souche Brésil et la souche Guadeloupe.....	p 90
Tableau 7: Comparaison entre les résultats de ChIP-seq et les résultats de ChIP-PCRq entre la souche Brésil et la souche Guadeloupe pour 11 <i>loci</i> .....	p 91
Tableau 8: Synthèse des différents profils d'expression observés dans la condition allopatrique par rapport à la condition sympatrique.....	p 105
Tableau 9: Différentes classes définies en fonction du nombre de FPKMs.....	p 111
Tableau 10: Résumé des résultats obtenus par Blast et Censor pour catégoriser les séquences répétitives de <i>S. mansoni</i> .....	p 118
Tableau 11: Caractéristiques des 14 nouveaux microsatellites de <i>S. mansoni</i> .....	p 120

# **CHAPITRE 1**

## **Introduction**

## 1. Contexte scientifique

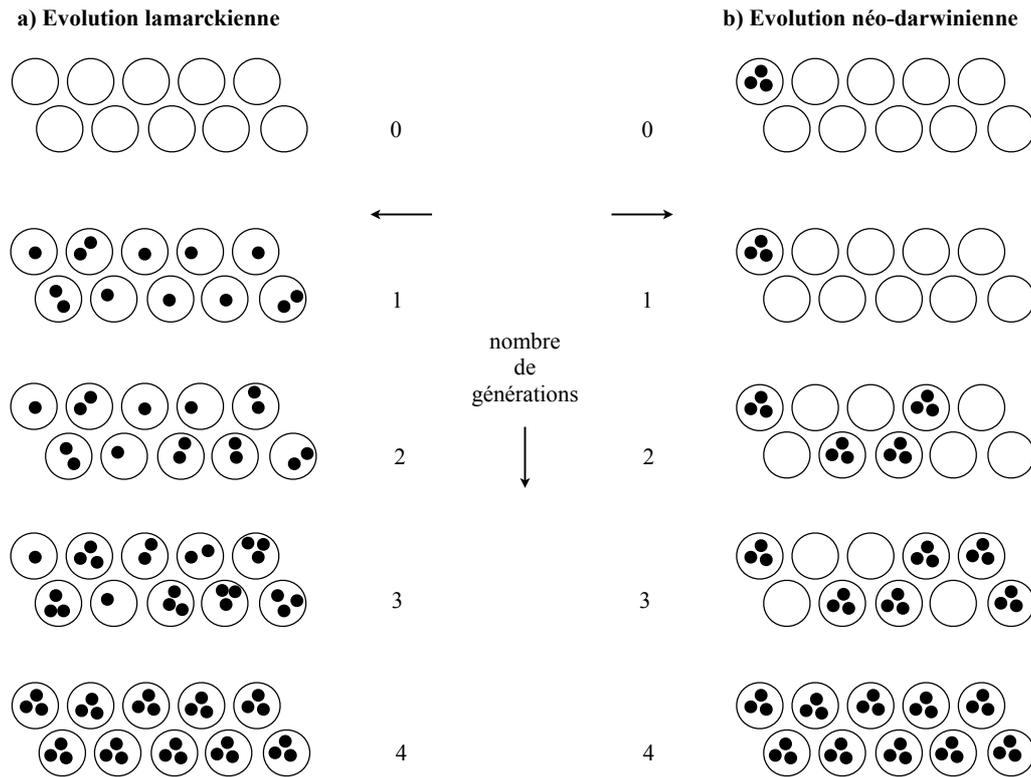
### 1.1. Relation entre l'adaptation et l'épigénétique

#### 1.1.1. Historique

Au 19<sup>e</sup> siècle, deux grandes théories de l'évolution sont apparues : l'évolution selon Lamarck (1744-1829) et l'évolution selon Darwin (1809-1882).

Les mécanismes de l'évolution lamarckienne peuvent être définis en trois points : (i) le changement d'un caractère peut être influencé par l'environnement, (ii) ce changement peut être transmis aux générations suivantes même si l'environnement inducteur n'est plus présent, (iii) ces phénotypes induits sont adaptatifs par rapport à cet environnement inducteur (revue par (Pal 1998)). Ainsi, Lamarck suggère que l'apparition de ces phénotypes, suite à un changement environnemental, est effective chez tous les individus d'une même population (Figure 1).

La théorie néo-darwinienne, quant à elle, énonce: (i) l'évolution est le fruit d'une modification progressive et continue des êtres vivants au cours des générations, (ii) la reproduction implique une hérédité : le matériel héréditaire subit des modifications par mutations aboutissant à une grande diversité, (iii) le mécanisme central est la sélection naturelle qui opère au niveau des populations en sélectionnant les individus les mieux adaptés à leur environnement (Darwin 1859). Cette théorie est expliquée dans l'introduction du livre de Charles Darwin *l'Origine des espèces* : « *Comme il naît beaucoup plus d'individus de chaque espèce qu'il n'en peut survivre, et que, par conséquent, il se produit souvent une lutte pour la vie, il s'ensuit que tout être, s'il varie, même légèrement, d'une manière qui lui est profitable, dans les conditions complexes et quelquefois variables de la vie, aura une meilleure chance pour survivre et ainsi se retrouvera choisi d'une façon naturelle. En raison du principe dominant de l'hérédité, toute variété ainsi choisie aura tendance à se multiplier sous sa forme nouvelle et modifiée* » (Darwin 1859). Ainsi, un individu donné présente un phénotype adaptatif afin de lui permettre d'évoluer dans un nouvel environnement et de faire persister ce trait favorable à travers les générations (Jablonka & Lamb 1999) (Figure 1).



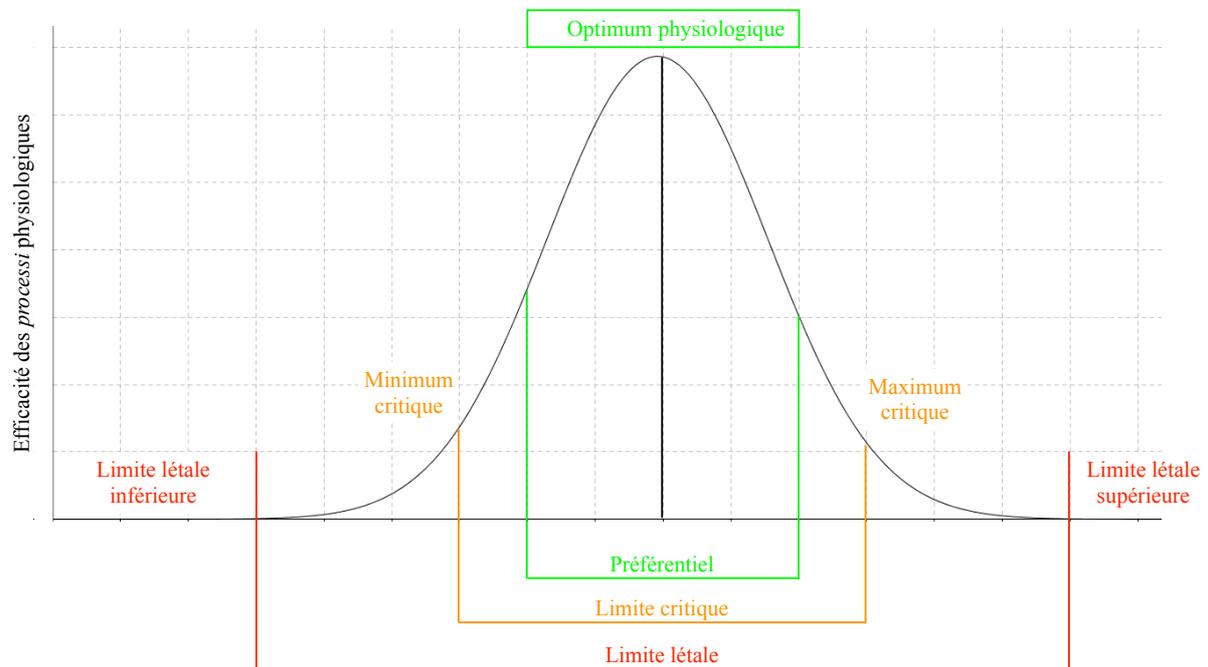
**Figure 1: Évolution adaptative lamarckienne ou néo-darwinienne.** Les individus présentant une adaptation sont représentés par des cercles à pois. La quantité de pois montre l'intensité de l'adaptation. La flèche indique le moment où le changement environnemental apparaît. Ce changement environnemental apparaît et persiste sur plusieurs générations. Dans l'évolution lamarckienne (a), tous les individus acquièrent l'adaptation et cette adaptation persiste graduellement durant plusieurs générations en réponse à ce changement environnemental. Dans l'évolution néo-darwinienne (b), un individu possède déjà la capacité à répondre à un changement environnemental, avant que celui-ci n'intervienne. Cette adaptation est transmise de façon graduelle aux générations suivantes et permet d'augmenter le nombre d'individus présentant cette capacité. Adapté de (Jablonka & Lamb 1999).

Dans ces deux grandes théories du 19<sup>e</sup> siècle, l'adaptation est basée sur l'apparition de variants phénotypiques. Ces variants sont réversibles et ne sont pas aussi stables que des mutations génétiques (Jablonka *et al.* 1995). L'origine de ces variants peut résulter d'un génotype hérité et/ou d'une multitude d'influences environnementales (Campbell & Reece 2007). Par exemple, une lignée pure d'organismes grandissant dans un environnement défavorable, va présenter des changements héréditaires de certains caractères qui persistent même si ces organismes se retrouvent dans leur environnement stable (Jablonka *et al.* 1995).

La variation intraspécifique (« individual variability » (Darwin 1859)) est omniprésente dans le monde du vivant. Cette variation est généralement référencée comme la variabilité phénotypique. Mayr introduit en 1963, le terme *polyphénisme*, plus tard qualifié d'épigénétique, afin de distinguer la variation phénotypique environnementale (l'apparition de différents phénotypes, dans une population, qui ne sont pas le résultat de changements

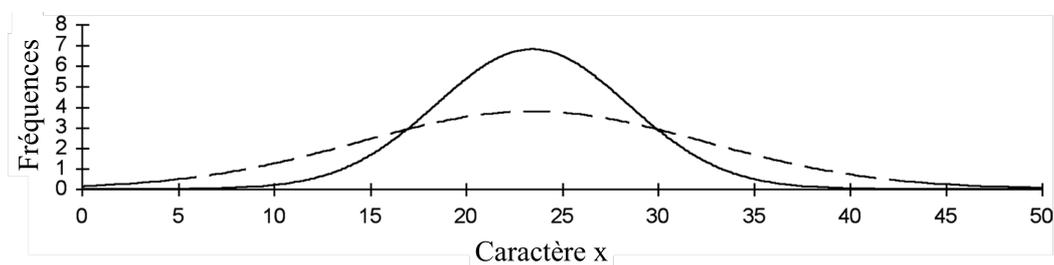
génétiques) de la variation phénotypique génétique (Mayr 1963). Les organismes possèdent une variabilité génétique qui ne contribue pas à la variation phénotypique dans des conditions habituelles, mais disposent d'un potentiel leur permettant de modifier les phénotypes dans un contexte environnemental. Ce phénomène est appelé variation cryptique.

Au vu de nos travaux, nous avons considéré que les organismes ont la capacité d'exprimer des phénotypes cryptiques, en conditions environnementales défavorables. Ces phénotypes sont codés par la séquence d'ADN mais non exprimés en conditions habituelles. Ils apparaissent lors d'un changement environnemental sans pour autant perdre l'expression de leurs phénotypes en conditions stables (Mc Guigan & Sgro 2009; Fusco & Minelli 2010). En effet, pour tout facteur environnemental il existe un gradient dans lequel tout processus physiologique ou écologique peut s'effectuer (Shelford 1913). La vie d'un organisme est possible seulement dans cet intervalle qui est limité par des bornes supérieures et inférieures au-delà desquelles la mortalité survient. L'étendue entre ces deux bornes létales varie selon les espèces, plus cet intervalle préférentiel est grand, plus l'espèce a la capacité de s'acclimater ou de s'adapter à des fluctuations environnementales grandes (Figure 2).



**Figure 2: Illustration de la loi de Shelford.** Il existe une relation générale entre l'efficacité des processus physiologiques d'un individu ou d'une population par rapport au gradient d'un facteur environnemental. Les processus physiologiques ou écologiques peuvent s'effectuer dans un intervalle limité par des bornes supérieures et inférieures. Au delà de ces bornes, la mortalité survient (Shelford 1913).

Tous les êtres vivants ont la capacité d'exprimer des phénotypes différents afin de maintenir une valeur sélective (*fitness*) optimale. En 1999, Pal & Miklos proposent un modèle théorique suggérant que la variabilité phénotypique augmente autour de l'optimum physiologique, sous influence environnementale (Pal & Miklos 1999). Dans ce modèle, les phénotypes se répartissent autour d'un optimum physiologique en conditions normales, mais sont limités par des bornes létales supérieures et inférieures. Sous conditions environnementales stressantes, la variabilité phénotypique évolue autour de l'optimum physiologique, avec une fréquence moindre, augmentant l'intervalle entre les bornes létales inférieures et supérieures (Figure 3).



**Figure 3: Modèle hypothétique de la variabilité phénotypique.** En conditions normales (ligne pleine) il y a une répartition des phénotypes autour de l'optimum physiologique alors qu'en conditions fluctuantes (ligne pointillée) l'intervalle entre les bornes supérieures et inférieures est agrandie, augmentant la variabilité phénotypique. Selon le modèle de Pal & Miklos (Pal & Miklos 1999).

Cette variabilité phénotypique transmise aux générations suivantes a été nommée « *epigenetic inheritance systems* » (Maynard Smith 1990; Jablonka *et al.* 1995). Elle joue un rôle essentiel dans l'évolution puisqu'elle permet aux organismes de s'adapter à un environnement sans changement génétique.

### 1.1.2. Adaptation vs Acclimatation

Lucien Cuénot distingue deux types d'adaptations successives: (i) l'acclimatation ou adaptation physiologique ponctuelle et réversible de l'individu à un nouveau milieu; (ii) l'accommodation ou adaptation irréversible d'un groupe établi de manière durable dans un milieu (Cuénot 1925).

L'acclimatation ou l'adaptation physiologique constitue la première expression de la plasticité phénotypique des espèces. Elle est réversible et résulte de l'existence de mécanismes de régulation physiologique, provoquant des modifications métaboliques pour que les êtres vivants maintiennent une valeur sélective maximale face aux fluctuations environnementales.

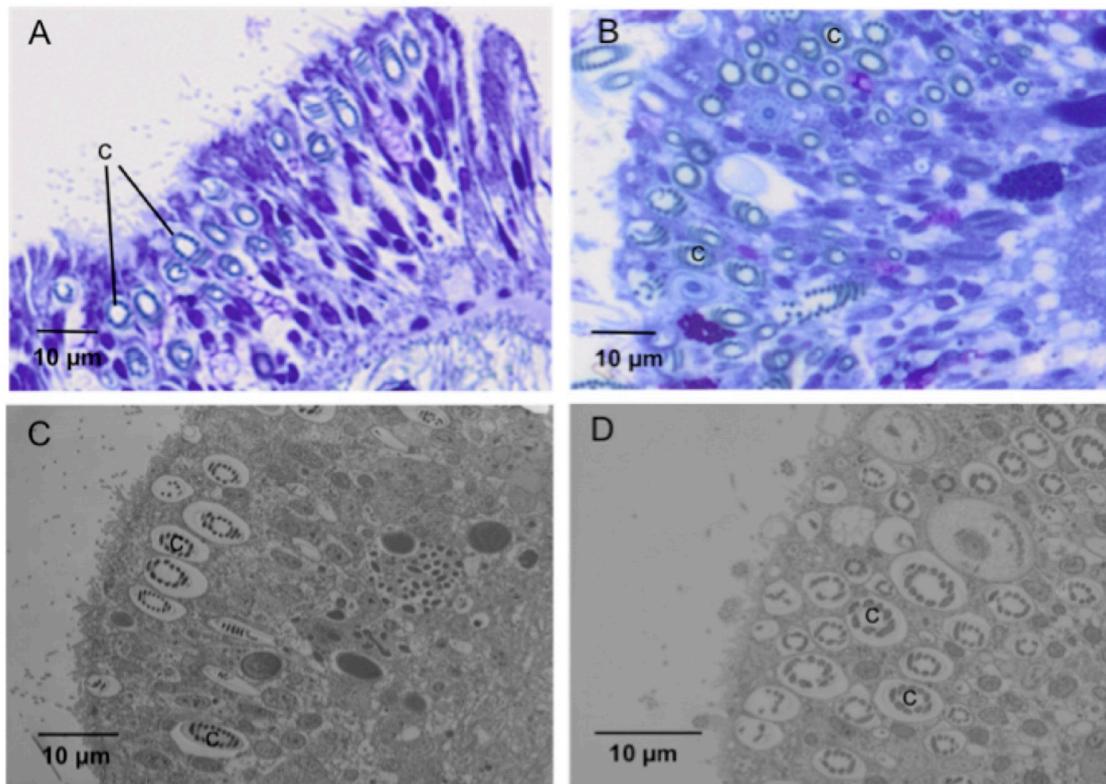
En biologie de l'évolution, l'adaptation ou l'accommodation est constatée dès lors qu'il existe une modification d'un caractère anatomique, d'un processus physiologique et/ou d'un trait comportemental dans une population d'individus, sous l'effet de la sélection naturelle. Charles Darwin a écrit : « *C'est à cette conservation des variations favorables [...] et à la destruction de celles qui sont nuisibles, que j'ai appliqué le nom de « sélection naturelle » ou de « survivance du plus apte »* » (Darwin 1859). Le nouvel état de ce caractère vise à améliorer la survie et le succès reproductif des individus qui en sont porteurs. Selon Ernst Mayr, la théorie darwinienne pourrait se découper en cinq points: (i) quelle que soit l'espèce, l'effectif peut croître de manière exponentielle si tous les descendants engendrés sont viables ; (ii) en dehors des fluctuations saisonnières, la taille des populations reste généralement stable ; (iii) les ressources (tant espace que nourriture) sont limitées ; (iv) les caractéristiques individuelles varient énormément pour qu'il n'existe pas deux individus identiques ; (v) les variations sont en grande partie héréditaires (Mayr 1982).

Dans ces travaux de thèse, nous allons parcourir les capacités d'adaptation et/ou d'acclimatation d'un parasite humain, *Schistosoma mansoni*, lorsqu'il est confronté à différents environnements. Mais au préalable, pour illustrer cette variabilité phénotypique des êtres vivants, prenons les exemples des coraux, des pinsons des Galápagos et des tardigrades.

#### 1.1.2.1. Le corail comme exemple d'acclimatation

Les coraux scléactiniaires sont capables de vivre dans des eaux tropicales oligotrophes grâce à leur association endosymbiotique de type phototrophique (zooxanthelle, dinoflagellés du genre *Symbiodinium*). Ce symbiote transfère au corail une large fraction des produits issus de la photosynthèse. Cependant, la phototrophie n'est pas leur unique mode de nutrition. Dans le cas d'un blanchissement, suite à un stress thermique, la dégradation voire la perte des zooxanthelles et/ou des pigments photosynthétiques peut apparaître. Quand le corail est privé de son symbiote, il perd des ressources nutritives conséquentes. L'hétérotrophie va devenir alors une source d'énergie essentielle et vitale. Dans ce cas, les coraux blanchis peuvent présenter 4 à 40 fois plus de cnidocytes (Figure 4) leur permettant de chasser plus de proies pour compenser le manque d'énergie initialement fourni par les zooxanthelles. Les coraux sont donc capables de s'acclimater à une modification de leur environnement (stress thermique) afin de passer d'un rapport de 60% phototrophie / 40 % hétérotrophie à une hétérotrophie complète pour palier au manque de ressources nutritives et donc de disposer

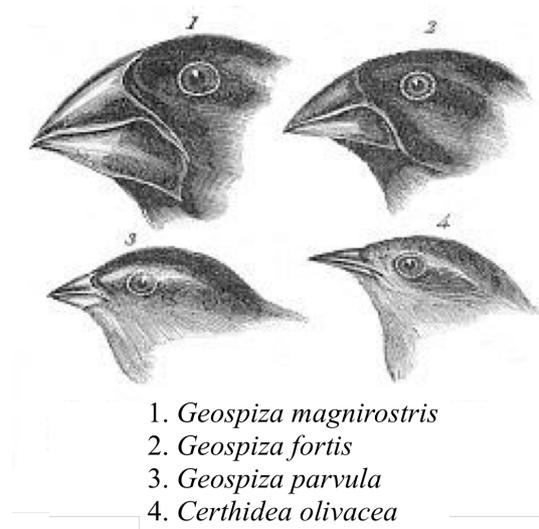
d'une meilleure résilience dans l'attente de conditions environnementales plus propices ou de revenir à un état symbiotique stable (Ladrière & Roberty in press).



**Figure 4: Observations microscopiques de coupes transversales de corail.** Observation en microscopie à lumière à transmission (A-B) et en microscopie électronique à transmission (C-D) d'ectoderme avec cnidocytes (c): chez un corail sain (A-C) et chez un corail blanchi (B-D) (Ladrière & Roberty in press).

#### 1.1.2.2. Les pinsons des Galápagos comme exemple d'adaptation

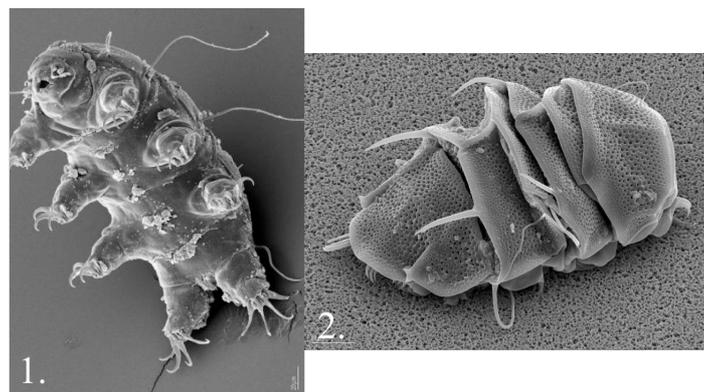
Les géospizes (pinsons) des Galápagos sont un excellent exemple d'adaptation à un environnement donné. Leur physiologie est adaptée pour mieux survivre, selon les spécificités de leur environnement immédiat. Ainsi *Geospiza magnirostris* possède un gros bec capable de casser des gros fruits à graines, lesquels tombent au sol pour qu'il puisse se nourrir. *Geospiza fortis*, à bec intermédiaire moyen, a une préférence alimentaire pour des petites graines quand celles-ci sont abondantes, sauf en cas de sécheresse où ce type de graine est plus rare et la consommation des plus grosses est favorisée. *Geospiza parvula* possède un bec intermédiaire plus fin ce qui lui permet de se nourrir essentiellement de petites graines et d'insectes. *Certhidea olivacea* utilise son bec étroit et pointu pour attraper exclusivement les insectes (Campbell & Reece 2007) (Figure 5).



**Figure 5: Différents phénotypes de bec chez les géospizes des Galápagos.** Chaque espèce a modifié la physiologie de son bec pour mieux survivre et s'adapter à son environnement direct (dépendant de la nutrition). 1. *Geospiza magnirostris* (géospize à gros bec), 2. *Geospiza fortis* (géospize à bec intermédiaire moyen), 3. *Geospiza parvula* (géospize à bec fin), 3. *Certhidea olivacea* (géospize olive à bec fin).

### 1.1.2.3. Les tardigrades comme exemple d'adaptation

Les tardigrades (oursins d'eau) sont de minuscules animaux, très proches des arthropodes qui sont capables de résister aux radiations, à de nombreux produits toxiques et à la déshydratation en faisant varier la proportion d'eau de 80% et à moins de 3%. Ils ont également la faculté d'entrer en cryptobiose ce qui leur permet de survivre pendant plus de dix ans sans eau, dans des conditions de vide absolu ou à de très hautes pressions (1200 atmosphères), à des températures extrêmes de  $-272.8^{\circ}\text{C}$  à plus de  $150^{\circ}\text{C}$  (Lam 2009) (Figure 6).



**Figure 6: Observation au microscope électronique à transmission d'un tardigrade.** En conditions stressantes, le tardigrade est capable de cryptobiose (2) pour survivre à des conditions environnementales différentes des conditions normales (1) (Lam 2009).

Ces exemples ciblés montrent que les mêmes êtres vivants, qui sont soumis à des milieux particuliers, expriment différents phénotypes pour répondre au mieux aux variations susceptibles de porter atteinte à leur survie.

### 1.1.3. La relation entre génotype, épigénotype, environnement et phénotype

L'hypothèse classique est que l'expression des phénotypes résulte de l'interaction entre le génotype et l'environnement. L'hypothèse émergente est que les phénotypes sont les résultats de l'expression du génotype et de l'épigénotype sous influences environnementales.

Le génotype d'un organisme fait référence à sa séquence d'ADN et sa combinaison allélique, alors que l'épigénotype fait référence aux modifications chimiques de cet ADN et de ses protéines associées, comme par exemple la modification de la structure de la chromatine pouvant affecter l'expression de certains gènes. Contrairement au génotype, l'épigénotype peut être influencé directement par l'environnement (Pembrey 2002; Liu *et al.* 2008).

Sewall Wright a été un des premiers à proposer l'existence de statuts cellulaires alternatifs (Wright 1945). Il suggérait que la persistance de statuts cellulaires alternatifs, nommée « l'hérédité cellulaire », est la résultante des changements dans les gènes nucléaires ou des changements cytoplasmiques des constituants dans les cellules (Wright 1945). Une des possibilités considérée était la présence, dans le cytoplasme, de plasmagènes (« auto-répliquants », sous-unités microscopiques). Une autre possibilité était relative aux propriétés d'auto-régulation des constituants du cytoplasme (Jablonka & Lamb 1999). À tous les statuts de différenciation, la persistance de ces statuts peut être basée sur des interactions à travers les constituants, qui font de la cellule un système de régulation à part entière. À partir de ces études, l'origine des statuts cellulaires différenciés dans la cellule est pensée pour une condition locale précise. Ces statuts favoriseraient certaines chaînes de contrôle des gènes, qui pourraient causer un déploiement des constituants pour passer d'un niveau basal stable à un autre statut en conditions variables (Wright 1945).

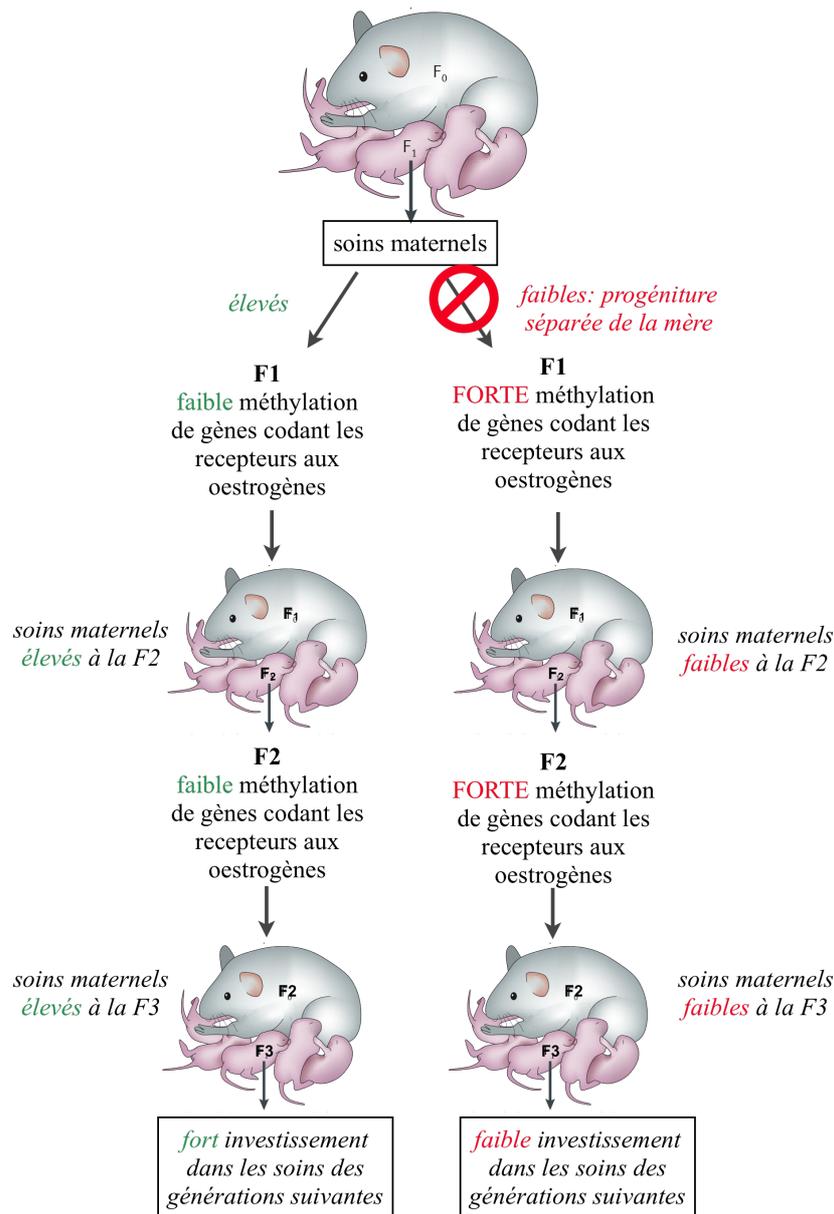
À partir des années 40-50, Hinshelwood et ses collègues ont argumenté que la persistance de réponses adaptatives dans les cultures bactériennes était causée par des changements de statuts métaboliques plutôt que liée à la préexistence de variants génétiques (Hinshelwood 1946; Hinshelwood 1953; Dean & Hinshelwood 1963). Une idée similaire a été suggérée en 1949 par Delbrück pour expliquer la persistance de statuts spécifiques, aux travers des

divisions cellulaires successives : une fois initié, un statut particulier se maintient grâce à une activité qui lui est propre (Delbrück 1949; Jablonka & Lamb 1999) .

L'épigénétique consiste en l'étude des changements de l'expression de certains gènes ainsi que de leur fonction, qui ne pourraient être expliqués par la séquence primaire d'ADN (Jablonka & Lamb 1999; Richards 2006; Bird 2007). Ces changements épigénétiques peuvent alors entraîner des changements phénotypiques sans impliquer de mutations génétiques (Gottsching *et al.* 2005). Par exemple, les jumeaux homozygotes partagent la même séquence d'ADN et leur ressemblance est souvent utilisée pour souligner la puissance génétique. Mais les jumeaux peuvent exprimer différents phénotypes impactés par des modifications épigénétiques induites lors de la vie de l'individu. Dans les stades précoces de développement, aucune différence génétique ou épigénétique n'est observable, mais chez des individus plus âgés des modifications épigénétiques (telles que les profils de méthylation de l'ADN) apparaissent et donnent lieu à des différences phénotypiques accentuées tout au long de la vie. Ces divergences peuvent conduire à une plus grande susceptibilité face à la maladie, à des caractères morphologiques ou comportementaux (Fraga *et al.* 2005).

Des études récentes indiquent que les variations épigénétiques dans des populations naturelles, peuvent être indépendantes de la variation génétique, et dans certains cas l'environnement peut induire des changements épigénétiques hérités par les générations futures (Bossdorf *et al.* 2008).

Le cas le plus frappant chez les animaux est celui des soins maternels chez les rongeurs. Habituellement, les mères (F0) s'occupent intensément de leur progéniture avec de haut taux de « licking/grooming » (LG) c'est-à-dire de toilettage. Si pour des raisons environnementales, les soins maternels ne sont pas apportés à la descendance suivante (F1), on constate que le taux de LG devient faible et les nouvelles femelles arrivées à l'âge adulte délaissent leurs jeunes (F2). Il a été démontré que cette modification de comportement résulte d'une forte méthylation dans la région promotrice de gènes codant pour les récepteurs aux œstrogènes. Cette méthylation entraîne une diminution significative d'expression de ces gènes dans le cerveau des jeunes descendantes (F2) issues de femelles (F1), les ayant très peu soignées. Arrivées à l'âge adulte, ces femelles de 2<sup>e</sup> génération (F2) ayant des gènes codant faiblement les récepteurs aux œstrogènes, sont insensibles à leurs propres œstrogènes circulants et de ce fait ne développent pas un comportement maternel adapté (revue par (Champagne 2008)) (Figure 7).



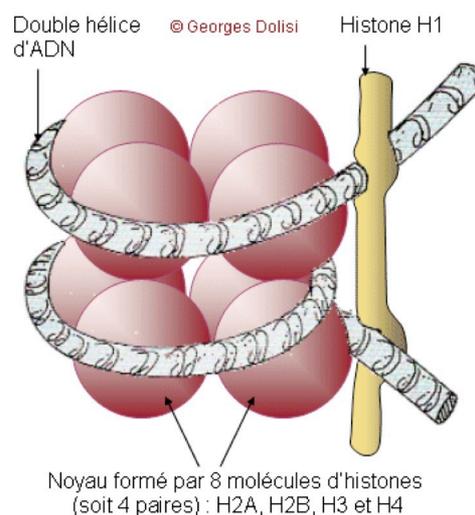
**Figure 7: Héritéité épigénétique des soins maternels aux jeunes chez les rongeurs.** Des conditions environnementales variables au cours du développement peuvent mener à des changements neuroendocriniens qui entraînent des modifications dans le comportement des progénitures suivantes. Quand les soins maternels sont prodigués à la génération F1, une faible méthylation du promoteur du gène codant pour les récepteurs aux œstrogènes est constatée. Ce faible taux de méthylation est transmis de génération en génération ainsi les générations F1, F2 et F3 apporteront des soins maternels à leurs progénitures. Si, au contraire, les soins maternels sont faibles, une forte méthylation du promoteur du gène codant pour le récepteur aux œstrogènes sera présente, entraînant une insensibilité de l'organisme aux hormones produites et amenant ainsi à un investissement très diminué. Ces modifications épigénétiques et comportementales sont héréditaires sur plusieurs générations. Adapté de Champagne et al. (revue par (Champagne 2008)).

Ces modifications épigénétiques sont partiellement héréditaires à travers la mitose et la méiose, même si les conditions initiales qui ont causées ce changement ont disparu (Gottsching *et al.* 2005; Champagne 2008). Dans le cas des expériences de Champagne *et al.*, seule la première génération est empêchée de fournir des soins à la progéniture. Les profils de méthylation et

les comportements de non-soins aux jeunes perdurent pendant les générations F2 et F3 alors que les mères sont au contact de leur progéniture et qu'aucun stress extérieur n'intervient.

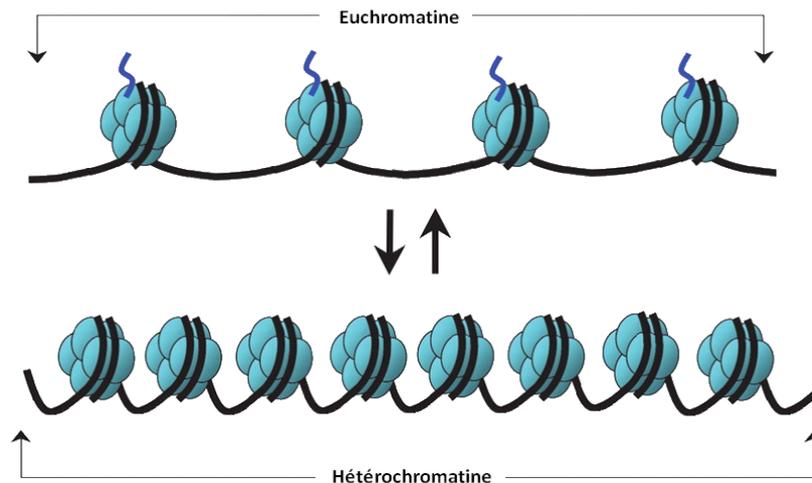
#### 1.1.4. Les supports de l'information épigénétique

L'épigénétique est basée sur un jeu de *processi* moléculaires qui peut activer, réduire ou totalement inhiber l'activité de certains gènes. Ces *processi* sont : (i) la méthylation des résidus cytosine dans la séquence d'ADN ; (ii) les changements dans la structure de la chromatine à travers des modifications chimiques, en particulier l'acétylation ou la méthylation des protéines histones ; (iii) des *processi* régulateurs médiés par les molécules microARN (miRNA) (Bossdorf *et al.* 2008) ; (iv) la topologie d'un locus dans le noyau (Kurz *et al.* 1996). Ces différents types de modifications épigénétiques ne sont pas indépendants les uns des autres et peuvent interagir de façon à réguler les gènes (Grant-Dawton & Dickinson 2005; Berger 2007). La plupart de ces modifications sont réversibles (Allis *et al.* 2007). Ces divers *processi* sont liés au fait que l'ADN n'est pas « nu », mais qu'il existe un intime complexe avec des protéines spécialisées qui, ensemble, forment la chromatine. Ainsi l'ADN est enroulé autour des unités nucléosomales constituées de protéines histones (Kornberg & Thomas 1974) (Figure 8). La structure de la chromatine est dépendante des modifications d'un noyau composé de 8 histones formant le nucléosome (Strahl & Allis 2000; Jenuwein & Allis 2001; Rice & Allis 2001)



**Figure 8: Schéma de la structure de la chromatine.** L'ADN (gris) est associé à des molécules histones (rouge) formant un collier de perles. Un nucléosome est composé de huit molécules d'histones (rouge). Un type d'histone, H1 (jaune), peut se lier à l'ADN à côté d'un nucléosome où il contribue à condenser davantage la fibre.

Dans sa forme la plus simple, la chromatine est initialement regardée comme plusieurs molécules présentes pour enrouler et organiser l'ADN, sous deux formes : l'euchromatine et l'hétérochromatine. L'euchromatine est à l'état décondensé. Elle est déspiralisée et correspond à des zones de gènes actifs, c'est-à-dire que les gènes peuvent être transcrits. L'hétérochromatine est à l'état condensé ; ses histones sont hypoacétylées et/ou hyperméthylées par rapport à l'euchromatine. Elle correspond à la partie de la chromatine inactive au niveau transcriptionnel (Figure 9).



**Figure 9: Schéma de la structure de l'euchromatine et de l'hétérochromatine.** L'euchromatine est à l'état décondensé, plus permissif à la transcription, alors que l'hétérochromatine est à l'état condensé, transcriptionnellement moins permissive (Luong 2009).

Deux types de modification de la chromatine, régulant la transcription de certains gènes, ont été listées et classifiées, selon qu'elles soient directement associées à la répression ou l'induction de l'expression (Tableau 1). Cette classification reste toutefois provisoire et des marques classées comme responsables de l'activation peuvent être parfois associées à la répression de gènes particuliers (Berger 2007). Le rôle des marques épigénétiques est encore en cours de classification et certaines de ces modifications sont conservées entre plusieurs espèces alors que d'autres sont différentes. Par exemple la marque H3K4me3 peut être associée soit à une activation, soit à une répression, soit à une combinaison de ces deux (Berger 2007).

**Tableau 1: Modifications de la chromatine sur les histones.** Ces modifications sont classées en fonction de leur corrélation avec l'induction et/ou la répression de gènes mais cette classification est provisoire. Adapté de Berger et al. (Berger 2007).

Marques	Histones	Rôles transcriptionnels
Lysine acétylée (Kac)	H3 (9, 14, 18, 56), H4 (5, 8, 13, 16), H2A, H2B	Induction
Sérine/Thréonine phosphorylées (S/Tph)	H3 (3, 10, 28), H2A, H2B	Induction
Arginine méthylée (Rme)	H3 (17, 23), H4 (3)	Induction
Lysine méthylée (Kme)	H3 (4, 36, 79)	Induction
	H3 (9, 27), H4 (20)	Répression
Lysine ubiquitinée (Kub)	H2B (123 chez champignons, 120 chez mammifères)	Induction
	H2A (119 chez mammifères)	Répression
Lysine SUMOylée (Ksu)	H2B (6/7), H2A (126)	Répression
Proline isomérisée (Pisom)	H3 (30-38)	Induction / Répression

Ces formes distinctives de la chromatine apparaissent suite à des mécanismes covalents ou non covalents, qui incluent une quantité de modifications post-traductionnelles des histones, des étapes de remodelage énergie-dépendantes qui mobilisent ou altèrent la structure des nucléosomes, ou la dynamique des variants d'histones (dans ou en dehors du nucléosome) et le rôle de microARNs non codants. A ce stade de ma thèse, une définition simplifiée des microARNs est nécessaire car je n'en ferai pas état par la suite. Les microARNs sont des ARNs simples brins d'une longueur d'environ 21 à 24 nucléotides et qui sont des répresseurs de transcription: en s'appariant à des ARNs messagers, ils guident leur dégradation ou la répression de leur traduction en protéines. Tous ces mécanismes permettent de présenter un réseau inter-relié qui crée la variation des polymères de la chromatine. Ces changements peuvent se produire de façon spontanée suite à l'influence de l'environnement. Plusieurs composantes diététiques et d'autres facteurs maternels (allaitement, soins aux petits, etc...) sont maintenant connus pour affecter l'expression de certains gènes *via* la méthylation de l'ADN, sans pour autant modifier la séquence (Weaver *et al.* 2004; Franklin & Mansuy 2010). Les fréquences des variants épigénétiques sont déterminées par la plasticité et l'efficacité de la transmission, et les évolutions adaptatives apparaissent suite à la sélection de variants héréditaires (Pal 1998).

#### 1.1.5. L'information épigénétique chez *S. mansoni* : État de l'art

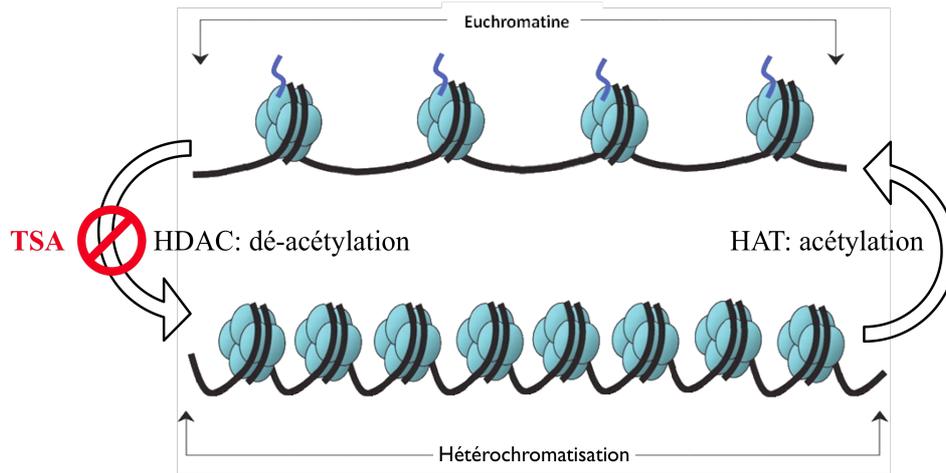
Avant mon arrivée au laboratoire, l'équipe a mis au point des outils permettant d'étudier l'épigénétique sur *S. mansoni* et son implication dans la variabilité phénotypique.

Les techniques d'immunoprécipitation de la chromatine associée à de la PCR quantitative (ChIP-qPCR) ou du séquençage massif (ChIP-seq) ont été établies précédemment. Certaines

marques spécifiques de l'euchromatine ou de l'hétérochromatine ont été ciblées. Ces marques utilisées sont: Histone H3 acétylée sur la Lysine 9 (H3K9ac, marque de promoteur actif), Histone H3 tri-méthylée sur Lysine 4 (H3K4me3, marque euchromatique), Histone H3 tri-méthylée sur la Lysine 9 et Histone H3 tri-méthylée sur la Lysine 27 (H3K9me3 et H3K27me3, marques hétérochromatiques). Ces différentes marques étant conservées chez *S. mansoni*, des anticorps commerciaux ont pu être utilisés (Cosseau *et al.* 2009). La spécificité et l'efficacité des anticorps ont été préalablement testées.

Par la méthode de Western Blot, nous avons vérifié que les anticorps n'interagissent pas avec d'autres protéines que celle d'intérêt. Une fois l'anticorps validé, le ratio anticorps-chromatine est déterminé par titration. Pour une certaine quantité de chromatine, l'anticorps doit être en excès afin que la quantité de produits immunoprécipités ne dépendent pas de la concentration en anticorps utilisée. La titration est réalisée en utilisant une quantité croissante d'anticorps pour une même quantité de chromatine. A partir d'un certain seuil d'anticorps, la quantité d'ADN immunoprécipité reste constante, ainsi la quantité d'anticorps utilisée est supérieure au seuil de saturation.

Lors de mes recherches, je me suis focalisée sur les modifications de la structure de la chromatine, associées à des modifications chimiques des histones, par exemple la méthylation ou l'acétylation de certaines histones, car des travaux antérieurs de l'équipe ont montré que, dans notre modèle *S. mansoni*, des modifications de la structure de la chromatine étaient associées à des modifications de transcription de certains gènes et à l'apparition de phénotypes cryptiques. Ces résultats ont été obtenus en modulant l'épigénotype de notre espèce modèle (présenté ci dessous) en traitant des larves avec un inhibiteur d'histones déacétylases (HDAC). Les HDACs sont présentes chez *S. mansoni* (Oger *et al.* 2008) et permettent de passer de l'état euchromatique à l'état hétérochromatique en désacétylant les histones. En inhibant l'enzyme responsable du passage de l'euchromatine à l'hétérochromatine, la chromatine reste à l'état décondensé, c'est-à-dire sous la forme d'une structure permissive pour la transcription de certains gènes. Un traitement à la Trichostatine A (TSA) (Yoshida *et al.* 1995), a été utilisé pour fortement réduire l'activité des HDACs chez cet organisme (Dubois *et al.* 2009; Cosseau *et al.* 2010) (Figure 10).



**Figure 10: Effet inhibiteur de la trichostatine A (TSA) sur les histones déacétylases (HDAC).** Pour passer de l'état décondensé (euchromatine) à l'état condensé (hétérochromatine), une enzyme intervient : une Histone DéAcétylase (HDAC). Cette dernière est la cible de la TrichoStatine A (TSA). En venant inactiver cette enzyme, la TSA bloque le passage de l'euchromatine à l'hétérochromatine, en laissant donc la chromatine à son état décondensé.

Ces travaux ont montré que le traitement à la TSA augmente la variabilité phénotypique sur des traits morphologiques et comportementaux, ainsi que sur le taux de transcription de gènes spécifiques impliqués dans le phénomène de compatibilité hôte/parasite amenant à des changements du succès infectieux du parasite. Ces données sont en accord avec le modèle théorique de Pal & Miklos, dans lequel la génération de variants phénotypiques épigénétiques est une première étape dans l'évolution adaptative (Pal & Miklos 1999).

#### 1.1.6. Vers une caractérisation approfondie de l'épigénome de *S. mansoni*

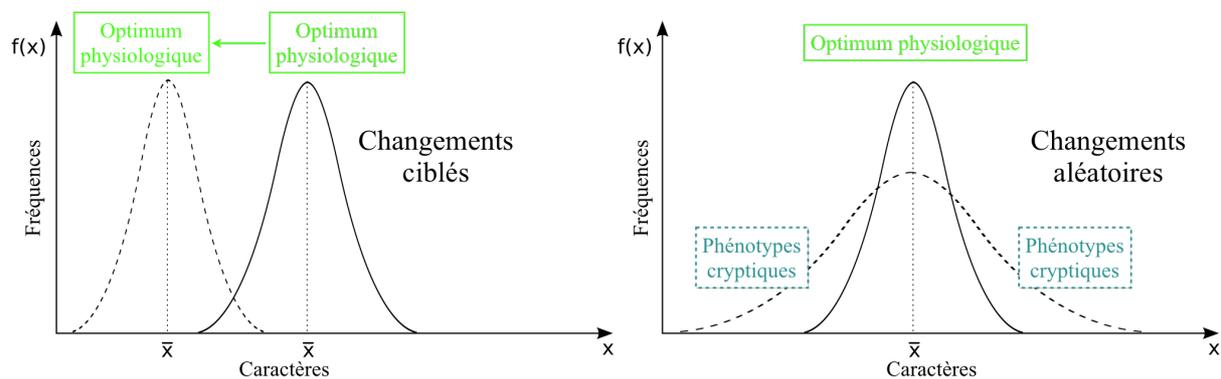
Au regard des travaux antérieurs, je me suis penchée sur l'influence de l'environnement direct sur l'épigénotype de *S. mansoni*, en me basant sur l'étude des modifications de la structure de la chromatine.

Durant cette thèse, les résultats issus des analyses de structure chromatinienne ont été confrontées à des données de transcriptomique (RNA-seq) acquises, afin de répondre aux questions suivantes: (i) Suite à un stress environnemental, y-a-t'il apparition de modifications épigénétiques de nature chromatinienne? (ii) Le changement environnemental peut-il entraîner des modifications d'ordre transcriptomique? (iii) Ces modifications, tant épigénétiques que transcriptomiques, sont-elles ciblées et/ou aléatoires dans le génome? (iv) Existe-il une corrélation directe entre les modifications épigénétiques et les modifications des profils de transcription? (v) Ces modifications moléculaires sont-elles corrélées à des

changements phénotypiques visibles ? (vi) Les changements engendrés, tant phénotypiques que moléculaires, sont-ils hérités par les générations suivantes?

D'après le modèle de Pal & Miklos, deux hypothèses non exclusives sont envisageables (Pal & Miklos 1999) :

- La première est que les modifications apparentes sont aléatoirement dirigées dans le génome. On observe alors une augmentation de la variabilité phénotypique associée à l'apparition de phénotypes cryptiques, avec une fréquence moindre mais restant centrée sur un optimum physiologique. Ce type de réponse augmente l'intervalle de vie des organismes en repoussant les bornes létales inférieures et supérieures (Figure 11).
- La deuxième hypothèse est que ces changements épigénétiques sont ciblés dans le génome et que dans ce cas la variabilité phénotypique reste inchangée alors que l'optimum physiologique n'est pas le même (Figure 11).



**Figure 11: Modèle théorique de réponse adaptative selon un changement environnemental.** Ce modèle présente deux hypothèses de réponse possible suite à un changement d'environnement (ligne pointillée) comparé à un environnement stable (ligne pleine). Sur le graphe de gauche, c'est la représentation de changements ciblés dans le génome avec une variabilité phénotypique inchangée mais un optimum physiologique déplacé. Sur le graphe de droite, c'est la représentation de changements aléatoires dans le génome suite à un stress environnemental (ligne pointillée) augmentant la variabilité phénotypique grâce à l'apparition de phénotypes cryptiques et un optimum physiologique inchangé. Adapté de (Pal & Miklos 1999).

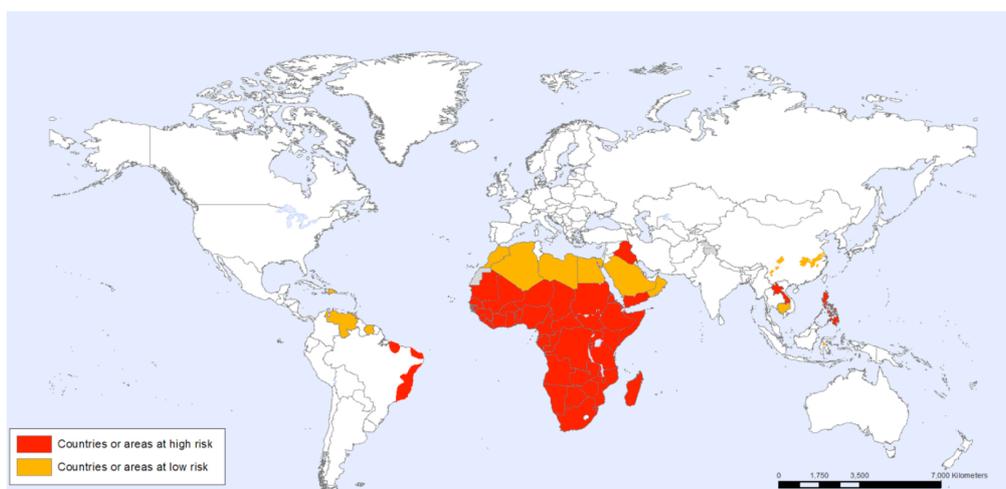
Le modèle d'étude choisi est un parasite de l'homme, *S. mansoni*, qui dans son cycle naturel est soumis à des conditions environnementales variables et des environnements différents. L'interaction hôtes/parasites est un excellent modèle en biologie de l'évolution puisque cette interaction subit des pressions sélectives fortes et une évolution rapide, entraînant souvent une dynamique co-évolutive : les hôtes évoluent de sorte à limiter et se protéger de l'infestation, et les parasites évoluent en parallèle pour contrer les défenses des

hôtes. On assiste à une véritable course aux armements expliquée par la théorie de la reine rouge (Van Valen 1973) qui énonce qu'il existe une probabilité constante d'extinction au cours du temps, même sans variation du milieu. Cette théorie suggère que cette évolution résulte des attaques parasitaires : les parasites cherchent sans cesse de nouveaux moyens d'attaquer leurs espèces-hôtes qui doivent donc obligatoirement s'adapter pour trouver la parade au risque de disparaître si elles ne parviennent pas à contourner assez vite la virulence de leurs assaillants.

## 1.2. Modèle d'étude

### 1.2.1. Zones de répartition

Les schistosomes sont des parasites appartenant à l'embranchement des Plathelminthes et à la classe des Trématodes. Le genre *Schistosoma* compte une vingtaine d'espèces dont cinq (*S. haematobium*, *S. intercalatum*, *S. japonicum*, *S. mansoni* and *S. mekongi*) sont les principaux agents de la schistosomiase (ou bilharziose) humaine (Chitsulo *et al.* 2004). Les espèces du genre *Schistosoma* sont réparties sur toute la ceinture tropicale et intertropicale (WHO 2011) (Figure 12) : *S. haematobium* (responsable de la bilharziose urinaire) est présent en Afrique et au Moyen Orient, *S. japonicum* (responsable de la bilharziose spléno-hépatique) est endémique de l'Asie, et *S. mansoni* (responsable de la bilharziose intestinale) est largement répandu en Afrique et constitue la seule espèce en Amérique du Sud et dans les Caraïbes où elle fut importée récemment (au cours des 16<sup>e</sup> et 17<sup>e</sup> siècles lors du déplacement massif des populations africaines vers le nouveau monde durant la traite négrière) (Morgan *et al.* 2001).



**Figure 12: Répartition de la schistosomiase en 2011.** Les pays à haut risque (rouge) et à faible risque (orange) sont concentrés dans l'hémisphère sud (WHO 2011).

### 1.2.2. Bilharziose

La bilharziose est une maladie parasitaire qui se trouve au second rang, après la malaria, en termes de morbidité et de mortalité, avec plus de 200 millions de personnes infestées dans 74 pays et plus de 200 000 personnes décédant par an (Chitsulo *et al.* 2004; Gryseels *et al.* 2006). Cette maladie est un fardeau économique important (Croce *et al.* 2010) car les personnes impactées par la maladie sont malheureusement difficilement soignables et les pays concernés voient leur développement local stagner du fait de ce fléau.

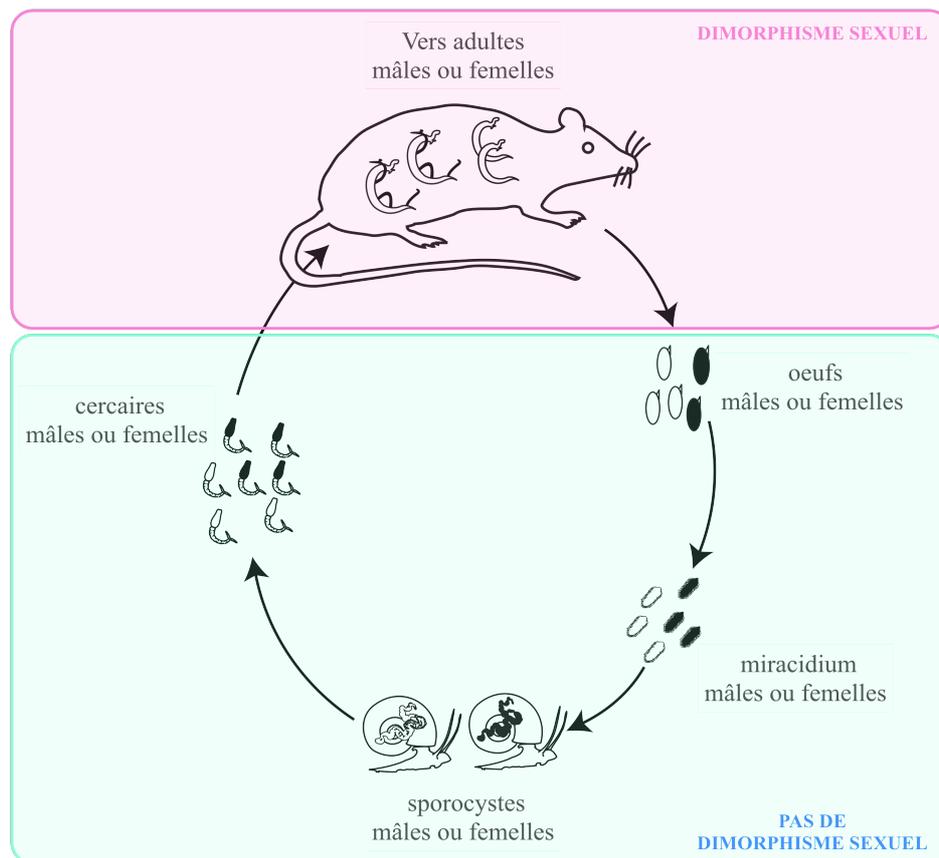
Plus particulièrement, la variabilité phénotypique du parasite *S. mansoni* est un réel problème médical tant pour les campagnes de vaccination que pour la multiplication de nouveaux foyers et la colonisation de nouveaux milieux. Les changements environnementaux liés à la fluctuation de points d'eau ainsi qu'à la croissance et la migration des populations peuvent faciliter l'émergence des schistosomiasés (Patz *et al.* 2000). Par sa grande capacité à s'acclimater et/ou s'adapter aux différents environnements auxquels le parasite est soumis, l'apparition de phénotypes cryptiques plus résistants ou plus infestants est envisageable. Il est donc d'une importance majeure de comprendre les mécanismes moléculaires sous-jacents à la réponse adaptative du parasite *S. mansoni* sous influences environnementales.

### 1.2.3. Cycle de vie et reproduction

*S. mansoni* est un parasite digène possédant un cycle de vie complexe (Figure 13). Il passe par un hôte intermédiaire mollusque dulçaquicole, *Biomphalaria glabrata* (ou d'autres espèces de *Biomphalaria* selon la localisation géographique), dans lequel a lieu une multiplication asexuée. Le parasite est ensuite relâché sous la forme d'une larve nageuse, la cercaire, qui infeste l'hôte définitif humain ou murin au contact de l'eau au sein duquel la reproduction sexuée a lieu. Seul le corps de la cercaire traverse l'épiderme (la queue se détache et reste dans le milieu extérieur) pour migrer vers le cœur de l'hôte définitif *via* le système sanguin (Standen 1953).

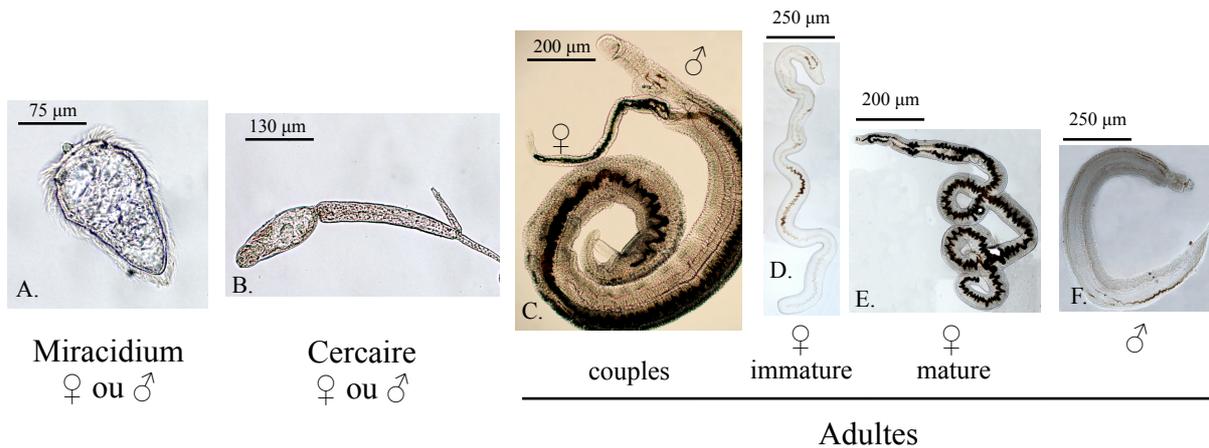
Vingt-quatre heures après la pénétration, la cercaire arrive dans le ventricule droit du cœur de l'hôte et se transforme en schistosomule qui met entre 5 et 7 jours post-pénétration pour arriver au foie. La schistosomule devenue adulte atteint alors sa maturité sexuelle. Ensuite, elle migre à contre courant le long de la veine porte hépatique afin que la ponte des œufs se fasse dans les veinules du système mésentérique.

Enfin, les œufs vont traverser les veinules à l'aide d'un éperon jusqu'aux intestins, pour être excrétés *via* les fèces de l'hôte et éclore au contact de l'eau pour laisser échapper une seconde larve appelée miracidium qui infectera l'hôte intermédiaire mollusque et boucler ainsi le cycle parasitaire (Figure 13). Certains œufs du parasite peuvent aussi s'accumuler dans le foie de l'hôte définitif occasionnant une forte réaction inflammatoire qui est un des symptômes de cette maladie parasitaire.



**Figure 13: Cycle de vie et dimorphisme sexuel de *S. mansoni*.** Le dimorphisme sexuel (rose) est restreint au stade adulte, alors qu'aux stades larvaires les phénotypes mâles et femelles sont indistinguables.

Une fois dans l'hôte définitif, 25 jours post pénétration, le dimorphisme sexuel entre le mâle et la femelle apparaît (Basch 1990). Même si aux stades larvaires, les individus sont génétiquement sexués, le dimorphisme sexuel est restreint au stade adulte (Figure 14) car si les femelles ne rencontrent pas de mâles ou si elles viennent à être séparées du mâle, elles restent ou redeviennent immatures (plus fines et sans ovaires). Le phénotype « femelle immature » va dans le sens d'un déterminisme sexuel non seulement génétique mais aussi environnemental, puisque les femelles adultes ont besoin d'un accouplement avec un mâle pour finir leur développement et devenir matures.



**Figure 14: Photos en microscopie des différents stades de développement de *S. mansoni*.** Les individus miracidia (A), cercaires (B) ou adultes (C-F) sont gonochoriques. Les miracidia (A) et cercaires (B) ne présentent aucun dimorphisme sexuel alors qu'au stade adulte un dimorphisme sexuel est présent. La femelle plus fine reste dans le canal gynécophore du mâle plus musculeux (C). Lorsque la femelle n'est pas accouplée elle est, ou redevient, immature (D), alors qu'accouplée au mâle (F) elle est mature (E), avec une quantité l'hémozoïne importante présente dans le tube digestif.

Des analyses de transcriptomique ont mis en évidence des produits de gènes spécifiques à chaque stade de développement et pour chaque sexe. Par la technique de *microarrays* et des approches statistiques, 973 gènes présentent un différentiel d'expression et peuvent décrire les trois caractéristiques principales du développement du schistosome : des *processi* asexués au sein de l'hôte intermédiaire mollusque, la maturation sexuelle au sein de l'hôte définitif humain ou murin, et le dimorphisme sexuel au stade adulte (Fitzpatrick *et al.* 2009). Il existe donc des mécanismes génétiques liés aux différents stades de développement et à la mise en place du dimorphisme sexuel mais ni ces mécanismes génétiques, ni les mécanismes liés à l'environnement ne sont encore étudiés.

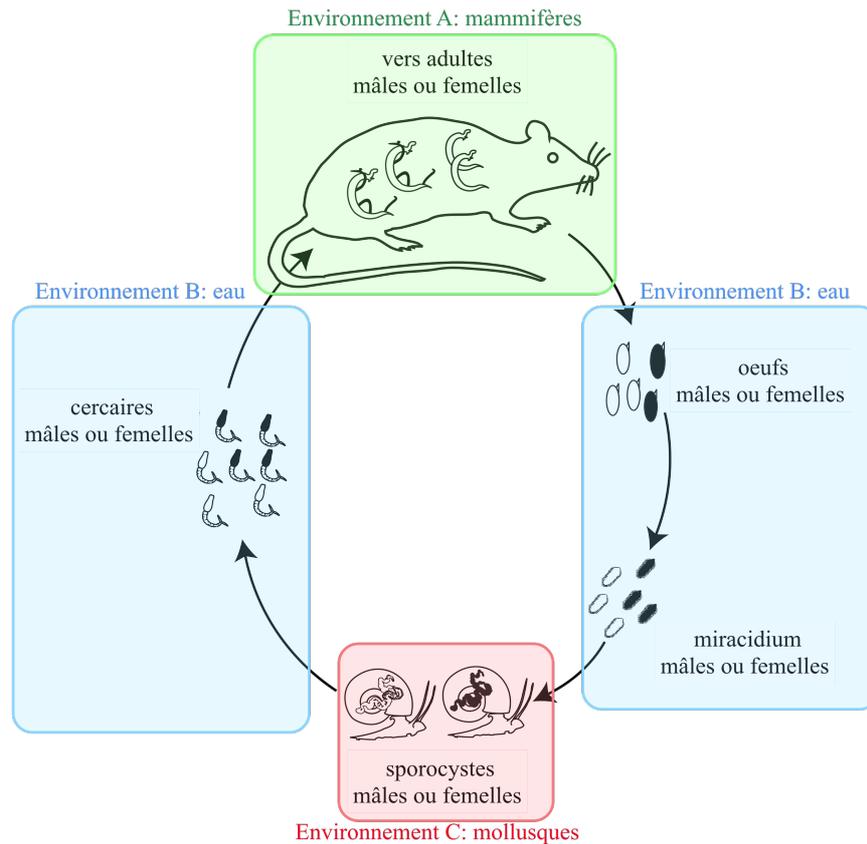
Au contraire des schistosomes, la plupart des parasites plathelminthes sont des espèces hermaphrodites (Loker & Brant 2006). Cette fréquence de l'hermaphrodisme suggère que cet état est un avantage pour ces parasites car il réduit probablement au minimum le risque de ne pas trouver de partenaire à l'intérieur de l'hôte (Desprès & Maurice 1995). D'un point de vue évolutif, l'acquisition des sexes séparés chez les Schistosomatidae, est concomitante avec l'invasion des animaux à sang chaud (Loker & Brant 2006). Ceci pourrait expliquer les bénéfices en terme de diversité génétique, agissant contre le système immunitaire sophistiqué des vertébrés à sang chaud, et/ou la spécialisation de chaque sexe est limitée à des « tâches ménagères ». Le mâle musculeux maintient la femelle dans son canal gynécophore. Le couple se déplace à contre courant du flux sanguin jusqu'aux veines mésentériques afin que la femelle relargue ses œufs. Le mâle maintient le couple grâce à ses ventouses, alors que la

femelle peut se déplacer seule dans les petites veinules, là où les œufs ont une chance de pénétrer à travers le tissu intestinal sans être lavés par le système circulatoire. En effet, ces veinules ne sont pas accessibles aux mâles (Read & Nee 1990).

Plus généralement, l'évolution de la reproduction sexuée chez les espèces gonochoriques les rend plus efficaces pour éliminer des mutations délétères en cassant des combinaisons non favorables et en augmentant la variabilité génétique *via* la recombinaison. Sans elle, chaque mutation néfaste ne pourrait pas être purgée du génome et serait transmise à la génération suivante (Felsenstein 1974). Ce phénomène est appelé *Cliquet de Müller* et serait présent aux stades asexués de *S. mansoni*. Pour la partie sexuée du cycle, on constate par contre un brassage génétique *via* la recombinaison, agissant contre les mutations irréversibles qui ont été accumulées pendant les stades asexués (stades larvaires).

#### 1.2.4. Environnement-hôte

Le parasite, *S. mansoni* étant un endoparasite, est soumis directement à chaque environnement-hôte à part entière, qu'il soit intermédiaire ou définitif, pour réussir à se maintenir et se développer. Aux deux stades larvaires, miracidium et cercaire, le parasite est sous forme d'une larve libre nageuse évoluant dans un environnement aqueux (Figure 15 B). Mais dès qu'elle entre en contact avec l'hôte intermédiaire ou définitif, elle passe d'un milieu aqueux à un milieu directement corrélé à l'hôte qui devient son environnement propre nécessaire à sa croissance, son maintien et sa reproduction (Figure 15 A et C).



**Figure 15: Différents environnements auxquels *S. mansoni* est confronté.** Tout au long de son cycle de développement, le parasite est soumis à différents environnements : environnement-hôte définitif humain ou murin (A), environnement aqueux (B), environnement-hôte intermédiaire mollusque (C).

Tout au long de son cycle de développement, les traits d'histoire de vie du parasite, qui sont des traits biologiques contribuant à la valeur sélective et en particulier aux composantes fondamentales que sont la survie et la reproduction, sont mesurables. Ces traits sont variables d'une espèce à l'autre même si elles sont très proches phylogénétiquement. Cette variabilité existe aussi entre les populations à l'intérieur d'une même espèce ou entre individus à l'intérieur d'une même population (Thomas *et al.* 2010). Ainsi, pour l'espèce *S. mansoni*, deux souches isolées géographiquement n'auront pas des traits d'histoire de vie identiques. Une souche pourra présenter une forte infectivité au niveau de l'hôte intermédiaire et une faible infectivité au niveau de l'hôte définitif, alors qu'une souche de la même espèce mais d'un isolat géographique différent, montrera un profil inverse.

Les traits d'histoire de vie du parasite *S. mansoni* ont été listés afin de comparer les performances des mâles et des femelles ainsi que leur pathogénicité contre le mollusque *B. glabrata* (Boissier *et al.* 1999).

Au niveau de l'hôte intermédiaire *B. glabrata*, il est possible d'assimiler une valeur de prévalence pour chaque infestation qui correspond au pourcentage de mollusques infestés par rapport au nombre total de mollusques exposés au parasite. Ce trait de vie est quantifiable en observant au travers des coquilles, la présence ou non de sporocystes développés au niveau de l'hépatopancréas. Cette prévalence est de plus confirmée par l'émission de cercaires. À cette valeur de prévalence, une estimation d'intensité d'infestation peut être assimilée en comptant le nombre de sporocystes implantés et développés dans le mollusque par rapport au nombre de parasite mis au contact du mollusque. Lors de mes travaux de thèse, ce trait n'a pas pu être mesuré car nous aurions été contraints de disséquer des mollusques parasités alors qu'ils sont nécessaires pour la suite de nos expériences.

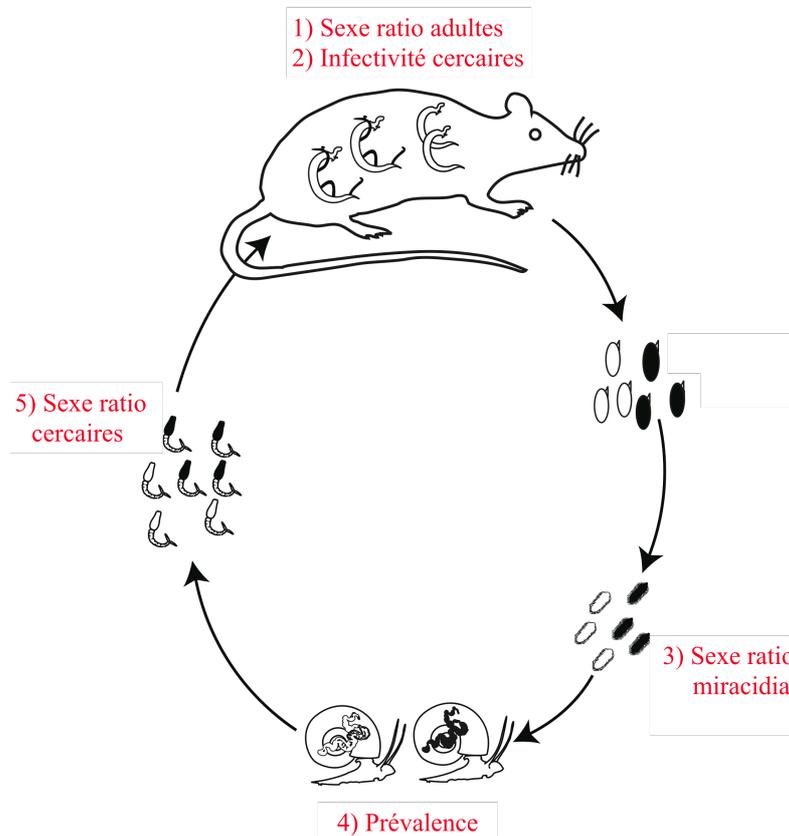
Au niveau de l'hôte définitif, l'infectivité des cercaires est mesurée par le ratio du nombre de vers développés dans l'hôte sur le nombre total de larves de parasite mis au contact de la souris lors de l'infestation. Nous avons aussi évalué la conséquence de l'infestation sur l'hôte en quantifiant la réaction immunopathologique de la souris, en calculant le ratio « poids de la rate sur le poids total de la souris ». Une souris saine a une rate qui représente environ 0,42% du poids total alors qu'une souris présentant des symptômes liés à l'infestation parasitaire a une rate aux alentours des 2% du poids total (Boissier, communication personnelle).

A tous les niveaux du cycle du parasite, le ratio ♀/♂ (sexe-ratio) est estimé. Aux deux stades larvaires (miracidia et cercaires), le sexe-ratio est évalué par PCR en utilisant des marqueurs spécifiques du chromosome W (chromosome sexuel spécifique des femelles) (Portela *et al.* 2010). Au stade adulte, le sexe-ratio est calculé en dénombrant le nombre de mâles et de femelles présents dans la souris (Figure 16). Le sexe-ratio d'une population est un paramètre démographique clé car il représente un concept central d'évolution sous sélection sexuelle prédisant l'intensité et la direction de la compétition à trouver un partenaire (Kvarnemo & Ahnesjö 1996).

Selon le principe de Fisher, la plupart des espèces ont un sexe-ratio à l'équilibre pour produire autant de mâles que de femelles. Chez *S. mansoni*, le sexe-ratio fluctue autour d'un équilibre 1 : 1 (un mâle : une femelle), cependant des sexes-ratios non équilibrés, biaisés vers les mâles ou femelles, sont retrouvés dans des populations animales (Hardy 2002).

Dans le modèle schistosome, de nombreuses publications montrent que les scientifiques sont intrigués par le sexe-ratio biaisé en faveur des mâles. Un grand nombre d'études de terrain, de travaux analytiques et de travaux en laboratoire ont été menés dans cette optique.

Tant les souches de laboratoire que celles issues de conditions naturelles présentent ce sexe-ratio biaisé mâle. Dans le cadre de cette thèse, ce trait devient dès lors primordial à mesurer puisqu'il reflète les conditions naturelles et si ce trait vient à changer, nous pourrions confronter nos résultats aux données bibliographiques sur le sujet.



**Figure 16: Traits d'histoire de vie mesurables tout au long du cycle de développement de *S. mansoni*.** (1) sexe-ratio évalué en dénombrant le nombre de vers mâles et femelles, (2) infectivité des cercaires par le ratio du nombre de vers développés sur le nombre de cercaires présentées à l'infestation, (3) sexe-ratio des miracidia quantifié par PCR, (4) prévalence d'infestation mesurée par le nombre de mollusques infestés sur le nombre total de mollusques exposés au parasite, (5) sexe-ratio des cercaires évalué par PCR.

### 1.2.5. *S. mansoni* comme modèle en adaptation évolutive

*S. mansoni* a su survivre alors qu'il est en permanence confronté à différents environnements comme le milieu extérieur ou les différents hôtes obligatoires à sa reproduction. Il s'est adapté pour résister au système immunitaire de l'hôte vertébré et depuis plusieurs dizaines d'années à des facteurs abiotiques comme les drogues utilisées dans les campagnes de contrôle de la bilharziose. Il s'est également acclimaté à de nouveaux environnements et de nouveaux hôtes lorsqu'il a été exporté du continent africain vers l'Amérique du Sud pendant la traite des esclaves aux 16<sup>e</sup> et 17<sup>e</sup> siècles.

*S. mansoni* est donc un modèle extraordinaire pour étudier les phénomènes de réponses adaptatives à différents environnements, et ce d'autant plus que tous les stades de développement de ce parasite sont expérimentalement accessibles en laboratoire. Par ailleurs, le génome de cet organisme est entièrement séquencé, assemblé et annoté (Berriman *et al.* 2009). Ces ressources génomiques sont actuellement disponibles dans des banques de données (SchistoDB, <http://schistodb.net/schisto/>) et pourront être exploitées afin de réaliser des analyses moléculaires approfondies. *S. mansoni* est donc un modèle privilégié pour tenter d'établir un lien entre les mécanismes moléculaires et l'évolution adaptative.

### 1.2.6. Traitements

#### 1.2.6.1. Traitements à succès

Le praziquantel (PZQ), constitue le traitement actuel pour le contrôle de la bilharziose. Cette médication administrée depuis les quarante dernières années, par voie orale en une dose unique, agit 1h après l'ingestion en paralysant les vers matures et en endommageant le tégument, mais a peu ou pas d'effet sur les œufs, les schistosomules et les vers immatures (Aragon *et al.* 2008). Les œufs peuvent donc être encore excrétés pendant plusieurs semaines suivant le traitement. (Gryseels *et al.* 2006). Le PZQ reste tout de même le traitement de choix avec certains avantages comme l'administration orale en une unique dose, une faible toxicité pour l'hôte et un coût peu élevé (Fenwick *et al.* 2003; Utzinger & Keiser 2004).

Cependant, certaines souches de schistosomes ont une sensibilité moindre au praziquantel avec la possibilité d'une apparition de résistance, déjà identifiée dans plusieurs pays d'Afrique (Ismail *et al.* 1999; Gryseels *et al.* 2001), ce qui est inquiétant au regard des centaines de millions de personnes à soigner. Aussi, depuis une dizaine d'années, des efforts ont été faits pour trouver de nouvelles thérapies (Dong *et al.* 2010 ; Laurent *et al.* 2008; Sayed *et al.* 2008; Abdulla *et al.* 2009; Rai *et al.* 2009), afin de tenter d'éradiquer la bilharziose à différents stades de développement du parasite (ex : l'artémésinine semble efficace contre le stade schistosomules). Des remèdes déjà utilisés pour d'autres parasitoses ont également été testés (Xiao *et al.* 2007; Boissier *et al.* 2009a; Keiser *et al.* 2009; Utzinger *et al.* 2010; Eissa *et al.* 2011), sous réserves toutefois de ne pas provoquer de résistances pour d'autres maladies co-endémiques comme le paludisme dans plusieurs pays (Keiser *et al.* 2009). Ainsi, des études sur les effets des trioxaquines ont été engagées au laboratoire (Boissier *et al.* 2009b; Pradines *et al.* 2011; Portela *et al.* 2012). Ce sont des drogues hybrides de deux principes actifs (le trioxane et l'aminoquinoline) qui ont été initialement développées comme anti-

malarique pour cibler l'hème des parasites, mais qui n'ont jamais été utilisées dans le traitement et le contrôle du paludisme. Les premiers résultats ont montré leur efficacité à tous les stades de développement du schistosome. Cette médication agit mieux sur les stades larvaires que le PZQ mais est nettement moins bonne sur les stades adultes. L'association de ces deux thérapies est actuellement en cours d'essais cliniques pour optimiser leurs effets à tous les stades de développement du schistosome (Portela *et al.* 2012).

#### 1.2.6.2. Traitements en échec

Certaines molécules utilisées pour soigner la bilharziose n'ont pas été convaincantes pour divers motifs tels que la toxicité, un effet moindre sur certain stade de développement (Tableau 2).

**Tableau 2: Historique des composés présentant des propriétés antischistosomias.** Ces traitements ont été démontrés comme ayant une activité chez l'animal et/ou l'homme (Cioli 1998; Laurent 2006).

Dates de première utilisation	Composés	Remarques
1900 av. JC	Antimoine	D'après d'anciens papyri
1918	Dérivés de l'antimoine	Très toxique
1920	Éméline	Moyennement active et toxique
1955	Métrifonate	Utilisé en Afrique
1962	Lucanthone	Remplacé par l'hycanthone
1964	Niridazole	Oral et moyennement actif
1965	Hycanthone	Mutagène et apparition de résistance
1969	Oxamniquine	Utilisée en Afrique mais apparition de résistance
1976	Amoscanate	Testé en Chine et toxique
1978	Ro-11 3128	Benzodiazépine, actif sur vers immatures
1978	Oltipraz	Toxique
1981	Cyclosporine A	Testé en laboratoire

Ces traitements ont été supprimés parce ce qu'ils étaient trop toxiques ou avaient des effets secondaires non négligeables pour l'homme (ex: antimoinés et dérivés, éméline, lucanthone, hycanthone, amoscanate, benzodiazépine et oltipraz); d'autres parce qu'ils n'étaient pas assez actifs contre le parasite ou ne présentaient pas de spécificité propre à la maladie (ex: le niridazole; la cyclosporine qui a une activité antiparasitaire contre le paludisme, la leishmaniose, la toxoplasmose); d'autres enfin en raison de l'existence de résistances qui peuvent être acquises ou innées (ex: métrifonate, hycanthone, oxamniquine).

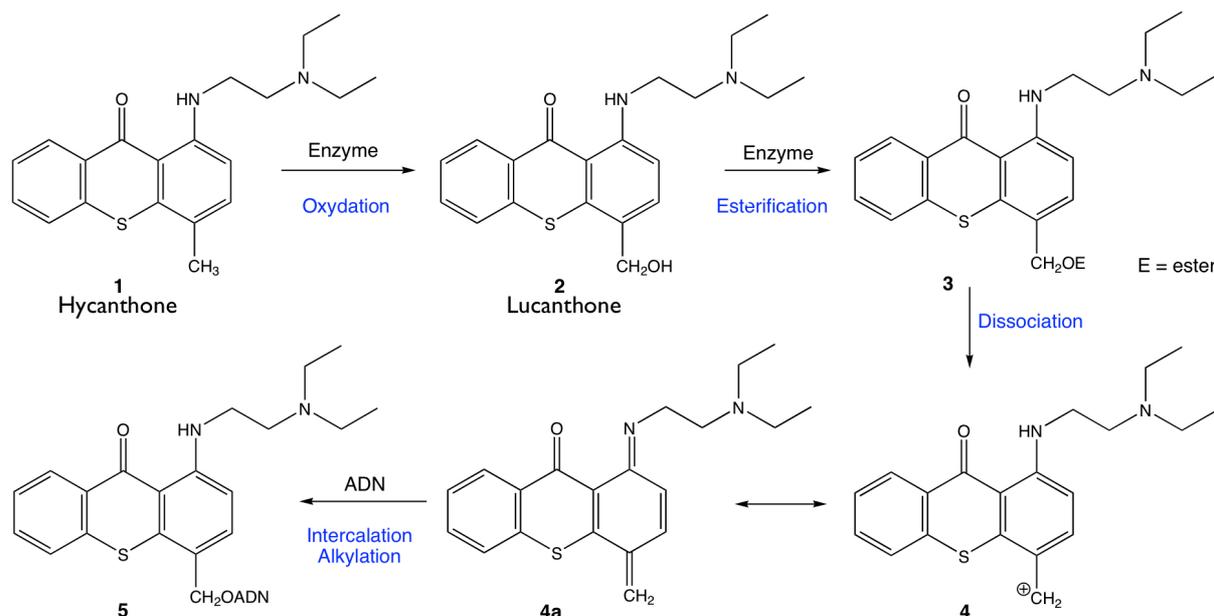
La résistance acquise à une molécule résulte de la perte de la sensibilité des parasites précédemment sensibles qui se transmet de manière génétique (Cioli *et al.* 1993). Elle est

consécutives à une ou plusieurs mutations sélectionnées par la molécule et transmises aux générations suivantes. La résistance innée est l'insensibilité des parasites à une molécule alors qu'ils n'ont jamais été exposés à celle-ci. Cette résistance se traduit par une différence de sensibilité de souches géographiquement éloignées soumises à la seule sélection naturelle.

Dans ces travaux de thèse, nous avons retenu les molécules hycanthane et lucanthane pour étudier la réponse adaptative du parasite *S. mansoni* à une drogue entraînant un phénomène de résistance dès la première génération.

En 1971, Rogers *et al.* relatent le premier cas de résistance à l'hycanthane obtenu en laboratoire (Rogers & Bueding 1971). Ces études présentent cette résistance comme une induction liée au traitement, sans sélection de mutations dans une population. Cette résistance est stable et peut perdurer sur plusieurs générations, en passant ainsi de 69% à 100% à travers 7 générations (Jansma *et al.* 1977).

Le lucanthane est une molécule active contre *S. mansoni* et *S. haematobium*. Le mode d'administration de ce médicament se fait uniquement par voie orale, puisqu'aucune activité n'est notée par injection. L'oxydation du lucanthane par *Aspergillus sclerotiorum* en hycanthane restaure l'activité. La biooxydation de l'hycanthane (Figure 17, **1**) donne le lucanthane (Figure 17, **2**) qui est converti par une enzyme (une kinase ou une sulfotransférase) en un ester (Figure 17, **3**) (E étant un groupement phosphate ou sulfate, respectivement). Dans tous les cas, l'ester (Figure 17, **3**) possède un groupement partant : une dissociation non enzymatique génère l'ion **4** (Figure 17) ou le composé **4a** (Figure 17) qui s'intercale et alkyle l'ADN pour donner le composé covalent **5** (Figure 17). Par la suite, un mécanisme similaire a été proposé pour l'oxamniquine (Archer 1985; Laurent 2006) (Tableau 2).



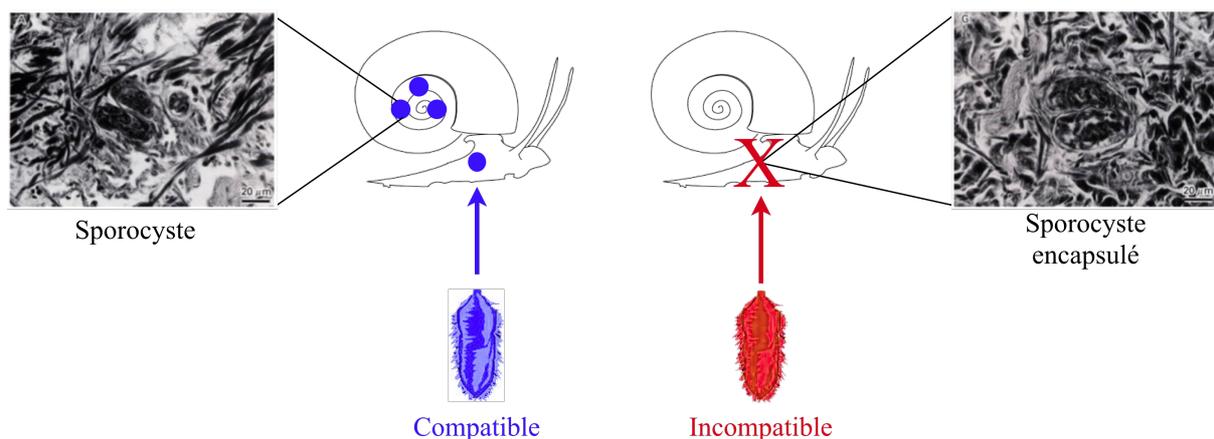
**Figure 17: Proposition d'un mécanisme d'action pour l'hycanthone et le lucanthone.** La biooxydation de l'hycanthone (1) donne le lucanthone (2) qui est converti par une enzyme en ester (3). L'ester (3) génère l'ion (4) ou le composé (4a) pour donner le composé covalent (5) (Cioli *et al.* 1985).

Pour être actif, l'hycanthone a besoin d'être métabolisé par des enzymes du schistosome (Cioli *et al.* 1985). La résistance serait due à l'absence d'activité enzymatique d'une enzyme de type estérase, qui serait active chez les parasites sensibles, mais inactive chez les individus résistants. Cette hypothèse a été validée en soumettant des vers adultes résistants à la forme ester de la molécule hycanthone, qui deviennent sensibles à la molécule estérifiée. Elle conduit à admettre que l'enzyme responsable de l'estérification de l'hycanthone est indispensable pour rendre la molécule active, son absence entraînant chez certains individus une résistance (Cioli *et al.* 1985). Ces résultats sont en accord avec les travaux sur la récessivité de ce trait. Lorsque des vers adultes résistants sont croisés avec des adultes sensibles, le trait « résistance » apparaît comme un trait récessif (Cioli & Pica-Mattoccia 1984). Tous ces travaux sont en adéquation avec les observations pharmacologiques et biochimiques (Cioli *et al.* 1985; Pica-Mattoccia *et al.* 1988; Archer *et al.* 1990). Des études plus récentes concluent que cette résistance est due à un seul gène récessif autosomal dans la mesure où les croisements (adultes sensibles x adultes résistants) ont montré que cette résistance n'est pas directement liée au sexe des parents (mâle ou femelle résistants) et que de ce fait ce trait est récessif et autosomal. Tous les croisements et *backcrosses* réalisés corrèlent avec les valeurs attendues dans la loi de la génétique mendélienne (Cioli *et al.* 1992).

Une question persiste toujours : Est ce que cette résistance résulte de l'apparition de nouveaux variants phénotypiques ou ces variants sont-ils déjà présents mais non exprimés en absence de traitement (phénotypes cryptiques)? Cette problématique constitue notre première énigme à résoudre au cours de cette thèse (**énigme 1 « Adaptation rapide de *Schistosoma mansoni* à une drogue anti-helminthe** »). Pour cela nous émettons l'hypothèse que cette résistance peut être liée à une hétérochromatisation, autour de certains gènes, entraînant leur inactivation et la capacité à résister à la drogue, après le traitement. Cette résistance pourrait être due à une épimutation.

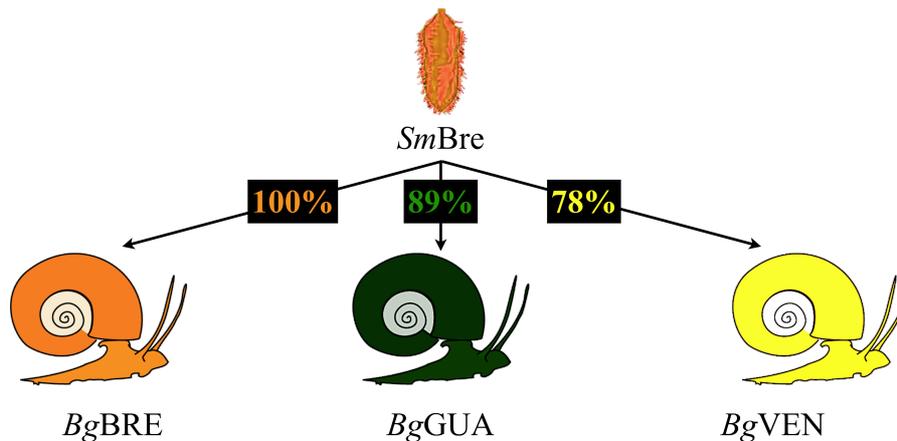
### 1.2.7. Phénomène de compatibilité

C'est au niveau de l'interaction avec l'hôte intermédiaire mollusque, que des études antérieures ont mis en avant une caractéristique particulière : le polymorphisme de compatibilité (Théron *et al.* 1997; Roger *et al.* 2008b; Roger *et al.* 2008c). Lors de cette interaction, les mécanismes d'infectivité du parasite et de défense de l'hôte sont à l'équilibre, ce qui conduit à la coexistence de phénotypes d'hôtes et de parasites compatibles et incompatibles. Le parasite soumis à un « environnement-hôte » va pouvoir exprimer différents phénotypes compatibles ou incompatibles vis-à-vis de cet environnement-hôte mollusque. Ainsi, certains individus parasites (compatibles) pourront se développer normalement dans l'hôte mollusque, alors que d'autres (incompatibles) seront immédiatement reconnus, encapsulés et tués par les cellules immunitaires du mollusque (Figure 18).



**Figure 18: Schéma du polymorphisme de compatibilité *S. mansoni*/*B. glabrata*.** Dans une combinaison compatible (bleue) le parasite se développe dans son hôte, et dans une combinaison incompatible (rouge) le parasite est reconnu, encapsulé et détruit pas les cellules immunitaires de l'hôte (Mitta *et al.* 2012).

Ce phénomène de compatibilité vis à vis d'une souche mollusque est observé pour plusieurs souches de parasite présentes au laboratoire. Certaines souches de parasites seront plus compatibles avec une souche allopatrique particulière. Ainsi, la souche de parasite *S. mansoni* issue du Brésil (*SmBre*) est 100% compatible avec son hôte sympatrique *B. glabrata* Brésil (*BgBRE*), alors qu'elle présente différents seuils d'incompatibilité avec d'autres souches allopatriques: 89% de *SmBre* infeste une souche de mollusque issue de Guadeloupe (*BgGUA*) alors que seuls 78% infeste une souche de mollusque issue du Venezuela (*BgVEN*) (Figure 19).



**Figure 19: Phénomène de compatibilité pour une souche de parasite *SmBre* (*S. mansoni* Brésil).** Cette souche est capable d'infester à 100% la souche de mollusque sympatrique *BgBRE* (*B. glabrata* Brésil), à 89% et 78% une souche allopatrique *BgGUA* (*B.a glabrata* Guadeloupe) et une souche allopatrique *BgVEN* (*B. glabrata* Venezuela), respectivement.

La compatibilité entre *S. mansoni* et *B. glabrata* repose sur l'expression par le parasite, d'épitopes qui pourraient être identifiés ou non par les molécules de reconnaissance immunitaire de l'hôte (Roger *et al.* 2008b; Roger *et al.* 2008c). Ce modèle implique qu'il existe chez les hôtes et les parasites, des répertoires polymorphes et/ou diversifiés d'épitopes et de molécules de reconnaissance, avec concordance ou non entre ces deux répertoires constituant la base fonctionnelle du polymorphisme de compatibilité (Moné *et al.* 2010).

Dans le cadre des thèses d'Emmanuel Roger et Yves Moné, au sein de notre laboratoire, des études comparatives entre souches ont permis de mettre en évidence des molécules potentiellement engagées dans cette compatibilité hôte/parasite. Les molécules principales sont des protéines polymorphes de type mucines, nommées *S. mansoni* Polymorphic Mucins (*SmPoMucs*) (Roger *et al.* 2008c) qui sont connues pour être impliquées dans les relations hôte-parasite à divers niveaux (pénétration dans l'hôte, évasion immunitaire). Ces mucines

spécifiques ne sont exprimées qu'aux stades interagissant avec le mollusque (miracidium et sporocystes) et sont localisées dans la glande apicale du parasite (un organe sécréteur). Ces protéines présentent un polymorphisme d'expression individuel, sous la forme d'une combinaison propre à chaque individu. Les *SmPoMucs* sont donc des candidats de choix pour être des acteurs du polymorphisme de compatibilité en tant qu'antigènes polymorphes susceptibles d'être identifiés par les molécules de reconnaissance du système immunitaire de l'hôte *B. glabrata*.

Des incubations de plasma de *B. glabrata* avec des antigènes de trématodes ont amené à la formation de complexes moléculaires (Adema *et al.* 1997; Hertel *et al.* 2005), dans lesquels les molécules de *B. glabrata* ont été caractérisées et nommées FREPs (*Fibrinogen Related Proteins*) (Adema *et al.* 1997). Les *SmPoMucs* peuvent interagir avec les ligands des FREPs (Roger *et al.* 2008b). Pour mieux comprendre ce phénomène, des expériences ont été menées pour identifier les protéines présentes dans le plasma de l'hôte au contact avec le parasite, ainsi que pour identifier, par co-immunoprécipitation, l'interaction FREPs/*SmPoMucs* (Moné *et al.* 2010).

L'analyse des complexes immunoprécipités suggère que les *processi* de reconnaissance entre *S. mansoni* et *B. glabrata* sont multifactoriels et impliquent divers récepteurs immunitaires de l'hôte et différentes protéines du parasite. D'autres études ont montré que chaque individu *S. mansoni* exprime un profil particulier de *SmPoMucs* (Roger *et al.* 2008b) qui peut être reconnu ou non par des profils particuliers de FREPs exprimés par les mollusques infestés (Moné *et al.* 2010). Suite à ces études, il a été émis comme hypothèse qu'un des *processi* de reconnaissance serait issu des combinaisons FREPs/*SmPoMucs* exprimés par les individus *S. mansoni* et *B. glabrata* au cours de leur interaction. Cette combinaison pourrait aboutir aux statuts compatibles ou incompatibles présentés précédemment (Moné *et al.* 2010).

Toutes ces expérimentations expliquent comment la souche de *S. mansoni* importée en Amérique du Sud a réussi à s'adapter, se développer et se maintenir. Elle a été confrontée à des environnements changeants de celui initial en Afrique, et à des souches de mollusques différentes. Pour survivre *S. mansoni* a été capable d'infester des hôtes intermédiaires mais comment a-t-il réussi à s'adapter? Cette problématique constitue notre deuxième énigme à résoudre au cours de cette thèse (**énigme 2 « Adaptation rapide de *Schistosoma mansoni* à l'Amérique du Sud »**).

### 1.3. Évolutions expérimentales sous pression de sélection : drogues anti-helminthes et hôte intermédiaire allopatrique

Dans le cadre de cette étude, différentes conditions biotiques (hôte sympatrique vs hôte allopatrique) et abiotiques (drogues anti-helminthes) ont été retenues afin d'évaluer leurs influences sur l'épigénome du parasite, confrontées à des traits d'histoire de vie et des profils de transcription globaux et/ou ciblés. Pour résoudre nos deux questions principales : « Adaptation rapide de *S. mansoni* à une drogue anti-helminthe » (énigme 1) et « Adaptation rapide de *S. mansoni* à l'Amérique du Sud » (énigme 2), le choix de la souche de parasite est primordial. Nous avons donc opté pour une souche de laboratoire issue du Brésil (*SmBre*) car elle présente un fond génétique très homogène sur 14 marqueurs microsatellites neutres.

Pour la première énigme à résoudre « Adaptation rapide de *S. mansoni* à une drogue anti-helminthe », nous avons choisi l'hycanthone et le lucanthone (drogues anciennement utilisées dans le contrôle de la bilharziose), car elles entraînent une forte résistance dès la première génération traitée et cette résistance est transmissible aux générations suivantes (Jansma *et al.* 1977). Des études antérieures ont démontré que cette résistance était due à un seul gène récessif autosomal (Cioli *et al.* 1992). Notre problématique de travail a donc été de déterminer si l'usage de cette drogue induit l'apparition de nouveaux variants phénotypiques résistants ou si ceux-ci préexistent dans la population initiale.

Ce modèle d'étude est à notre avantage puisque nous avons la capacité d'obtenir des populations clonales par la reproduction asexuée qui a lieu dans le mollusque.

En nous basant sur les recherches de Luria & Delbrück (Luria & Delbrück 1943), ainsi que sur les études de Lederberg & Lederberg (Lederberg & Lederberg 1952), nous avons imaginé une évolution expérimentale nous permettant de discriminer entre une induction de nouveaux variants dans une population ou une prédisposition (variants déjà présents dans la population) de certains individus, en utilisant la souche de parasite *S. mansoni* Brésil (*SmBre*), le mollusque sympatrique Brésil (*BgBRE*) et l'hôte définitif souris. En observant plusieurs populations clonales de *SmBre* après traitement à l'hycanthone ou au lucanthone, nous avons pu établir que les phénotypes différaient entre les clones et au sein de chaque clone, en comparaison aux lots témoins sans traitement. Nous avons également réalisé une étude épigénétique sur les vers adultes résistants par la technique de ChIP-seq, afin de comparer les profils chromatiniens entre des individus résistants et des individus non traités.

Pour notre deuxième énigme « Adaptation rapide de *S. mansoni* à l'Amérique du Sud », nous avons choisi une souche de parasite spécifique d'une souche de mollusque (souche de mollusque sympatrique) qui est tout de même capable d'infester une autre souche de mollusque (souche de mollusque allopatrique) avec une valeur sélective moindre et deux souches de laboratoire issues d'isolats géographiques différents. Ainsi, nous avons retenu le parasite *S. mansoni* Brésil (*SmBre*) pour son fond génétique stable, le mollusque sympatrique *B. glabrata* de Brésil (*BgBRE*), le mollusque allopatrique *B. glabrata* de Guadeloupe (*BgGUA*) provenant d'un autre isolat géographique et présentant des différences phénotypiques et génétiques du mollusque sympatrique.

Une population de parasite *SmBre* est confrontée à des mollusques sympatriques (*BgBRE*) ainsi qu'à des mollusques allopatriques venant d'un isolat autre géographique (*BgGUA*). Les cycles sont maintenus pendant trois générations successives. Durant ces 3 générations, plusieurs traits d'histoire de vie sont mesurés (sexe-ratio, infectivité, prévalence) ainsi que l'effet immunopathologique du parasite sur l'hôte définitif (mesure de la réaction inflammatoire de la rate). Toutes ces mesures ont pu être confrontées à des données épigénétiques (issues de CHIP-PCRq et CHIP-seq) et à des données transcriptomiques (RNA-seq et PCRq) afin de caractériser la réponse du parasite lorsqu'il est exposé à des environnements différents.

Les études de réponse au stress biotique (changement d'hôte) ont fait ressortir un trait de vie variable entre la condition sympatrique et la condition allopatrique : le sexe-ratio des cercaires suite au développement au sein de l'hôte allopatrique.

Ainsi, lorsque le parasite se multiplie dans un mollusque sympatrique, le sexe-ratio des cercaires émises est équilibré voire biaisé en faveur des mâles (Boissier & Moné 2000) alors que dans la condition allopatrique, un enrichissement en femelles est observé dans la population de cercaires émises lorsque le parasite a changé d'environnement. Ce sexe-ratio biaisé en faveur des femelles est retrouvé au sein de l'hôte définitif mais ensuite à la reproduction sexuée au sein de cet hôte vertébré, les sexes-ratios des miracidia excrétés sont de nouveau équilibrés voire biaisés en faveur des mâles.

Une problématique a ainsi émergé de ces expériences et a mené à la troisième question scientifique: Comment le sexe est-il déterminé chez *S. mansoni*? (**énigme 3 « Différenciation sexuelle sans présence de gène W-spécifiques dans un système ZW »**).

Nous savons que le parasite *S. mansoni* possède des chromosomes sexuels et fait partie du système ZZ/ZW avec une femelle hétérogamétique. Il présente également, tout au long de son cycle de vie, des sexes séparés mais ce dimorphisme sexuel n'apparaît qu'au stade adulte (aux stades larvaires, miracidia et cercaires, aucun dimorphisme sexuel n'est observable). De ce fait, lors de l'étude de l'impact de l'environnement-hôte intermédiaire sur le parasite, nous avons décidé d'étudier les bases moléculaires du déterminisme sexuel chez *S. mansoni*, afin de proposer un modèle pour l'émergence des chromosomes sexuels dans les systèmes ZW.

Toutes ces évolutions expérimentales nous ont permis de quantifier et qualifier certains mécanismes moléculaires impliqués dans la réponse adaptative de *S. mansoni* tant à un stade larvaire (cercaire) qu'au stade adulte et sous plusieurs influences environnementales (traitement anti-helminthes, hôte intermédiaire allopatrique), afin de donner des éléments de réponse à nos trois énigmes : **(1)** « Adaptation rapide de *S. mansoni* à une drogue anti-helminthes », **(2)** « Adaptation rapide de *S. mansoni* à l'Amérique du Sud », **(3)** « Différenciation sexuelle en l'absence de gène W-spécifiques dans un système ZW »

## 2. Objectifs de la recherche doctorale

Cette recherche doctorale - dont l'enjeu rappelons le est primordial dans la lutte contre la bilharziose intestinale qui cause plus de 200 000 décès par an - a permis d'évaluer la réponse adaptative du parasite *S. mansoni* lorsqu'il est confronté à des environnements différents.

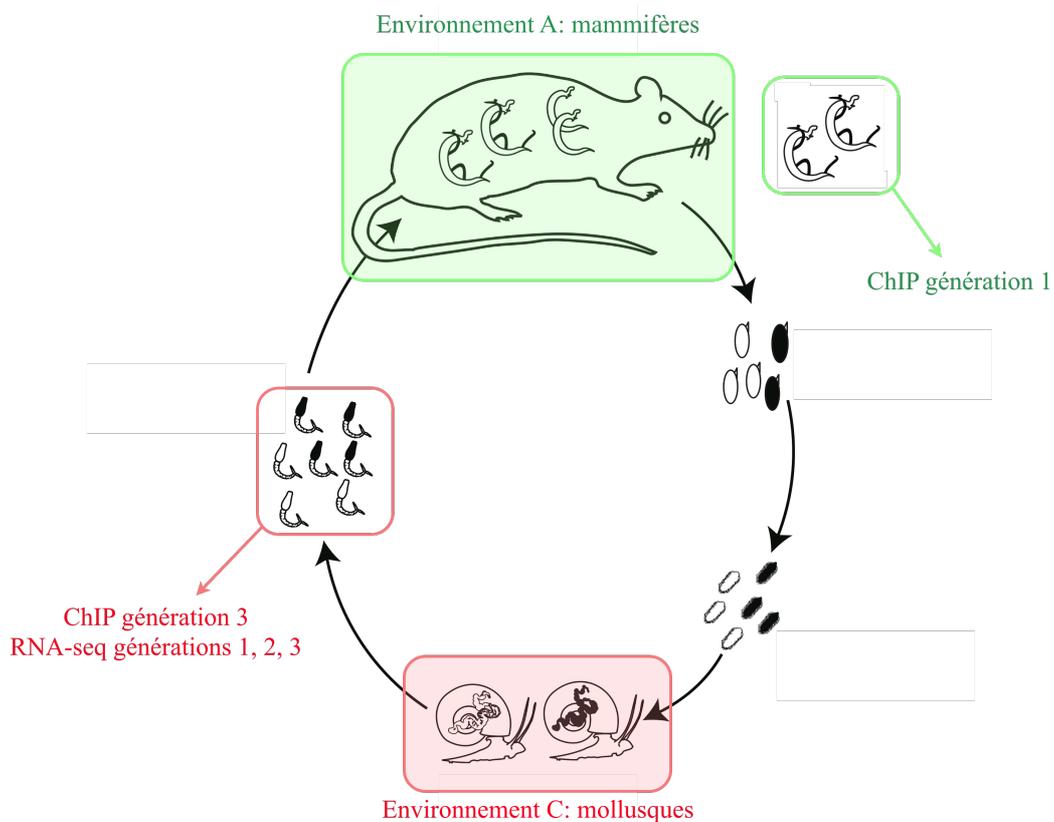
Pour la première fois depuis l'apparition des techniques de séquençage massif, la réponse adaptative d'un parasite humain a pu être quantifiée, qualifiée et analysée par des études de transcriptomique et épigénomique globales. Nous avons ainsi généré les premiers transcriptomes sur trois générations successives et le premier épigénome d'un parasite métazoaire.

Grâce à ces avancées techniques, nous pouvons maintenant raisonner à l'échelle globale du génome, et non par expériences ciblées gènes par gènes. Par ailleurs, toutes ces données moléculaires ont pu être confrontées à des données concrètes de traits de vie, accessibles tout au long du cycle de développement du parasite dans différents environnements. Enfin, nous avons pu proposer un modèle sur l'émergence des chromosomes sexuels chez les espèces ZW (Lepesant *et al.* 2012a) (Annexe 2), lorsque nous avons étudié les bases moléculaires du

déterminisme sexuel chez cette espèce présentant une sexualité particulière, pour résoudre l'énigme « Adaptation rapide de *S. mansoni* à l'Amérique du Sud ».

Pour les expériences de traitements anti-helminthes (Figure 20 A), seules les analyses épigénétiques de la première génération de vers adultes résistants ont pu être menées.

Lors des expériences allopatriques (Figure 20 C) : le séquençage et l'analyse des 3 générations au niveau transcriptomique ont pu être réalisés mais pour l'étude de la structure chromatinienne, seule la 3<sup>e</sup> génération a pu être séquencée et analysée.



**Figure 20: Synthèse des analyses réalisées au cours de cette thèse.** Différentes conditions environnementales ont été choisies (environnement A : l'hôte définitif traité ou non à l'hycanthone/lucanthone, environnement C : l'hôte intermédiaire sympatrique ou allopatrique). Dans le cadre des expériences avec le traitement anti-helminthe, les analyses de ChIP-seq ont porté sur la première génération d'individus résistants et sensibles. Dans le cadre des expériences hôte sympatrique vs allopatrique, les analyses de ChIP-seq ont été menées sur la troisième génération alors que les analyses RNA-seq sont mises en œuvre sur les trois générations pour les deux conditions.

## **CHAPITRE 2**

### **Matériel et Méthodes**

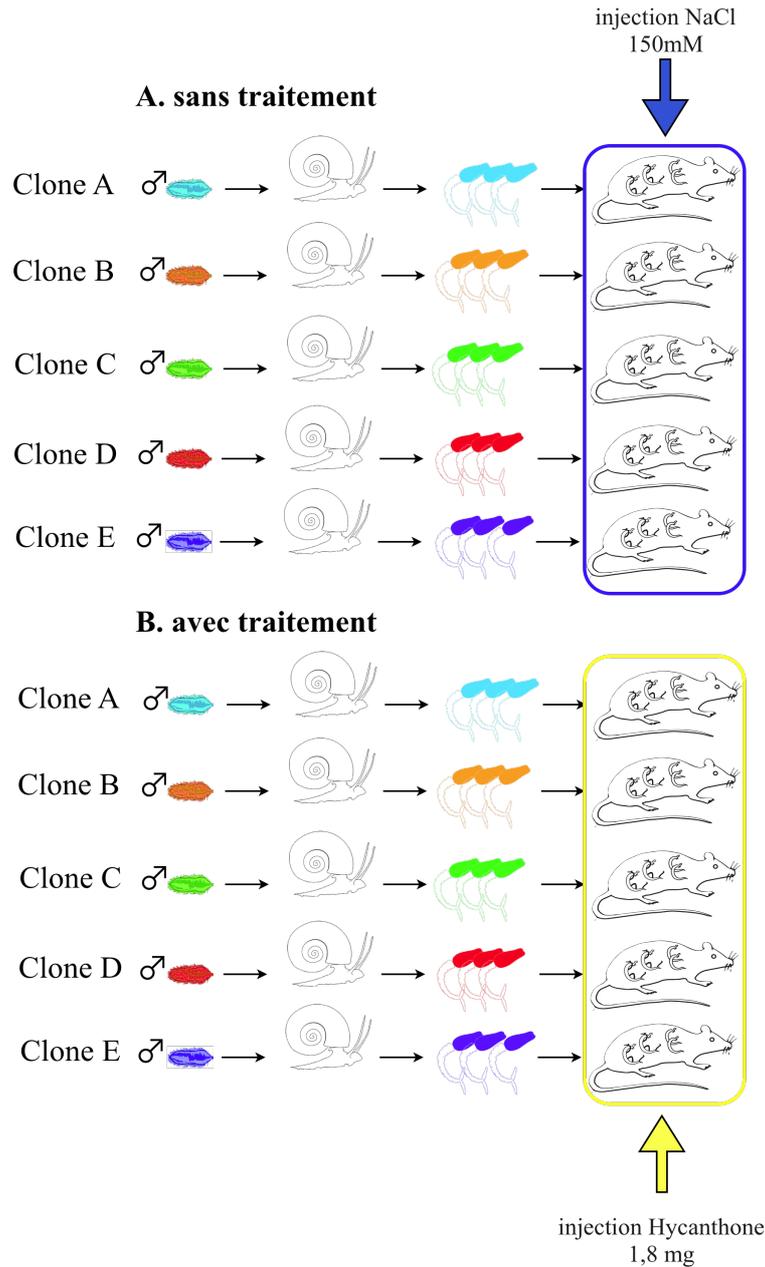
## 1. Matériel biologique

### 1.1. Énigme 1 « Adaptation rapide de *S. mansoni* à une drogue anti-helminthe »

Pour répondre à cette interrogation, nous avons opté pour la souche de laboratoire *S. mansoni* issue d'un isolat brésilien (*SmBre* pour *S. mansoni* Brésil), car elle présente un fond génétique homogène sur 14 marqueurs microsatellites neutres (Bech *et al.* 2010). Cette souche de parasite a été maintenue par des passages successifs à travers son hôte intermédiaire spécifique, *B. glabrata* Brésil (*BgBRE* pour *B. glabrata* Brésil), et un hôte définitif murin, *Mus musculus*.

Pour le traitement avec l'hycanthone, nous avons préalablement procédé à des infestations monomiracidiales (un miracidium par mollusque) afin d'obtenir des populations clonales de cercaires. En effet, un seul miracidium est capable de se développer au sein de l'hôte, se métamorphoser en sporocyste primaire puis secondaire et se multiplier de manière asexuée en une population de cercaires clonales unisexuées (Boissier *et al.* 2003b).

Suite à ces infestations, le sexe de chaque clone a été testé par analyse de marqueurs spécifiques (technique expliquée ci-après) afin de produire cinq populations clonales de cercaires de sexe mâle. Les infestations de souris ont été mises en œuvre de façon contrôlée, en les anesthésiant une heure, puis en leur rasant le ventre pour que les cercaires (140) pénètrent plus facilement. Pour chaque clone, dix souris ont été infestées. Après 62 jours post-infestation, cinq ont été traitées à l'hycanthone en une seule injection intramusculaire de 60 mg/kg (environ 1,8 mg/individu pour un poids moyen de 30g) et les cinq autres ont reçu un traitement contrôle (NaCl 150 mM) (Figure 21). Ensuite, les souris ont toutes été euthanasiées à 57 jours post-traitement (cf. les conditions utilisées dans les études de (Jansma *et al.* 1974; Jansma *et al.* 1977)).



**Figure 21: Protocole expérimental utilisé pour les traitements à l’hycanthon.** Suite à des infestations monomiracidiales, cinq clones mâles ont été choisis pour infester deux lots de souris : un lot non traité et un lot traité avec la molécule hycanthon.

Pour le traitement au lucanthone, nous avons préalablement réalisé des infestations monomiracidiales afin d’obtenir des populations clonales de cercaires. Ensuite, le sexe de chaque clone a été testé par analyse de marqueurs spécifiques (technique expliquée également par la suite), afin de sélectionner cinq populations clonales de cercaires de sexe mâle et cinq populations clonales de cercaires femelles. Les infestations des souris se sont faites de façon contrôlée avec 70 cercaires mâles et 70 cercaires femelles, pour former des couples de clones. Dix souris ont été infestées avec un couple de clones. À 62 jours post-infestation, cinq ont été

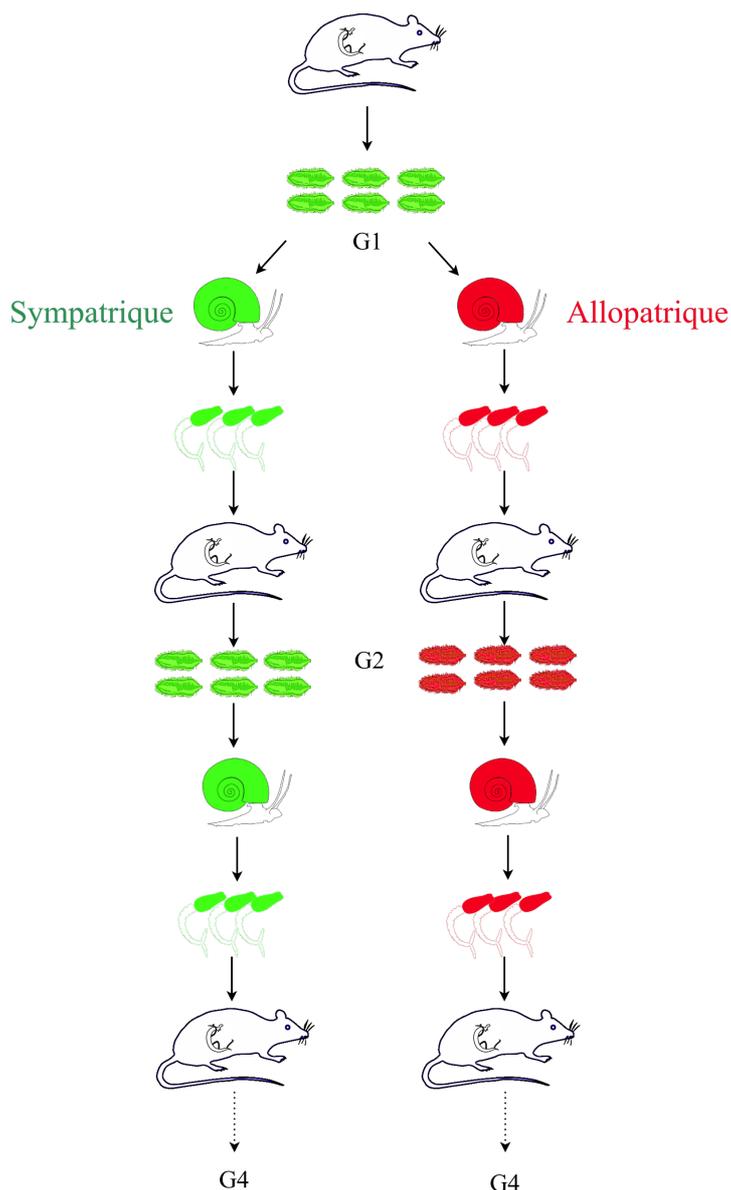
traitées au lucanthone par ingestion journalière de 20 mg/kg (soit 0,3 mg/individu de poids moyen 30g) pendant trois jours consécutifs et les cinq autres souris, ont été traitées avec un traitement contrôle NaCl (150 mM) (cf. les conditions utilisées dans les études de (Jordan & Randall 1962)). L'euthanasie des souris est intervenue 13 jours post-traitement en raison d'une forte mortalité, ce qui n'influence pas les résultats puisque le temps de demie-vie du lucanthone est de 24 heures (Blair 1958). L'évolution expérimentale est sensiblement la même que pour le traitement à l'hycanthone (Figure 21), à la différence que les souris sont infestées avec des couples de clones (deux sexes).

Pour chaque expérience, les vers adultes sont récupérés par perfusion de la veine porte, dans une solution dosée à 0.8 % (w/v) NaCl et à 0.8 % (w/v) Trisodium Citrate (Théron *et al.* 1997). Ensuite, une vérification du système digestif est effectuée afin de trouver les adultes non emportés par la perfusion. Puis les adultes sont dénombrés et stockés à sec à -80°C, en prévision des analyses de biologie moléculaire.

### 1.2. Énigme 2 « Adaptation rapide de *S. mansoni* à l'Amérique du Sud »

Pour répondre à cette question, nous avons utilisé la souche parasite de laboratoire *S. mansoni* issue d'un isolat brésilien (*SmBre* pour *S. mansoni* Brésil) en raison de son fond génétique homogène. Pour les hôtes intermédiaires, nous avons opté pour deux souches de mollusques *B. glabrata*, la première provenant du même isolat géographique brésilien que la souche de parasite choisie qui représente la condition sympatrique (*BgBRE* pour *B. glabrata* Brésil), la deuxième provenant d'un isolat géographique différent, Guadeloupe (*BgGUA* pour *B. glabrata* Guadeloupe) qui représente la condition allopatrique.

Dans la combinaison *SmBre/BgBRE* (sympatrique), le parasite *SmBre* est capable d'infester, de se développer et de se maintenir dans *BgBRE* alors que dans l'association *SmBre/BgGUA* (allopatrique), un certain nombre de parasites *SmBre* sont reconnus, encapsulés et détruits par les cellules immunitaires de l'hôte *BgGUA*. Seule une proportion d'individus *SmBre* arrive à infester, se développer et se maintenir dans la souche allopatrique. Chaque condition a été perdurée sur trois générations successives en conservant toujours la même souche d'hôte intermédiaire (Figure 22) et en passant par un hôte définitif murin femelle pour normaliser tous les cycles, le sexe de l'hôte définitif pouvant influencer la réponse inflammatoire liée à l'infestation par *S. mansoni* (Boissier *et al.* 2003a).



**Figure 22: Protocole expérimental de changements d’hôte intermédiaire.** Une même population initiale de miracidia, issue de la souche *S. mansoni* Brésil (*S<sub>m</sub>Bre* en vert), a servi à infester une population sympatrique de mollusque de la souche *B. glabrata* Brésil (*B<sub>g</sub>BRE* en vert) et une population allopatrique de mollusques de la souche *B. glabrata* Guadeloupe (*B<sub>g</sub>GUA* en rouge). Ces cycles sont maintenus dans les hôtes intermédiaires définis sur trois générations successives.

Toutes les modalités précises d’infestation, de maintien de cycle et de récupération des échantillons sont décrites dans la publication en préparation (Lepesant *et al.* in prep.) (Annexe 1).

### 1.3. Énigme 3 « Différenciation sexuelle en l'absence de gènes W-spécifique dans un système ZW »

Pour résoudre cette interrogation, les différents stades de développement de *S. mansoni* (miracidia, cercaires, adultes) ont été collectés selon les protocoles établis (Théron *et al.* 1997; Boissier *et al.* 2003b). Des infestations monomiracidiales ont été faites comme précédemment, afin d'avoir des populations cercariennes unisexes. Tous les détails d'échantillonnage et maintien de cycle pour cette étude sont décrits dans la publication ci-jointe (Lepesant *et al.* 2012a) (Annexe 2).

## 2. Analyses statistiques et de génétique des populations

Les analyses statistiques et de génétique des populations ont été réalisées, en collaboration avec J. Boissier, sous les logiciels SPSS 18.0 et Fstat 2.9, respectivement, pour les résultats obtenus dans les expériences de changement d'hôte (sympatrique *vs* allopatrique).

Afin d'étudier les variances, un test paramétrique de Levene a été utilisé. En revanche, pour comparer deux populations ou plus de deux populations indépendantes, nous nous sommes servis des tests non-paramétriques de Mann-Whitney et Kruskal-Wallis, respectivement.

En ce qui concerne les analyses de génétique des populations, nous avons vérifié que les populations sont à l'équilibre d'Hardy-Weinberg, et que les *loci* ne sont pas en déséquilibre de liaison. La différenciation génétique entre deux populations a été estimée avec les indices de Wright ( $F_{st}$ ). La diversité génétique a été mesurée par le pourcentage d'hétérozygotes ( $H_e$ ) et la richesse allélique ( $A_n$ ).

Pour les tests statistiques et de génétique des populations nous avons choisi un seuil de significativité  $p > 0.05$ .

## 3. Traits d'histoire de vie

Les différents traits d'histoire de vie mesurables dans un cycle de développement de *S. mansoni* sont : les sexes-ratios à tous les stades, l'infectivité des cercaires, la prévalence des mollusques infestés, l'intensité de l'infestation, l'émission cercarienne et sa chronobiologie.

Lors des travaux constituant cette thèse, nous avons quantifié les principaux de traits d'histoire de vie, notamment les sexes-ratios à tous les stades, l'infectivité des cercaires et la prévalence de mollusques infestés. Nous avons également évalué l'effet du parasite sur la

réponse inflammatoire de l'hôte vertébré en calculant le ratio du poids de la rate sur le poids total de la souris.

Le sexe-ratio au stade adulte, se fait par le dénombrement des adultes développés au sein de l'hôte murin, alors que le sexe-ratio des deux populations larvaires (miracidia et cercaires) est estimé par une analyse par PCR utilisant un marqueur sexe spécifique et un marqueur autosomal (Portela *et al.* 2010; Lepesant *et al.* 2012a). Le sexage des populations larvaires est expliqué plus en détail ci-après.

L'infectivité des cercaires est quantifiée en calculant le ratio du nombre d'adultes développés au sein de l'hôte vertébré sur le nombre de cercaires mises à son contact dans le cadre de l'infestation contrôlée.

La prévalence des mollusques infestés est évaluée en observant au travers des coquilles, les sporocystes développés au niveau de l'hépatopancréas, afin de déterminer le pourcentage de mollusques parasités par rapport au nombre total de mollusques exposés au parasite. Cette proportion est vérifiée par l'observation de l'émission cercarienne.

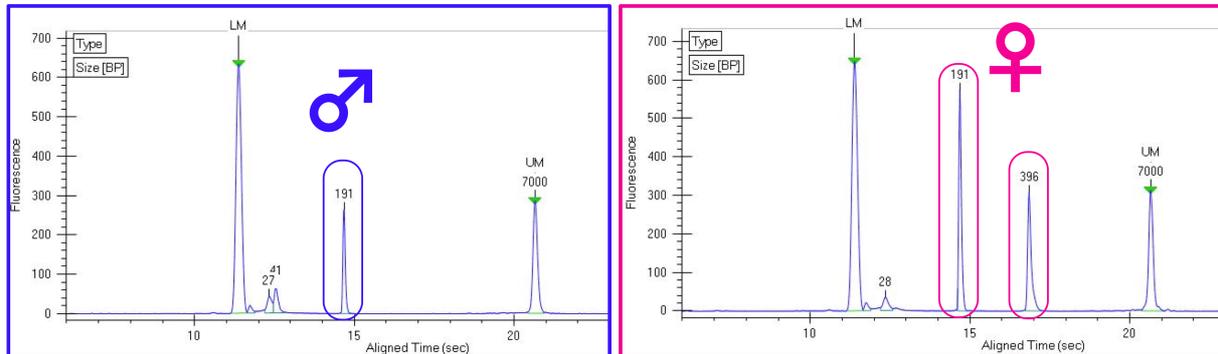
Au stade adulte, les effets immunopathologiques du parasite sur la souris sont établis en mesurant sa réaction inflammatoire qui est quantifiable en calculant le ratio du poids de la rate sur le poids total corporel. Ainsi, une souris saine a une rate d'environ 0,40% de son poids total (Boissier, communication personnelle) alors qu'une souris malade de la bilharziose a une rate équivalant à environ 2% de son poids total, voire plus.

#### 4. Techniques de biologie moléculaire et cellulaire

##### 4.1. Sexage et génotypage des stades larvaires

Le sexage est réalisé sur 22 à 34 individus extraits individuellement par une technique d'extraction d'ADN rapide mise au point au laboratoire (Beltran *et al.* 2008). Les amplifications PCR sont ensuite effectuées avec un marqueur femelle spécifique et un marqueur autosomal présent chez le mâle et la femelle. Ces PCRs sont réalisées en multiplex et sont ensuite analysées sur gel ou séparées par capillaires (LabChip) donnant des électrophérogrammes ou des profils de bandes précisant les tailles des produits amplifiés pour discriminer les individus mâles des individus femelles. Ainsi, les mâles présentent un profil pic unique (gène autosomal commun mâle/femelle) alors que les femelles présentent un profil

avec deux pics, un pic correspondant à la séquence femelle spécifique alors que l'autre correspond au marqueur autosomal (Lepesant *et al.* 2012a) (Figure 23) .



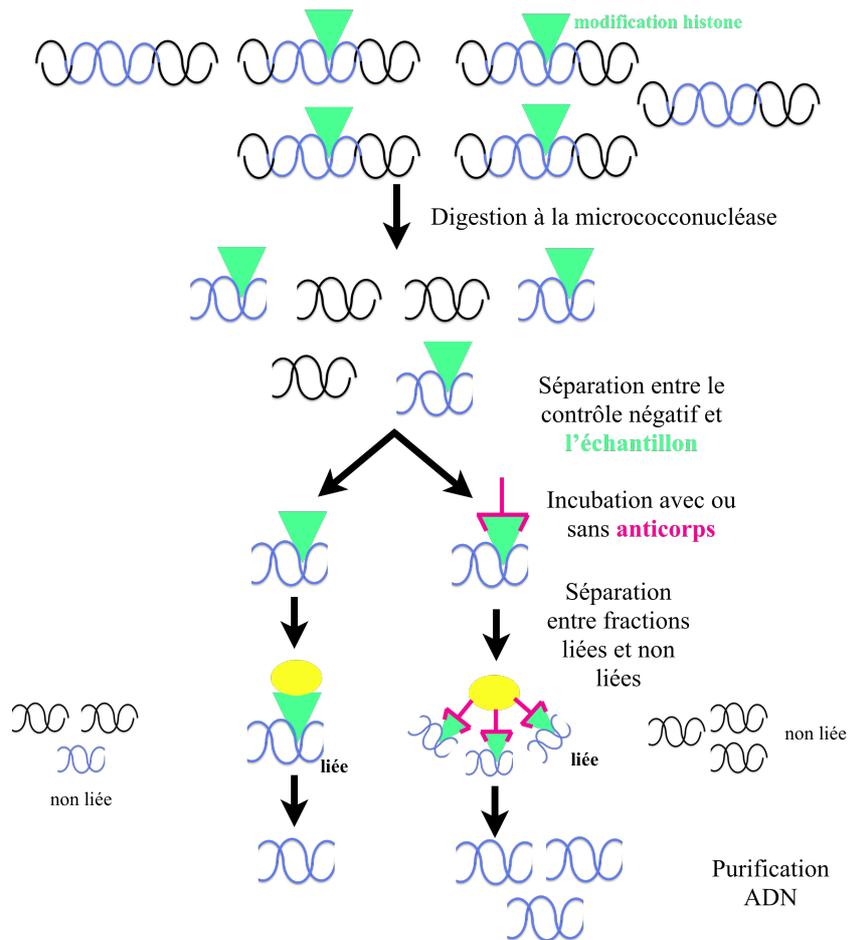
**Figure 23: Électrophérogrammes pour un génotype mâle et un génotype femelle.** Les individus mâles (bleu) se caractérisent par un profil avec un seul pic (191pb), correspondant au marqueur autosomal des mâles et des femelles. Les individus femelles (rose) présentent des profils avec deux pics (191 pb et 396 pb) correspondant respectivement au marqueur autosomal et au marqueur femelle spécifique.

Le génotypage est réalisé à partir de la même extraction d'ADN que celle prévue pour le sexage. Quatorze marqueurs microsatellites neutres sont utilisés : SMC1, SMDO11, Smbr16, R95529, SMD57, SMDA28, SmBr1, SMD28, SCMSMOXII, SMS7-1, Smbr10, Smbr13, SMD25, L46951 (Durand *et al.* 2000; Blair *et al.* 2001; Curtis *et al.* 2001; Rodrigues *et al.* 2002; Rodrigues *et al.* 2007). Les fragments amplifiés sont ensuite analysés grâce à un séquenceur automatique Beckman Coulter (Bech *et al.* 2010).

#### 4.2. ImmunoPrécipitation de la Chromatine (ChIP)

Le protocole d'immunoprécipitation de la chromatine appliqué à *S. mansoni* a été au laboratoire (Cosseau *et al.* 2009) et est disponible sur le serveur MethDB (<http://methdb.univ-perp.fr/epievo/>).

La technique d'immunoprécipitation de la chromatine (ChIP) (Figure 24) est une technique de choix à privilégier pour étudier la relation entre la modification des histones et l'expression de certains gènes, car elle repose sur les interactions antigènes-anticorps.



**Figure 24: Principe d'immunoprécipitation de la chromatine.** Une première digestion à la nucléase micrococciale permet de digérer entre les nucléosomes (en noir ADN sans modification des histones, en bleu ADN présentant une modification d'histone schématisée par le triangle vert). Une séparation entre le contrôle négatif et l'échantillon est ensuite réalisée. Une incubation avec ou sans anticorps spécifique de marque d'histone modifiée (triangle vert) a lieu, et une séparation sur billes de sépharose (jaune) permet la séparation entre les fractions liées et les fractions non liées. Une dernière étape d'extraction d'ADN a lieu avant d'envoyer les échantillons au séquençage massif ou pour l'analyse par PCRq.

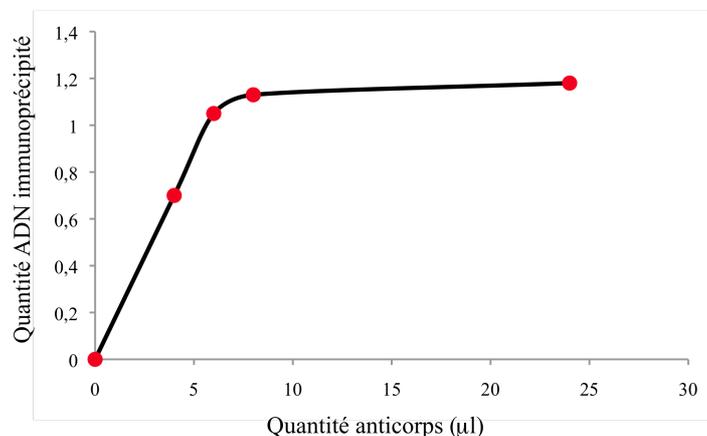
Les modifications des histones sont des marques épigénétiques importantes qui influencent la structure de la chromatine et jouent un rôle dans la régulation de la transcription. La chromatine est présente sous deux formes :

- l'euchromatine qui est à l'état décondensé. Elle est despiralisée et correspond à des zones de gènes actifs, pouvant être transcrits;
- l'hétérochromatine qui est à l'état condensé. Ses histones sont hypoacétylées et hyperméthylées par rapport à l'euchromatine. Elle correspond à la partie de la chromatine inactive au niveau transcriptionnel.

La spécificité et la force de l'interaction antigène-anticorps déterminent la qualité de l'immunoprécipitation. Les histones sont des protéines hautement conservées donc beaucoup d'anticorps commerciaux sont disponibles.

La spécificité et l'efficacité des anticorps doivent préalablement être testées. Pour avoir une bonne indication de la spécificité, des expériences de Western Blot sont réalisées pour vérifier que les anticorps ne vont pas interagir avec d'autres protéines que celles d'intérêt. Si des bandes apparaissent à différentes tailles de celles attendues, d'autres anticorps pour la même cible sont alors testés. Idéalement une unique bande de la bonne taille est détectée, mais si elle n'est pas visible en Western Blot, l'anticorps sera tout de même testé en immunoprécipitation.

Quand l'anticorps est validé par Western Blot, le ratio anticorps-chromatine est déterminé par titration. Pour une certaine quantité de chromatine, l'anticorps doit être en excès afin que la quantité de produits immunoprécipités ne dépende pas de la quantité d'anticorps utilisée. La titration est réalisée en utilisant une quantité croissante d'anticorps pour une même quantité de chromatine. A partir d'une certaine quantité d'anticorps, la quantité d'ADN immunoprécipité reste constante (Figure 25).



**Figure 25: Quantité d'ADN immunoprécipité en fonction d'une quantité croissante d'anticorps anti H3K4me3.** Cette titration nous permet de déterminer le seuil de saturation où la quantité d'ADN ne dépend pas de la quantité d'anticorps, nous permettant de déterminer la quantité d'anticorps à utiliser.

Pour la suite des expériences, la quantité d'anticorps choisie est supérieure au seuil de saturation. Si aucun seuil n'est déterminé lors de la titration, l'anticorps n'est plus utilisé (Tableau 3). Cette titration est coûteuse en anticorps et en matériel biologique, mais une fois ces validations réalisées la reproductibilité de l'immunoprécipitation est assurée.

**Tableau 3: Anticorps testés pour les analyses d'immunoprécipitation de la chromatine.** *Le nom des fournisseurs, numéro de lots, cibles, structure de la chromatine associée et quantités d'anticorps utilisées.*

Fournisseurs	# lot	Cibles	Quantités d'anticorps	Etats chromatinien
Upstate	DAM16924924	H3K9ac	8 µl	Euchromatine/Transcription active
Upstate	NG1680351	H3K4me3	4 µl	Euchromatine
Abcam	733951	H3K9me3	4 µl	Hétérochromatine
Diagenode	A29900242	H3K27me3	8 µl	Hétérochromatine
Abcam	786671	H3	8 µl	Histone H3

Dans le cadre de cette thèse, deux marqueurs hétérochromatiques (H3K9me3 et H3K27me3) et deux marqueurs euchromatiques (H3K9ac et H3K4me3) ont été utilisés, ainsi qu'un anticorps anti histone H3 pour la normalisation des données mais celui-ci n'a pas précipité assez de matériel (Cosseau, communication personnelle), d'où une absence de résultats suite au séquençage massif de ces échantillons.

La technique d'immunoprécipitation de la chromatine peut être associée à du séquençage massif type Illumina (ChIP-seq) (Cosseau *et al.* 2009) ou à de la PCR quantitative (ChIP-qPCR) (Caby & Pierce 2009). Cette technique de ChIP-seq a été appliquée sur les trois générations des deux conditions, mais le séquençage massif n'a pu être mené que sur la troisième génération de la condition sympatrique et allopatrique.

### 4.3. Extractions des ARNs totaux

#### 4.3.1. Pour les analyses de transcriptomique par RNA-seq

En vue du séquençage massif de l'ARN des cercaires (RNA-seq), pour l'analyse du transcriptome, un protocole d'extraction d'ARN total de cercaires a été mis en place après l'avoir adapté à partir de méthodes classiques telles que les extractions Trizol-Chloroforme couramment utilisées, et de procédés issus de kits d'extraction d'ARN, comme les kits QIAGEN (Rneasy® mini kit).

La mise au point de ce protocole a été une étape cruciale car nous devons fournir, au Centre de séquençage avec lequel nous collaborions (Genomix Montpellier, MGX), une quantité minimale de bonne qualité d'ARN total malgré le peu du matériel biologique dont nous disposions.

Nous avons donc optimisé toutes les étapes, du broyage des individus, en passant par les premières étapes d'extraction et d'élimination de contaminants génomiques, à la purification

finale. Ce protocole est disponible sur le serveur MethDB (<http://methdb.univ-perp.fr/epievo/>) pour une quantité maximale de 5000 cercaires à extraire.

#### 4.3.2. Pour les quantifications relatives ciblées par PCRq

L'extraction d'ARN total, en vue de la quantification de la transcription de ces séquences répétées femelles spécifiques par PCR quantitative, est mise en œuvre selon un protocole classique d'extraction Trizol-Chloroforme. Une élimination des contaminants d'ADN génomique est obligatoire en utilisant une enzyme DNase. Les ARNs totaux, une fois extraits et purifiés, sont reverse-transcrits selon le protocole Superscript II RT de chez Invitrogen, puis les ADNc complémentaires (ADNc) sont purifiés sur colonne.

La quantification relative de la transcription est réalisée par PCR quantitative pour chaque séquence répétée identifiée spécifique femelle, à chaque stade de développement du parasite (miracidia, cercaires, adultes) par rapport à un gène de référence 28S (Annexe 2).

#### 4.4. FISH (*Fluorescent In Situ Hybridization*)

Pour localiser les différentes séquences répétitives présentes sur le chromosome W, nous avons travaillé sur des métaphases issues de sporocystes tardifs (Hirai & LoVerde 1995). Pour chaque séquence répétée, des sondes spécifiques sont dessinées et validées, puis hybridées sur les cellules en métaphase (Hirai & LoVerde 1995). Les chromosomes sont ensuite colorés à l'iodure de propidium et observés avec un microscope à épifluorescence. Pour chaque séquence répétitive, nous avons étudié entre 7 et 34 métaphases femelles. Le protocole est détaillé dans la publication ci-jointe (Lepesant *et al.* 2012a) (Annexe 2).

### 5. Techniques d'analyses bioinformatiques

#### 5.1. Epigénome global : ChIP-seq

Pour les expériences de ChIP-seq, la technique de séquençage à haut débit Illumina est retenue. Le séquençage des différents échantillons a été réalisé avec la plateforme de séquençage génomique Montpellier GenomiX (<http://www.mgx.cnrs.fr/>) ou de l'Université d'Oregon (<http://www.cgrb.oregonstate.edu/faculty/freitag/>). Le séquençage des échantillons d'immunoprécipitation de chromatine est réalisé en condition *single-end* et 50 paires de bases (50 pb). Le séquençage porte sur 4 conditions avec 4 anticorps spécifiques : H3K9Me3,

H3K27Me3, H3K9ac, H3K4me3 au stade adulte pour l'énigme 1 et au stade cercaire pour l'énigme 2.

Le séquençage commence par la préparation d'une banque comprenant cinq étapes : (1) réparations des extrémités des fragments d'ADN ou d'ARN, (2) ajout d'une base A aux extrémités, (3) ligation des adaptateurs aux deux extrémités, (4) purification sur gel pour une sélection de taille (400 pb +/- 40 dans notre cas), (5) amplification par PCR permettant aussi une sélection puisque seuls les fragments portant les deux adaptateurs seront amplifiés.

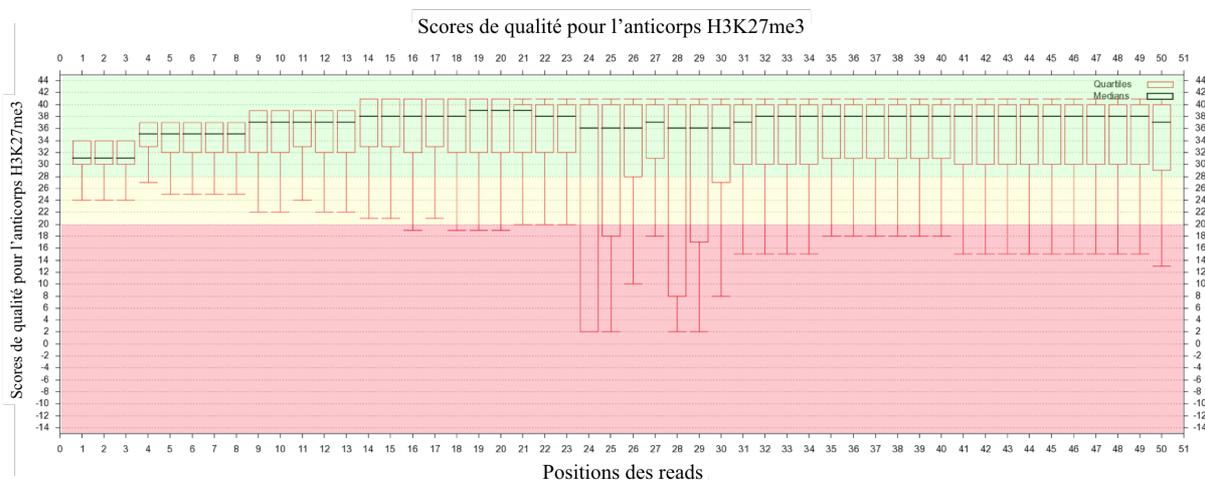
La validation de la banque se fait par une quantification (concentration et taille des fragments) de l'ADN sur une puce Agilent DNA1000.

Suite à la construction et la validation de la banque, une génération de clusters et l'hybridation d'amorces de séquençage sont effectuées. Les fragments d'ADN sont hybridés à une *flow-cell* qui est une lame en verre sur laquelle sont fixées de façon covalente des séquences d'ADN complémentaires à celles des adaptateurs. Chaque *flow-cell* comprend 8 pistes pour le traitement de 8 banques en même temps, de façon indépendante. Une fois les fragments d'ADN fixés à la *flow-cell*, une amplification isotherme est réalisée puis rendue possible par la formation de ponts ADN et la génération de populations clonale de fragments d'ADN (clusters). Les ponts sont ensuite linéarisés et l'extrémité libre des fragments d'ADN est fixée à la *flow-cell* pour l'hybridation de l'amorce de séquençage sur l'extrémité libre des fragments d'ADN fixés à la *flow-cell*.

Le séquençage est réalisé sur un *HiSeq 2000* d'Illumina grâce à la technique SBS (*Sequence By Synthesis*). Cette dernière consiste en l'incorporation et la détection séquentielle de nucléotides par polymérisation en se servant des clusters d'ADN fixés à la *flow-cell* comme matrice. Chaque type de nucléotide est marqué par un fluorochrome (émettant après excitation à différentes longueurs d'ondes) et par un agent bloquant qui empêche l'incorporation de plusieurs nucléotides lors d'un même cycle de séquençage. A la fin de chaque cycle, un tampon de clivage est injecté dans chaque piste, permettant la coupure du fluorochrome et de l'agent bloquant pour chaque nucléotide incorporé. Cette étape sert à incorporer un nouveau nucléotide lors du cycle suivant. Un cycle de séquençage peut être découpé en plusieurs étapes répétées de façon cyclique : lavage, incorporation d'un nucléotide, lavage, détection, lavage, clivage.

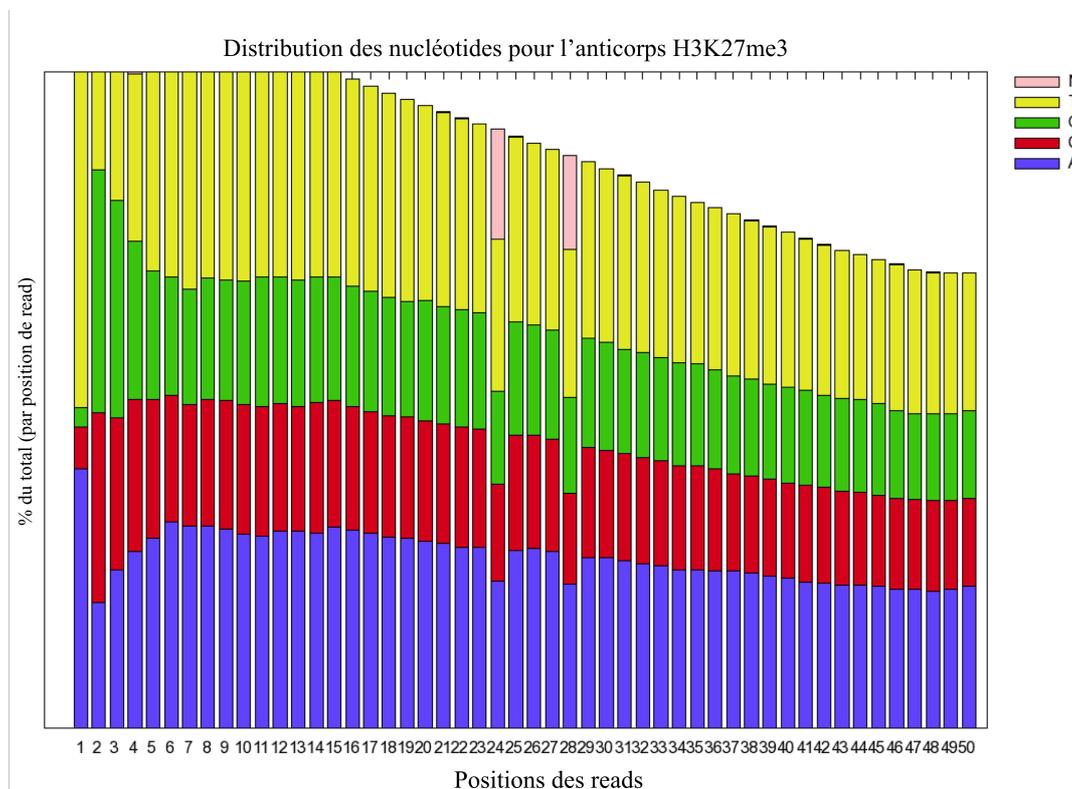
Le centre de séquençage vérifié la qualité dès le premier cycle. Pour cela, un rapport est généré sur l'analyse de la première base incorporée pour estimer : la densité en cluster sur chaque piste de la *flow-cell*, l'intensité du signal généré et la qualité du focus de l'appareil.

La première étape informatique consiste à ôter les séquences des adaptateurs, par l'outil Cutadapt (Martin 2011) car elles sont toujours présentes dans les données issues du séquençage massif. Nous avons ensuite procédé à des analyses de qualité des séquences au format FastQ Sanger (Cock *et al.* 2010) par FASTX-Toolkit ([http://hannonlab.cshl.edu/fastx\\_toolkit/](http://hannonlab.cshl.edu/fastx_toolkit/)) (Figure 26).



**Figure 26: Exemple de boîte à moustaches des scores de qualité pour l'anticorps H3K27me3.** Les éléments de ce boxplot sont les suivants : la ligne noire dans chaque boxplot représente la médiane des scores de qualité, le cube rouge représente l'écart interquartile entre le premier quartile (25%) et le troisième quartile (75%), les traits rouges supérieurs et inférieurs représentent 10% et 90% des points de la distribution. L'arrière plan du graphique divise l'axe des ordonnées en 3 parties colorées : en vert une très bonne qualité, en jaune une qualité raisonnable et en rouge une mauvaise qualité. Les seuils de qualité bons, moyens ou mauvais sont fixés par l'expérimentateur.

En parallèle des analyses de qualité, la répartition des différents nucléotides selon les positions des reads (pb) est observée pour les différentes conditions (Figure 27).



**Figure 27: Exemple de graphique de distribution des nucléotides par position des séquences pour l'anticorps H3K27me3.** En bleu le nucléotide A, en rouge le nucléotide C, en vert le nucléotide G, en jaune le nucléotide T, en rose N lorsque le séquençage n'identifie pas la base nucléotidique.

Une fois les adaptateurs ôtés, nous avons observé que l'analyse de la qualité des séquences était suffisamment bonne pour la suite du traitement des données. L'alignement des séquences, contre la version 5.2 du génome *S. mansoni* NMRI (Protasio et al. 2012), est réalisé grâce au logiciel Bowtie version 0.12.7 (Langmead et al. 2009) avec les paramètres `-p 4 ; -n 2 ; -m 1 ; --best ; --strata ; --chunkmbs 200`.

Toutes les données traitées dans les analyses ChIP-seq sont résumées dans le Tableau 4 avec le nombre de séquences obtenues par le centre de séquençage (Montpellier GenomiX), le nombre de séquences alignées et le nombre de séquences non alignées car répétées plusieurs fois dans le génome.

**Tableau 4: Nombre de séquences utilisées à chaque étape de l'analyse ChIP-seq.** Les analyses ont porté sur les échantillons issus de la condition sympatrique ou allopatrique la génération 3.

	Sympatrique génération 3				Allopatrique génération 3			
	H3K9ac	H3K4me3	H3K9me3	H3K27me3	H3K9ac	H3K4me3	H3K9me3	H3K27me3
<b>total</b>	62 998 112	45 327 579	52 905 182	64 834 197	46 668 170	23 711 850	46 500 180	44 631 970
<b>alignées</b>	39 497 891	31 037 746	27 591 501	37 784 224	27 634 887	14 760 582	23 786 774	22 378 829
<b>non alignées (répétées)</b>	7 682 441	7 525 525	7 461 839	9 905 804	4 895 145	5 456 241	5 536 397	5 432 449

Pour l'étude de ChIP-seq, une détection de pics et une visualisation graphique a pu être possible grâce aux outils : BayesPeak (Spyrou *et al.* 2009), PeakRanger (Feng *et al.* 2011) et T-PIC (Hower *et al.* 2011) afin de comparer les résultats obtenus avec différents logiciels et les intégrer ensuite dans un navigateur de génome (*Genome Browser*, GBrowse v2.42) (Stein *et al.* 2002), afin de les visualiser les résultats de façon graphique. PeakRanger est le logiciel le plus simple permettant la quantification des pics majoritaires (c'est-à-dire proportionnellement au nombre de séquences alignées à une certaine position) alors que les autres logiciels détectent des régions enrichies, sans assimiler une valeur d'intensité.

BayesPeak (Spyrou *et al.* 2009) fonctionne selon un algorithme qui modélise les positions des fragments séquencés et détermine les positions de zones enrichies, basé sur une méthode statistique bayésienne et un modèle HMM (Hidden Markov Model). Le modèle probabiliste prend en compte les caractéristiques naturelles des données et les intègre dans un modèle HMM. Cette méthode divise le génome en régions équidistantes dont la taille n'est pas inférieure à la moitié de la longueur moyenne des fragments, ceci dépendant de l'expérience, de la longueur des sites d'intérêt et de la résolution attendue (Spyrou *et al.* 2009).

PeakRanger (Feng *et al.* 2011) propose une stratégie dans laquelle la taille de fragment est fournie par l'utilisateur et non pas estimée par le nombre de séquences alignées, simplifiant ainsi la conception du logiciel tout en améliorant les performances. L'algorithme commence par rechercher le premier sommet de la région enrichie. Ce point maximal est calculé en multipliant la valeur du signal par Delta, qui est un facteur choisi par l'expérimentateur. Delta varie entre 0 et 1. Si le bruit de fond est conséquent, l'algorithme permet de faire des traitements supplémentaires avant de définir les maximums (Feng *et al.* 2011).

T-PIC (Hower *et al.* 2011) (*Tree shape Peak Identification for ChIP-seq*) est basé sur un test statistique robuste. Plus précisément, ce logiciel définit et évalue la forme des pics pour distinguer les positions aléatoires de celles non aléatoires sur le génome. Le fichier d'entrée

correspond aux séquences alignées pour chaque condition. Chaque séquence contenant chaque paire de bases est alignée et le génome est divisé en plusieurs régions pour lesquelles des arbres sont construits.

Tous les tests de logiciels ont été réalisés dans le cadre d'un stage de Master qui a mené à une thèse (David Roquis). Suite à la comparaison des différents logiciels, nous avons décidé de travailler avec le logiciel PeakRanger v1.16 puisqu'il est définitivement le plus simple d'application et le plus rapide. Ce logiciel a été utilisé avec les paramètres par défaut.

## 5.2. Transcriptome global : RNA-seq

Pour les expériences de RNA-seq, la technique de séquençage à haut débit Illumina est retenue. Le séquençage des échantillons d'ARN, stade cercaire, est effectué en condition *single end* et 100 paires de bases (pb).

Préalablement, une banque d'ARN messagers (ARNm) est préparée. Celle-ci comprend huit étapes : (1) sélection des ARN polyadénylés sur bille magnétique oligo(dT), (2) fragmentation chimique des ARNs sélectionnés, (3) synthèse du premier brin d'ADN complémentaire (ADNc), (4) synthèse du second brin d'ADNc, (5) réparation des extrémités des ADNc double brin, (6) ajout d'une base A aux extrémités, (7) ligation des adaptateurs à ces extrémités adénylées, (8) amplification par PCR permettant aussi une sélection puisque seuls les fragments portant les deux types d'adaptateurs seront amplifiés.

Toutes les étapes qui suivent la construction des banques sont identiques au séquençage des échantillons de ChIP présenté ci-dessus : les étapes de validation des banques, la génération des clusters, l'hybridation des amorces de séquençage, le séquençage et le contrôle de qualités finales sont réalisés selon le même ordre et les mêmes protocoles que pour le ChIP-seq.

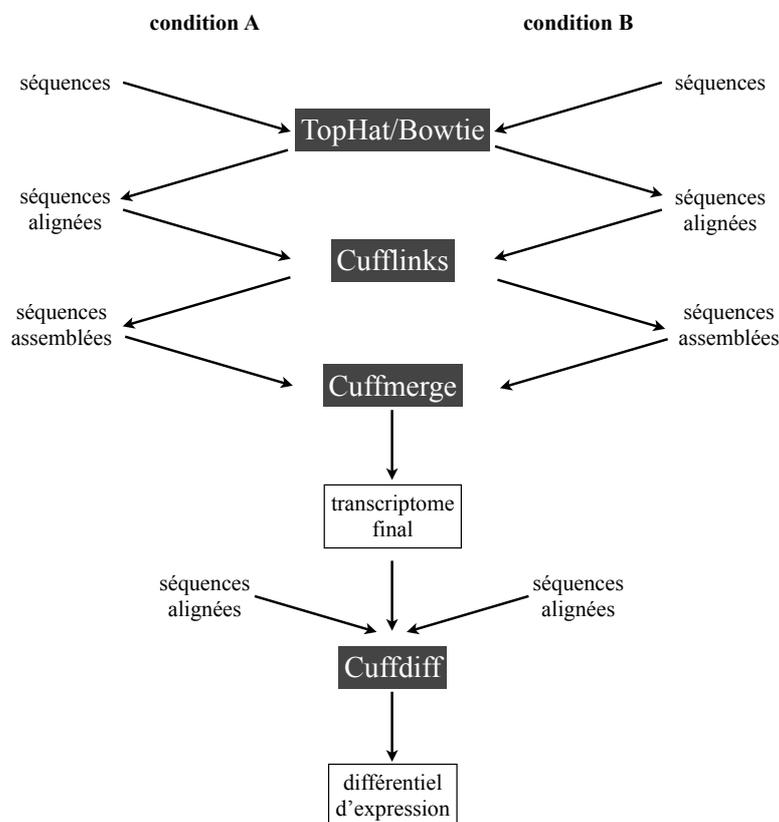
Les premières étapes informatiques sont aussi semblables à l'analyse du ChIP-seq. Nous avons ôté les séquences des adaptateurs, apprécié la qualité et la répartition de différents nucléotides et enfin réalisé l'alignement contre le génome de référence NMRI (v5.2) grâce à TopHat/Bowtie version 1.4.1 (Trapnell *et al.* 2012). Les paramètres sont : *max intron length* = 30000 ; *min isoform fraction* = 0.05 ; *Pre mRNA fraction* = 0.05. Bowtie est un des logiciels les plus populaires car il utilise une structure économique des données intégrant une indexation du génome permettant le stockage du génome de référence et une recherche rapide

des similitudes (<http://bowtie-bio.sourceforge.net/index.shtml>). Cependant Bowtie présente deux limites : il ne prévoit ni l'alignement entre les *reads* et un génome contenant des grands vides (gaps), ni l'alignement des séquences portant sur des introns. Pour pallier ces limitations, le logiciel TopHat a été développé. TopHat utilise Bowtie comme « moteur » d'alignement et décompose les séquences que Bowtie ne peut aligner, en petites séquences appelées segments. Souvent ces segments, quand ils sont traités de façon indépendante, peuvent s'aligner sur le génome. Lorsque plusieurs segments s'alignent à des localisations éloignées les unes des autres sur le génome (par exemple 100 pb et plus de 100 kb), TopHat déduit que ces séquences sont soumises à l'épissage et estime les sites d'épissage. Par le traitement de chaque séquence initialement non alignée, TopHat peut construire un index des sites d'épissage dans le transcriptome sans a priori d'annotation. Cette capacité est indispensable car de nombreuses études RNA-seq ont montré des informations incomplètes pour ces régions d'épissage (Trapnell *et al.* 2012).

Après l'exécution de TopHat v1.4.1, les fichiers d'alignement sont fusionnés par Cuffmerge, qui est inclus dans le logiciel Cufflinks afin de constituer une base uniforme pour les calculs des niveaux d'expression de chaque transcrit, pour chaque condition. Les séquences et les fichiers fusionnés sont ensuite assimilés par Cuffdiff, inclus dans Cufflinks, pour calculer les taux d'expression et effectuer les tests statistiques (Trapnell *et al.* 2012).

Cufflinks/Cuffdiff (Trapnell *et al.* 2012) est choisi pour la détection des gènes différentiellement exprimés et pour les validations statistiques. La quantification précise des niveaux d'expression des gènes, nécessite l'identification des différentes isoformes de produits de gènes pour chaque séquence. Cufflinks assemble les transcrits à partir des séquences alignées sur le génome. Un échantillon peut contenir différents variants mais Cufflinks est en mesure de déduire la structure d'épissage de chaque gène. Toutefois, les gènes ont parfois plusieurs événements d'épissage entraînant diverses reconstructions possibles. Cufflinks fait état d'un assemblage de transcriptomes parcimonieux. Après l'étape d'assemblage, Cufflinks quantifie le niveau d'expression de chaque transcrit. Cette comptabilisation est effectuée à l'aide d'un modèle statistique rigoureux afin d'identifier les artefacts (Trapnell *et al.* 2012). Cufflinks contient un programme séparé, Cuffdiff, qui calcule l'expression dans deux ou plusieurs conditions et teste statistiquement les différences observées entre toutes les conditions. Le modèle statistique retenu suppose que les séquences produites pour chaque transcrit sont proportionnelles à leur abondance mais qu'elles fluctuent en raison de la variabilité, lors de la préparation des banques et du séquençage de chaque

échantillon biologique. Cufflinks et Cuffdiff peuvent automatiquement modéliser et soustraire une fraction importante de ces biais dans la distribution des séquences pour chaque transcrit, cela permettant d'estimer l'abondance (Trapnell *et al.* 2012). Le protocole utilisé est schématisé dans la Figure 28.



**Figure 28: Représentation schématique du protocole expérimental utilisé pour les analyses RNA-seq.**

Toutes les données traitées dans les analyses RNA-seq sont résumées dans le Tableau 5 avec le nombre de séquences: obtenues par le centre de séquençage (Montpellier GenomiX), après le filtre de qualité, une fois les adaptateurs retirés, exploitées pour l'alignement et enfin le nombre de séquences alignées.

**Tableau 5: Nombre de séquences utilisées à chaque étape de l'analyse RNA-seq.** Les analyses ont porté sur les échantillons issus de la condition sympatrique ou allopatrique pour les trois générations

	Sympatrique			Allopatrique		
	Génération 1	Génération 2	Génération 3	Génération 1	Génération 2	Génération 3
<b>total</b>	29 526 547	19 650 926	25 733 992	39 930 524	33 790 715	27 852 056
<b>alignées</b>	2 431 977	4 674 800	29 416 286	42 767 340	20 162 160	29 662 598
<b>répétées</b>	1 251 492	3 587 089	15 305 902	12 599 139	8 901 133	12 911 902

Tous les outils de bioinformatiques mis en œuvre dans les différentes analyses de ChIP-seq et RNA-seq sont, pour la plupart, intégrés dans Galaxy, permettant d'avoir une interface graphique facile d'utilisation.

### 5.3. Génome global : gDNA-seq

Le séquençage massif de l'ADN génomique est réalisé en *single-end* et 36 paires de bases (pb) et selon le protocole du fabricant.

SOAP2/SOAP (Li *et al.* 2009) est utilisé pour l'alignement des séquences nous permettant ainsi de distinguer les séquences uniques des séquences répétées. Toutes les séquences s'alignant en plusieurs endroits du génome sont ensuite assemblées par le logiciel Velvet 0.7.01 (Zerbino & Birney 2008). Pour l'étude de ces séquences répétées, Blast (Altschul *et al.* 1990), Censor/Repbase (Kohany *et al.* 2006), TEclass (Abrusan *et al.* 2009) et Tandem Repeats Finder (Benson 1999) servent à une première annotation (Lepessant *et al.* 2012b). Tous les détails techniques sont présentés dans la publication ci-jointe (Lepessant *et al.* 2012b) (Annexe 3).

# **CHAPITRE 3**

## **Résultats**

### 1. Enigme 1 « Adaptation rapide de *S. mansoni* à une drogue anti-helminthe »

Comment le parasite *S. mansoni* est-il capable de générer des phénotypes résistants à certaines drogues en un temps de génération très court?

Dans les études précédentes (Jansma *et al.* 1974; Jansma *et al.* 1977), les auteurs ont utilisé une souche de laboratoire issue de Puerto Rico pour montrer l'apparition de résistance à la molécule hycanthane, chez *S. mansoni*, dès la première génération traitée. Les infestations sur souris ont été réalisées avec des couples et après 62 jours post-infestation, un traitement avec l'hycanthane par une unique dose de 60mg/kg en injection intramusculaire a été administré aux hôtes infestés.

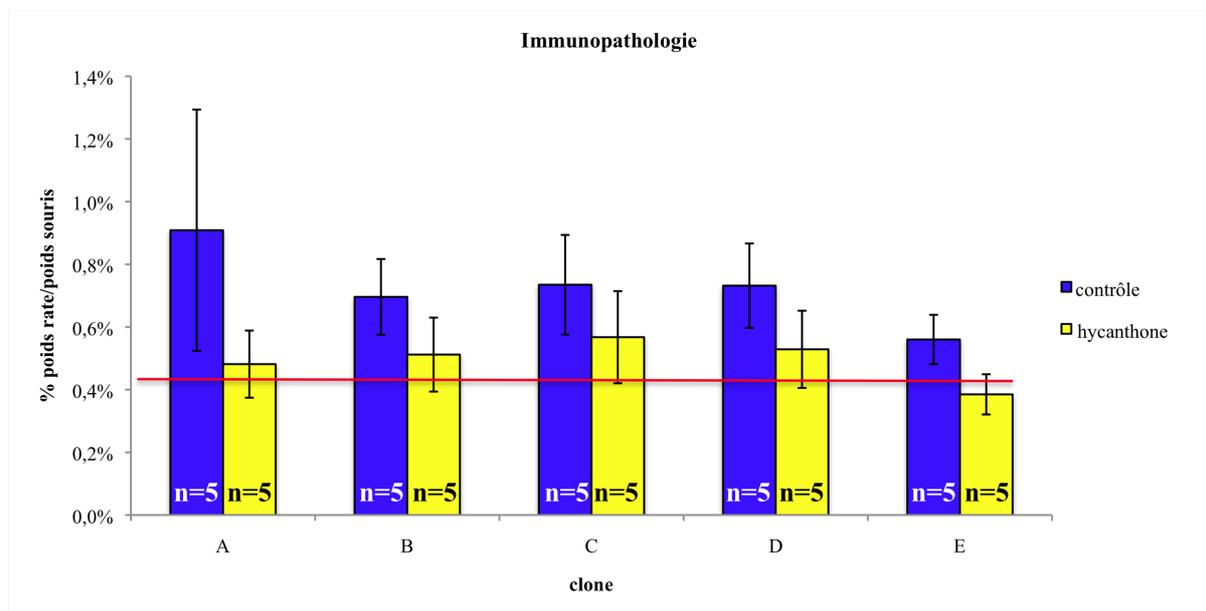
Au sein de notre laboratoire, cette souche de parasite n'est pas maintenue. Nous avons donc réalisé des expériences préalables afin de choisir une de nos souches, en procédant à des infestations multiples avec trois souches de laboratoire : une souche brésilienne (*SmBre*), une souche humaine guadeloupéenne (*SmGH2*) et une souche issue d'un isolat guadeloupéen différent (*SmDFO*). Les souris sont infestées simultanément avec les trois souches. Le traitement par la drogue est administré selon les mêmes conditions que les études précédentes (Jansma *et al.* 1974; Jansma *et al.* 1977), après 62 jours post-infestation, par une injection intramusculaire de 60mg/kg et le sacrifice des souris à 57 jours post-traitement (Jansma *et al.* 1974; Jansma *et al.* 1977). Tous les individus résistants sont récupérés individuellement et génotypés sur cinq marqueurs microsatellites neutres, spécifiques de chaque souche.

Cette analyse nous permet d'identifier la souche à laquelle appartient chaque individu. Ainsi, nous avons constaté que :

- la souche humaine de Guadeloupe (*SmGH2*) présente une sensibilité totale, aucun individu résistant n'est retrouvé après traitement ;
- la souche brésilienne (*SmBre*) et la deuxième souche Guadeloupéenne (*SmDFO*) montrent l'apparition d'une résistance : 30% des individus pour chaque souche survivent, après traitement.

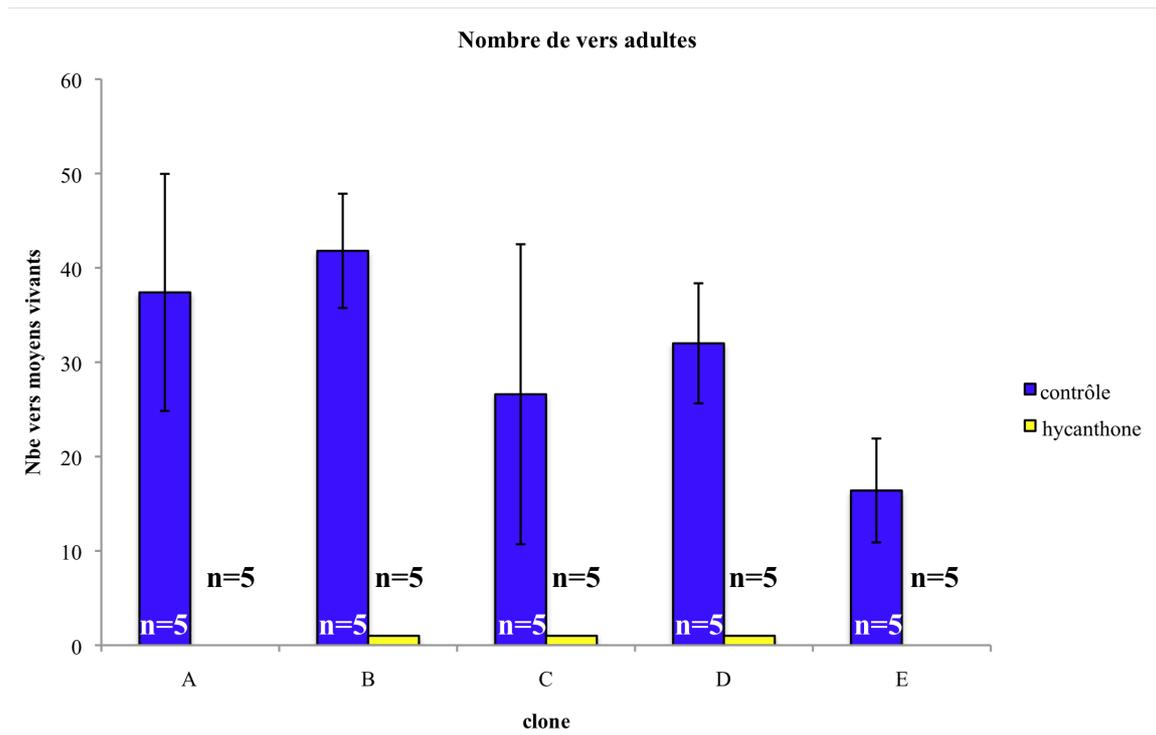
Au vu de ces résultats, deux souches (*SmBre* et *SmDFO*) peuvent être retenues pour l'étude de l'apparition de résistance à l'hycanthane. Notre choix s'est porté sur la souche brésilienne (*SmBre*) car elle présente un fond génétique très homogène, sur 14 marqueurs microsatellites neutres.

Le fait d'utiliser cinq clones différents nous permet de comparer la réponse du parasite au sein de chaque clone mais aussi entre les clones. En plus du dénombrement des adultes résistants, nous avons quantifié la réponse immunopathologique de la souris suite à l'infestation et au traitement. Cette mesure est réalisée en même temps que les sacrifices puisque ce trait est quantifié par le ratio du poids de la rate sur le poids de la souris, afin d'estimer la réponse inflammatoire. Pour chaque clone, traité et non traité, il y a cinq réplicats (n=5). Ce trait d'histoire de vie n'a présenté aucune différence significative ( $p > 0.05$ ) entre les différents clones, au sein de chaque clone et entre les conditions traitées/non traitées. Mais les profils entre les clones semblent montrer la même tendance, avec une baisse de la réponse inflammatoire après le traitement (Figure 29).



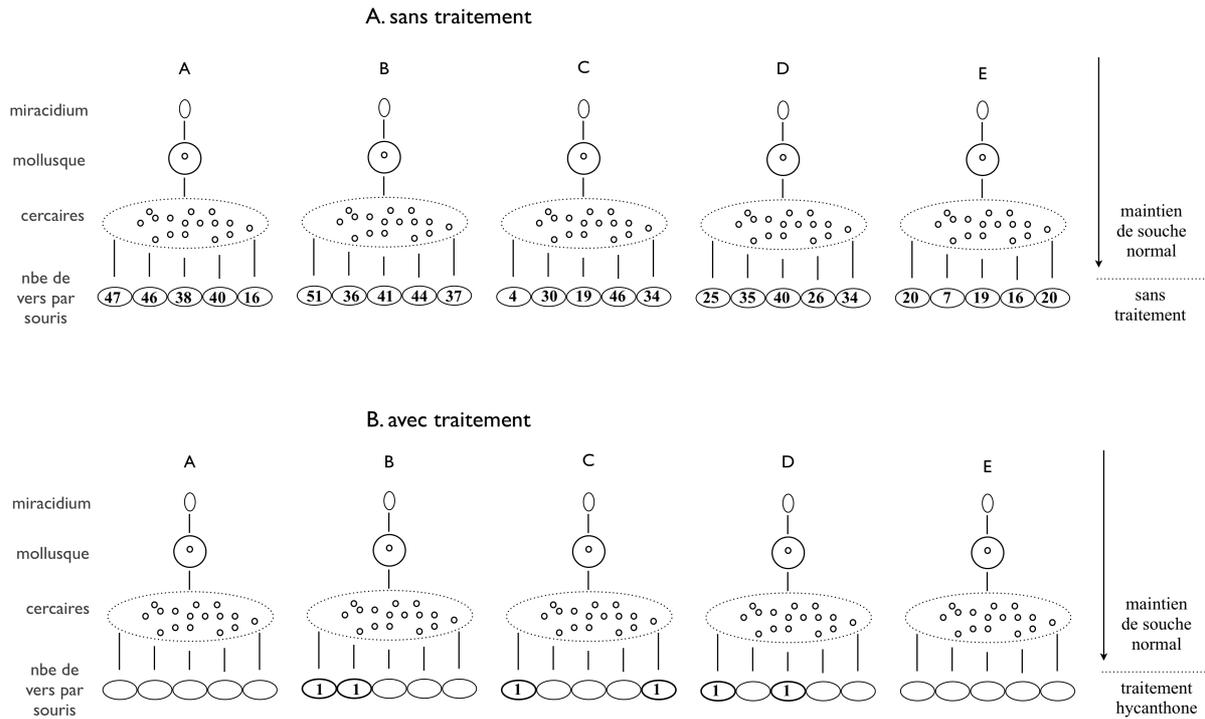
**Figure 29: Traits d'histoire de vie : réponse immunopathologique de l'hôte définitif.** Ce trait est quantifié par le ratio poids de la rate/poids de la souris afin d'estimer la réponse inflammatoire suite à l'infestation et au traitement. Les individus traités avec un traitement contrôle (en bleu) et les individus traités à l'hycanthone (en jaune) ne présentent aucune différence significative dans la réponse inflammatoire. Le trait rouge représente le ratio poids rate/poids souris pour une souris saine.

Le dénombrement des vers développés au sein de chaque condition et au sein de chaque clone, montre que les individus résistants sont distribués entre trois clones sur cinq. Seuls deux clones (A et E) ne présentent pas de vers résistants d'où une sensibilité totale à l'hycanthone, alors que dans les trois autres clones (B, C, D) on distingue au moins un individu résistant (Figure 30).



**Figure 30: Nombre de vers adultes moyens développés suite au traitement hycanthone.** Dans la condition contrôle (bleu) les vers dénombrés sont répartis dans tous les clones. Dans la condition traitée à l'hycanthone (jaune) les individus résistants (jaune) sont présents dans trois clones (B, C, D) sur cinq. Seuls deux clones (A, E) présentent une sensibilité totale à l'hycanthone, aucun ver résistant retrouvé.

Les résultats issus du traitement à la molécule hycanthone peuvent être schématisés selon les expériences de Luria & Delbrück (Luria & Delbrück 1943) ainsi que celles de Lederberg & Lederberg (Lederberg & Lederberg 1952). Adapté de Jablonka & Lamb (Jablonka & Lamb 1999) (Figure 31).



**Figure 31: Synthèse des résultats de l'expérience de traitement à l'hycanthone.** Les individus résistants à l'hycanthone sont répartis entre différents clones. Les valeurs représentent le nombre de vers survivants retrouvés. Adapté de Jablonka & Lamb (Jablonka & Lamb 1999).

Après traitement à l'hycanthone, beaucoup de vers adultes sont retrouvés enkystés dans le foie. Ce constat s'explique par le fait que nous avons utilisé uniquement des clones mâles, qui ne migrent pas puisqu'ils ne sont pas accouplés.

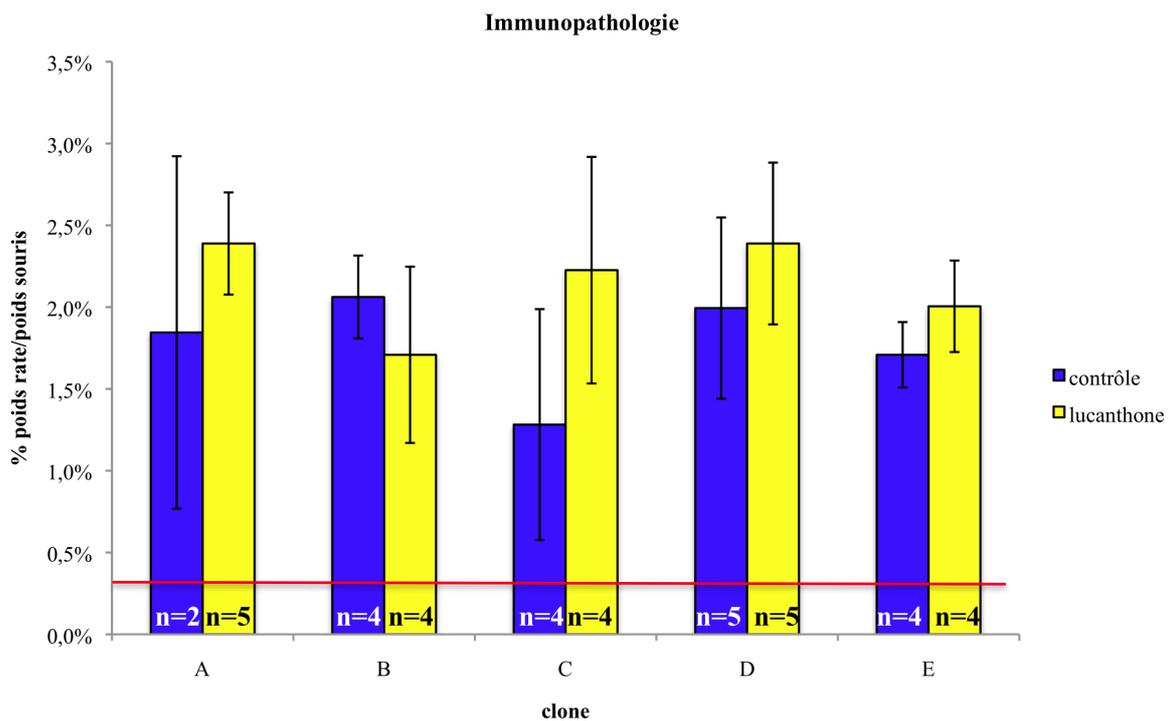
En effet, lors d'expériences préliminaires avec des couples, une résistance de 30% a été notée. Des études précédentes supposent que les femelles résistantes semblent protéger les mâles sensibles dont la mortalité devient moindre (Jansma *et al.* 1977) alors que l'accouplement d'un mâle résistant avec une femelle sensible ne diminue pas la mortalité des femelles (Jansma *et al.* 1977).

Pour compléter cette approche expérimentale, nous avons immunoprécipité l'ADN des adultes survivants, pour la condition traitée et la condition non traitée, avec des anticorps spécifiques de marques euchromatiques (H3K9ac et H3K4me3) et hétérochromatiques (H3K9me3 et H3K27me3). Le séquençage massif a été réalisé par la technique Illumina en collaboration avec l'Université d'Oregon. Dans le temps qui m'était imparti pour ma thèse, les données n'ont pu être acquises que tardivement et de ce fait les analyses de structure chromatinienne sont en cours d'analyse. Cette étude vise à mettre en évidence les modifications dans la structure de la chromatine dans des régions ciblées et/ou aléatoires du

génomique après traitement, afin d'observer si elles peuvent être corrélées à l'apparition de la résistance.

Nous avons poursuivi nos recherches avec la molécule lucanthone résultant de la biactivation de l'hycanthone (Archer 1985; Laurent 2006) car cette dernière n'est plus commercialisée. L'administration du lucanthone se fait par voie orale, en trois doses réparties sur trois jours, afin d'atteindre la même dose finale que l'hycanthone (60mg/kg).

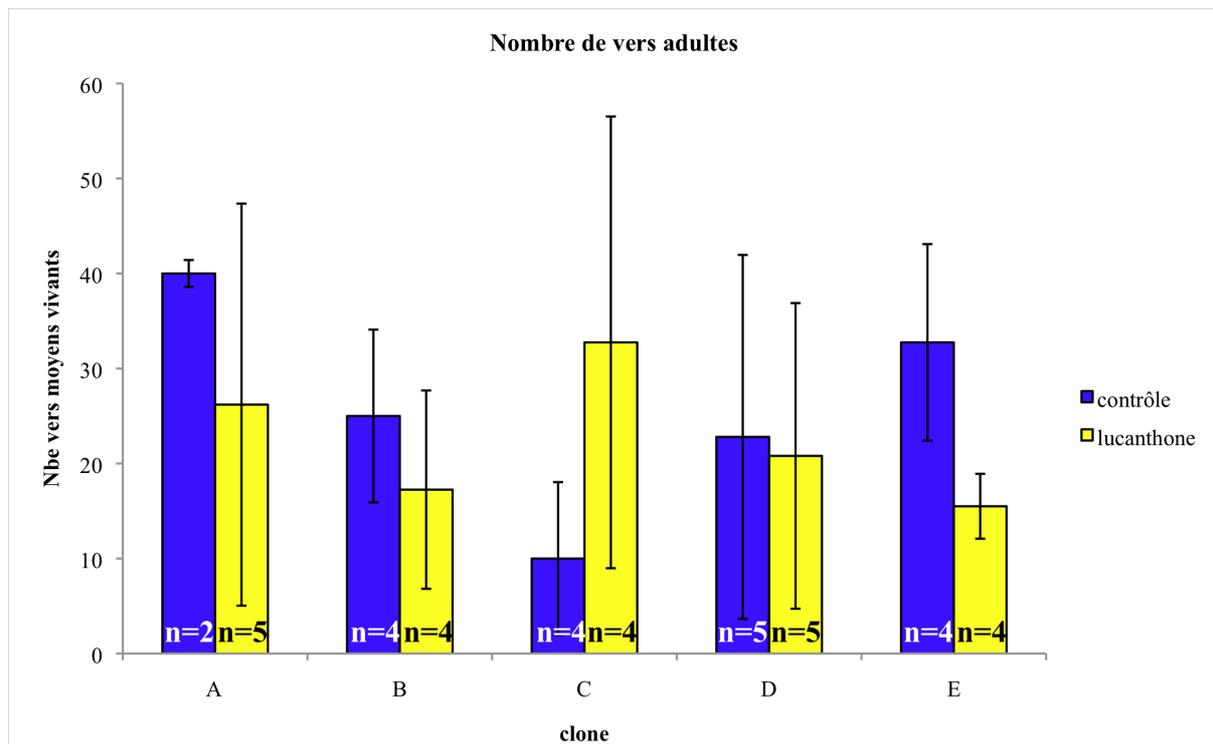
Les expériences avec le lucanthone sont reproduites dans les mêmes conditions que celles avec l'hycanthone mais en utilisant des couples. Le nombre d'individus varie entre 2 et 5. Suite aux infestations monomiracidiales, cinq couples (un clone mâle/un clone femelle) sont formés. La mesure de la réaction inflammatoire (poids de la rate/poids de la souris) ne montre aucune différence significative entre la condition contrôle et la condition traitée, ni entre chaque couple de clones (Figure 32).



**Figure 32: Nombre de vers adultes moyens développés suite au traitement lucanthone.** Ce trait est quantifié par le ratio poids de la rate/poids de la souris afin d'estimer la réponse inflammatoire suite à l'infestation et au traitement. Les individus traités avec le contrôle (en bleu) et les individus traités au lucanthone (en jaune) ne présentent aucune différence significative dans la réponse inflammatoire. Le trait rouge représente le ratio poids rate/poids souris pour une souris saine.

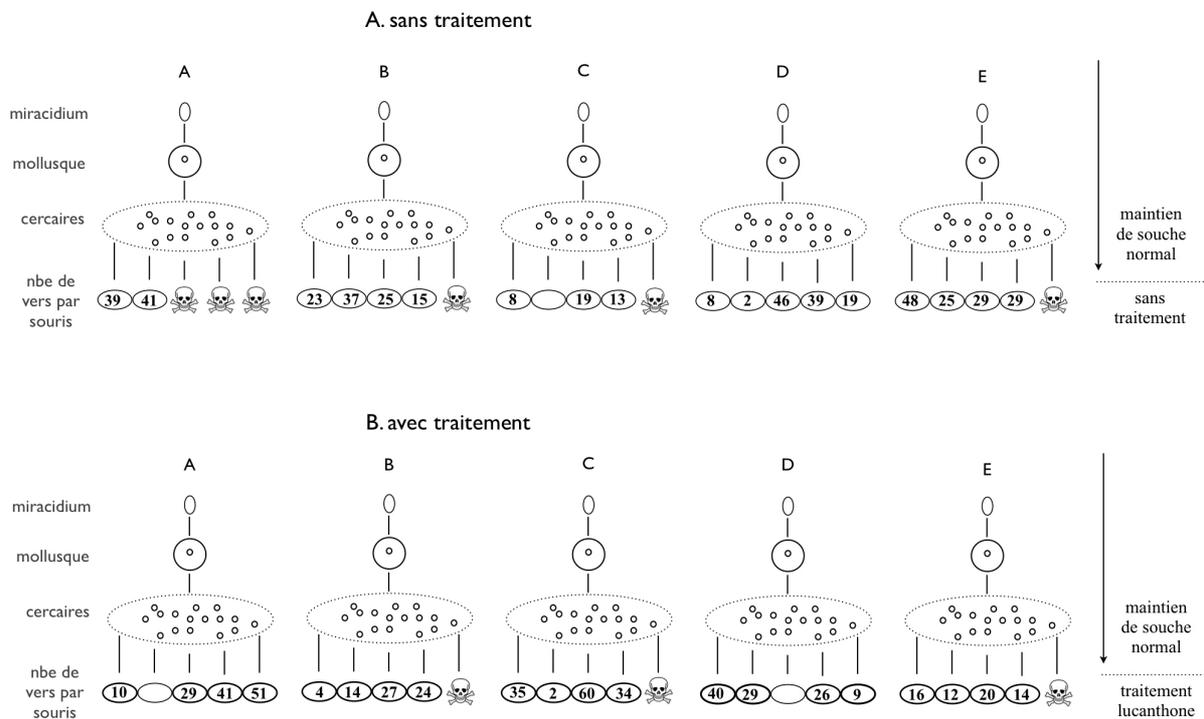
La réaction inflammatoire a été plus forte lors du traitement au lucanthone. Ce résultat s'explique par le fait que les infestations ont été réalisées avec des couples, d'où la migration des individus adultes et l'accumulation des œufs dans le foie.

Les résultats du traitement au lucanthone sont identiques à ceux notés pour l'hycanthone : les individus résistants sont distribués entre les différents clones et ne sont pas restreints à un couple de clones. Par contre, l'on constate un plus grand nombre de vers résistants et aucun des clones ne montre une sensibilité totale au lucanthone (Figure 33).



**Figure 33: Résultats de l'expérience de traitement au lucanthone.** Dans la condition contrôle (bleu) les individus dénombrés sont présents dans tous les clones et dans la condition traitée au lucanthone (jaune) les individus résistants se répartissent entre les cinq couples de clones.

Les résultats issus du traitement avec la molécule hycanthane sont schématisés dans la figure 34.



**Figure 34: Synthèse des résultats de l'expérience de traitement au lucanthone : génération 1.** Les individus résistants au lucanthone sont répartis entre différents clones. Adapté de Jablonka & Lamb (Jablonka & Lamb 1999).

Nous avons ensuite cherché à savoir si cette résistance est héréditaire. Nous avons donc récupéré les œufs issus de la reproduction sexuée des vers adultes pour les faire éclore sous la lumière blanche. Puis, des mollusques *B. glabrata* Brésil (*BgBRE*) ont été infestés avec 10 miracidia. Les cercariae émises 35 jours post-infestation ont servi à infester des souris, afin qu'elles soient traitées au lucanthone. Seul le clone C n'a pu être utilisé pour les infestations sur souris, car les mollusques infestés n'émettaient aucune cercaire.



résultats peuvent donc être remis dans le contexte des expériences de Luria & Delbrück (1943) pour déterminer si la résistance est latente avant le traitement et est sélectionnée par celui-ci, ou si c'est le traitement qui entraîne l'apparition de résistance dans la population.

## 2. Énigme 2 « Adaptation rapide de *S. mansoni* à l'Amérique du Sud »

Comment une souche de parasite *S. mansoni* peut-elle s'adapter rapidement à un nouvel environnement ?

Dans un premier temps, nous avons procédé à une étude approfondie de deux souches issues d'isolats géographiques différents afin d'étudier des candidats génétiques et/ou épigénétiques distincts entre la souche Brésil et la souche Guadeloupe.

Puis, dans un deuxième temps, nous avons analysé les séquences promotrices des gènes impliqués dans le polymorphisme de compatibilité, les *SmPoMucs*, pour ces deux souches de laboratoire.

Enfin, nous avons mis en œuvre une évolution expérimentale visant à exposer une souche de parasite donnée à différentes souches de mollusques, afin d'observer la réponse adaptative.

### 2.1. Étude des candidats génétiques et épigénétiques différents entre deux souches de parasites issues d'isolats géographiques distincts

Pour se faire, un séquençage massif des échantillons issus de CHIP, précipités avec la marque H3K9ac, est réalisé sur deux isolats différents (Brésil et Guadeloupe). Cette marque est privilégiée car elle est représentative de promoteur actif (Lee *et al.* 2010) et d'une structure chromatinienne euchromatique. La technique Illumina est utilisée pour le séquençage massif de l'ADN génomique et a généré environ 120 Mb de données, alors que le génome de *S. mansoni* en fait 370 Mb. La stratégie choisie a pour but de répertorier des candidats génétiques et épigénétiques dans les deux souches. La comparaison *in silico* permet l'identification d'enrichissements différents dans des régions génomiques pour ces deux souches. Cette méthode, combinée à de la PCR quantitative (approche ciblée) met en avant un polymorphisme génétique ou une différence d'acétylation sur certains *loci*.

Le logiciel CHIPDiff souligne les différences dans le nombre de séquences alignées entre les deux souches.

Les deux isolats analysés et la souche de référence (Puerto Rico), représentent trois isolats géographiques de *S. mansoni* distincts. Si le polymorphisme génétique, entre les souches

Brésil et Guadeloupe est grand, alors les différences observées dans l'alignement peuvent apparaître à plusieurs *loci* dans le génome. Pour cela, ChIPDiff peut identifier un polymorphisme génétique entre l'isolat Brésil et l'isolat Guadeloupe. Si, dans un deuxième cas, le nombre de copies pour certains *loci* est différent entre ces deux souches, l'enrichissement sera lui aussi différent puisque le nombre de séquences alignées est dépendant du nombre de copies présentes dans le génome. Mais si aucun polymorphisme génétique ou différence dans le nombre de copies n'est établi, les différences identifiées par ChIPDiff sont dues à des modifications épigénétiques, telles qu'une différence d'acétylation entre les deux souches. Il n'est pas exclu que les différences génétiques et épigénétiques soient accumulées.

Suite à l'utilisation de ChIPDiff, 321 différences ont été relevées entre les deux souches. Ce logiciel identifie le nombre de séquences alignées entre les deux souches en analysant des blocs de taille 1kb. Il existe 128 blocs de 1kb enrichis dans la souche Brésil, et 193 blocs de 1kb enrichis dans la souche Guadeloupe. Pour la suite des analyses, nous avons exclu les blocs de 1kb isolés et nous avons seulement conservé des régions contenant plusieurs groupes de 1kb physiquement proches. En utilisant ce critère, nous avons retenu 9 régions, certaines contenant des gènes putatifs (Tableau 6).

**Tableau 6: Régions contenant des différences dans le nombre de séquences alignées entre la souche Brésil et la souche Guadeloupe** (Lepesant *et al.* 2011) (Annexe 4).

Scaffold	Positions	ORF prédite	Annotation	Expression
<b>Guadeloupe &gt; Brésil</b>				
Smp_scaff000234	0-310 kb	Smp_159800	famille MEG2 (ESP15)	œufs
		Smp_159810	famille MEG2 (ESP15)	œufs
		Smp_159820	famille MEG2 (ESP15)	œufs
		Smp_159830	famille MEG2 (ESP15)	œufs
		Smp_159840	Protéine putative	œufs, schistosomules, adultes
Smp_scaff000421	10-70kb	Smp_171100	zinc métalloprotéase	œufs, miracida, femelles
Smp_scaff000496	0-2.3 kb 10-15 kb			
<b>Brésil &gt; Guadeloupe</b>				
Smp_scaff000131	0-70 kb	Smp_149750	Protéine putative	œufs, cercaires
		Smp_051690	Protéine putative	pas d'ESTs
		Smp_149760	Protéine putative	pas d'ESTs
Smp_scaff000490	8-16 kb			
Smp_scaff000538	0-24 kb	Smp_093470	Protéine putative	pas d'ESTs
Smp_scaff000539	0-35 kb	Smp_173370	M13 peptidase non définie	pas d'ESTs
Smp_scaff019017	10-40 kb	Smp_121530	Protéine putative	œufs, adultes
<b>Guadeloupe &gt; Brésil ou Brésil &gt; Guadeloupe</b>				
Smp_scaff000154	449-500 kb	Smp_152600	Protéine pax-6	œufs, miracidia
		Smp_152610	Protéine putative	pas d'ESTs
		Smp_152620	Homologue éperons	œufs, schistosomules
		Smp_152630	Protéine putative	pas d'ESTs

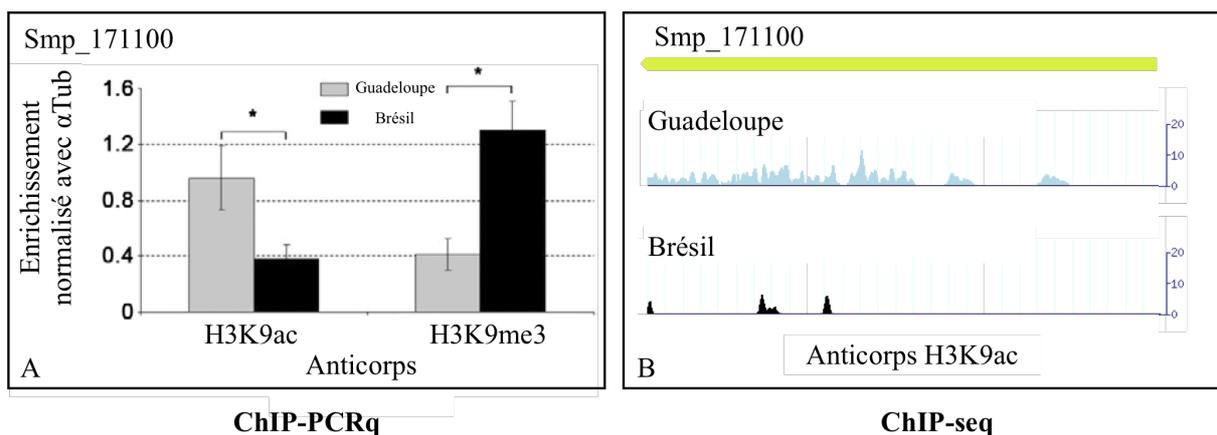
Nous avons analysé 11 *loci* de ces 9 régions par ChIP-qPCR avec deux anticorps (H3K9Ac et H3K9Met3). Cette analyse ciblée nous permet de distinguer le polymorphisme du à un nombre de copies et le polymorphisme génétique et/ou épigénétique (Tableau 7).

**Tableau 7: Comparaison entre les résultats de ChIP-seq et les résultats de ChIP-PCRq entre la souche Brésil et la souche Guadeloupe pour 11 *loci*** (Lepesant *et al.* 2011) (Annexe 4).

Cibles	Nbe de <i>loci</i>	PCR sur Guadeloupe	PCR sur Brésil	Différences entre Brésil et Guadeloupe ChIP-PCRq	Différences entre Brésil et Guadeloupe ChIP-seq	Comparaison du nombre de copie	Génétique ou Épigenétique?
Smp173370	2	+	+	Guadeloupe > Brésil	Brésil > Guadeloupe	Guadeloupe > Brésil	Non confirmé
Smp171100	2	+	+	Guadeloupe > Brésil	Guadeloupe > Brésil	Guadeloupe = Brésil	Épigénétique
Smp159800	2	+	-	nd	Guadeloupe > Brésil	nd	Génétique
Smp159810	4	+	-	nd	Guadeloupe > Brésil	nd	Génétique
Smp159820	2		-	nd	Guadeloupe > Brésil	nd	Génétique
Smp159830	1	+	-	nd	Guadeloupe > Brésil	nd	Génétique
Smp159840	1	+	-	nd	Guadeloupe > Brésil	nd	Génétique
Scaff000131	1	+	+	Guadeloupe = Brésil	Brésil > Guadeloupe	Guadeloupe > Brésil	Non confirmé
Scaff000154. 2	1	+	-	nd	Guadeloupe > Brésil	nd	Génétique
Scaff000154. 3	1	+	+	Guadeloupe > Brésil	Guadeloupe > Brésil	Guadeloupe = Brésil	Épigénétique
Scaff000154. 5		-	+	nd	Brésil > Guadeloupe	nd	Génétique

Parmi les 9 régions, 2 ont particulièrement attiré notre attention. La première est une région de 300 kb située sur le *scaffold* Smp\_scaff000234 avec un nombre de séquences alignées plus élevé dans la souche Brésil que dans la souche Guadeloupe. Cette région contient 4 gènes codant pour des protéines secrétées dans les œufs (famille ESP15, *Egg Secreted Protein*). Ces gènes ont été récemment classifiés dans la famille des gènes à microexons (MEG) (DeMarco *et al.* 2010). Ces derniers sont des gènes contenant de petits exons (< 36 pb). Il est supposé que cette structure atypique pourrait permettre de créer des variants protéiques hautement polymorphes à travers des évènements d'épissage alternatif. Ces ESP15 constituent donc de très bons candidats pour des mécanismes d'évasion immune. Pour cette région, l'analyse par ChIP-qPCR n'a pas été possible car les amorces destinées à l'amplification de la cible n'ont marché que dans la souche Guadeloupe, ce qui laisse supposer qu'il s'agit ici d'un polymorphisme d'origine génétique. (Lepesant *et al.* 2011) (Annexe 4).

Le deuxième est le locus Smp\_171100 situé sur le *scaffold* Smp\_scaff000421 où une différence épigénétique est clairement identifiée et confirmée par ChIP-qPCR avec les 2 anticorps H3K9ac et H3K9me3 (Figure 36). Nous montrons ainsi que Smp\_171100 présente un fort taux d'acétylation dans la souche Guadeloupe, et inversement un fort taux de méthylation dans la souche Brésil ce qui conforte les résultats de ChIP-Seq. Ce gène se trouve donc dans une région très euchromatique dans la souche Guadeloupe et inversement dans une région très hétérochromatique dans la souche Brésil (Lepesant *et al.* 2011) (Figure 36).



**Figure 36: Profil d'enrichissement pour les souches Guadeloupe et Brésil.** (A) Enrichissement autour du gène Smp\_171100 pour les marques H3K9ac et H3K9me3. Le gène est présent dans une région euchromatique dans la souche Guadeloupe, et inversement présent dans une région hétérochromatique dans la souche brésil. Ces données sont normalisées par rapport au gène  $\alpha$ Tub. (B) Profils de la marque H3K9ac obtenu en ChIP-seq pour le gène Smp\_171100. Les données issues de ChIP-PCRq et ChIP-seq présentent une réelle cohérence, avec un enrichissement plus fort pour la marque H3K9ac autour du gène Smp\_171100 dans la souche Guadeloupe (Lepesant *et al.* 2011 (Annexe 4)).

Ce gène, Smp\_171100, est annoté comme une métalloprotéase d'une famille d'endopeptidase neutre (NEP). Ces enzymes dégradent les protéines de façon métal-dépendante. Des travaux ont montré que ce gène modulerait la réponse immunitaire de son hôte intermédiaire *B. glabrata* (Duvaux-Miret *et al.* 1992). Son expression dans les œufs et au stade miracidia (données issues des banques ESTs) confirment son importance à ces stades.

Ainsi, cette analyse comparative a permis d'identifier de nouveaux candidats pour le polymorphisme de compatibilité, dont certains présentent un polymorphisme génétique, d'autres un polymorphisme épigénétique ou encore une différence dans le nombre de copies. Deux des candidats semblent coder pour des protéines potentiellement impliquées dans des mécanismes d'évasion immune, ce qui est intéressant dans la mesure où le polymorphisme de compatibilité concerne l'interaction avec un hôte.

Tous les résultats de cette analyse sont publiés, en 2011, dans *Memorias do Instituto Oswaldo Cruz* (IF 2.058) (Lepesant *et al.* 2011). L'article est proposé en Annexe 4.

## 2.2. Analyse des séquences promotrices des gènes impliqués dans le polymorphisme de compatibilité, les SmPoMucs

Nous avons ensuite étudié les séquences promotrices des gènes *SmPoMucs* pour la souche Brésil et la souche Guadeloupe. Ce travail est réalisé en collaboration avec les co-auteurs de la publication (Annexe 5), ma contribution portant plus particulièrement sur les analyses épigénétiques des séquences promotrices.

Les mucines polymorphes de *S. mansoni*, les *SmPoMucs*, sont les marqueurs moléculaires du polymorphisme de compatibilité. Ces protéines ont été identifiées par des approches de protéomique comparative entre les souches Guadeloupe et Brésil. En effet, l'étude du protéome de ces deux souches a permis de montrer que ces protéines *SmPoMucs* constituent une des principales différences entre ces deux souches. Ce sont des glycoprotéines hautement polymorphes qui interagissent avec les FREPS (*Fibrinogen Related Protein*) du mollusque d'eau douce *B. glabrata*, lesquelles sont impliquées dans la réponse immunitaire du mollusque après infestation par le parasite trématode.

Les protéines *SmPoMucs* sont codés par une famille multigénique composée de 10 membres organisés en 4 groupes dans le génome : groupe 1, groupe 2, groupe 3 et groupe 4. Chaque groupe peut posséder des sous groupes (ex : dans le groupe 3 il existe les sous groupes 3.1,

3.2, 3.3 et 3.4). Ces protéines *SmPoMucs* présentent une partie carboxy-terminale conservée et une partie amino-terminale variable, composée d'unités répétées. Il existe 3 types d'unités répétées (r1, r2 et r3), leur quantité et nature sont variables suivant les différents groupes (Roger *et al.* 2008a).

Bien qu'il existe un polymorphisme entre les souches Guadeloupe et Brésil, on ne distingue aucune différence génétique dans la zone codante entre ces deux souches. Les gènes *SmPoMucs* sont uniquement transcrits au stade interagissant avec le mollusque, avec un pic d'expression au stade infestant, le miracidium. Par ailleurs, au niveau intra-populationnel, on révèle un fort polymorphisme d'expression individuelle. En effet, chaque individu miracidia exprime une combinaison spécifique de gènes *SmPoMucs* (Roger *et al.* 2008b). Au niveau inter-populationnel, une comparaison des transcrits des souches Guadeloupe et Brésil a mis en avant une seule différence entre elles, par la présence d'un transcrit du groupe 3.1 possédant une alternance d'unités répétées r1 et r2 exclusivement dans la souche incompatible Guadeloupe.

De plus, de récentes études montrent que la trichostatine A (TSA), un inhibiteur d'HDAC (histone déacétylase), influence les profils de transcription de ces gènes. Le traitement à la TSA induit, dans la population, la génération de nouvelles combinaisons de transcrits sans pour autant supprimer les anciennes (Cosseau *et al.* 2010). Elles suggèrent une participation de mécanismes épigénétiques de nature chromatiniennne pour contrôler la transcription des *SmPoMucs*.

Afin d'obtenir des évidences expérimentales de cette considération, nous avons choisi de nous focaliser sur les changements génétiques et épigénétiques apparaissant durant l'évolution du polymorphisme de compatibilité entre les deux souches Brésil et Guadeloupe de parasite *S. mansoni*. Nous nous sommes concentrés sur la séquence des promoteurs des gènes actifs *SmPoMucs* et nous avons cherché d'éventuelles différences dans les séquences des promoteurs entre ces deux souches qui pourraient expliquer ces différences de transcription.

Dans un premier temps, une analyse transcriptionnelle des différents groupes a montré que les profils de transcription changent entre les souches Brésil et Guadeloupe. En effet, les groupes 1, 2 et 3.1(r1-r2) sont approximativement 2.9, 3.5 et 25.8 (resp.) fois plus transcrits dans la souche Guadeloupe que dans la souche Brésil. Le groupe 3.1(r1-r2) est quasiment indétectable dans la souche Brésil et présente une forte transcription dans la souche

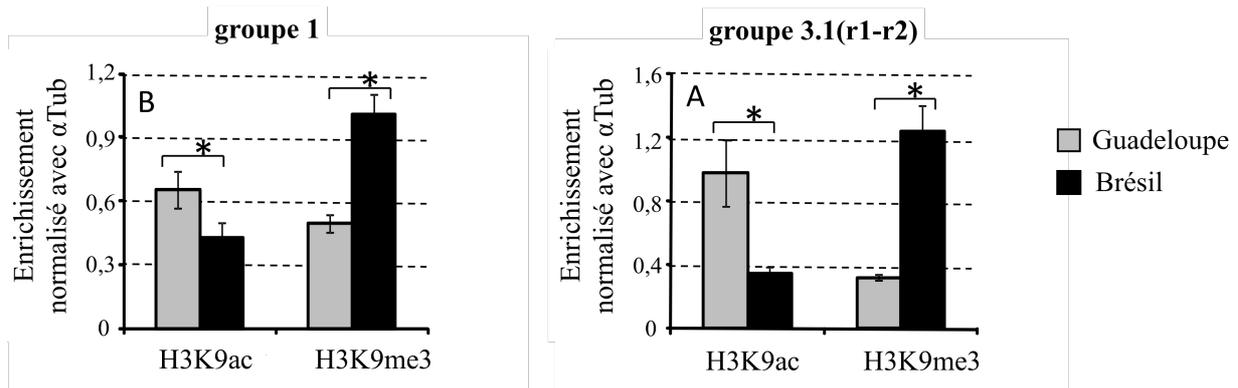
Guadeloupe (Perrin *et al.* under revision). Ces résultats sont cohérents avec ceux obtenus dans des études précédentes (Roger *et al.* 2008a).

Dans un deuxième temps, des analyses génétiques ont démontré que la séquence minimale du promoteur est de 1kb pour avoir une activité transcriptionnelle. De plus, on sait que les gènes sont très conservés et qu'aucune substitution n'a été décelée dans la boîte TATA et le TSS (*Transcription Start Site*). La boîte TATA est une séquence d'ADN présente au niveau de la région promotrice et se situe en amont du premier nucléotide transcrit. Le TSS quant à lui, est le premier nucléotide de l'ADN qui est transcrit en ARN. L'étude de la diversité nucléotidique a mis en avant un faible nombre d'allèles et de génotypes pour les deux souches, et tous les individus sont homozygotes.

Sur la base génétique, les promoteurs *SmPoMucs* des souches Brésil et Guadeloupe ne présentent pas de différences pouvant expliquer les différences de transcription. Pour plus de détails sur cette partie, l'article en cours est en Annexe 5.

Au vu des résultats obtenus dans une étude précédente, établissant l'effet de la TSA sur la génération de nouvelles combinaisons de transcrits (Cosseau *et al.* 2010), nous avons décidé de comprendre l'implication de l'épigénétique, de nature chromatinienne, sur ces gènes *SmPoMucs*. C'est à ce niveau que ma contribution est intervenue.

L'influence de la TSA a confirmé l'augmentation de la transcription pour les groupes 1 et 2 mais pas pour le groupe 3.1(r1-r2). Nous avons étudié, par une approche ciblée en PCR quantitative, la structure chromatinienne autour des séquences promotrices de différents groupes, pour les deux souches. Pour cela, nous avons immunoprécipité l'ADN avec un anticorps spécifique de l'euchromatine (H3K9ac) et de l'hétérochromatine (H3K9me3). L'analyse ciblée a montré un fort enrichissement pour la marque H3K9ac dans les groupes 1 et 3.1(r1-r2) dans la souche Guadeloupe, ce qui signifie que les séquences promotrices de ces deux groupes sont plus euchromatiques que dans la souche Brésil. Inversement, la souche Brésil présente un fort enrichissement pour la marque H3K9me3 pour ces mêmes groupes, d'où une hétérochromatisation plus forte que dans la souche Guadeloupe (Figure 37).



**Figure 37: Enrichissement, obtenu en ChIP-PCRq, pour les marques H3K9ac et H3K9me3 des régions promotrices des gènes *SmPoMucs* (1, 3.1(r1-r2)) dans la souche Guadeloupe et la souche Brésil.** Ces données sont normalisées par rapport au gène  $\alpha$ Tub. Les profils chromatiniens sont exactement l'inverse d'une souche à l'autre, avec un fort taux d'euchromatisation dans la souche Guadeloupe (gris) et un fort taux d'hétérochromatisation dans la souche Brésil (noir) (Perrin *et al.* under revision).

Il est à noter que la séquence promotrice du groupe 3.1(r1-r2) présente une euchromatisation très forte, ce qui est cohérent avec les données obtenues lors du traitement à la TSA. Ce dernier ne semble pas influencer le niveau transcriptionnel. Cela peut être expliqué par le fait que l'acétylation de ce promoteur est saturée (Perrin *et al.* under revision).

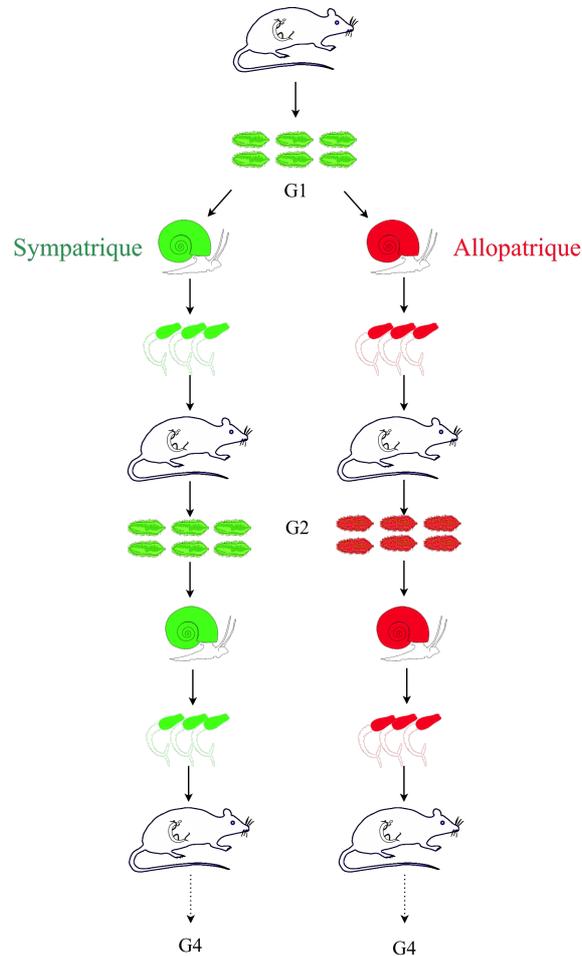
Cette étude, réalisée en collaboration avec les co-auteurs de la publication, a mis en avant l'absence de différence génétique entre les séquences promotrices des différents groupes de *SmPoMucs* pour les souches Brésil et Guadeloupe. Cependant, une forte différence épigénétique a été constatée, avec une forte euchromatisation des promoteurs des groupes 1 et 3.1(r1-r2) dans la souche Guadeloupe, et inversement une forte hétérochromatisation de ces groupes dans la souches Brésil. Ces résultats semblent suggérer une implication épigénétique, de type chromatinien, sur les gènes *SmPoMucs* et le polymorphisme de compatibilité (Perrin *et al.* under revision). Tous les détails sont dans la publication en cours (Annexe 5).

### 2.3. Évolution expérimentale de l'adaptation d'une souche de parasite *S. mansoni* à différentes souches d'hôte intermédiaire

#### 2.3.1. Traits d'histoire de vie : mesures phénotypiques

Afin de répondre à l'énigme 2 « Adaptation rapide de *S. mansoni* à l'Amérique du Sud », une évolution expérimentale est mise en place. Pour cela, une même souche de parasite *S. mansoni*, issue d'un isolat géographique brésilien (*SmBre*), est exposée à deux souches de mollusque *B. glabrata*, une souche sympatrique issue du même isolat géographique brésilien (*BgBRE*) et une souche allopatrique issue d'un isolat géographique différent, la Guadeloupe (*BgGUA*). Ces deux conditions d'infestation sont maintenues sur trois générations

successives afin de mieux comprendre les mécanismes adaptatifs de ce parasite lorsqu'il est soumis à un environnement pour lequel il n'est pas spécifique (Figure 38).



**Figure 38: Protocole expérimental de changement d'hôte intermédiaire (rappel).** Une même population initiale de miracidia, issue de la souche *S. mansoni* Brésil (*SmBre* en vert), infeste une population sympatrique de mollusque issue de la souche *B. glabrata* Brésil (*BgBRE* en vert) et une population allopatrique de mollusques issue de la souche *B. glabrata* Guadeloupe (*BgGUA* en rouge). Ces cycles sont maintenus dans les hôtes intermédiaires définis sur trois générations successives.

Pendant le cycle de vie de *S. mansoni*, différents traits de vie peuvent être quantifiés :

- Au stade adulte : le sexe-ratio des individus développés, l'infectivité des cercaires et la réponse inflammatoire de l'hôte définitif (par exemple le poids de la rate) ;
- Aux différents stades larvaires (miracidia et cercaires) : les sexes-ratios estimés par la technique de marqueurs PCR mis en place au laboratoire (Portela *et al.* 2010) ;
- Au stade cercaire : des traits de vie supplémentaires comme l'émission cercarienne et sa chronobiologie ;

- Au stade sporocyste (stade de développement au sein du mollusque, hôte intermédiaire) : la prévalence de mollusques infestés et l'intensité d'infestation.

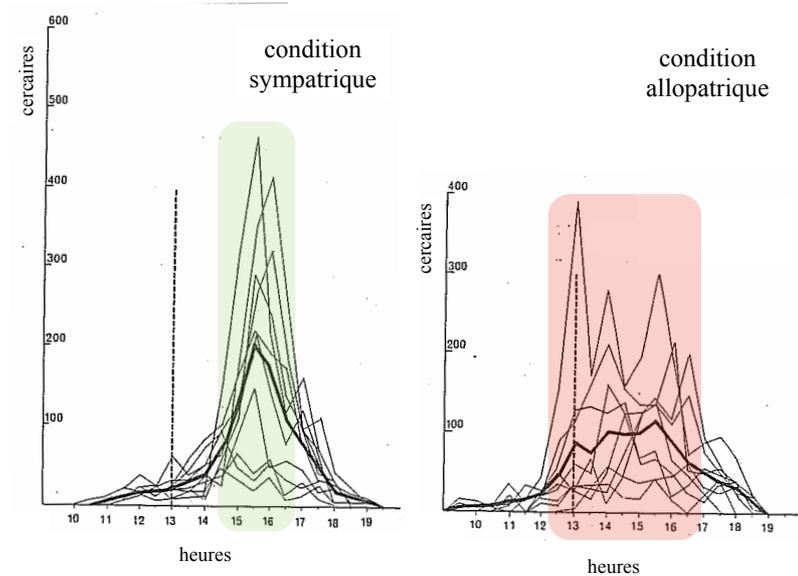
Durant ce projet, un maximum de traits de vie est quantifié précisément: les sexes-ratios (miracidia, cercaires, adultes), l'infectivité des cercaires, la prévalence de mollusques infestés et la réponse immunopathologique de l'hôte vertébré. Les traits de vie émission cercarienne et chronobiologie ne sont pas évalués mais des observations microscopiques sont faites.

Certains traits de vie ne montrent aucun changement entre les deux conditions allopatrique et sympatrique. Ces différents traits sont : l'infectivité des cercaires et la réponse immunopathologique de l'hôte vertébré et ce, entre les deux conditions sur les trois générations (Lepesant *et al.* in prep.) (Annexe 1).

En comparant les moyennes des infectivités entre la condition sympatrique et la condition allopatrique, il n'y a aucune différence significative entre les deux groupes et ce pour les trois générations. Mais en confrontant les variances, on distingue une seule différence significative entre les deux groupes, à la génération 2, avec une augmentation de la variance dans le groupe expérimental.

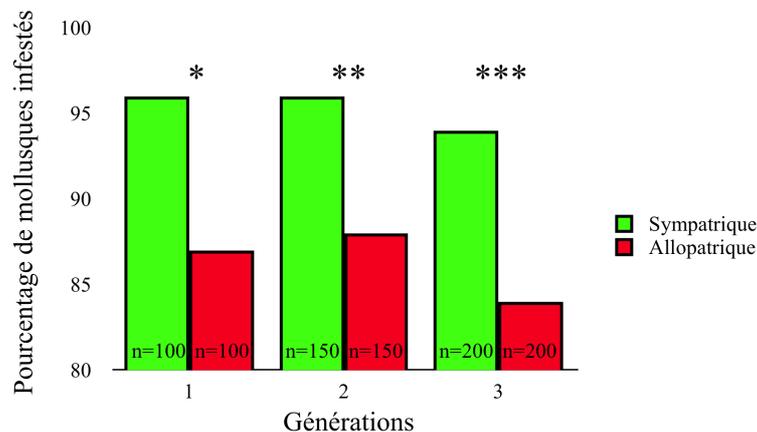
Aucune différence significative, dans la réponse inflammatoire, n'a été trouvée dans les valeurs moyennes et dans les variances (Lepesant *et al.* in prep.) (Annexe 1).

D'autres traits d'histoire de vie mesurés présentent, quant à eux, des différences entre la condition sympatrique et la condition allopatrique : la prévalence de mollusques infestés, les sexes-ratios, l'émission cercarienne et la chronobiologie. Ces deux derniers traits n'ont pas été quantifiés précisément mais nous avons noté que l'émission cercarienne de la condition allopatrique est beaucoup plus conséquente que la condition sympatrique (environ 100 fois plus) et que la chronobiologie d'émission montre un profil d'émission particulier si le parasite se développe au sein de l'hôte allopatrique. Lorsqu'une souche de parasite est confrontée à son hôte sympatrique, le profil d'émission est caractérisé par un unique pic sur une durée restreinte. Mais quand une souche de parasite se développe dans un hôte allopatrique, le profil est caractérisé par un aplatissement, avec la formation d'un plateau, confirmant une durée d'émission étalée dans le temps (Figure 39). Ces résultats sont en adéquation avec des études précédentes réalisées au laboratoire (Théron 1980).



**Figure 39: Profil d'émission cercarienne selon le développement du parasite au sein d'un hôte intermédiaire sympatrique ou allopatrique.** Dans la condition sympatrique (*SmBre/BgBRE*), le profil présente un pic distinct alors que dans la condition allopatrique (*SmBre/BgGUA*) le profil est caractérisé par un aplatissement. Adapté de (Théron 1980).

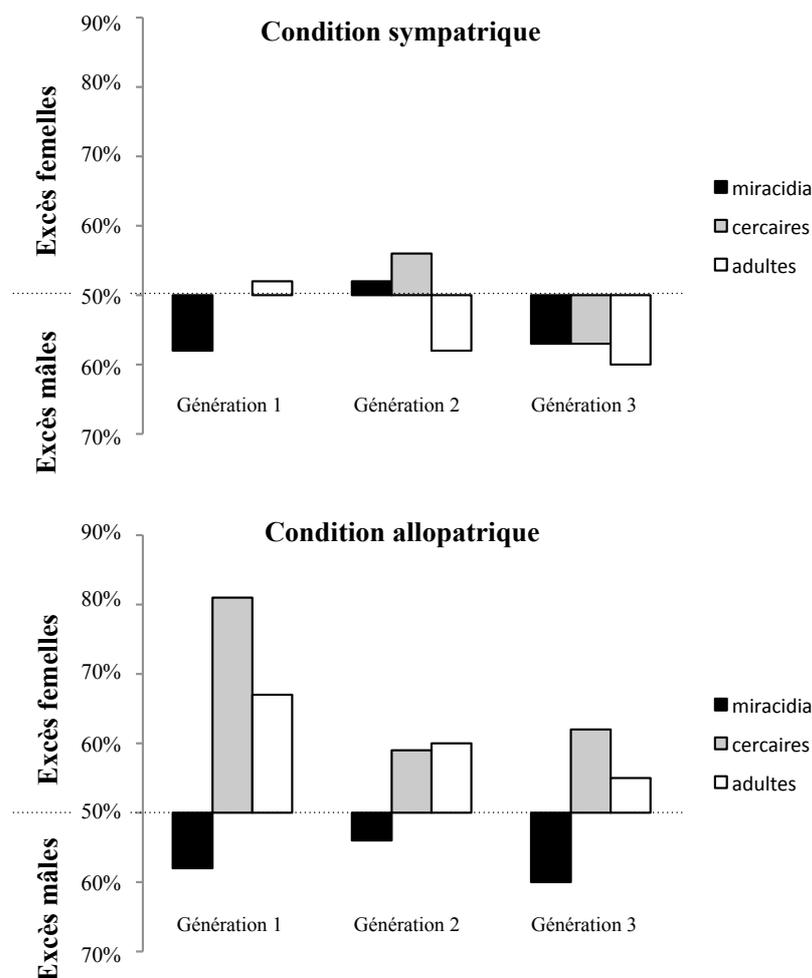
Pour le trait d'histoire de vie de prévalence des mollusques infestés, une valeur inférieure dans la condition allopatrique par rapport à la condition sympatrique est constatée sur les trois générations (Lepesant *et al.* in prep.) (Figure 40).



**Figure 40: Prévalences de mollusques infestés dans la condition sympatrique et la condition allopatrique.** La prévalence pour chaque groupe est stable pour les deux conditions mais la condition allopatrique (rouge) présente une prévalence plus faible que la condition sympatrique (vert) (Lepesant *et al.* in prep.). \* $p < 0.05$ .

Cette différence est significative pour les trois générations (G1 :  $p = 0.04$ ; G2:  $p = 0.0176$ ; G3:  $p = 0.0021$ ). Pour chaque condition, la prévalence reste inchangée entre les trois générations de la même condition (sympatrique  $p = 1, 0.55$  et  $1$  ; allopatrique  $p = 0.97, 0.366$  et  $0.61$ )

Le trait de vie le plus modifié, lorsque le parasite est soumis à un nouvel environnement, est le sexe-ratio à différents stades. Ainsi, lorsqu'une population de parasites est soumise à un environnement qui lui est propre, le sexe-ratio des individus est équilibré ou en faveur des mâles (Beltran & Boissier 2009). Mais lorsque cette même population de parasites (avec un sexe-ratio équilibré) est exposé à un environnement-hôte différent (souche allopatrique), on constate que le sexe-ratio est biaisé en faveur des femelles, et ce pour les stades cercaires et adultes. Au stade miracidia, post-reproduction sexuée, le sexe-ratio se rééquilibre ou retourne en faveur des mâles (Figure 41).



**Figure 41: Sexes-ratios des différents stades de développement dans la condition sympatrique et la condition allopatrique.** La condition allopatrique présente un sexe-ratio biaisé en faveur des femelles, aux stades cercaires et adultes (Lepasant et al. in prep.) (Annexe 1).

La représentation graphique montre les pourcentages de mâles et de femelles dénombrés. Ces valeurs sont calculées par un ratio de chaque sexe par rapport au nombre d'individus total. Il est alors possible de déterminer un excès en mâles ou en femelles lorsque le ratio n'est pas de 1 : 1 (un mâle pour une femelle), soit un sexe-ratio de 50%. Par exemple, dans la condition allopatrique, première génération, les miracidia présentent un excès de mâles, c'est-à-dire un sexe-ratio biaisé en faveur des mâles (60% de mâles et 40% de femelles). Lorsque cette population se développe au sein de l'hôte allopatrique, on constate que le sexe-ratio des cercaires et des adultes est biaisé femelle (80% et 65% de femelles contre 20% et 35% de mâles, respectivement).

La population initiale du parasite servant à infester deux populations de mollusques (sympatrique vs allopatrique), possède un sexe-ratio initial biaisé mâle. Lorsque cette population *SmBre* est confrontée à un hôte sympatrique (*BgBRE*), le sexe-ratio reste inchangé ou est biaisé mâle (1.1 fois plus de femelles), et ce pour les trois générations.

En revanche, si cette même population initiale est soumise à un environnement mollusque allopatrique (*BgGUA*), le sexe-ratio est dramatiquement changé avec environ 1.8 fois plus de femelle dans la population de cercaires issue de cette infestation allopatrique. Ce sexe-ratio biaisé en faveur des femelles est retrouvé au stade adulte avec une baisse du nombre de femelles, puisque l'inféctivité des cercaires femelles est moindre que celle des mâles (Boissier *et al.* 1999). Suite à la reproduction sexuée au sein de l'hôte définitif, le sexe-ratio de la population de miracidia retrouve un équilibre ou est biaisé en faveur des mâles. De ce fait, l'augmentation du nombre de femelle est présent après l'infestation par *S. mansoni* d'un hôte intermédiaire allopatrique. Ce biais est retrouvé pour les trois générations successives de la condition allopatrique et peut persister sur sept générations (Théron, communication personnelle).

Les changements dans certains traits d'histoire de vie nécessitent un questionnement sur les modifications génétiques. Pour investir cela, nous avons génotypé entre 24 et 32 individus, aux stades cercaires et miracidia, pour chaque condition et ce pour les trois générations successives. Ce génotypage, réalisé avec 14 marqueurs microsatellites, montre que les populations sont à l'équilibre d'Hardy Weinberg, que l'indice de différenciation entre les populations est faible ( $F_{st}$  moyen 1.11%), qu'il n'existe pas de différence de diversité génétique entre les différentes populations (pourcentage d'hétérozygotie entre 1.94% et 8.1%,

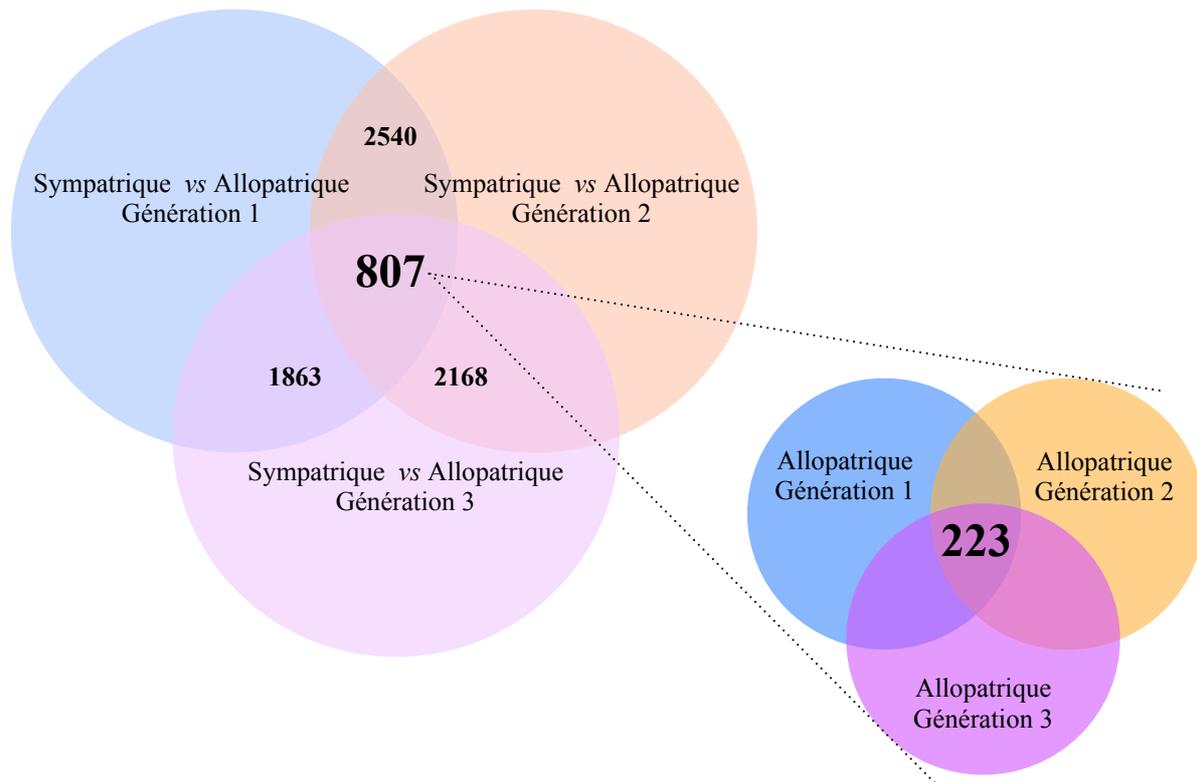
H=5.35, 11 ddl, p=0.94) et que le nombre d'allèles n'est pas significativement différent entre les populations analysées (au maximum 3 allèles, H=5.48, 11 ddl, p=0.94).

Tous ces résultats vont être prochainement publiés (Lepesant *et al.* in prep.). L'article en préparation est en Annexe 1.

### 2.3.2. Analyse transcriptomique (RNA-seq) et épigénétique (ChIP-seq)

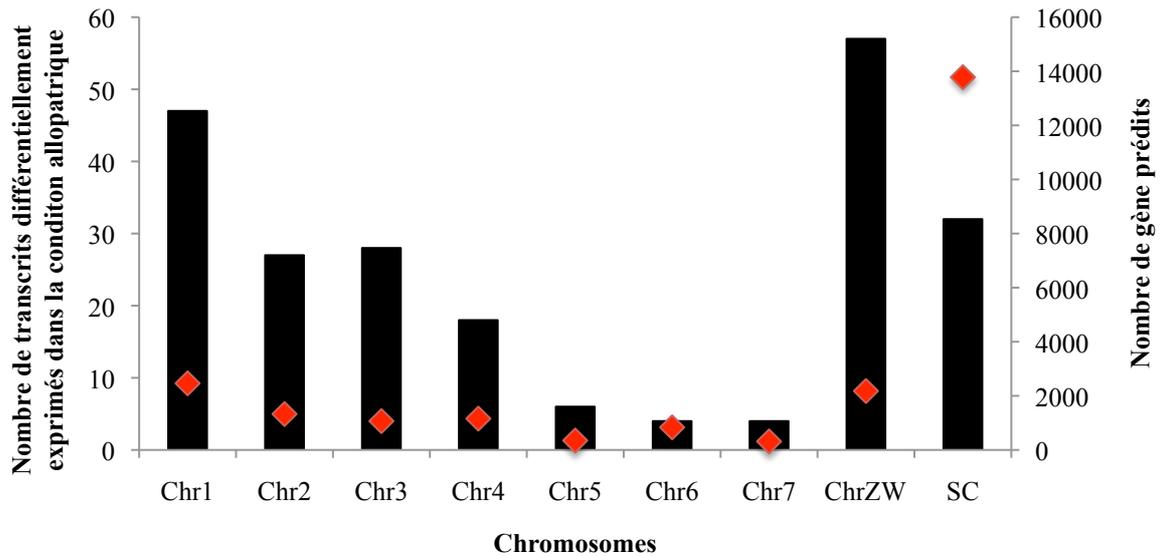
Suite aux résultats obtenus lors de l'évolution expérimentale montrant la modification de traits phénotypiques (prévalence, sexe-ratio, chronobiologie d'émission) sans modification génétique, nous avons décidé d'investir des analyses de transcriptomique, au stade cercaire. Pour cela, une extraction d'ARN total a été mise au point dans le but d'effectuer un séquençage massif. Cette technique a été optimisée pour obtenir une quantité et une qualité suffisantes pour les analyses de RNA-seq. Le séquençage massif a été réalisé sur le stade cercaire, pour les trois générations de chaque condition. Après un nettoyage des séquences et l'alignement de celles-ci sur le génome de référence, des analyses de différentiel d'expression ont été menées grâce au logiciel CuffDiff. Ces analyses préliminaires sont encore en cours d'étude mais elles nous ont d'ores et déjà permis d'identifier des transcrits différentiellement exprimés entre les conditions et au sein de chaque condition, pour les trois générations.

Au vu des résultats obtenus par le logiciel, un grand nombre de transcrits sont différentiellement exprimés entre la condition sympatrique et la condition allopatrique pour chaque génération, autour de 2000 transcrits par génération. Seuls 807 transcrits sont communs et différentiellement exprimés entre les deux conditions, toutes générations et conditions confondues. Parmi ces 807 transcrits, seuls 223 transcrits sont communs aux trois générations et présentent un différentiel d'expression dans la condition allopatrique (Figure 42).



**Figure 42: Diagramme de Venn des transcrits différentiellement exprimés entre la condition sympatrique et allopatrique.** Un grand nombre de transcrit ( $\sim 2000$  transcrits) est différentiellement exprimé entre les deux conditions pour chaque génération mais seuls 807 sont communs aux trois générations, toutes générations confondues. Parmi ces 807 transcrits, seuls 223 sont différentiellement exprimés dans les trois générations.

Les transcrits différentiellement exprimés dans la condition allopatrique, par rapport à la condition sympatrique, sont répartis entre les différents chromosomes. Une certaine proportion (environ 14%) est identifiée sur des *scaffolds* non chromosomiques (SC). Ces *scaffolds* font partie de l'assemblage du génome v5.2 mais n'ont pas pu être rattachés à un chromosome particulier. De ce fait, ils ne sont pas identifiés parmi les 7 chromosomes autosomaux ou les chromosomes sexuels regroupés dans le groupe ZW (Figure 43).



**Figure 43: Répartition des transcrits différentiellement exprimés dans la condition allopatrique par rapport à la condition sympatrique.** Ces transcrits sont répartis sur les différents chromosomes. Les chromosomes sexuels sont regroupés dans un même groupe (ZW) et une proportion de scaffolds est identifiée comme non chromosomiques (SC) dans l'annotation v5.2 du génome de *S. mansoni*. L'histogramme (noir) représente le nombre de transcrits différentiellement exprimés dans la condition allopatrique, pour chaque chromosome et la partie non identifiée (SC). Les points rouges représentent le nombre de gènes prédits par chromosome et dans la région non annotée (SC).

La répartition, des transcrits différentiellement exprimés dans la condition allopatrique, ne semble pas être dépendante du nombre de gènes prédits par chromosome. Il est à noter que le nombre de gènes prédits pour les chromosomes sexuels regroupe les gènes Z et le gènes W, puisque ces deux chromosomes sont identifiés dans un même groupe.

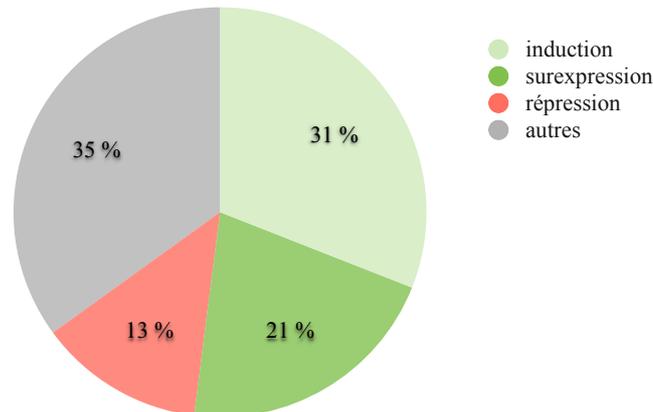
Différents profils d'expression sont retrouvés selon les générations. Nous avons observé que les niveaux d'expression peuvent être les mêmes à travers les trois générations (145 sur 223). Les analyses sont réalisées en comparant l'activité transcriptionnelle de la condition allopatrique par rapport à la condition sympatrique, pour chaque génération. Dans les profils constants sur les trois générations, nous avons identifié différents phénomènes : (i) d'induction, le transcrit n'est jamais exprimé dans la condition sympatrique mais l'est dans la condition allopatrique ; (ii) de surexpression, le transcrit est surexprimé dans la condition allopatrique ; (iii) de répression, les transcrits sont réprimés dans la condition allopatrique par rapport au niveau de transcription retrouvé dans la condition sympatrique (Tableau 8). D'autres profils présentent des niveaux d'expression variables entre les trois générations. Nous avons observé 15 profils différents. Par exemple un transcrit peut être réprimé à la génération 1, puis induit à la génération 2 et enfin surexprimé à la génération 3, dans la condition allopatrique par rapport à la condition sympatrique. Ou encore, le transcrit est

réprimé jusqu'à la génération 2 incluse puis est surexprimé à la génération 3 allopatrique. Les différents profils observés sont synthétisés dans le Tableau 8.

**Tableau 8: Synthèse des différents profils d'expression observés dans la condition allopatrique par rapport à la condition sympatrique.** *L'induction a lieu dans la condition allopatrique quand aucune transcription n'est retrouvée dans la condition sympatrique. La surexpression dans la condition allopatrique signifie que le transcrit est exprimé dans la condition sympatrique et surexprimé dans la condition allopatrique. La répression est présente dans la condition allopatrique en comparaison au taux de transcription observé dans la condition sympatrique.*

Génération 1	Génération 2	Génération 3	Nombre de transcrits concernés	Pourcentages
Niveau d'expression sur les 3 générations				
Induction	Induction	Induction	69	65 %
Surexpression	Surexpression	Surexpression	29	
Répression	Répression	Répression	47	
Niveaux d'expression différents selon la génération				
Induction	Induction	Surexpression	16	35%
Induction	Surexpression	Induction	4	
Induction	Surexpression	Répression	2	
Induction	Surexpression	Surexpression	4	
Répression	Induction	Induction	3	
Répression	Induction	Répression	1	
Répression	Induction	Surexpression	1	
Répression	Surexpression	Surexpression	1	
Répression	Surexpression	Répression	5	
Répression	Surexpression	Surexpression	3	
Surexpression	Surexpression	Induction	8	
Surexpression	Induction	Surexpression	10	
Surexpression	Répression	Répression	10	
Surexpression	Surexpression	Répression	3	
Surexpression	Répression	Surexpression	7	
			223	

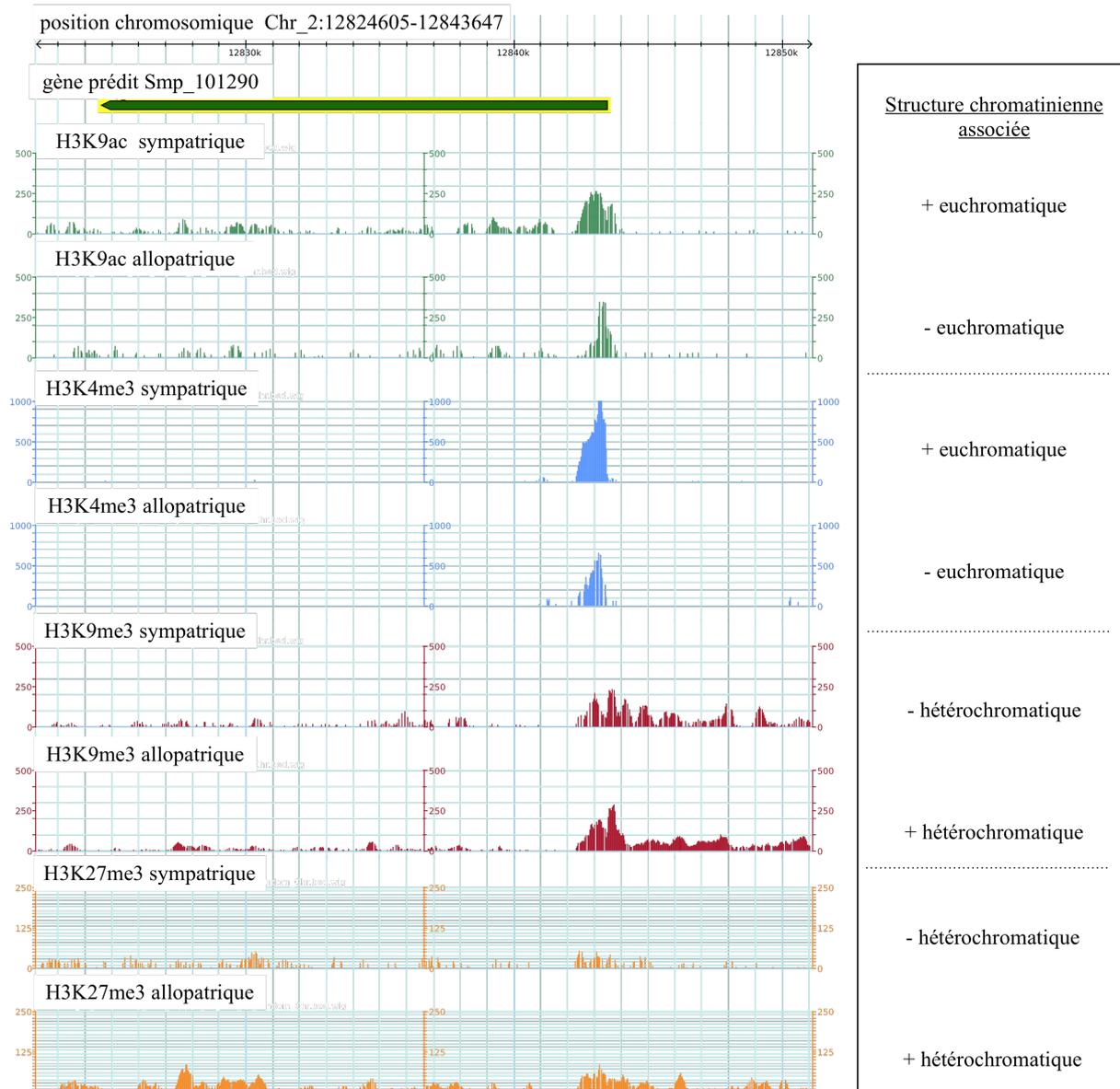
Parmi ces 223 transcrits, 65% (145 gènes) ont un profil d'expression similaire au cours des trois générations : 31% sont induits, 21% sont surexprimés et 13% sont réprimés dans la condition allopatrique par rapport à la condition sympatrique. Les 35% restants (soit 78 transcrits) ne présentent pas un niveau d'expression constant entre les trois générations et nous avons pu noter 15 différents profils (Figure 44).



**Figure 44: Diagramme de répartitions des profils d'expression observés dans la condition allopatrique pour 223 transcrits différentiellement exprimés dans les trois générations.** L'induction, la surexpression et la répression sont présentes dans la condition allopatrique sur les trois générations. La catégorie « autres » comprend tous les profils d'expression n'étant pas constants sur les trois générations allopatriques.

Pour ces transcrits ayant une expression stable à travers les générations, nous avons observé visuellement les profils chromatiniens associés, pour les marques euchromatiques et hétérochromatiques. Certaines différences transcriptionnelles entre les deux conditions semblent associées à des modifications de certaines marques. Parmi les 145 transcrits ayant un profil d'expression similaire sur les trois générations, 27 (soit ~18%) présentent une différence pour au minimum une marque et au maximum les quatre (H3K9ac, H3K4me3, H3K9me3, H3K27me3).

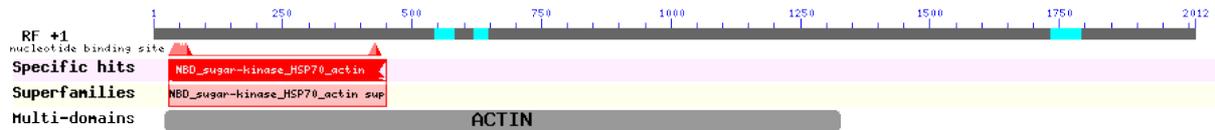
Prenons l'exemple du gène Smp\_101290 qui est exprimé approximativement trois fois moins dans la condition allopatrique que dans la condition sympatrique. Il présente un enrichissement plus fort pour les marques dites euchromatiques (H3K9ac et H3K4me3) dans la condition sympatrique, et inversement un enrichissement plus fort pour les marques dites hétérochromatiques (H3K9me3 et H3K27me3) dans la condition allopatrique. Ce gène semble donc plus euchromatique dans la condition où il est le plus transcrit, et plus hétérochromatique dans la condition où il est réprimé. Les données de transcription semblent montrer une corrélation avec les données épigénétiques obtenues en ChIP-seq (Figure 45).



**Figure 45: Exemple de transcrite présentant un différentiel d'expression associé à des modifications de la structure chromatinienne dans la condition allopatrique.** Le gène *Smp\_101290* est approximativement trois fois moins exprimé dans la condition sympatrique que dans la condition allopatrique. Il présente un enrichissement pour les marques H3K9ac (vert) et H3K4me3 (bleu) plus fort dans la condition sympatrique. Et un enrichissement pour les marques H3K9me3 (rouge) et H3K27me3 (orange) dans la condition allopatrique. Le gène est donc plus euchromatique dans la condition sympatrique, où il est transcrit trois fois plus que dans la condition allopatrique (hétérochromatisation plus forte).

Ce gène, *Smp\_101290*, est annoté comme étant codant pour une protéine liée à l'actine (ARP2) et une analyse des domaines conservés a mis en avant la présence d'un domaine (Figure 46):

- Le domaine NBD de liaison nucléotidique de kinase, protéine HSP (*Heat Shock Proteins*) et de la superfamille de l'actine, incluent les liaisons à l'actine et les molécules chaperones HSP70.



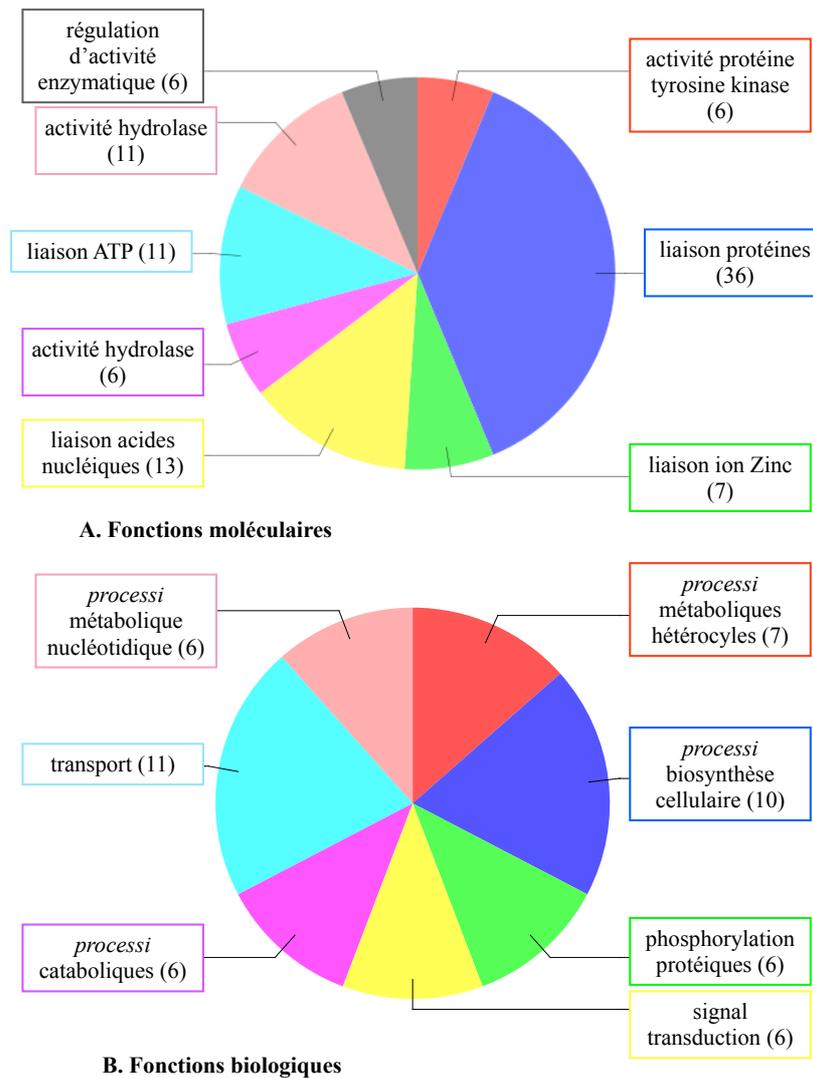
**Figure 46: Domaines conservés d'un gène différentiellement exprimé dans la condition allopatrique, associé à des modifications de la structure chromatinienne.** Le domaine conservé est le domaines NBD (rouge).

De plus, cette analyse a mis en avant des similarités de séquences avec un homologue de ACTR2 chez les bactéries ( $p = 2.8^e-157$ ) et *Bos Taurus* ( $p = 1.9^e-156$ ). Ce gène code pour une protéine connue pour être une des principales constituantes du complexe ARP2/3 qui est généralement situé à la surface cellulaire et est indispensable à la motilité cellulaire à travers le mouvement d'actine. Ce gène pourrait être impliqué dans la mobilité cellulaire.

En revanche, 21 transcrits (soit ~14%) différentiellement exprimés (parmi les 145 ayant un profil d'expression similaire entre les générations), ne présentent pas de différence dans les profils chromatinien et pour une grande proportion (soit ~67%) nous n'avons pas de données sur la structure chromatinienne.

Ces observations visuelles seront reprises ultérieurement par la technique de ChIP-PCRq sur les mêmes échantillons que ceux séquencés. Cette analyse ciblée nous permettra de confirmer les structures chromatinien autour des transcrits identifiés dans la condition allopatrique.

Lors des analyses préliminaires des 223 transcrits identifiés comme différentiellement exprimés dans la condition allopatrique vis à vis de la condition sympatrique, nous avons utilisé le logiciel d'annotation automatique Blast2Go pour mettre en évidence que ces transcrits sont impliqués dans différentes grandes fonctions moléculaires (régulation d'activité enzymatique, activité hydrolase, activité ligase...) (Figure 47 A) ou de grandes fonctions biologiques (*processi* métaboliques, transport, *processi* de biosynthèse cellulaire...) (Figure 47 B).



**Figure 47: Grandes fonctions moléculaires et biologiques dans lesquelles sont impliqués les transcrits différentiellement exprimés dans la condition allopatrique.**

Cette annotation nous a permis d'identifier un gène transcrit approximativement 2 fois plus transcrit dans la condition allopatrique que dans la condition sympatrique, sur les trois générations. Ce gène, Smp\_140390, est annoté comme un gène codant pour une enzyme N-méthyltransférase agissant sur les lysines des histones (histone-lysine N-méthyltransférase).

Une analyse des domaines conservés a mis en avant la présence de deux domaines (Figure 48):

- Le domaine RRM qui est hautement conservé chez les eucaryotes et qui est impliqué dans les *processi* d'expression post-transcriptionnelle des gènes incluant les ARN messagers, les ARN ribosomiaux, l'exportation et la stabilité de l'ARN ;
- Le domaine SET qui est impliqué dans les interactions protéines-protéines, ainsi que les activités de protéine-lysine méthyltransférase.



**Figure 48: Domaines conservés d'un gène différentiellement exprimé dans la condition allopatrique.** Les domaines conservés sont les domaines RRM (rouge) et SET (bleu).

De plus, cette analyse a établi des similarités de séquences avec la N-lysine-méthyltransférase ( $p = 5.9^{\circ}-94$ ) et le gène SETD1B (SET domain containing 1B) ( $p = 6.8^{\circ}-92$ ). De nombreuses histones méthyltransférases contiennent un domaine SET et catalysent l'addition de groupement méthyl sur des résidus lysine. Ce gène SETD1B code pour une méthyltransférase spécifique de la lysine 4 sur l'histone H3 (H3K4) (Lee *et al.* 2007). La marque H3K4 méthylée représente un indice d'activation épigénétique de la transcription. Cette étude doit être approfondie mais elle met tout de même en avant la surexpression d'un gène prédit, similaire au gène SETD1B, chez les individus issus de la condition allopatrique, sur trois générations.

Une étude ciblée de la transcription et de la structure chromatinienne autour de ce gène nous permettrait de confirmer les résultats issus du séquençage massif.

Ce gène semble être un bon candidat pour appréhender l'influence des changements environnementaux sur la réponse adaptative rapide du parasite, caractérisée par des changements transcriptomiques et épigénétiques de nature chromatinienne.

Les données issues du RNA-seq sur les trois générations sont encore en cours d'analyses. Il serait intéressant de vérifier les changements transcriptionnels au sein de chaque condition. Lorsqu'un transcrite est induit puis surexprimé dans la condition allopatrique, il est possible de vérifier si le taux de transcription augmente réellement ou si l'augmentation de la transcription est liée à la condition sympatrique. En effet, nos analyses préliminaires sont basées sur la comparaison un à un des transcrits entre la condition allopatrique et la condition sympatrique, pour chaque génération. Si la transcription diminue, d'une génération à l'autre dans la condition sympatrique, mais reste inchangée dans la condition allopatrique, la différence sera identifiée comme une expression plus forte entre les générations allopatriques. De plus, des validations ciblées par PCR quantitative seront mises en œuvre pour les transcrits identifiés comme ayant une activité transcriptionnelle dramatiquement changée par rapport à la condition sympatrique, ou codant pour des gènes impliqués dans de grandes fonctions.

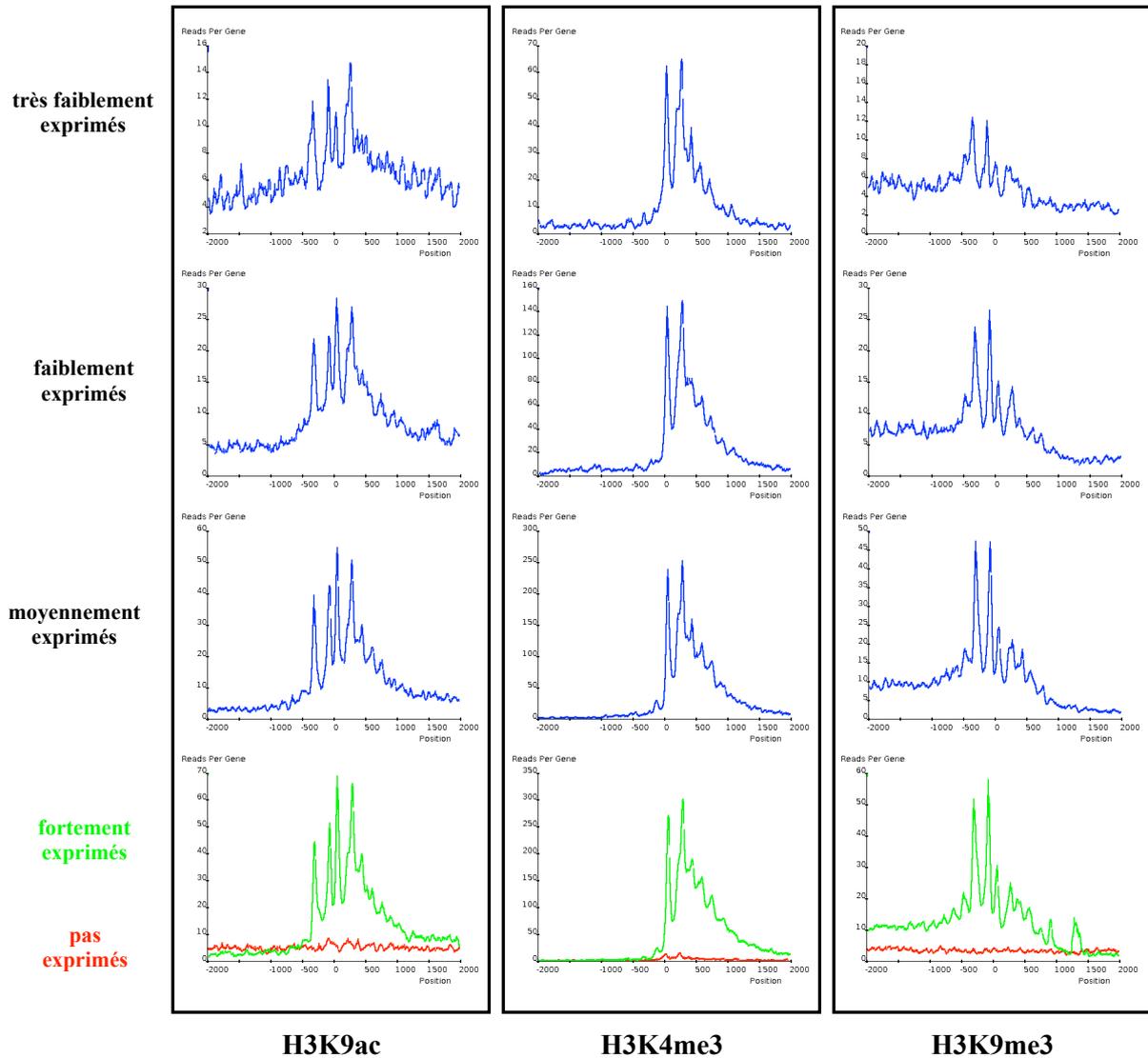
Pour les analyses épigénétiques basées sur une approche globale, nous nous sommes focalisés sur les marques d’histones H3K9ac, H3K4me3, H3K9me3 et H3K27me3. L’étude s’est portée sur ces différentes marques car elles sont conservées chez *S. mansoni*. De ce fait, nous avons pu utiliser des anticorps commerciaux, après validation. L’ADN immunoprécipité pour chaque marque a été séquencé par la technique Illumina, en collaboration avec Montpellier GenomiX, pour les deux conditions, uniquement pour la génération trois. Les analyses bioinformatiques menées nous ont permis d’identifier différents profils, pour chaque marque, dépendant du taux de transcription. Les profils sont classés en fonction de leur niveau transcriptionnel. Sur la base des données obtenues par le logiciel CuffDiff, cinq classes sont formées selon la valeur des FPKMs (*Fragment Per Kilobase of exons per Million reads mapped*). Les FPKMs correspondent au nombre de séquences alignées par gène, normalisé par le nombre de séquences totales dans chaque condition et par la taille des transcrits. Les différentes classes formées, en fonction du nombre de FPKMs sont synthétisées dans le Tableau 9.

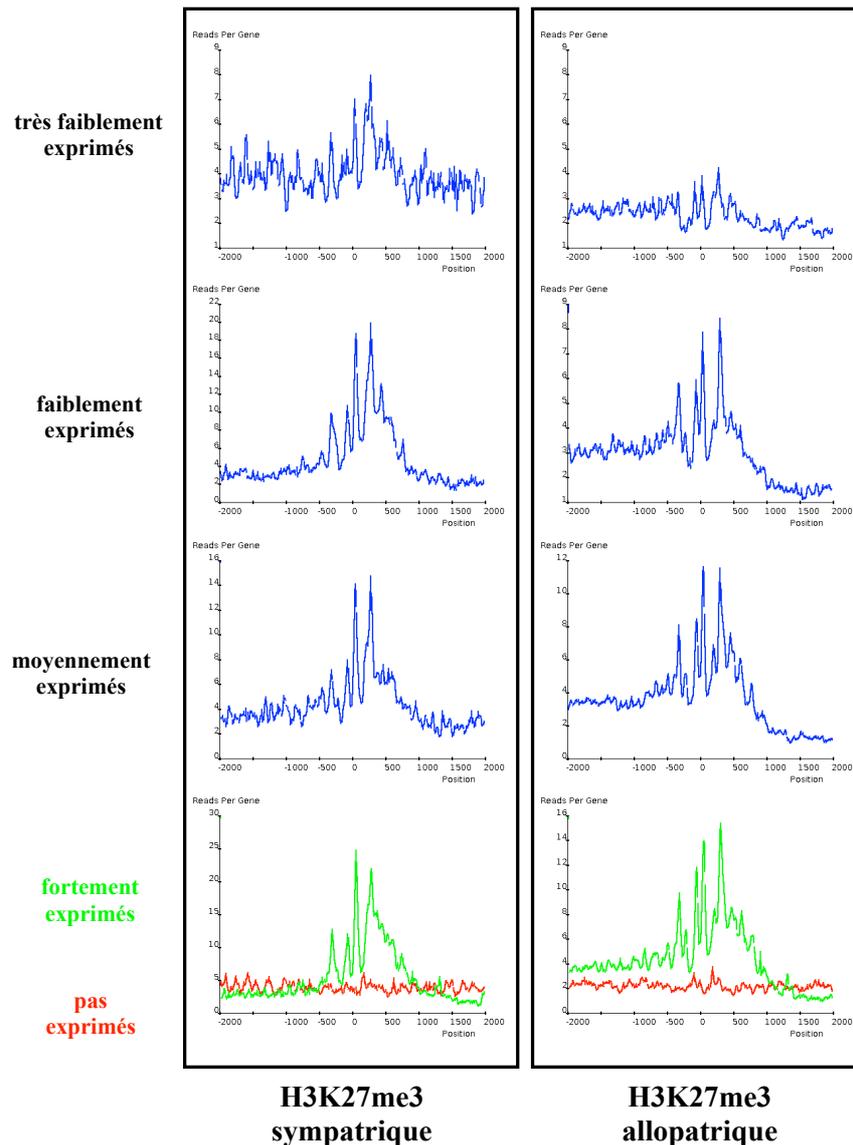
**Tableau 9: Différentes classes définies en fonction du nombre de FPKMs.**

Niveau d’expression	FPKMs
Pas d’expression	0
Expression très faible	1-5
Expression faible	5-10
Expression moyenne	10-1000
Expression forte	1000-10000

L’analyse de ces différentes marques nous permet de définir des profils types pour chaque marque, en fonction de la transcription à l’échelle du génome. Ces profils ont été construits autour de la région du TSS (*Transcription Start Site*) pour une moyenne de 500 gènes. Cependant, nous ne pouvons pas comparer les taux d’enrichissement entre les marques. En effet, les profils ne sont pas normalisés entre chaque marque mais uniquement entre chaque condition. Les gènes étudiés sont d’une taille supérieure à 2 kb, pour que les profils ne soient pas plus grands que la taille du gène, et ils n’ont pas d’autres gènes présents 3kb en amont et en aval afin d’éviter la superposition des marques. Suite à la normalisation des données entre les deux conditions, nous avons constaté que les profils pour les marques H3K9ac, H3K4me3 et H3K9me3 sont identiques entre les deux conditions. Seule la marque H3K27me3 présente

un profil différent pour chaque condition. Ceci est dû à la qualité des séquences obtenues pour la condition sympatrique est moins bonne que celle de la condition allopatrique. De ce fait, les profils ne sont pas identiques et sont organisés par marque selon le taux de transcription globale (Figure 49).



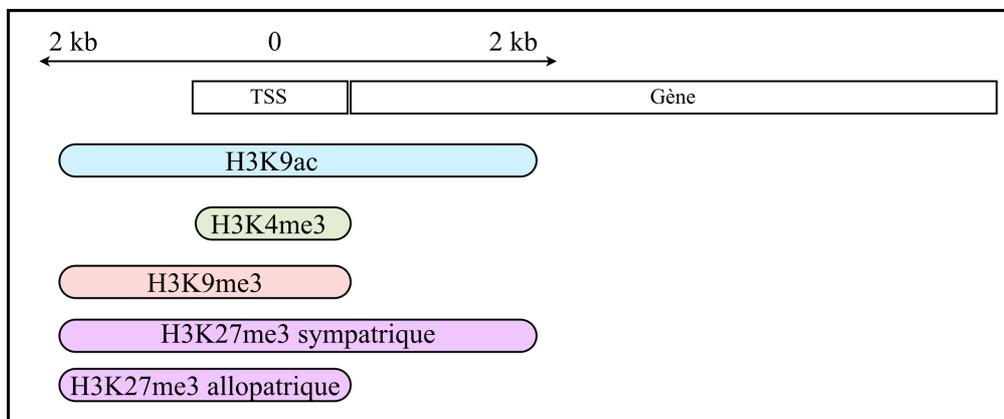


**Figure 49: Profils types des marques d'histones (H3K9ac, H3K4me3, H3K9me et H3K27me3), dans la région promotrice, en fonction du niveau d'expression à l'échelle du génome.** Les profils des marques H3K9ac, H3K4me3 et H3K9me3 sont les mêmes dans la condition sympatrique et la condition allopatrique, alors que la marque H3K27me3 présente des différences entre les deux conditions, dues à des différences de qualité dans les séquences.

Il est possible de constater que le profil des marques dépend du niveau transcriptionnel global associé. En effet, pour les gènes faiblement exprimés, l'enrichissement pour chaque marque est plus faible que pour les gènes fortement exprimés. Les gènes non exprimés présentent une absence d'enrichissement pour toutes les marques, mais à partir d'une très faible expression on peut relever l'apparition d'un pic d'enrichissement pour les quatre marques. De plus, l'enrichissement pour chaque marque plafonne rapidement à partir d'une expression moyenne (à partir de 10 FPKMs).

Les marques ont différents profils d'enrichissement (Figure 50):

- H3K9ac : pic d'enrichissement en amont du TSS, qui perdure dans les introns et les exons du gène ;
- H3K4me3 : pic d'enrichissement uniquement au niveau du TSS, qui ne se maintient pas dans les introns et les exons du gène ;
- H3K9me3 : pic d'enrichissement en amont du TSS, au niveau du TSS, mais qui ne perdure pas dans les introns et les exons du gène ;
- H3K27me3 : pic d'enrichissement en amont, dans le TSS et dans les introns/exons du gène dans la condition sympatrique ; alors que dans la condition allopatrique la marque ne perdure pas dans la région introns/exons.



**Figure 50: Schéma de la répartition des différentes marques d'histones (H3K9ac, H3K4me3, H3K9me3 et H3K27me3) autour du Transcription Start Site (TSS).** Toutes les marques sont enrichies au niveau du TSS mais H3K9ac est présent en amont et en aval du TSS, tout comme H3K27me3 dans la condition sympatrique. H3K4me3 est identifiée uniquement dans la région du TSS. H3K27me3 dans la condition allopatrique est observée en amont du TSS.

Dans notre modèle, les marques H3K9me3 et H3K27me3 ne sont pas forcément associées à des profils de transcription pour les gènes faiblement exprimés. En effet, plus le gène est transcrit, plus l'enrichissement pour ces marques augmente.

Les analyses des données issues du ChIP-seq pour la génération trois sont encore en cours d'analyse mais les résultats préliminaires nous ont permis d'ores et déjà de mettre en avant une corrélation entre les marques d'histone et le niveau transcriptionnel à l'échelle du génome. Des analyses RNA-seq et ChIP-seq seront, par la suite, validées par PCR quantitative sur des gènes cibles.

Cette évolution expérimentale montre une influence directe de l'environnement hôte sur certains phénotypes du parasite (ex : prévalence et sexe-ratio). L'apparition des nouveaux phénotypes ne semble pas être corrélée à des modifications génétiques basées sur une analyse de marqueurs neutres. Cependant, le changement d'environnement entraîne un différentiel de transcription entre les individus issus de la condition sympatrique et ceux issus de la condition allopatrique. Les analyses sont encore en cours, mais il nous est tout de même possible de constater l'influence de l'environnement sur l'induction ou la répression de certains transcrits dans la condition allopatrique vis à vis de la condition sympatrique. La problématique principale de ces travaux était de savoir si l'environnement influence de manière aléatoire ou ciblée dans le génome. Au vu des résultats obtenus par le RNA-seq, nous avons mis en évidence qu'un grand nombre de transcrits est différentiellement exprimé à chaque génération (~10% des gènes totaux) mais qu'une petite proportion est similaire entre les trois générations (~1% des gènes totaux). Nous pouvons donc conclure que les changements, liés à une modification de l'environnement, sont principalement aléatoires dans le génome, avec un faible pourcentage de modifications ciblées et que les différences de transcription, à l'échelle du génome, peuvent être associées à des profils chromatiniens puisque les marques d'histones semblent être dépendantes des niveaux transcriptionnels.

Enfin, lors de nos travaux nous avons identifié un gène candidat Smp\_140390 qui peut être impliqué dans la modulation de la structure chromatinienne, plus particulièrement sur la méthylation des lysines des histones. Ce gène est deux fois plus transcrit dans la condition allopatrique que dans la condition sympatrique. Une analyse plus approfondie de ce gène candidat, ainsi que des autres transcrits différentiellement exprimés, sera indispensable pour corréler les données phénotypiques, transcriptomiques et épigénétiques.

### 3. Énigme 3 « Différenciation sexuelle en l'absence de gène W-spécifique dans un système ZW »

La dernière énigme est : comment existe-t-il une différenciation sexuelle, en l'absence de gène W-spécifique, dans un système ZW ?

Dans un premier temps, suite au séquençage du génome de deux souches de laboratoire, nous avons réalisé une étude approfondie des séquences répétitives. Nous avons ainsi proposé un nouvel assemblage et une base de données classifiant ces séquences.

Dans un deuxième temps, nous nous sommes focalisés sur l'étude de séquences répétitives spécifiques du chromosome sexuel femelle, le W. En effet, après analyse de séquences issues de séquençage massif, nous avons identifié les différences entre les deux chromosomes sexuels. Ainsi, la région femelle spécifique est composée principalement de séquences répétitives.

### 3.1. Analyse des séquences répétitives et *de novo* assemblage de ces séquences

Les séquences répétitives étaient anciennement considérées comme des séquences poubelles. Depuis quelques années, des études prennent en compte ces séquences pour répondre à des questions biologiques ou évolutives. Suite au séquençage du génome de *S. mansoni* (v3.1), le logiciel RepeatScout a été utilisé afin d'identifier les séquences répétitives présentes dans le génome. La prédiction de ces séquences donne une estimation de 40% de séquences répétées, dont 20% correspondent à des éléments transposables (Berriman *et al.* 2009).

Depuis une trentaine d'années, plusieurs méthodes ont été développées pour identifier les séquences répétitives. Quand le génome de *S. mansoni* a été publié, des programmes conventionnels de prédiction ont permis de mettre en avant des séquences répétitives additionnelles. Certaines ont été identifiées par l'institut J. Craig Venter (1225) et cinquante cinq séquences répétées sont disponibles sur GenBank.

En procédant au séquençage massif du génome de deux souches (Brésil et Guadeloupe) de notre laboratoire, nous avons décidé de raisonner différemment pour identifier les séquences répétitives présentes dans le génome. Pour assembler le pseudogénome de séquences répétées, nous avons utilisé un total de 38 004 342 séquences de 36 pb générés par le Genome Analyser II (Illumina). Les séquences sont disponibles sur NCBI SRA (NCBI Sequence Read Archive, SRA012151.6 et SRA043796.1). Les séquences sont alignées contre le génome de référence v3.1 avec les logiciels SOAP2/SOAPaligner (Li *et al.* 2009). Ces logiciels n'alignent que les séquences avec une unique occurrence et nous donne la possibilité d'isoler celles qui sont répétitives. Les séquences présentant plus d'une occurrence dans les génomes, représentent 12 535 613 séquences soit environ 33% des séquences totales. Ces séquences répétitives sont assemblées par le logiciel Velvet 0.7.01 (Zerbino & Birney 2008) et un logiciel d'assemblage de séquences est utilisé pour générer au final 8594 *contigs*. Cela nous permet de concevoir un pseudogénome analysé indépendamment de l'assemblage des séquences uniques. Nos résultats révèlent un grand nombre de séquences répétitives additionnelles.

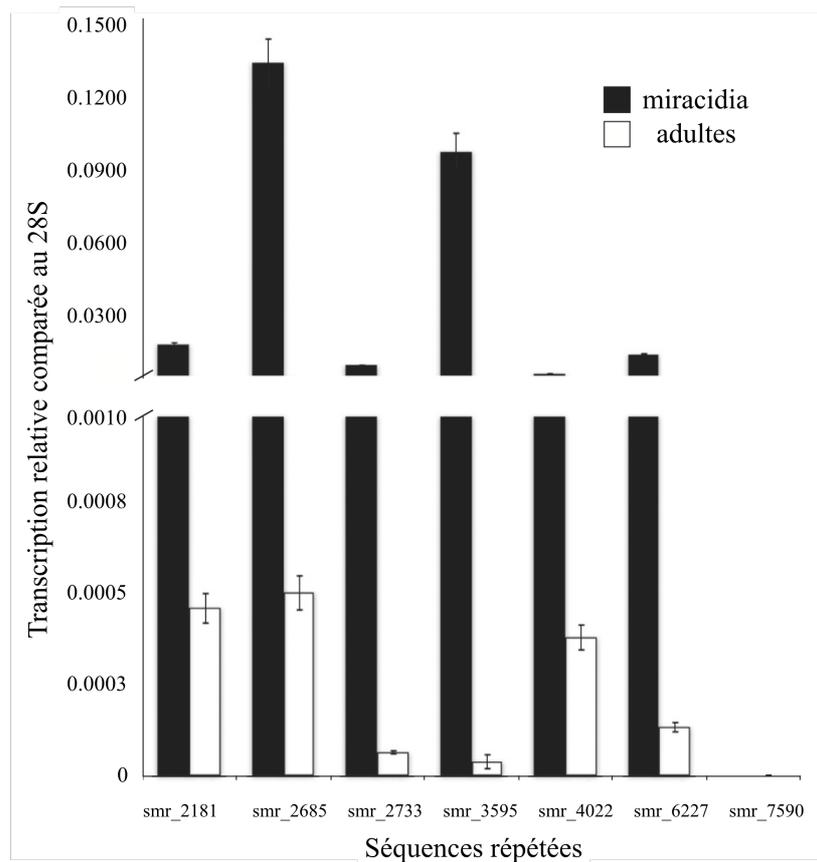
Chaque séquence répétitive est assemblée individuellement et une première annotation de ces *contigs* est réalisée par Blast (Altschul *et al.* 1990), Censor/Repbase (Kohany *et al.* 2006), TEclass (Abrusan *et al.* 2009) et Tandem Repeats Finder (Benson 1999). Des validations par PCR quantitative sont menées sur 9 *contigs* choisis pour : leur absence d'annotation avec le logiciel Censor, le fait qu'elles soient exclusives et abondantes dans le génome de *S. mansoni*. De plus, basés sur une analyse des banques ESTs, ces 9 *contigs* semblent présenter des niveaux d'expression différents. Sur 7 des 9 *contigs*, le résultat obtenu confirme notre assemblage. En effet, smr\_2181, smr\_2685, smr\_2733, smr\_3000, smr\_3595, smr\_4022, and smr\_6227 présentent une bande à la taille attendue. Pour deux *contigs* (smr\_3000 and smr\_3826) une bande supplémentaire apparaît, à un poids moléculaire double de la taille attendue, suggérant des répétitions en tandem. Grâce à des analyses *in silico*, nous avons confirmé 6531 séquences répétitives (76% des séquences identifiées comme répétées) comme ayant deux ou plusieurs occurrences dans le génome et aucune annotation de gènes ou protéines.

En effectuant des comparaisons de séquences, par l'outil Blast, nous avons identifié 306 *contigs* comme correspondant à des gènes ou des familles de gènes, dont 40 gènes mitochondriaux. Par ailleurs, 1332 séquences (autres que celles identifiées comme gènes) sont des séquences uniques et 230 séquences sont absentes du génome de référence (v3.1). En plus de ces séquences, nous en avons ôté 10 supplémentaires suspectées d'être des contaminants issus du contact avec les hôtes intermédiaire et définitif ou avec l'eau des aquariums. Suite à l'annotation avec le logiciel Tandem Repeats Finder, 516 séquences supplémentaires, contenant des répétitions en tandem, sont identifiées. Parmi les séquences répétitives trouvées dans notre analyse, 2 445 séquences répétées s'alignent entièrement ou partiellement avec des séquences déjà identifiées par RepeatScout lors du séquençage du génome de référence (v3.1). Notre analyse nous a donc permis d'identifier 4 143 nouvelles séquences répétitives. Grâce aux résultats combinés des logiciels d'annotation Censor et de comparaison de séquences Blast, nous avons classé les 6531 séquences en 9 groupes. Mais, selon les résultats de TEclass, les séquences répétitives peuvent être classées hiérarchiquement en 5 groupes. La majorité des séquences répétitives annotées appartiennent aux classes I et II des (retro)transposons (Tableau 10).

**Tableau 10: Résumé des résultats obtenus par Blast et Censor pour catégoriser les séquences répétitives de *S. mansoni*.**

Classes	Nombre	%
Classe I (rétrotransposons)	565	8.6
Classe I (rétrotransposons non-LTR)	720	11
Classe II (transposons ADN)	239	3.7
SINEs	31	0.5
Rétrovirus endogènes	18	0.3
Séquences répétées dispersées	21	0.3
Pseudogènes, ARNsn, ARNr, ARNt	38	0.6
Inconnus	4899	75
Total	6531	

Pour 1 921 des 6 531 (29%), des ESTs sont retrouvées suggérant une transcription de ces séquences. Nous avons vérifié le taux de transcription de 7 séquences répétées, présentant entre 0 et 52 ESTs, aux stades miracidia et adultes. Six (smr\_2181, smr\_2685, smr\_2733, smr\_3595, smr\_4022 and smr\_6227) montrent une faible mais significative transcription, relative à un gène de référence alors que pour une séquence répétée, aucune transcription n'est quantifiée (smr\_7590) (Figure 51).



**Figure 51: Expression relative de 7 séquences répétées aux stades miracidia et adultes.** Ces données sont normalisées par rapport au 28S (Lepesant et al. 2012b) (Annexe 3) .

Aucune corrélation n'est apparue entre le nombre d'ESTs et le taux de transcription mesuré par PCR quantitative.

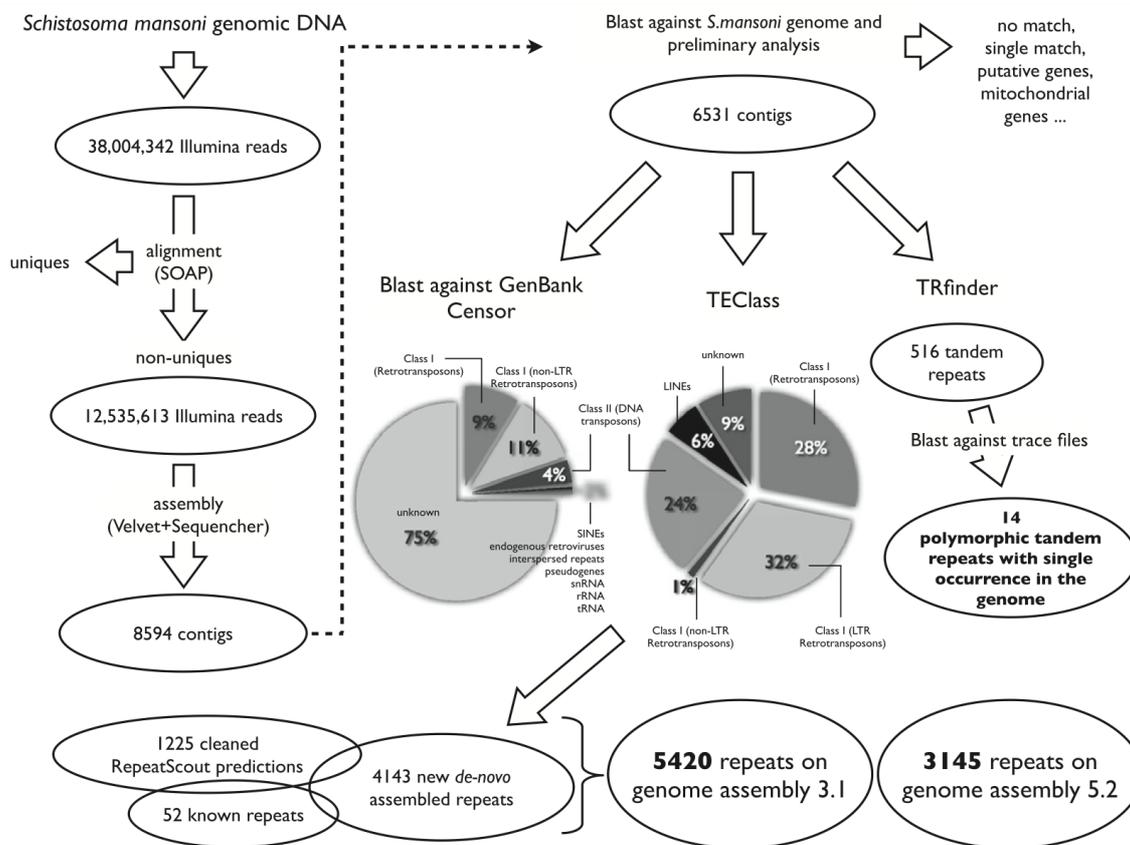
L'annotation par Tandem Repeats Finder et la comparaison de séquences nous a permis d'identifier 14 nouveaux marqueurs microsatellites polymorphes. Cinq d'entre eux sont localisés dans des régions codantes de gènes prédits, neuf sont probablement des marqueurs neutres (Tableau 11).

**Tableau 11: Caractéristiques des 14 nouveaux microsatellites de *S. mansoni*.** Ces microsatellites ont été identifiés lors de l'assemblage et annotation du pseudogénome de séquences répétitives (Lepesant et al. 2012b) (Annexe 3).

Cibles	Position génomique	Gènes prédits
smr_93	Smp_scaff000078:1883819..1884094	
smr_534	Smp_scaff008778:843..259	
smr_1025	Smp_scaff018982:4667..4845	smp_194060
smr_1715	Smp_scaff001981:160389..160587	
smr_2261	Smp_scaff002707:1010..917	smp_106750
smr_2501	Smp_scaff006696:1005..1124	
smr_2770	Smp_scaff000617:14517..14293	smp_098420
smr_3069	Smp_scaff011538:822..633	
smr_3778	Smp_scaff015860:304..150	
smr_3838	Smp_scaff008307:1204..419	
smr_6232	Smp_scaff000018:21638..21810	smp_127980
smr_6375	Smp_scaff000018:21638..21810	smp_186020
smr_6449	Smp_scaff008307:205..393	
smr_7342	Smp_scaff008269:1329..1531	

Afin de générer une base de données combinée, nous avons utilisé les 1 225 séquences répétées prédites. Suite à l'analyse des 55 séquences disponibles sur GenBank, nous avons identifié 3 duplicats : Sm\_SR-AB2 (AF025674.1), Sm\_SR1\_pol\_3 (U66331.1) et Sm\_Sinbad\_iS4-T (AY965073.1). Pour Sm\_salmonid (AY834402.1), il a été montré précédemment que cette séquence n'est pas présente chez *S. mansoni* (Grunau & Boissier 2010). En conséquence, nous avons ôté ces 4 séquences de la base de données, à laquelle nous avons ajouté 51 séquences disponibles sur Genbank ainsi qu'une répétition en tandem identifiée par notre laboratoire (TandemRepeat\_266) : 1225 prédits + 55 GenBank – 3 duplicats – 1 faux positif + 1 précédemment identifié = 1277. Au total notre base de données de séquences répétées contient 5 420 séquences.

Notre pipeline de travail est résumé dans la figure ci dessous (Figure 52).



**Figure 52: Représentation schématique du pipeline utilisé pour l'identification et l'annotation des séquences répétées dans le génome de *S. mansoni*** (Lepesant *et al.* 2012b) (Annexe 3).

Nous nous sommes servis de notre base de données pour annoter l'assemblage 3.1 du génome de *S. mansoni*. Nous avons utilisé RepeatMasker (Smit *et al.* 1996) pour estimer que 47,40% du génome est constitué de séquences répétitives. Tous les résultats et détails de cette analyse sont dans l'article en Annexe 3. Durant la préparation de la publication, la version 5.2 du génome de *S. mansoni* a été publiée, dans laquelle la redondance des *scaffolds* est réduite (Protasio *et al.* 2012). Nous avons donc décidé de vérifier, par un logiciel de comparaison de séquences, si certaines séquences identifiées comme répétées dans notre base de données ne sont pas absentes du nouvel assemblage. Nous avons ôté les séquences qui ne s'alignaient pas, d'où une proportion de séquences répétées maintenant estimée à 47,73% pour la version v5.2 de génome (valeur très proche de la version 3.1).

Tous les résultats sont publiés en 2012 dans *Experimental Parasitology* (IF 2.122) (Lepesant *et al.* 2012b). L'article est en Annexe 3.

### 3.2. Analyse des changements chromatinien autour des séquences répétitives satellites suggérant un possible mécanisme pour l'émergence des chromosomes sexuels

Chez les schistosomes, le sexe est déterminé par des chromosomes sexuels, avec une femelle hétérogamétique (ZW) et un mâle homogamétique (ZZ) (Liberatos & Short 1983). Sur la base cytogénétique, il existe une différence flagrante entre le chromosome Z et le chromosome W dans l'espèce *Schistosoma*. Cependant pour *S. mansoni*, la discrimination entre le chromosome Z et le chromosome W est uniquement basée sur des différences de structure chromatinienne (Grossman *et al.* 1981). *S. mansoni* est donc un modèle de choix pour étudier l'implication de la structure chromatinienne sur la détermination du sexe chez les espèces ayant un système ZW.

Le génome de *S. mansoni* est séquencé et partiellement assemblé (version 3.1 avec 19 033 *scaffolds*) (Berriman *et al.* 2009). Récemment, une version d'assemblage par chromosome a été publiée (version 5.2 composée de 882 *scaffolds*) (Protasio *et al.* 2012). En 2009, Criscione *et al.* proposent une carte génétique pour 210 *scaffolds* (version 3.1) utilisant des marqueurs microsatellites. Ils ont ainsi identifié 8 *linkage groups* correspondant aux 7 autosomes et aux chromosomes sexuels (positionnés dans le même groupe), indiquant que ces derniers peuvent recombiner (Criscione *et al.* 2009). Malgré cela, leurs études ont montré qu'une petite région des chromosomes sexuels, environ 18 Mb, présente une répression de recombinaison.

Plusieurs questions restent en suspens : (i) quelles sont les différences génétiques entre les chromosomes Z et W chez *S. mansoni* ? (ii) quel est le mécanisme de répression de recombinaison dans cette région particulière ?

Suite au séquençage massif de l'ADN génomique mâle et femelle, nous avons aligné les séquences sur le génome de référence (v3.1). En calculant un ratio mâle sur femelle (ZZ/ZW), nous avons pu identifier les séquences mâles, puisqu'elles présentent un ratio autour de 2 (deux chromosomes Z chez le mâle contre un seul chez la femelle). Cela nous a permis d'identifier 15 *scaffolds* spécifiques pour le mâle. Les ratios obtenus par l'analyse des données de séquençage massif ont été validés par PCR quantitative sur 13 *loci* choisis arbitrairement. Cette validation confirme les résultats issus de séquençage et montre une réelle cohérence entre les analyses massives et ciblées. Récemment, l'assemblage total des chromosomes sexuels a été publié (Protasio *et al.* 2012). De ce fait, nous avons réitéré notre analyse et avons pu observer que 70% des séquences uniques sont identiques entre le chromosome Z et le chromosome W, dans la version 5.2 qui estime la région Z-spécifique à

18Mb avec 782 gènes prédits (Protasio *et al.* 2012). Lors de l'analyse sur la version 3.1, 90% des séquences uniques étaient assignées comme pseudoautosomales et la région Z-spécifique était estimée à 6.5Mb.

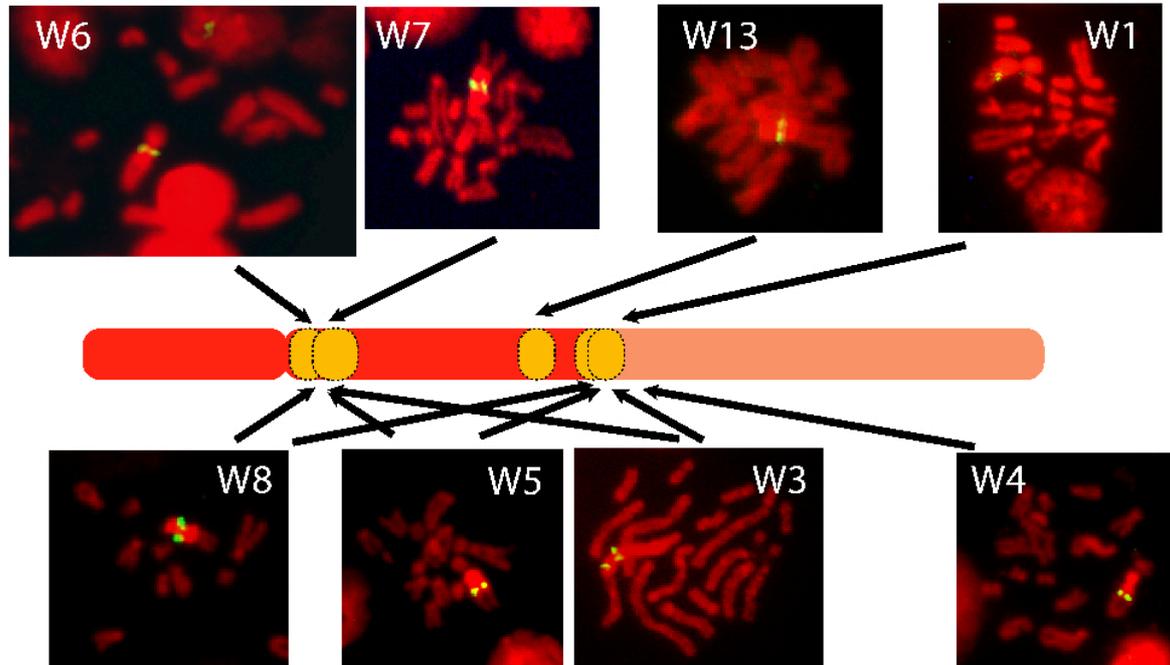
Parmi les 15 *scaffolds* spécifique du chromosome Z, 205 gènes sont prédits (en accord avec les prédictions de SchistoDB). Cependant, seulement 4 gènes semblent coder pour des protéines putatives impliquées dans la spermatogénèse ou pour des gènes paralogues présentant une expression spécifique dans les testicules. Ainsi, nous avons identifié les séquences pseudoautosomales et les séquences spécifiques du chromosome Z ou du chromosome W. En nous basant sur la carte génétique proposée par Criscione *et al.* (Criscione *et al.* 2009), nous avons observé que tous les *scaffolds* Z-spécifiques sont localisés dans une région du chromosome Z qui présente une répression de recombinaison. Dans cette région, les séquences uniques sont entrecoupées de séquences répétées. Ces dernières sont pseudoautosomales. En conséquence, la répression de recombinaison dans cette région, n'est pas uniquement due à l'absence de séquences chromatides sœurs. Ces résultats ont été confirmés avec l'assemblage de la version 5.2.

Dans des études précédentes il a été montré que 105 *scaffolds* sont spécifiques du chromosome W uniquement présent chez les femelles (ZW) (Portela *et al.* 2010). Nous disposons maintenant d'indications suggérant que les séquences femelles spécifiques sont composées de séquences répétitives. En utilisant la technique de WISH (Whole In Situ Hybridization), développée au laboratoire (Portela *et al.* 2010), 33 nouvelles séquences répétitives femelles ont été trouvées, en plus de celles précédemment identifiées (W1, W2, SMAphafem-1). La validation de ces 36 séquences répétitives femelles a été menée par PCR classique (Lepesant *et al.* 2012a) (Annexe 2). Une analyse des banques et les résultats de PCR quantitative montrent que 8 de ces séquences répétées sont transcrites. Pour un certain nombre de ces séquences, le nombre de copies a été estimé par PCR quantitative et est modéré (100-400 copies), à l'exception de SMAphafem-1 (plusieurs milliers de copies), confirmé par des études antérieures (DeMarco *et al.* 2004).

Par ailleurs, la base de données SchistoDB a été utilisée pour identifier 8 gènes putatifs pouvant être localisés dans cette région composée de séquences répétitives. Une inspection manuelle de tous les *loci* a montré 4 gènes putatifs présents dans les données de séquençage massif du génome femelle, mais 3 sont identiques et la région codante prédite est très petite.

Aucune similarité avec des protéines connues n'a pu être trouvée avec les logiciels de comparaison de séquences.

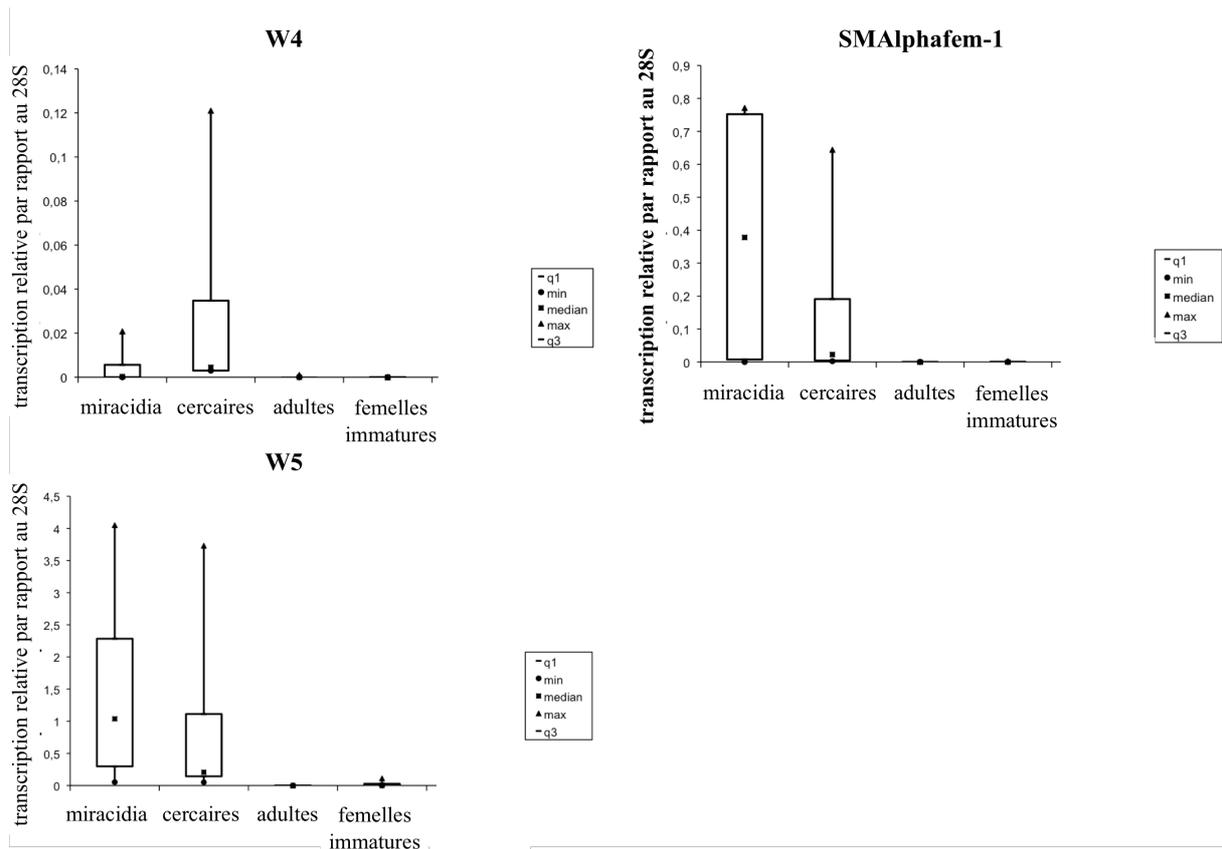
Suite à toutes ces analyses *in silico*, nous avons localisé les séquences répétitives les plus abondantes. La technique de FISH (Fluorescent In Situ Hybridization) est utilisée sur des métaphases tardives de sporocystes secondaires (Figure 53).



**Figure 53: Localisation, par FISH (Fluorescent in situ Hybridization), sur le chromosome W des 8 séquences répétées les plus abondantes.** Photos en microscopie à fluorescence de métaphases tardives de sporocystes secondaires, réparties autour d'un chromosome W schématisé. Les séquences étudiées sont présentes dans la région hétérochromatique (rouge) du chromosome W. Aucune séquence n'est retrouvée dans la région euchromatique (rose) (Lepesant et al. 2012a) (Annexe 2).

Toutes les séquences répétées étudiées (les 8 plus abondantes), sont (i) arrangées en large blocs satellites ; (ii) localisées dans la région hétérochromatique du chromosome W : soit à la région péricentromérique, soit à la frontière euchromatine/hétérochromatine, soit aux deux localisations ; (iii) présentes uniquement sur le bras long (Lepesant et al. 2012a) (Annexe 2).

Les données ESTs suggèrent que certaines de ces séquences répétées sont transcrites. La transcription de W1 et SMA $\alpha$ phém-1 a déjà été démontrée dans des études précédentes (Grevelding 1999). Suite à l'extraction d'ARN des différents stades de développement du parasite *S. mansoni*, nous avons quantifié le taux de transcription de W3, W4, W5 et SMA $\alpha$ phém-1 (Lepesant et al. 2012a) (Figure 54).

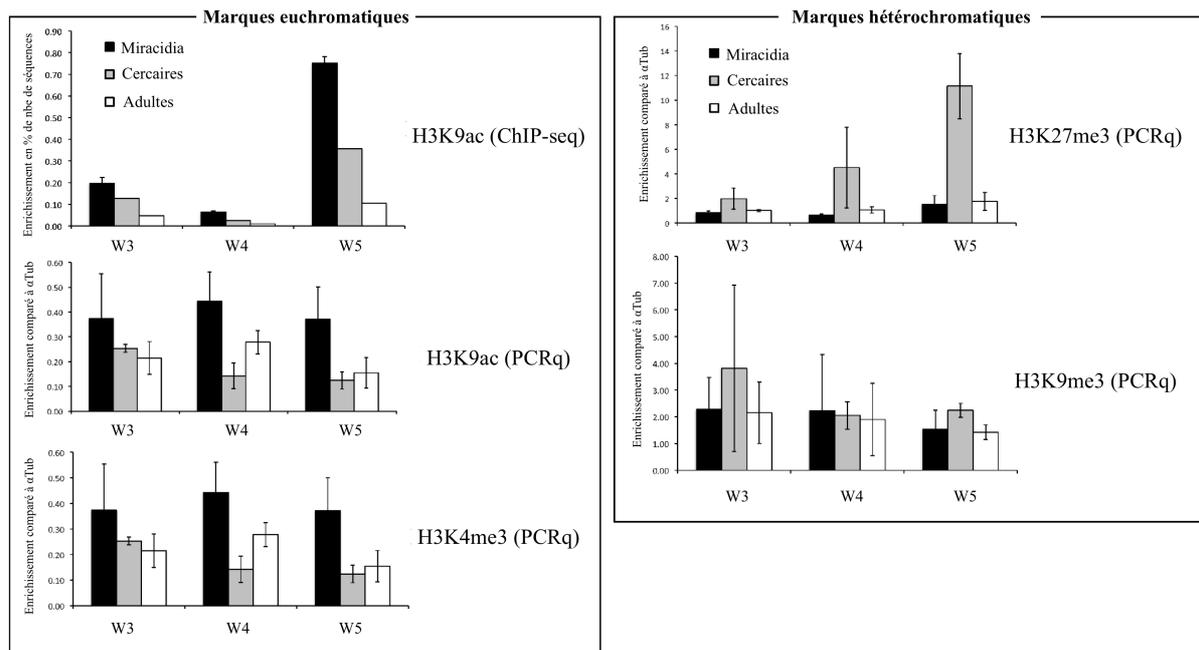


**Figure 54: Représentation des taux de transcription des séquences répétées W4, W5 et SMA1phafem-1 aux différents stades de développement de *S. mansoni*.** Les données sont normalisées par rapport au gène de référence 28S. Les expériences ont été répétées entre 3 et 5 fois. Maximum (triangle), minimum (cercle), quartile 2 à 4 (large rectangle). La transcription des séquences répétitive est restreinte aux stades larvaires (miracidia et cercaires) (Lepesant et al. 2012a) (Annexe 2).

Pour la séquence répétée W3, aucune transcription significative n'a été notée. Les séquences répétitives W4, W5 et SMA1phafem-1 sont transcrites uniquement aux stades larvaires. Aucun transcrit n'a été détecté au stade adulte (couples et adultes femelles immatures). Au niveau génomique, nous avons noté moins de 20% de différence dans le nombre de copies (mesure par PCR quantitative) entre plusieurs échantillons biologiques, et nous n'avons pas observé de décroissance du nombre de copies durant le cycle de vie du parasite. L'absence de transcription chez les adultes n'est donc pas due à une absence de ces séquences à ce stade.

Dans certains cas, la transcription des séquences répétées est corrélée à des changements de structures chromatiniennes (Reinhart & Bartel 2002). Nous avons donc décidé d'analyser les isoformes histones pouvant être potentiellement associées à des séquences répétées femelles spécifiques. L'immunoprécipitation de la chromatine, associée au séquençage massif (ChIP-seq), a été retenue pour évaluer l'abondance de la marque H3K9ac autour des séquences répétées aux stades miracidia, cercaires et adultes. Les 36 séquences répétées

féelles spécifiques montrent une décroissance graduelle de cette marque, des stades larvaires au stade adulte. Pour vérifier ces données de séquençage massif, nous avons utilisé la technique d'immunoprécipitation de la chromatine (avec des anticorps spécifiques de marques euchromatiques (H3K9ac et H3K4me3) ou hétérochromatiques (H3K9me3 et H3K27me3)) associée à de la PCR quantitative, pour les séquences répétées W3, W4 et W5 (Lepesant *et al.* 2012a) (Figure 55).



**Figure 55: Structure chromatinienne, évaluée par ChIP-seq et ChIP-PCRq, autour des séquences répétées femelles spécifiques W3, W4 et W5 aux différents stades de développement.** Moyenne de trois expériences indépendantes pour la PCRq, deux expériences indépendantes pour le séquençage de H3K9ac stade miracidia, et une expérience ChIP-seq pour cercaires et adultes. L'enrichissement euchromatique autour des régions répétées décroît des miracidia (noir) aux cercaires (gris) puis aux adultes (blanc). Les données de ChIP-seq et ChIP-PCRq se corrént. Le statut hétérochromatique reste sensiblement le même à tous les stades (Lepesant *et al.* 2012a) (Annexe 2).

Les deux marques euchromatiques (H3K9ac et H3K4me3) sont enrichies pour ces séquences répétées (W3, W4 et W5), au stade miracidia, où la transcription apparaît. Au contraire, chez les adultes, les marques euchromatiques sont beaucoup moins enrichies. Basé sur les expériences de PCR quantitative, les cercaires occupent une position intermédiaire. En combinant les expériences de ChIP-seq et ChIP-PCRq, on peut observer une baisse de la marque H3K9ac des miracidia aux cercaires puis aux adultes. La marque H3K4me3 présente le même profil que H3K9ac entre les stades miracidia et cercaires mais reste ensuite constante entre cercaires et adultes. Les marques hétérochromatiques, H3K9me3 et H3K27me3, sont abondantes au stade cercaires mais faibles aux stades miracidia et adultes.

En d'autres termes, autour de ces séquences répétées femelles spécifiques, nous observons trois types distincts de structures chromatinienne :

- au stade miracidia, les séquences répétées sont clairement euchromatiques ;
- au stade cercaire, une large proportion de ces séquences est hétérochromatique ;
- au stade adulte, nous avons trouvé une structure particulière qui n'est ni euchromatique ni hétérochromatique mais associée à l'absence de transcription.

Nous avons ensuite testé si les changements observés dans la structure de chromatine étaient le résultat ou la cause des changements de transcription. Si l'hypoacétylation des histones est la cause de l'inactivation de la transcription, alors l'inactivation des histones déacétylases entraîne une répression de transcription. Mais, si la transcription des séquences répétées est à l'origine des changements de structure chromatinienne, alors l'inhibition des histones déacétylases ne changera pas le taux de transcription, car chaque augmentation de transcription va renforcer la déacétylation et donc contourner l'inhibition.

Nous avons traité des vers adultes avec de la trichostatine A (TSA) qui est un inhibiteur des histones déacétylases (HDAC). Après deux heures de traitement, des changements de comportements sont apparus avec des mouvements de convulsion. La transcription a été de nouveau quantifiée par PCR quantitative pour W4, W5 et SMA $\alpha$ fem-1, mais pour aucune des séquences répétées, un taux de transcription n'a été trouvé. En revanche, nous avons observé une augmentation de la transcription de retrotransposons présents sur les chromosomes autosomaux. Afin de confirmer l'absence de toxicité de la TSA, des tests de cytotoxicité ont été réalisés.

Toutes ces analyses nous ont permis d'identifier les séquences spécifiques du chromosome Z, du chromosome W et les séquences pseudoautosomales, ainsi que la région Z-spécifique qui est composée de séquences uniques entrecoupées de séquences répétées pseudoautosomales, et la région W-spécifique qui est composée de 36 familles de séquences répétitives, parmi lesquels aucun gène W-spécifique n'a pu être identifié. Les huit séquences les plus abondantes sont toutes localisées dans la région hétérochromatique du bras long sur le chromosome W et certaines sont transcrites aux stades larvaires. Cette transcription est associée à des changements dans la structure chromatinienne, autour de ces séquences répétées, tout au long du cycle de développement de *S. mansoni*.

Tous les détails et résultats de ces analyses sont publiés, en 2012, dans *Genome Biology* (IF 9.036). L'article est en Annexe 2.

# **CHAPITRE 4**

## **Discussion**

Ces travaux de thèse ont pour principal objectif de mieux comprendre l'influence de l'environnement sur le parasite humain *S. mansoni*, responsable de la bilharziose intestinale, maladie qui se trouve au second rang mondial (après la malaria) en terme de morbidité et de mortalité induites par le parasite, avec plus de 200 millions de personnes infestées dans 74 pays et plus de 200 000 décès par an (Chitsulo *et al.* 2004; Gryseels *et al.* 2006).

Lors du cycle de développement, les différents stades sont exposés à des environnements particuliers. Ainsi, les stades larvaires libres (miracidia et cercaires) se retrouvent dans un milieu aqueux, dépendant des conditions biotiques et abiotiques. Puis ces larves peuvent être exposées à différents hôtes intermédiaires (souche de mollusque) et définitifs (humain ou murin). *S. mansoni* étant un endoparasite, chaque hôte dans lequel il va se maintenir devient son environnement direct. Pour survivre, ce parasite est contraint de lutter contre les défenses immunitaires des hôtes respectifs, et depuis quelques années il doit trouver la parade, essentiellement au stade adulte, aux traitements thérapeutiques anti-helminthes.

Notre hypothèse de travail est que ces environnements fluctuants peuvent influencer, à différents moments du cycle, l'épigénome de *S. mansoni* par des modifications phénotypiques héréditaires mais réversibles. Ces changements pourraient lui conférer un réel avantage pour s'acclimater et s'adapter rapidement à des environnements non stables.

Afin de vérifier cette supposition, nous avons décidé de changer l'environnement du parasite, principalement l'environnement-hôte.

Pour les modifications environnementales abiotiques, nous avons choisi de travailler sur l'apparition rapide de résistance à une drogue antihelminthe, *S. mansoni* étant capable de générer des phénotypes résistants à l'hycanthone et au lucanthone, en un temps de génération très court (Rosi *et al.* 1967; Rogers & Bueding 1971; Jansma *et al.* 1974; Jansma *et al.* 1977; Cioli & Pica-Mattocchia 1984; Pica-Mattocchia *et al.* 1988; Cioli *et al.* 1992).

Pour les modifications environnementales biotiques, nous avons établi une évolution expérimentale soumettant une même population initiale de parasites à deux souches d'hôtes intermédiaires différents. Chaque souche de parasite est spécifique de son hôte intermédiaire (souche sympatrique) mais est tout de même capable d'infester, à un taux plus faible et un coût certain, d'autres souches de mollusques (souche allopatrique). Ainsi, certains individus parasites (compatibles) pourront se développer normalement au sein de l'hôte mollusque alors que d'autres (incompatibles) seront (à une proportion plus ou moins grande) reconnus,

encapsulés et tués par les cellules immunitaires du mollusque. Seuls certains individus arriveront à se développer et se maintenir au sein du mollusque qui ne leur est pas spécifique (souche allopatrique).

Ces deux évolutions expérimentales ont mené à deux énigmes : (1) « Adaptation rapide de *S. mansoni* à une drogue anti-helminthe » et (2) « Adaptation rapide de *S. mansoni* à l'Amérique du Sud ». Les résultats obtenus pour cette dernière ont généré une troisième énigme subsidiaire: (3) « Différenciation sexuelle en l'absence de gène W-spécifique dans un système ZW ».

Afin de résoudre notre hypothèse principale, il nous a fallu répondre à ces trois questions. Pour cela, des expérimentations ont été mises en place ainsi que des outils techniques nécessaires à l'analyse allant de la population (traits de vie) à la molécule (génomique, épigénomique et transcriptome).

Ainsi, nous avons : (i) élaboré un protocole de biologie moléculaire pour l'extraction d'ARN de haute qualité (<http://methdb.univ-perp.fr/epievo/>); (ii) mis en place des marqueurs spécifiques du sexe pour identifier les individus mâles des individus femelles aux stades où le dimorphisme sexuel n'est pas établi (Lepesant *et al.* 2012a) (Annexe 2) et (iii) utilisé des outils bioinformatiques afin d'analyser les grandes quantités de données issues de séquençage massif d'échantillons de différentes natures (ADN, ARN et ADN immunoprécipité) et de classifier la totalité des séquences répétitives du génome de *S. mansoni* (Lepesant *et al.* 2012b) (Annexe 3).

**Comment *S. mansoni* est-il capable de générer rapidement des phénotypes résistants à un traitement anti-helminthe, et ce en un temps de génération court?**

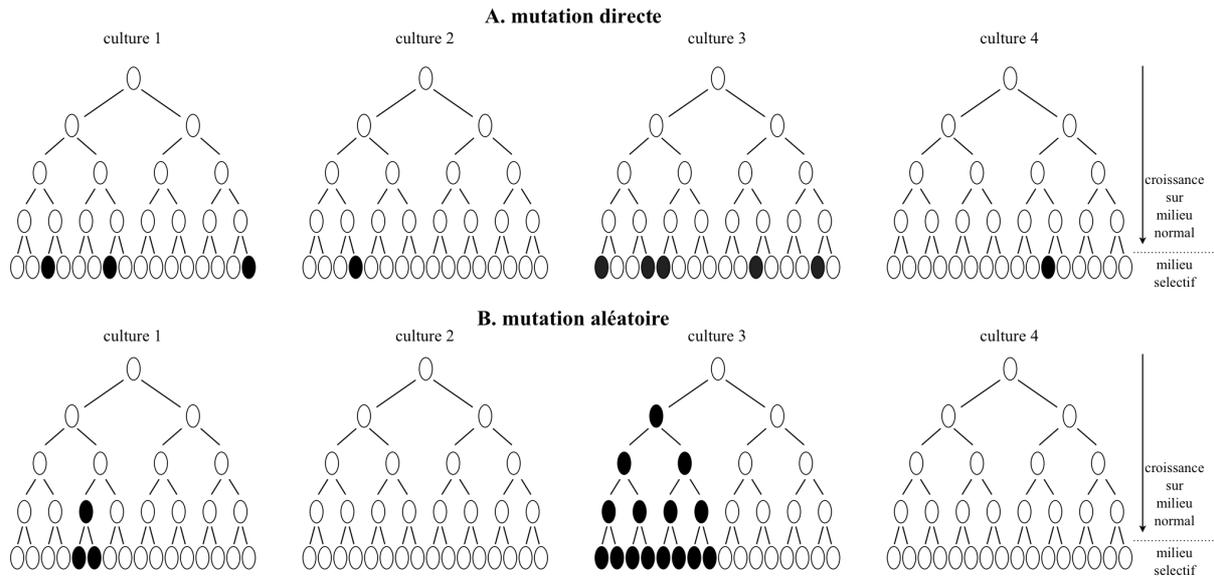
Pour répondre à cette première question scientifique, un protocole expérimental a été mis en œuvre en utilisant comme traitement anti-helminthe, tout d'abord l'hycanthone puis le lucanthone, deux molécules qui visent à changer l'environnement-hôte définitif. Les modèles expérimentaux des deux protocoles sont basés sur des travaux de Luria & Delbrück (Luria & Delbrück 1943) ainsi que Lederberg & Lederberg (Lederberg & Lederberg 1952).

L'acceptation universelle selon laquelle les variations sont aléatoires, résulte d'une série de travaux sur des bactéries, rapportée par Luria & Delbrück en 1943 (Luria & Delbrück 1943).

Chez les bactéries, tous les organismes se divisent, la distinction entre les cellules somatiques et les cellules germinales n'a pas d'importance, et les influences environnementales sur l'organisme peuvent passer aux générations suivantes. Quelques adaptations à des agents antibactériens sont connues pour apparaître rapidement (Jablonka & Lamb 1999). Le point essentiel des expériences de Luria & Delbrück est de montrer que leurs mutations adaptatives peuvent apparaître spontanément en absence d'un agent sélectif.

Pour étudier l'événement de résistance aux phages chez les bactéries *Escherichia coli*, deux hypothèses ont été retenues : (i) les mutants phages-résistants sont induits par la présence des phages ; (ii) les mutants phages-résistants apparaissent spontanément et sont sélectionnés par la présence des phages. Ces deux hypothèses donnent différentes prédictions sur la distribution des colonies mutantes dans des cultures parallèles.

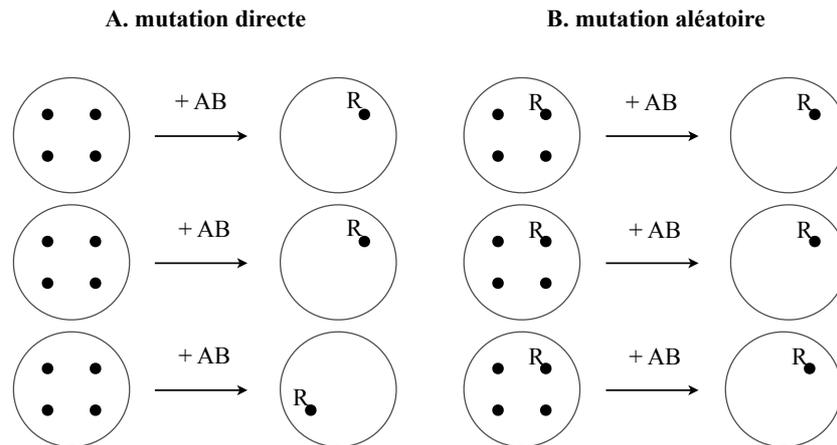
Dans le type d'expérience de Luria & Delbrück, des séries de cultures identiques, avec un fond génétique très semblable, sont maintenues sur des milieux de culture non sélectifs. Quand chaque culture atteint une densité spécifique, elles sont transférées sur des milieux contenant les phages (milieu sélectif). Les cellules sensibles meurent, alors que les résistantes survivent et forment des colonies qui sont ensuite dénombrées (Luria & Delbrück 1943). La distribution des colonies survivantes dépend de mutations directes ou de mutations aléatoires. Si le nombre de mutants augmente en réponse à la présence du phage, le nombre de colonies résistantes dans chaque culture ne montre pas de variations. Puisqu'il est peu probable que la présence des phages induise des changements héréditaires, le nombre de colonies mutantes dans les différentes cultures doit suivre une loi de Poisson (loi des événements rares) avec une variance égale à la moyenne (Jablonka & Lamb 1999). Cette distribution est schématisée dans la Figure 56 (A). Toutefois, si les mutations ne sont pas induites mais apparaissent spontanément avant que les bactéries ne soient confrontées aux phages, le nombre de colonies mutantes doit varier fortement entre chaque culture, selon les stades de développement. Les cultures dans laquelle la mutation apparaît tôt dans le développement ont plus de colonies mutantes alors que les cultures où la mutation se fait tardivement présentent moins de colonies résistantes (Figure 56 B). Ainsi la variance de la distribution du nombre de colonies mutantes sera plus élevée que la moyenne (Jablonka & Lamb 1999).



**Figure 56: Hypothèses des expériences de mutation directe ou aléatoire de Luria & Delbrück** (Luria & Delbrück 1943). (A) Distribution des colonies mutantes si toutes les mutations sont directes. (B) Distribution si toutes les mutations apparaissent aléatoirement et sélectionnées par le milieu. Modifié de (Jablonka & Lamb 1999).

Luria & Delbrück constatent que le nombre de colonies mutantes varie entre chaque culture. La distribution est bien celle prédite par l'hypothèse selon laquelle les mutations ne sont pas induites par l'environnement mais apparaissent spontanément (Figure 56 B). Les mutants phages-résistants existent avant l'exposition aux phages et sont simplement sélectionnés par la présence des phages (Luria & Delbrück 1943; Jablonka & Lamb 1999).

En 1952, Lederberg & Lederberg confirment (Lederberg & Lederberg 1952) les résultats obtenus par Luria & Delbrück (Luria & Delbrück 1943). En utilisant la technique de transfert de colonies, ils montrent clairement que les mutations sont présentes à de faibles fréquences avant l'application des conditions sélectives. Le transfert de colonies implique : la croissance de colonies bactériennes sur un milieu agar non sélectif puis le transfert exact des copies de colonies, à des positions identiques, sur des milieux contenant un agent sélectif (antibiotique streptomycine). Lorsqu'elles se trouvent sur les milieux sélectifs, seules les colonies mutantes croissent. Lederberg & Lederberg ont relevé que les patrons de colonies mutantes sont les mêmes sur chaque répliquat, montrant que la mutation est bien présente avant que les colonies ne soient transférées sur le milieu sélectif (Lederberg & Lederberg 1952) (Figure 57 B).



**Figure 57: Protocole expérimental de Lederberg & Lederberg.** (A) Distribution des colonies mutantes si toutes les mutations sont directes. (B) Distribution si toutes les mutations apparaissent aléatoirement. R, résistance ; AB, antibiotique (Lederberg & Lederberg 1952).

Dans les deux expériences, la résistance aux phages ou à l'antibiotique est présente avant l'exposition des bactéries au milieu sélectif. Les mutations n'apparaissent pas en tant que réponse adaptative à la pression de sélection (Jablonka & Lamb 1999).

En nous basant sur ces travaux, nous avons établi une évolution expérimentale pour répondre à la question : La résistance, à l'hycanthone ou au lucanthone, est-elle déjà présente dans la population et sélectionnée par le traitement ou, est-ce le traitement qui entraîne la résistance de certains individus parasites?

En 1971, Rogers *et al.* relatent le premier cas de résistance à l'hycanthone obtenu en laboratoire (Rogers & Bueding 1971). Cette résistance apparaît à la première génération, est stable et peut perdurer sur plusieurs générations. Cette résistance peut passer de 69% à 100% à travers 7 générations (Jansma *et al.* 1977). La résistance est due à une absence d'activité enzymatique. L'enzyme ciblée serait active chez les parasites sensibles, mais inactive chez les individus résistants. Par l'évolution expérimentale mise en place, nous avons cherché à comprendre si l'absence d'activité enzymatique existe avant le traitement ou si le traitement entraîne l'inactivation enzymatique et donc la résistance.

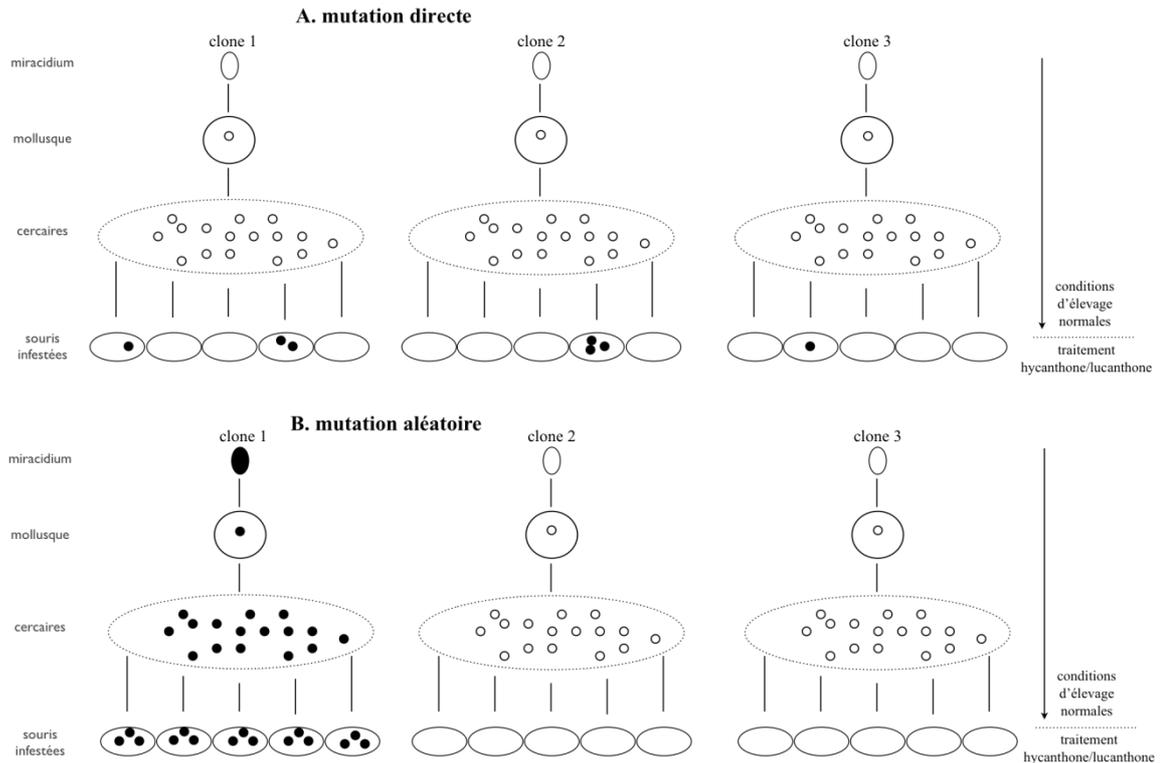
En cela, le parasite d'étude *S. mansoni* présente un fort avantage puisqu'à partir d'infestations monomiracidiales sur mollusque (un miracidia par mollusque), il nous est possible de produire des populations clonales de cercaires avec un fond génétique homogène. Les clones de parasite miment les colonies bactériennes et le milieu sélectif est le traitement avec l'hycanthone ou le lucanthone. Chaque clone est répliqué plusieurs fois. La distribution des

individus résistants entre les différents clones et au sein de chaque clone nous permet de déterminer si la résistance à ces deux drogues existe avant le traitement et sélectionnée par ce dernier (mutation aléatoire), ou si la résistance est induite par le traitement (mutation directe) selon Luria & Delbrück (1943).

Les résultats de résistance diffèrent entre les expériences hycanthone et lucanthone. En effet, après traitement à l'hycanthone, le nombre d'individus survivants est fortement diminué par rapport à la condition non traitée. Les individus résistants sont répartis entre plusieurs clones. Sur cinq clones, seuls deux présentent une sensibilité totale et aucun individu résistant n'a été retrouvé. Au contraire, avec le traitement au lucanthone, le nombre d'individus survivants n'est pas significativement différent de la condition non traitée. Les individus survivants sont retrouvés dans tous les clones. Cette divergence entre les deux traitements peut être expliquée par deux raisons :

- La première est que le lucanthone est moins efficace que l'hycanthone. Il est à noter que les deux drogues, n'étant plus commercialisées, ont été récupérées dans des laboratoires travaillant anciennement sur ces drogues. Il nous est donc impossible d'avoir des renseignements précis quant à leur date d'acquisition et leur conservation.
- La deuxième est que dans l'expérience hycanthone, les individus traités sont uniquement des individus mâles alors que dans l'expérience lucanthone les individus traités sont des couples. Dans cette dernière expérience, les infestations ont été réalisées avec des couples de clones afin de récupérer les œufs issus de la reproduction sexuée pour réaliser le traitement sur la deuxième génération. Cette différence peut entraîner une modification de l'efficacité de la drogue avec une protection plus forte des individus accouplés puisqu'il a été montré que les femelles résistantes à l'hycanthone ont la capacité de protéger le mâle avec lequel elles sont accouplées (Jansma *et al.* 1977). Ce phénomène peut également se reproduire dans le cas du traitement au lucanthone qui est issu de la bioactivation de l'hycanthone.

Suite aux deux traitements, la distribution des individus résistants est différente de celle proposée dans l'hypothèse de mutation aléatoire, par Luria & Delbrück (Luria & Delbrück 1943).



**Figure 58: Design expérimental de Luria & Delbrück adapté à *S. mansoni*.** (A) Distribution des colonies mutantes si toutes les mutations sont directes. (B) Distribution si toutes les mutations apparaissent aléatoirement. Adapté de Jablonka & Lamb (Jablonka & Lamb 1999).

Nos résultats sont en faveur de l’hypothèse d’une mutation directe proposée par Luria & Delbrück (Figure 58 A). En effet, le patron de répartition des individus résistants correspond à cette hypothèse: la résistance à ces drogues est apparue après traitement.

La définition de mutation a évolué depuis les études de Luria & Delbrück, lesquels proposaient le terme de mutation directe dans le cas d’une mutation complète d’un gène. Actuellement une mutation est considérée comme une modification de la séquence d’ADN. Les résultats issus des évolutions expérimentales mettent plutôt en avant un phénomène d’épimutation, qui serait cohérent avec les études démontrant l’absence d’une activité enzymatique entraînant la résistance (Cioli *et al.* 1992). L’enzyme est présente chez tous les individus mais est active chez les individus sensibles et inactive chez les individus résistants. Cette modification d’activité peut être expliquée par un phénomène d’hétérochromatisation, menant à une inhibition de cette enzyme. Cette hypothèse est confortée par les études de Jansma *et al.* qui montrent que les phénotypes résistants sont métastables (Jansma *et al.* 1977). Ils sont présents sur plusieurs générations successives mais parfois cette résistance

disparaît en absence de traitement. Les études, actuellement en cours, portent sur la modification de la structure chromatinienne chez les individus résistants à l'hycanthonne.

Lors du traitement avec le lucanthonne, nous avons montré que la deuxième génération traitée présente un grand nombre de vers résistants mais les résultats surprenants ne nous permettent pas de répondre à la question d'héritabilité. Les molécules hycanthonne et lucanthonne n'étant plus commercialisées, nous pourrions reproduire ces évolutions expérimentales avec la molécule oxamniquine qui est encore utilisée contre *S. mansoni*. Plusieurs études ont mis en avant la résistance croisée de cette molécule avec l'hycanthonne (Cioli *et al.* 1992) mais aussi avec le praziquantel (Fallon & Doenhoff 1994), actuellement utilisé dans le traitement général de la bilharziose. Il est donc d'une importance majeure de comprendre les mécanismes moléculaires sous-jacents à cette résistance et induits par ces traitements.

Au sein de chaque clone, plusieurs réplicats ont été réalisés afin de se baser, en parallèle, sur les expériences de Lederberg & Lederberg (Lederberg & Lederberg 1952). Cependant, les résultats obtenus au sein d'un clone, entre chaque réplicat, ne sont pas identiques. Cette différence, avec les expériences de transfert de plaques de Lederberg & Lederberg, pourrait être due à la grande variabilité d'infectivité de *S. mansoni* (Boissier *et al.* 1999).

En complément de nos travaux, des mesures de réponse immunopathologique de l'hôte définitif sont effectuées. Le ratio du poids de la rate sur le poids de la souris donne une estimation de la réponse inflammatoire. Pour les deux expériences, cette mesure n'a montré aucune différence entre les souris traitées et les souris non traitées. Tous les individus présentent un ratio entre 1% et 2% ce qui est en accord avec le fait que tous les individus sont infestés (Boissier, communication personnelle).

Cependant, l'expérience avec l'hycanthonne semble montrer une baisse dans la réponse inflammatoire chez les individus traités. L'absence de différences significatives, entre les lots contrôles et les lots traités, peut être corrélée au fait que les hôtes infestés mais traités conservent une trace de cette infestation. Le traitement ne réduit pas directement la réaction inflammatoire. De plus, les vers adultes sensibles à l'hycanthonne ont été retrouvés enkystés dans le foie, entraînant une réponse de l'organisme hôte.

Dans le cas de l'expérience lucanthonne, le ratio poids rate/poids souris est plus élevé (2% contre 1% avec l'hycanthonne) car les souris sont infestées avec des couples, d'où la reproduction sexuée et l'émission d'œufs. Ces derniers peuvent être excrétés par les fèces ou

sont accumulés dans le foie. Le traitement n'a aucun effet sur les œufs stockés dans le foie, donc la réponse inflammatoire de l'hôte reste inchangée.

- En résumé

Ces évolutions expérimentales nous ont permis de mettre en avant que la résistance du parasite à ces deux drogues est induite par le traitement, ce qui ne va pas dans le sens des expériences sur les bactéries de Luria & Delbrück (1943) et Lederberg & Lederberg (1952). Ces deux traitements anti-helminthes réduisent le nombre d'adultes, mais n'agissent pas directement sur la réponse inflammatoire de l'hôte.

Afin d'approfondir nos recherches sur les mécanismes moléculaires sous-jacents à cette résistance, nous avons immunoprécipité l'ADN d'individus sensibles et résistants avec des anticorps spécifiques de marques dites euchromatiques (H3K9ac et H3K4me3) ou dites hétérochromatiques (H3K9me3 et H3K27me3). Ces échantillons sont séquencés par la technique Illumina, en collaboration avec l'Université d'Oregon. Cependant, dans le temps qui m'était imparti pour ces travaux de thèse, les analyses sont en cours. En effet, les séquences ayant été reçues tardivement, nous sommes en train de procéder aux contrôles de qualité qui nous permettront ensuite d'engager les études bioinformatiques. Il semble d'ores et déjà que les séquences sont de bonne qualité pour les marques H3K9ac, H3K9me3 et H3K27me3, et d'une qualité moindre pour H3K4me3. Les analyses de ces séquences, entre la condition traitée et la condition non traitée à l'hycanthone, seront menées en suivant les protocoles bioinformatiques utilisés pour celles de la génération 3 de l'expérience de changement d'hôte intermédiaire sympatrique *vs* allopatrique.

**Comment *S. mansoni* s'est-il adapté rapidement à l'Amérique du Sud?**

Pour répondre à cette deuxième question scientifique, nous avons tout d'abord parcouru les génomes et les épigénomes de deux souches de laboratoire issues d'isolats géographiques différents, afin d'identifier des candidats génétiques et/ou épigénétiques. Ensuite, nous avons analysé les séquences promotrices des gènes impliqués dans le polymorphisme de compatibilité entre le parasite et l'hôte intermédiaire, les *SmPoMucs* (*S. mansoni* Polymorphic Mucins). Enfin, nous avons proposé une évolution expérimentale d'infestation d'hôte

intermédiaire sympatrique ou allopatrique afin d'observer la réponse adaptative du parasite lorsqu'il est soumis à un environnement inhabituel. Cette évolution expérimentale vise à mimer l'émigration de la bilharziose de l'Afrique vers l'Amérique lors de la traite des esclaves (Morgan *et al.* 2001).

Nous savons que *S. mansoni* a su s'adapter à un nouvel environnement et à de nouveaux hôtes lorsqu'il a été déplacé au cours des 16<sup>e</sup> et 17<sup>e</sup> siècles de l'Afrique vers l'Amérique du Sud (Morgan *et al.* 2001).

*S. mansoni* constitue la seule espèce présente en Amérique du Sud. Le polymorphisme de compatibilité décrit pour l'interaction *S. mansoni* / *B. glabrata*, est une des illustrations de l'adaptation affichée par le parasite envers son hôte invertébré (Théron & Coustau 2005). Il existe un équilibre entre les mécanismes de défense de l'hôte et la virulence et/ou infectivité du parasite. Les candidats moléculaires de ce polymorphisme de compatibilité ont été décrits précédemment (Roger *et al.* 2008a; Roger *et al.* 2008c). Ce polymorphisme se traduit par la présence de certains phénotypes hôtes et parasites compatibles. En effet, chaque souche de parasite est spécifique d'une souche particulière d'hôte intermédiaire, mais en conservant tout de même la capacité à infester des hôtes allopatriques avec une valeur sélective plus faible. Un miracidium peut être infestant ou non infestant si son phénotype est respectivement concordant ou non concordant avec le phénotype de l'hôte.

Le polymorphisme de compatibilité, qui caractérise l'interaction *S. mansoni* / *B. glabrata*, a permis de sélectionner au laboratoire deux souches de parasite *S. mansoni*, une compatible (Brésil) et une incompatible (Guadeloupe), vis-à-vis de la même souche de mollusques *B. glabrata* Brésil. Lors de ces travaux de thèse, nous avons décidé d'analyser le génome et l'épigénome de ces deux souches de façon globale (à l'échelle du génome entier par séquençage massif) et de manière ciblée (certains gènes d'intérêt).

Pour les analyses de génomique et d'épigénomique, nous avons utilisé un anticorps dirigé contre l'histone H3 acétylée sur la lysine 9 (H3K9ac), qui est une marque représentative de transcription active. L'ADN immunoprécipité avec cet anticorps, pour chaque souche, est séquencé par la technique Illumina, en collaboration avec Montpellier GenomiX. L'étude des données issues de ce séquençage nous permet de comparer les épigénomes des deux souches, mais aussi leurs génomes. Les différences d'alignement avec le génome de référence, ainsi que les validations par PCR quantitative vont nous servir à discriminer les différences génétiques de celles épigénétiques entre ces deux souches (Lepesant *et al.* 2011) (Annexe 4).

Pour les analyses ciblées, nous nous sommes essentiellement focalisés sur les gènes *SmPoMucs* impliqués dans le polymorphisme de compatibilité. Ce sont des protéines glycosylées extrêmement polymorphes. Le polymorphisme s'exprime, au niveau intra- et inter-souche mais également au niveau individuel, sur toute la molécule (Roger *et al.* 2008b).

Ainsi, la région N-terminale est composée d'un domaine hautement glycosylé caractérisé par un polymorphisme quantitatif et qualitatif du nombre de séquences répétées. La région C-terminale n'est, a priori, pas glycosylée et ne présente pas d'homologie de séquence ni de structure pour des domaines connus. Le polymorphisme de cette région est essentiellement le fruit de mécanismes d'épissages alternatifs ou aberrants (Roger *et al.* 2008b). Les *SmPoMucs* sont exprimés spécifiquement durant les stades en interaction avec le mollusque (miracidium et sporocyste primaire) (Roger *et al.* 2008a) avec une expression supérieure au stade infestant, le miracidium. De plus, au niveau individuel, vient s'ajouter un polymorphisme d'expression. Effectivement, chaque larve de *S. mansoni* est capable d'exprimer une combinaison de variant(s) *SmPoMuc* unique (Roger *et al.* 2008b). Les *SmPoMucs* sont regroupés en 4 groupes (1, 2, 3 et 4) et au sein de chaque groupe, il peut exister des sous groupes (ex : sous groupes 3.1, 3.2, 3.3 et 3.4). A cela, se rajoute des subdivisions dans les sous groupes (ex : au sein du groupe 3.1, un groupe 3.1(r1-r2)) qui comprend des répétitions en tandem de l'exon 2. Chaque individu miracidia exprime une combinaison unique de ces différents groupes et sous groupes. Entre les souches Brésil et Guadeloupe, le niveau de transcription de certains groupes peut être variable. Des analyses approfondies des différences génétiques et épigénétiques ont été menées pour ces deux souches de parasite (Perrin *et al.* under revision). Cette étude a été réalisée en collaboration avec les co-auteurs de la publication (Annexe 5) et ma contribution a porté essentiellement sur les analyses épigénétiques des séquences promotrices des *SmPoMucs*.

Lors de l'analyse globale, nous avons comparé les profils obtenus en ChIP-seq (séquençage de l'ADN immunoprécipité avec un anticorps dirigé contre H3K9ac). Cette méthode nous a permis d'identifier au moins 9 régions qui présentent des différences quantitatives dans les séquences alignées contre le génome de référence (NMRI) (Berriman *et al.* 2009), accentuant ainsi des différences génétiques ou épigénétiques. De ce fait, les différences génétiques peuvent être soulignées par l'absence d'alignement de séquences dans une des deux souches, sur le génome de référence ou alors par une différence dans le nombre de copies présentes dans chacun des deux génomes. Les différences épigénétiques sont caractérisées par un enrichissement plus ou moins fort pour la marque H3K9ac, dans la

souche Brésil ou dans la souche Guadeloupe. Pour confirmer et discriminer les différences génétiques de celles épigénétiques, une approche ciblée par PCR quantitative est utilisée. Nous avons supposé que l'absence d'amplification par PCR dans une souche (avec un résultat positif dans l'autre souche) signifie qu'il existe un polymorphisme génétique entre ces deux souches (Lepesant *et al.* 2011) (Annexe 4).

Par ces différentes méthodes, nous avons pu mettre en avant 11 *loci* distincts entre les deux souches et les analyses ciblées ont permis de recenser 7 *loci* présentant une différence génétique, 2 *loci* indiquant une différence dans le nombre de copies et 2 *loci* montrant une différence épigénétique associée à la marque H3K9ac (Lepesant *et al.* 2011) (Annexe 4).

Parmi les candidats génétiques, nous avons identifié une région de 300 kb, située sur le *scaffold* Smp\_scaff000234, qui contient 4 copies de gènes annotés comme MEG (*Micro Exon Gene*) (Lepesant *et al.* 2011) (Annexe 4). Cette région présente un nombre plus élevé de séquences alignées dans la souche Guadeloupe par rapport à la souche Brésil. Ces gènes de type MEG ont été décrits chez *S. mansoni* (DeMarco *et al.* 2010). Ils sont composés de 18 membres de familles différentes et servent à la création de variants protéiques par épissage alternatif des exons courts. MEG-2 est exprimé dans les œufs et les miracidia, et est sécrété dans les œufs. Son rôle potentiel dans le contexte de la stratégie de l'évasion immunitaire conférant un avantage au parasite, a déjà été discuté (DeMarco *et al.* 2010). Les schistosomes produisent simultanément et non pas séquentiellement, des variants protéiques.

Par ailleurs, il n'existe aucune évidence de la distribution allélique des variants des MEG dans les populations de schistosomes. Si la génération de variations protéiques, par l'expression des gènes MEG, est une stratégie pour l'évasion immunitaire, alors un mécanisme différent est présent chez ces parasites. Cependant, la transcription et la variation des protéines MEG suggèrent qu'une forme de pression par le système immunitaire est exercée pour les sélectionner (DeMarco *et al.* 2010). Les gènes *SmPoMucs* affichent des caractéristiques communes avec les gènes MEG et leur rôle potentiel dans l'hypothèse de la course aux armements a été discuté (Moné *et al.* 2010). Les gènes MEG-2 deviennent des candidats dès lors prometteurs pour un second déterminant moléculaire du polymorphisme de compatibilité *S. mansoni* / *B. glabrata* (Lepesant *et al.* 2011) (Annexe 4).

Parmi ceux ayant une différence épigénétique, nous avons identifié une région de 60 kb, sur le *scaffold* Smp\_scaff000421, qui contient un gène (Smp\_171100) annoté comme une métalloprotéase de la famille des endopeptidases neutres (NEP) (Lepesant *et al.* 2011)

(Annexe 4). Chez *S. mansoni*, ces NEPs peuvent jouer un rôle dans l'immunomodulation de l'hôte intermédiaire *B. glabrata* (Duvaux-Miret *et al.* 1992). Ce gène est exprimé dans les œufs et les miracidia. Nous avons montré, lors de cette analyse, qu'il est situé dans une région plus euchromatique dans la souche Guadeloupe (dite incompatible), comparativement à la souche Brésil (dite compatible). Par conséquent, ce gène constitue également un excellent candidat pour étudier les bases épigénétiques du polymorphisme de compatibilité, puisqu'il pourrait stimuler la réponse immunitaire innée de l'hôte ou la freiner, conférant ainsi un avantage certain au parasite lors de l'infestation (Lepesant *et al.* 2011) (Annexe 4).

Cette approche, combinant les données de ChIP-Seq, les analyses bioinformatiques et les confirmations par PCR quantitative, a mis en évidence des candidats génétiques et épigénétiques potentiellement impliqués dans le polymorphisme de compatibilité menant à des stratégies d'évasion immunitaire qui ont été largement décrits pour d'autres modèles parasites (Dixon *et al.* 2010; Merrick & Duraisingh 2010; Wang *et al.* 2010). Nos analyses nous ont permis d'étudier l'adaptation d'un parasite *S. mansoni* à un hôte invertébré *B. glabrata* (Lepesant *et al.* 2011) (Annexe 4).

Lors de l'analyse ciblée, nous avons procédé une étude approfondie des gènes *SmPoMucs* impliqués dans le polymorphisme de compatibilité. Tous les groupes existent dans les différentes souches de *S. mansoni* mais leur analyse a mis en avant la présence du groupe 3.1(r1-r2) uniquement dans la souche Guadeloupe. De plus, des différences de phénotypes sont observées entre la souche Guadeloupe et la souche Brésil, différences qui ne peuvent être expliquées uniquement par un différentiel génétique. C'est pourquoi, nous avons décidé d'analyser les différences apparentes au niveau des séquences promotrices.

Dès la première analyse des niveaux de transcription entre deux souches, il en est résulté un taux de transcription des *SmPoMucs* plus élevé dans la souche Guadeloupe (dite incompatible) que dans la souche Brésil (dite compatible). Nos résultats sont cohérents par rapport aux recherches précédentes qui montrent l'existence de différences qualitatives mais aussi quantitatives pour l'expression de ces gènes (Roger *et al.* 2008c).

Nous avons ensuite comparé les séquences génétiques et l'état chromatinien autour de ces gènes pour chacune des deux souches. Notre analyse comparative de la variation des séquences des promoteurs dans les différents groupes de famille de gènes, a révélé une forte conservation des séquences promotrices des *SmPoMucs* entre les deux souches.

Contrairement aux séquences codantes des *SmPoMucs*, pour lesquelles de nombreux événements de recombinaison ont été précédemment décrits (Roger *et al.* 2008b), un seul de ces événements est détecté dans les séquences promotrices (du groupe 2 au groupe 1). L'étude de Roger *et al.* suggère que les recombinaisons entre les gènes génèrent de la variabilité génétique (Roger *et al.* 2008b). Cependant, les résultats obtenus dans nos analyses indiquent qu'elles sont moins fréquentes dans les régions promotrices.

La différence de séquence, entre les promoteurs des souches Brésil, Guadeloupe et NMRI (génomme de référence, (Berriman *et al.* 2009)), est petite. De plus, basé sur une analyse microsatellites de 14 marqueurs neutres, un très faible nombre de substitutions est constaté entre les souches Brésil et Guadeloupe. Ainsi, seulement une à deux substitutions ont été observées entre ces deux souches, à l'exception du groupe 2 dans les régions non conservées entre les groupes de promoteurs des *SmPoMucs*. Mais aucune des substitutions entre la souche Brésil et la souche Guadeloupe n'apparaît dans la boîte TATA ou le TSS (*Transcription Start Site*) (Perrin *et al.* under revision).

Une analyse des différences nucléotidiques entre les deux souches, montre qu'il existe de 0 (groupe 2) à 8 substitutions (groupe 3.1(r1-r2)), ce qui entraîne des substitutions nucléotidiques par site similaire ou inférieur à ceux relevés dans les introns vraisemblablement neutres des *SmPoMucs* (Perrin *et al.* under revision).

Au niveau populationnel, nos analyses des promoteurs du groupe 1 des *SmPoMucs*, dans les deux souches d'intérêt, ont révélé une très faible variabilité allélique et nucléotidique au sein de chaque souche, et de hautes différences de fréquences alléliques entre la souche Brésil et la souche Guadeloupe en raison de substitutions fixes. Tous les individus sont homozygotes pour le groupe 1 des *SmPoMucs*, ce qui suggère que les souches de *S. mansoni* présentent, sur l'ensemble du génome, une homozygotie. L'entretien des deux souches en laboratoire mène à un coefficient de consanguinité élevé et à une clonalité conséquente pour ces deux souches (Bech *et al.* 2010). Malgré le manque de diversité génétique dans les deux souches, la différenciation génétique est élevée entre les deux souches puisque les allèles fixés dans chaque souche, pour le promoteur du groupe 1 des *SmPoMuc* et neuf microsatellites, sont différents. Par conséquent, avec la consanguinité élevée dans chaque souche, il est difficile d'estimer le niveau de variation entre les deux souches. Mais de faibles différences nucléotidiques apparaissent au sein et entre les souches, ce qui n'explique pas le niveau de variation d'ADN complémentaire observé, en particulier pour le groupe 2 des *SmPoMucs*

(Roger *et al.* 2008b). Au vu de nos résultats, il est possible que les séquences promotrices des gènes *SmPoMucs* et leurs fonctions soient conservées entre les deux souches étudiées. Ainsi, les différences génétiques entre les séquences au sein de chaque groupe de promoteurs *SmPoMucs*, ne peuvent pas expliquer seulement, à elles seules, la grande variation et le polymorphisme de compatibilité.

Des études récentes ont soulevé la question de l'implication de l'épigénétique dans la création de variabilité de ces gènes, suite à un traitement à la TSA (Trichostatine A) avec une étude de l'impact du statut d'acétylation des histones sur les larves miracidia (où les *SmPoMucs* sont exprimés) (Cosseau *et al.* 2010). L'inhibiteur d'HDAC influence les traits phénotypiques des larves *S. mansoni*.

Lors de nos travaux, nous avons analysé l'influence du traitement sur les niveaux d'expression des gènes pour les souches Brésil et Guadeloupe. Cette dernière présente une expression des *SmPoMucs* dose dépendante, alors que dans la souche Brésil aucun effet du traitement n'est observé. Ce résultat suggère que le statut d'acétylation des histones dans les séquences promotrices, est régulé différemment entre les deux souches. Nous avons tout de même vu que les HDACs jouent un rôle important dans la régulation du niveau d'acétylation dans la souche Guadeloupe, pouvant influencer à la fois le phénotype mais aussi le niveau de transcription. Plus précisément, un effet de la TSA est rapporté sur les groupes 1 et 2 dans cette souche alors qu'aucun effet n'est observé pour le groupe 3.1(r1-r2). Suite à des analyses de la structure chromatinienne, utilisant différents anticorps spécifiques de marques dites euchromatiques (H3K9ac) ou dites hétérochromatiques (H3K9me3), nous avons pu mettre en évidence que le taux d'acétylation du groupe 3.1(r1-r2) est très élevé, comparé aux autres groupes, dans la souche Guadeloupe. Cela suggère qu'il existe une régulation différentielle du niveau d'acétylation par les HDACs entre les groupes de *SmPoMucs* au sein d'une même souche (Perrin *et al.* under revision).

L'analyse de la structure chromatinienne démontre clairement des différences de marques épigénétiques dans les séquences promotrices des *SmPoMucs* du groupe 1 et du groupe 3.1(r1-r2) entre les deux souches, résultant en une configuration différente de la chromatine. En effet, les résultats montrent un enrichissement plus fort pour la marque H3K9ac (euchromatique) et un enrichissement plus faible pour la marque H3K9me3 (hétérochromatique) dans la souche Guadeloupe (incompatible) que dans la souche Brésil (compatible). Par conséquent, la structure de la chromatine locale diffère entre les deux

souches pour les groupes 1 et 3.1(r1-r2). Ces résultats sont en accord avec les données d'expression, puisqu'une acétylation plus forte est corrélée à une expression élevée dans la souche Guadeloupe (Perrin *et al.* under revision).

Nous avons montré que, chez *S. mansoni*, l'augmentation de la variation épigénétique conduit à une augmentation de la variation phénotypique (Cosseau *et al.* 2010). Il ressort de nos analyses que deux isolats de *S. mansoni*, avec différents traits d'histoire de vie, notamment leur compatibilité vis à vis d'un hôte intermédiaire, révèlent de faibles différences nucléotidiques des séquences promotrices des *SmPoMucs*, mais de grandes différences épigénétiques (Perrin *et al.* under revision). Dans les deux populations de parasites, la variation épigénétique a produit des variants phénotypiques, permettant leur adaptation à des environnements différents (des hôtes compatibles). Notre scénario est donc le suivant: au 15<sup>e</sup>-16<sup>e</sup> siècle, la souche ancêtre des souches contemporaines Guadeloupe et Brésil a migré *via* le commerce des esclaves de l'Afrique vers les îles des Antilles et de l'Amérique du Sud, respectivement (Morgan *et al.* 2005). Les deux souches ont dû s'adapter à un nouvel hôte intermédiaire. Le goulot d'étranglement initial, résultant de la migration d'un nombre limité de parasites, et la forte pression de sélection ont agi sur les deux variantes, génétiques et épigénétiques, des molécules clés du polymorphisme de compatibilité. Les variations génétiques et épigénétiques ont été fortement diminuées lors du maintien de ces deux souches en laboratoire (Perrin *et al.* under revision). Il est probable que la variation épigénétique conservée à partir de la souche ancestrale ainsi que le taux plus élevé des modifications épigénétiques plutôt que des variations génétiques dans les générations suivantes, a permis aux souches de parasite de s'adapter rapidement à leurs hôtes et au nouvel environnement. L'hypothèse est que l'information épigénétique concerne l'activité transcriptionnelle d'un gène et non pas son potentiel à coder. En d'autres termes, un gène peut être activé et désactivé en lien avec la structure chromatinienne, mais la protéine résultante reste inchangée. La perte de fonction des gènes peut être facilement imaginée par un mécanisme épigénétique, comme l'hétérochromatisation. L'hypothèse classique d'Ohno est que des duplications de gènes peuvent être impliquées dans l'évolution (Ohno *et al.* 1968) et des études ont montré que ces gènes dupliqués ont tendance à être hétérochromatiques (Rodin *et al.* 2005).

Il est intéressant de noter que les protéines *SmPoMucs* présentes dans le polymorphisme de compatibilité sont codées par des gènes dupliqués. Nous proposons que les gènes *SmPoMuc* dupliqués peuvent fournir un système supplémentaire pour la variation phénotypique (Perrin *et al.* under revision). Les études de Cosseau *et al.* (Cosseau *et al.* 2010) et nos résultats,

démontrent une modulation aléatoire de l'expression des gènes *SmPoMucs* à travers des changements de structure chromatinienne. Ce différentiel d'expression entraîne de nouvelles combinaisons de gènes *SmPoMucs* qui augmente la variation phénotypique, conférant ainsi un avantage au parasite puisque la probabilité que les phénotypes hôtes et parasites correspondent est forte, d'où une évolution adaptative plus rapide (Verstrepen & Fink 2009).

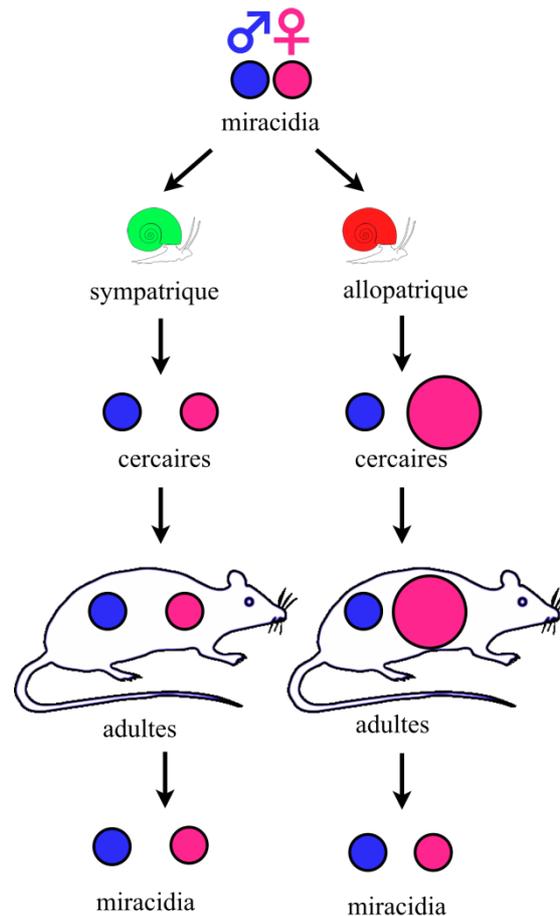
Afin d'investir plus profondément la question scientifique de l'adaptation de *S. mansoni* à l'Amérique du Sud, nous avons soumis une même souche de parasite à différentes souches d'hôtes intermédiaires mollusques. La souche de parasite choisie est *S. mansoni* Brésil (*SmBre*) car elle présente un fond génétique très homogène sur 14 marqueurs microsatellites neutres. Les souches d'hôtes intermédiaires utilisées sont : la souche *B. glabrata* Brésil (sympatrique vis à vis de la souche de parasite) et la souche *B. glabrata* Guadeloupe (allopatrique). Ces cycles sympatrique et allopatrique sont maintenus sur trois générations.

Lors du développement de *S. mansoni*, il est possible de mesurer différents traits de vie : au stade adulte (sexe-ratio, réponse inflammatoire de l'hôte, infectivité des cercaires), au stade miracidia (sexe-ratio), au stade sporocyste (prévalence, intensité d'infestation), au stade cercaire (sexe-ratio, chronobiologie et quantité de l'émission cercarienne). Durant ces travaux, nous avons mesuré précisément un maximum de ces traits, plus particulièrement : les sexes-ratios à tous les stades, la réponse inflammatoire de l'hôte définitif, l'infectivité cercarienne et la prévalence de mollusques infestés. D'autres traits, comme l'émission cercarienne et sa chronobiologie, n'ont pu être quantifiés mais des observations microscopiques montrent un différentiel entre la condition sympatrique et la condition allopatrique. En effet, les cercaires issues d'un développement au sein d'un hôte intermédiaire allopatrique présentent une émission de l'ordre de 100 fois plus forte que la condition sympatrique et la chronobiologie d'émission est modifiée. Nos observations sont en corrélation avec des études réalisées précédemment au sein du laboratoire (Théron 1980). Dans une condition sympatrique, la courbe d'émission est caractérisée par un pic d'émergence bien individualisé, traduisant une faible dispersion temporelle. Dans le cas de la condition allopatrique, les courbes d'émission subissent un important aplatissement, caractérisant une dispersion temporelle plus large (Théron 1980). En ce qui concerne l'émission drastiquement augmentée, elle s'explique par une augmentation de l'intensité d'infectivité du parasite au niveau du mollusque (plus de miracidia se sont implantés et développés) ou par le fait que chaque sporocyste développé produit plus de cercaires pour contrer ce changement environnemental.

Parmi les autres traits de vie quantifiés précisément, nous constatons que pour ceux relatifs à l'infectivité cercarienne, la réponse inflammatoire de l'hôte définitif et le sexe-ratio au stade miracidia, aucun ne présente une différence suite au développement du parasite au sein de l'hôte allopatrique. L'hôte intermédiaire n'affecte pas ces différents traits (Lepesant *et al.* in prep.) (Annexe 1). Par contre, il semble agir sur les prévalences de mollusques infestés et les sexes-ratios aux stades cercaires et adultes. En effet les prévalences de mollusques infestés sont significativement plus faibles dans la condition allopatrique (~ 84 %) que dans la condition sympatrique (~ 96 %). Cette différence est maintenue sur les trois générations et peut perdurer jusqu'à sept générations (Théron, communication personnelle). Dans le cas des sexes-ratios, un changement drastique est apparu dès le stade cercaire (stade suivant directement le développement du parasite dans la souche de mollusque allopatrique) et est retrouvé au stade adulte. La population initiale de parasites, servant à infester les deux hôtes intermédiaires (sympatrique et allopatrique) présente un sexe-ratio équilibré (1 :1). Lors du développement au sein de la souche de mollusque sympatrique puis au sein de l'hôte définitif, ce sexe-ratio reste inchangé ou se biaise en faveur des mâles tout au long du cycle et ce sur les trois générations. Ce qui est en accord avec les études précédentes établissant que l'espèce *S. mansoni* possède majoritairement un sexe-ratio biaisé en faveur des mâles (Morand *et al.* 1993; Boissier *et al.* 1999; Boissier & Moné 2000, 2001; Boissier *et al.* 2003a; Beltran & Boissier 2009). Inversement, lorsque la population de parasites (initialement sexes équilibrés) s'est développée au sein de l'hôte intermédiaire allopatrique, un biais en faveur des femelles apparaît avec un enrichissement de 1.8 par rapport à la population initiale. Ce biais est retrouvé après infestation de l'hôte définitif mais avec un enrichissement moindre, qui s'explique par une infectivité inférieure des cercaires femelles par rapport aux cercaires mâles (Boissier *et al.* 1999). Un tel changement dans les sexes-ratios de *S. mansoni* a déjà été montré plusieurs fois (Moné 1997; Zavodna *et al.* 2008), hormis les études de Moné *et al.* indiquant un sexe-ratio biaisé plus fortement en faveur des mâles lorsque le mollusque sympatrique infesté est en compétition avec une autre espèce de mollusque non vectrice de *S. mansoni* (Moné 1997). La présence de ce mollusque non vecteur stimule la croissance du mollusque infesté et augmente ainsi l'espace disponible pour le parasite (Moné 1991; Mouahid *et al.* 1992), favorisant le développement des sporocystes mâles d'où un sexe-ratio biaisé mâle au niveau de l'hôte définitif (plus de cercaires mâles produites que de femelles) (Moné 1997).

En revanche, dans les études de Zavodna *et al.*, il est démontré l'apparition d'un biais en faveur des femelles lorsqu'une population de parasites s'est développée au sein d'un hôte intermédiaire allopatrique (Zavodna *et al.* 2008). Il est également précisé que les mâles et les femelles schistosomes possèdent des différentiels d'expression génique ou protéique durant le cycle de développement (Fitzpatrick *et al.* 2004; Fitzpatrick *et al.* 2005; Wuhler *et al.* 2006). Les auteurs proposent que les variations de réponse immunologique entre les différentes souches, vis à vis des parasites mâles et femelles, peuvent filtrer les sexes qui seront ensuite transmis à l'hôte définitif. Ils concluent donc que la base génétique de l'hôte intermédiaire peut être d'une importance majeure dans l'établissement des individus parasites jusqu'à l'hôte définitif (Zavodna *et al.* 2008).

Les résultats de nos travaux sont cohérents avec ceux obtenus par Zavodna *et al.*. En effet, le sexe-ratio du parasite est biaisé en faveur des femelles aux stades cercaires et adultes, après son développement au sein d'un hôte intermédiaire allopatrique. Par contre, le sexe-ratio au stade miracidia est équilibré ou biaisé en faveur des mâles alors qu'après la reproduction sexuée, nous observons un retour « à la normale ». Cependant, si cette population de miracidia, issue d'un développement initial dans l'hôte allopatrique, sert à infester de nouveau un hôte allopatrique, le biais en faveur des femelles est retrouvé. L'environnement hôte peut influencer les sexes-ratios, sur 3 générations et peut perdurer jusqu'à 7 générations (Théron, communication personnelle). Nos résultats sur les sexes-ratios sont schématisés dans la figure ci-dessous (Figure 59).



**Figure 59: Schéma de synthèse des résultats obtenus sur le trait d’histoire de vie sexe-ratio dans les conditions sympatriques et allopatriques.** *Lorsqu’une population, sexe-ratio équilibré, est maintenue dans un cycle sympatrique, le sexe-ratio reste inchangé. Mais, quand cette même population est maintenue à travers un hôte intermédiaire allopatrique, le sexe-ratio est drastiquement biaisé en faveur des femelles. Suite à la reproduction sexuée, ce biais retrouve un ratio 1 :1.*

Par cette évolution expérimentale, nous avons établi l’influence de l’environnement (ici l’hôte intermédiaire) sur certains traits d’histoire de vie. Ces modifications phénotypiques ne sont pas dues à des modifications génétiques car les analyses sur 14 marqueurs microsatellites n’ont montré aucune différence entre les deux conditions, sur les trois générations.

Nous avons alors décidé d’observer les éventuels changements transcriptomiques qui peuvent refléter ces changements phénotypiques, en séquençant entièrement l’ARN des populations cercariennes pour les deux conditions sur les trois générations. Cette expérience a été possible grâce à l’optimisation du protocole d’extraction d’ARN afin d’avoir une quantité et une qualité suffisante. Le séquençage a été réalisé par la technique Illumina, en collaboration avec Montpellier GenomiX.

L'analyse préliminaire de ces résultats, nous a permis d'identifier 223 transcrits différentiellement exprimés dans la condition allopatrique, en comparaison à la condition sympatrique. Certains transcrits sont induits (31 %) ou surexprimés (21%) et d'autres sont réprimés (13 %), dans les trois générations. Enfin, une proportion de transcrits (35%) ont des profils transcriptionnels variables selon la génération étudiée. En observant les transcrits présentant une activité stable sur les trois générations (145 au total), certains sont associés à des marques d'histones distinctes (H3K9ac, H3K4me3, H3K9me3 et/ou H3K27me3). Pour ~18% d'entre eux, une différence pour au moins une marque est constatée entre les deux conditions alors que pour ~14 % des transcrits différentiellement exprimés on ne relève pas de différences chromatiniennes. Il est à noter que pour une majorité des transcrits (~68 %) nous n'avons pas de données sur la structure chromatinienne associée.

Suite à une annotation automatique des 223 transcrits différentiellement exprimés dans la condition allopatrique, nous avons pu identifier un transcrit, Smp\_140390, annoté comme un gène prédit qui code pour une enzyme N-méthyltransférase. De plus, ce gène montre de grandes similitudes avec le gène SETD1B codant pour une méthyltransférase agissant spécifiquement sur la lysine 4 de l'histone 3 (H3K4) (Lee *et al.* 2007). La marque H3K4 méthylée représente un indice d'activation épigénétique de la transcription (Pekowska *et al.* 2011). Il est intéressant de soulever le fait que ce transcrit est exprimé deux fois plus chez les individus issus du développement au sein de l'hôte intermédiaire allopatrique. Ce gène semble donc être un bon candidat pour étudier l'influence des changements environnementaux sur la réponse adaptative rapide du parasite, caractérisée par des modifications transcriptomiques et épigénétiques de nature chromatinienne.

De ce fait, nous avons mis en évidence l'influence de l'hôte intermédiaire sur des traits phénotypiques mais aussi sur l'expression de certains transcrits et avons également montré que ces différences transcriptomiques peuvent être associées à des modifications de la structure chromatinienne.

Pour pousser plus loin l'implication de l'épigénétique suite à un changement environnemental, nous avons analysé les profils de marques d'histones (H3K9ac, H3K4me3, H3K9me3 et H3K27me3) pour les deux conditions, à la génération 3, afin de proposer des profils types pour chaque marque, en fonction la transcription. Nous sommes les premiers à avoir généré l'épigénome de *S. mansoni*, sur quatre marques d'histone. Les analyses autour du TSS nous ont permis d'établir des profils, pour chaque marque, associés à un niveau de

transcription : plus le gène est exprimé, plus les marques (H3K9ac, H3K4me3, H3K9me3 et H3K27me3) sont enrichies. Dans tous les cas, si aucune expression n'est quantifiée, aucun enrichissement n'est observé. Si une très faible expression apparaît, un pic est constaté pour toutes les marques mais en revanche, les profils plafonnent rapidement à partir d'une expression moyenne (10 FPKMs). Nous avons aussi montré que toutes les marques ont un profil particulier et sont toutes présentes dans la région du TSS sauf certaines qui peuvent être prolongées en amont et/ou en aval de cette région. La particularité de *S. mansoni*, est que les marques H3K9me et H3K27me3 sont enrichies pour les gènes exprimés alors qu'elles sont classiquement associées à des marques hétérochromatiques, donc plutôt non permissives à la transcription (Melcher *et al.* 2000; Rea *et al.* 2000; Nakayama *et al.* 2001; Peters *et al.* 2003; Rice *et al.* 2003).

Nos analyses préliminaires soulignent l'existence de marques d'histones très conservées mais, chez *S. mansoni*, les marques dites hétérochromatiques (H3K9me3, H3K27me3) ne sont pas forcément associées à des régions de répression de la transcription. Plusieurs expérimentations au laboratoire ont été menées afin d'identifier des marques d'histones spécifiques de l'hétérochromatine et donc de la répression de la transcription, mais soit les anticorps commercialisés ne permettent pas un enrichissement suffisant, soit les marques sont associées à de la transcription. Il est donc envisageable que la biologie de l'organisme *S. mansoni* empêche l'utilisation d'anticorps commerciaux, en raison d'une structure hétérochromatique particulière et non classique.

- En résumé

Nos analyses globales et ciblées, ainsi que l'évolution expérimentale ont mis en évidence des différences génétiques et épigénétiques entre deux souches de parasites, issues d'isolat géographique distincts.

En étudiant les différences épigénétiques qui sont présentes dans le génome, nous avons pu identifier un gène potentiellement impliqué dans le polymorphisme de compatibilité, en plus des *SmPoMucs* dont leur analyse a montré que la distinction majeure entre les deux souches, est basée sur une différence dans la structure chromatinienne.

Les données de génétique acquises dans plusieurs études n'expliquent pas, à elles seules, ce polymorphisme alors que l'évolution expérimentale prouve l'influence directe de

l'environnement-hôte sur les phénotypes, mais également la transcription de certains gènes ainsi que leur structure chromatinienne associée.

Les études globales des transcriptomes et épigénomes des individus issus de la condition sympatrique et de la condition allopatriques sont encore en cours. Les résultats obtenus devront par la suite être confirmés par une analyse ciblée par PCR quantitative.

**Comment existe-t-il une différenciation sexuelle entre les mâles et les femelles, en l'absence présence de gène W-spécifique ?**

Avant de répondre à cette troisième question scientifique, il est nécessaire de rappeler qu'en dépit des progrès considérables, les éléments responsables de la mise en place des chromosomes sexuels restent encore énigmatiques. Selon le modèle du *cliquet de Müller*, la reproduction sexuée a évolué pour que les mutations délétères soient éliminées par recombinaison entre les autosomes parentaux (Felsenstein 1974). Afin de maintenir l'isolement des deux sexes différents, la recombinaison doit cependant être réprimée (au moins partiellement) entre les chromosomes sexuels.

L'accumulation de séquences répétitives sur le chromosome sexuel hétérogamétique a également été constatée dans de nombreux exemples, mais leurs rôles sont encore inconnus. Le fait que l'ADN répétitif soit considéré comme non fonctionnel est dorénavant contesté en raison de la découverte de la transcription de séquences répétées sur les autosomes ainsi que la production de petits ARNs qui peuvent être liés à des événements d'hétérochromatisation (Fukagawa *et al.* 2004). La présence de gros blocs hétérochromatiques est également une caractéristique commune des chromosomes sexuels. Jusqu'à présent, ces observations ont été faites, isolées les unes des autres, et généralement dans des espèces différentes, ce qui rend la construction d'un modèle hypothétique difficile. Malgré leur abondance, les séquences répétées sont souvent considérées comme de l'ADN poubelle mais des études démontrent que ces séquences sont nécessaires à l'établissement de la frontière entre une région euchromatique et une région hétérochromatique (Shapiro & Von Sternberg 2005).

Lorsque le génome de *S. mansoni* a été publié, 40% a été identifié comme étant composé de séquences répétitives (Berriman *et al.* 2009). A la suite du séquençage massif du génome d'une souche du laboratoire, nous avons constaté qu'un grand nombre de séquences

répétitives est présent. Nous avons donc décidé de mettre en place un assemblage *de novo* pour identifier toutes les séquences répétées du génome de *S. mansoni* sans a priori (Lepesant *et al.* 2012b) (Annexe 3). Cette stratégie nous a permis d'identifier 4143 nouvelles séquences répétées, amenant la proportion répétée du génome à environ 47%. Nous avons par la suite annoté ces séquences afin de les classifier. Grâce à cette nouvelle approche, nous avons authentifié 14 nouveaux marqueurs microsatellites dont 9 semblent être des marqueurs neutres (Lepesant *et al.* 2012b) (Annexe 3).

Nous avons décidé d'analyser complètement la composition en séquences, les gènes, le contenu des séquences répétées, ainsi que la transcription et la structure de la chromatine dans les régions chromosomiques sexes spécifiques entre les chromosomes Z et W. Dans ces régions d'intérêt, la répression de recombinaison a été décrite précédemment (Criscione *et al.* 2009). Nos données, en relation avec les études précédentes, ont affiné les modèles actuels pour la suite des événements qui ont conduit à la différenciation des chromosomes sexuels chez *S. mansoni*. Ils pourraient représenter un modèle général de ce processus chez les espèces ZW.

Criscione *et al.* ont identifié une région de 20 *scaffolds* où la répression de recombinaison est observée (Criscione *et al.* 2009). Ils suggèrent que ces séquences sont spécifiques du chromosome Z, présent chez les mâles (ZZ) et les femelles (ZW). Nous avons trouvé dans 13 de ces *scaffolds*, un ratio mâle/femelle  $\geq 1.5$  (en PCR quantitative) indiquant une surreprésentation dans le génome masculin. Ces résultats sont cohérents avec le fait que les mâles possèdent deux chromosomes Z alors que les femelles n'en ont qu'un seul. Cependant, 7 *scaffolds* n'ont montré aucun déséquilibre entre les mâles et les femelles (mâle/femelle  $\leq 1,4$ ), c'est-à-dire que ces séquences ne sont pas spécifiques du chromosome Z. En d'autres termes, si la recombinaison est réprimée ce n'est pas dû à l'absence de séquences homologues entre les chromosomes sexuels. Au regard de ces résultats, nous trouvons au moins 2 blocs de séquences qui sont partagés entre le chromosome Z et W, situés dans la grande région où la recombinaison est réprimée. Ce résultat a été confirmé par la version la plus récente de l'assemblage du génome (Protasio *et al.* 2012).

Nous voyons trois hypothèses possibles à cette répression de recombinaison: (i) les blocs de séquence Z / W sont inversés, (ii) les séquences sont hétérochromatiques empêchant ainsi la recombinaison, (iii) les *scaffolds* sont chimériques. En effet, sur les 48 *scaffolds* initialement trouvés dans le groupe Z/W (Criscione *et al.* 2009), quatre sont présents sur d'autres

chromosomes dans la version 5.2 du génome de *S. mansoni* (Protasio et al. 2012). De plus, nos analyses n'ont pas permis d'identifier des gènes paralogues de détermination du sexe, au sein des gènes prédits dans les *scaffolds* identifiés comme Z-spécifiques.

Suite à l'assemblage d'un pseudogénome de séquences répétées et à son annotation, nous avons pu mettre en évidence que la région spécifique du chromosome W est essentiellement composée de 36 familles de séquences répétées arrangées en grands blocs satellites (Lepesant *et al.* 2012a; Lepesant *et al.* 2012b) (Annexe 2 et 3). Ces séquences répétitives sont abondantes sur le chromosome W, toutefois notre analyse par PCR sur différents individus mâles indique que, parfois, ces séquences peuvent également être retrouvées sur d'autres chromosomes mais avec un nombre de copies faibles. Ces séquences répétées peuvent être occasionnellement transmises aux autosomes, comme pour le cas de la séquence répétitive femelle W1 (Grevelding 1999) et peuvent dépendre de la structure de la chromatine autour des séquences répétées et/ou des régions flanquantes.

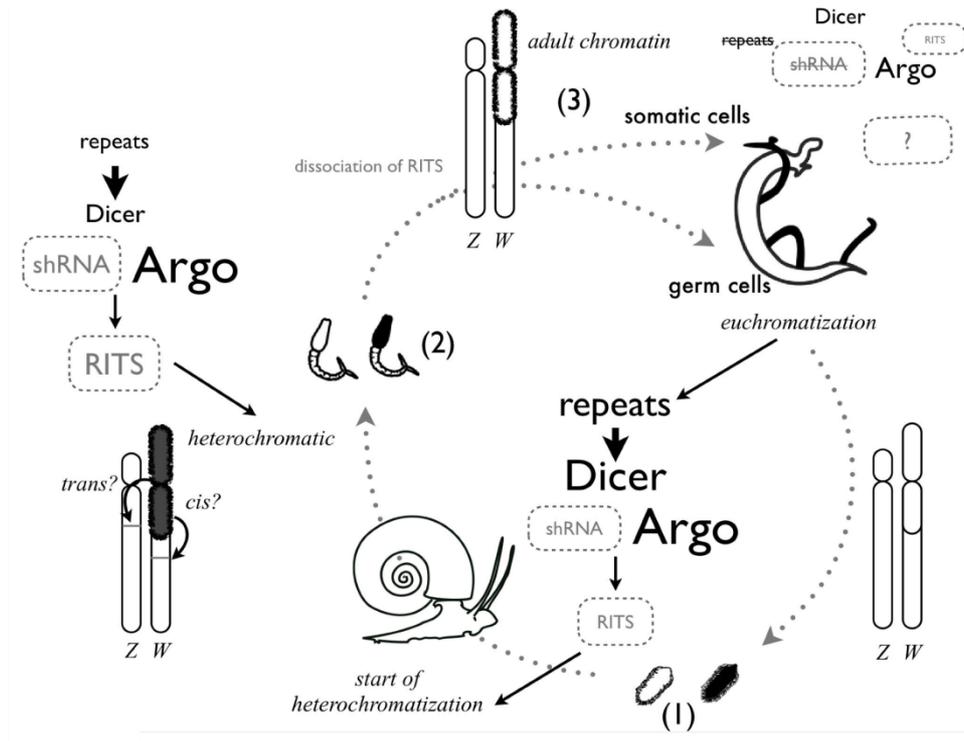
Dans la plupart des espèces qui possèdent des chromosomes sexuels Y ou W, il est constaté que : (i) des séquences répétitives s'accumulent sur ces chromosomes, (ii) de vastes régions sont hétérochromatiques et (iii) ces chromosomes se détériorent ou sont complètement absents dans un cas extrême. Nous montrons que, chez *S. mansoni*, le chromosome W ne fait pas exception à ces règles.

L'hétérochromatisation du chromosome W chez les schistosomes est connue depuis longtemps et a même été utilisée comme marqueur pour l'identification du sexe au stade cercaire, où les phénotypes mâles et femelles sont indiscernables (Grossman *et al.* 1980; Grossman *et al.* 1981; Liberatos & Short 1983). Sur la base de l'analyse cytogénétique, certains auteurs ont voulu montrer que l'hétérochromatisation du chromosome W commence au stade miracidia (Liberatos & Short 1983). Comme il est impossible de déterminer les bandes chromosomiques à ce stade puis réutiliser les larves pour l'infestation et la production d'adultes, ces résultats sont difficiles à vérifier. Nos résultats indiquent clairement que les séquences répétitives qui sont situées dans la région hétérochromatique du chromosome W portent une signature euchromatique au stade miracidia et la perdent progressivement au cours du développement jusqu'au stade adulte. Ce processus est accompagné d'une diminution de la transcription jusqu'à une inhibition complète des séquences répétées, au stade adulte présentant un dimorphisme sexuel. En effet, suite à une analyse de la transcription de certaines séquences répétées, nous avons mis en évidence que les séquences

femelles spécifiques ne sont transcrites qu'aux stades larvaires (miracidia et cercaires) et jamais au stade adulte. La transcription est donc restreinte aux stades où aucun dimorphisme sexuel n'est visible (Lepesant *et al.* 2012a) (Annexe 2).

Au cours de la transition miracidia/cercaires, c'est-à-dire au moment précis où le dimorphisme sexuel commence à se développer, les séquences répétitives s'hétérochromatisent. Les séquences répétées sexe-spécifique sont retrouvées dans de nombreuses espèces (Griffiths & Holland 1990a; Griffiths & Holland 1990b; Kejnovsky *et al.* 2009). Dans certains cas, la transcription a été décrite et l'on soupçonne que ces séquences répétitives jouent un rôle dans le processus de détermination du sexe (Shapiro & Von Sternberg 2005; Ugarkovic 2005).

La transcription des éléments répétitifs de type satellite chez *S. mansoni* est particulièrement intéressante suite à la découverte récente d'expression, stade dépendante, des éléments qui constituent l'ARN interférence (ARNi) chez les schistosomes (Krautz-Peterson & Skelly 2008; Luo *et al.* 2009). Dans de nombreux organismes, le RNAi et les changements de structure de chromatine sont liés (Volpe *et al.* 2002; Fukagawa *et al.* 2004; Verdel *et al.* 2004; Kanellopoulou *et al.* 2005). Il est tentant de spéculer que la transcription des séquences répétées W-spécifiques est effectivement à l'origine de compaction de la chromatine sur le chromosome W au cours du cycle de vie. Aussi, à partir de nos données et celles de la littérature, nous proposons un modèle hypothétique sur la différenciation sexuelle dans un système ZW (Figure 60).



**Figure 60: Modèle hypothétique du rôle des séquences répétées femelles spécifiques dans l'hétérochromatization du chromosome W** (Lepesant *et al.* 2012a) (Annexe 2). La transcription de séquences répétées spécifiques femelles entraînerait l'hétérochromatization du chromosome W, pouvant agir en *cis* ou en *trans* sur des régions proches.

Dans notre modèle, la structure chromatinienne autour des séquences répétitives est restaurée au moment de la reproduction sexuée (formation de miracidia). Dans les miracidia, les séquences répétitives sont euchromatiques et plusieurs d'entre elles sont transcrites. Les transcrits sont traités par une voie qui présente une similarité avec le RNA interférence. Ensuite, un processus de répression est initié pour former un complexe de séquences répétées induisant la formation de l'hétérochromatine autour de ces séquences. A ce stade, les miracidia ont infesté l'hôte mollusque et se développent. Au stade cercaire, la plupart des séquences répétées sont hétérochromatiques et non transcrites. Nous émettons l'hypothèse que l'hétérochromatization s'étend au-delà des frontières des séquences répétitives et que les *loci* proches sont aussi inhibés. Si parmi ces *loci*, un déterminant sexuel est présent, l'hétérochromatization conduirait à un effet dose qui pourrait être à l'origine de la formation du phénotype femelle adulte. Une fois l'inhibition de ce locus, en *cis* ou *trans*, les séquences répétées ne sont plus transcrites et la structure péricentromérique de la chromatine est fixée dans une configuration inconnue. La transcription reste réprimée. A partir de nos données, nous avons mis en avant que les marques euchromatiques (H3K9ac et H3K4me3) ou

hétérochromatiques (H3K9me3 et H3K27me3) ne sont pas abondantes (Lepesant *et al.* 2012a) (Annexe 2).

Ce modèle est conforté par nos résultats du traitement *in vitro* sur adultes. Suite au traitement avec un inhibiteur d'HDAC (la TSA), ne détectant aucune induction de transcription dans les séquences répétitives. Seule une induction de retrotransposons autosomaux a été observée.

Notre modèle théorique peut être complété par les récentes études précisant que le chromosome W a évolué après au moins 4 phénomènes d'inversion (Hirai *et al.* 2012) et l'hétérochromatisation. Les auteurs montrent que les régions homologues entre les deux chromosomes sexuels ont été déplacées du chromosome Z vers le chromosome différencié W par trois événements d'inversion, avant l'hétérochromatisation. Seule une inversion a été identifiée comme intervenant dans une région hétérochromatique, après l'hétérochromatisation du chromosome W (Hirai *et al.* 2012). L'hypothèse sur l'évolution des chromosomes Z et W serait alors :

- Une première inversion sur un des chromosomes homologues ancestraux de ZW, avant une hétérochromatisation ;
- Une seconde inversion au milieu du bras long du chromosome ancestral du W ;
- Une troisième inversion sur le bras court du chromosome Z, après la séparation des chromosomes Z et W ;
- Une quatrième inversion uniquement sur le chromosome W après l'évolution des chromosomes sexuels et l'hétérochromatisation du chromosome W.

○ En résumé

Nos expériences menées établissent d'une part l'absence de gène W-spécifique et d'autre part que la région spécifique du chromosome W est composée de séquences répétitives. Toutes ces séquences sont localisées en larges blocs satellites dans la région hétérochromatique. Cependant, uniquement aux stades larvaires, certaines de ces séquences répétitives sont transcrites et sont associées à une marque euchromatique au stade miracidia. Suite au traitement des adultes avec un inhibiteur de HDAC, aucune induction de transcription à ce stade n'a été montrée. De ce fait, en nous basant sur nos données et la littérature, nous avons proposé un modèle hypothétique impliquant la transcription de ces séquences répétitives femelles spécifiques menant à une hétérochromatisation de cette région sur le chromosome

W, pouvant agir en *cis* ou en *trans* afin de réguler les gènes à proximité. Ce modèle propose une suite d'évènements moléculaires amenant finalement à la différenciation sexuelle entre les mâles et les femelle et pourra être affiné grâce aux études récentes relatives à l'existence d'un phénomène d'inversion (Hirai *et al.* 2012).

## **CHAPITRE 5**

### **Conclusions et Perspectives**

L'objectif principal de ces travaux de thèse était de mettre en avant l'influence de l'environnement sur le parasite humain *S. mansoni* et d'appréhender les mécanismes moléculaires sous-jacents à une évolution adaptative rapide.

Les six questions posées au début de cette thèse étaient : (i) Suite à un stress environnemental, y-a-t'il apparition de modifications épigénétiques de nature chromatinienne ? (ii) Le changement environnemental peut-il entraîner des modifications d'ordre transcriptomique ? (iii) Ces modifications, tant épigénétiques que transcriptomiques, sont-elles ciblées et/ou aléatoires dans le génome ? (iv) Existe-il une corrélation directe entre les modifications épigénétiques et les modifications des profils de transcription ? (v) Ces modifications moléculaires sont-elles corrélées à des changements phénotypiques visibles ? (vi) Les changements engendrés tant phénotypiques que moléculaires sont-ils hérités par les générations futures ?

Pour répondre à ces interrogations, nous avons élaboré plusieurs évolutions expérimentales, en faisant varier des paramètres biotiques et abiotiques, considérés comme inhabituels pour le parasite. Ces différentes évolutions ont permis de poser deux énigmes : (1) « Adaptation rapide de *S. mansoni* à une drogue anti-helminthe » et (2) « Adaptation rapide de *S. mansoni* à l'Amérique du Sud ». À cette deuxième énigme, une troisième s'est ajoutée : (3) « Différenciation sexuelle en l'absence de gène W-spécifique dans un système ZW ».

Les deux premières grandes énigmes ont été appréhendées par la mise en place d'évolutions expérimentales, visant à donner des éléments de réponses à nos six questions posées dans l'introduction.

Lors de la première énigme « Adaptation rapide de *S. mansoni* à une drogue anti-helminthe », nous avons étudié la réponse adaptative de *S. mansoni* à un traitement à l'hycanthone ou au lucanthone. Ces drogues, anciennement utilisées, entraînent l'apparition de résistance dès la première génération. Nos travaux ont établi que la résistance est induite par le traitement. Ces résultats ne corrèlent pas avec ceux obtenus sur les bactéries par Luria & Delbrück (Luria & Delbrück 1943). À ce jour, nos expérimentations ne nous permettent pas de répondre précisément à nos six interrogations car les analyses épigénétiques et le traitement bioinformatique des données reçues tardivement sont toujours en cours, suite au séquençage massif de l'ADN immunoprécipité, avec des anticorps spécifiques de marques euchromatiques (H3K9ac et H3K4me3) et hétérochromatiques (H3K9me3 et H3K27me3), d'individus résistants et d'individus sensibles. Nous avons néanmoins constaté que la qualité

des séquences pour les marques H3K9ac, H3K9me3 et H3K27me3 semble bonne alors que celle concernant la marque H3K4me3 est moindre. Les analyses à suivre seront calquées sur le même protocole expérimental que celui utilisé pour le traitement des données du séquençage massif de l'expérience sympatrique *vs* allopatrique. Ces futures études viseront à mettre en avant les mécanismes épigénétiques, basés sur la structure chromatinienne, pouvant engendrer la résistance à la drogue antihelminthe hycanthone. Les résultats nous permettront de répondre à deux questions : (i) Suite à un stress environnemental, y-a-t'il apparition de modifications épigénétiques de nature chromatinienne ? (ii) Ces modifications moléculaires sont-elles corrélées à des changements phénotypiques visibles ?

Il nous est tout de même possible de proposer l'hypothèse d'une hétérochromatisation après traitement de l'hôte de *S. mansoni*, qui entraînerait l'absence d'activité enzymatique chez les individus résistants. Cette supposition est cohérente avec le fait que les phénotypes résistants sont métastables, c'est à dire présents sur plusieurs générations puis disparaissent à un moment donné de la population. La résistance à l'hycanthone peut être liée à une épimutation.

S'agissant de l'interrogation : « Les changements engendrés tant phénotypiques que moléculaires sont-ils hérités par les générations futures ? », nous avons maintenu notre évolution expérimentale, avec le lucanthone, sur deux générations successives. Les résultats obtenus sont en partie attendus et en partie étonnants. Ainsi, le nombre de vers résistants retrouvés après traitement est plus élevé dans la deuxième génération traitée, ce qui est conforme aux résultats publiés sur l'hycanthone (Jansma *et al.* 1977). Mais le fait d'observer un nombre très restreint dans la condition contrôle ne nous permet pas d'interpréter complètement nos résultats. Les molécules hycanthone et lucanthone n'étant plus commercialisées, la même évolution expérimentale avec la molécule oxamniquine peut être envisagée car plusieurs études établissent la résistance croisée avec l'hycanthone (Cioli *et al.* 1992) mais aussi avec le praziquantel utilisé actuellement (Fallon & Doenhoff 1994). Il serait donc intéressant de réaliser cette expérience afin de voir si les individus résistants de première génération présentent les mêmes profils épigénétiques que les individus résistants à l'hycanthone et de la maintenir sur plusieurs générations pour vérifier si les changements potentiellement observés sont transmis à la génération suivante.

Lors de la deuxième énigme « Adaptation rapide de *S. mansoni* à l'Amérique du Sud », nous avons étudié le génome et épigénome de deux souches issues d'isolats géographiques

différents. Ces analyses ont été menées à l'échelle globale mais également de manière ciblée sur des gènes d'intérêt impliqués dans le polymorphisme de compatibilité, les *SmPoMucs*.

Ces études mettent en avant des différences génétiques ainsi que des différences épigénétiques entre ces deux souches à l'échelle du génome. Nous avons pu identifier un gène candidat *Smp\_171100*, présentant des différences épigénétiques, potentiellement impliqué dans ce polymorphisme de compatibilité.

Par ailleurs, l'analyse approfondie des gènes *SmPoMucs* a mis en évidence une diversité génétique très faible entre les deux souches, qui ne peut pas expliquer à elle seule ce polymorphisme. Cependant, une région, *Smp\_scaff000234*, comprenant 4 copies de gènes MEG, a été identifiée comme différence génétique entre les deux souches Brésil et Guadeloupe. Cette région peut être impliquée dans un phénomène d'évasion immune, et présente un intérêt particulier puisque les gènes *SmPoMucs* appartiennent à la famille des MEG et sont d'ores et déjà établis pour jouer un rôle dans le polymorphisme de compatibilité. Cette région, récemment identifiée, peut jouer un rôle supplémentaire dans ce polymorphisme.

Cependant, d'importantes différences épigénétiques de nature chromatinienne ont été démontrées pour les gènes *SmPoMucs* entre les deux souches.

Enfin, une évolution expérimentale d'infestation d'hôte intermédiaire allopatrique montre que cet hôte peut influencer fortement certains traits de vie du parasite (prévalences de mollusques infestés et sexes-ratios) ainsi que le transcriptome et l'épigénome. En effet, suite au développement au sein de l'hôte allopatrique, la population cercarienne émise présente des différences d'expression pour un certain nombre de transcrits (223 au total). Ces différences de transcription peuvent être d'ordre induction, surexpression ou répression. Parmi ces transcrits différentiellement exprimés dans la condition allopatrique, certains sont associés à une modification de marque d'histone. L'analyse approfondie de la structure chromatinienne, nous a permis de mettre en évidence des profils types pour les marques H3K9ac, H3K4me, H3K9me3 et H3K27me3, selon l'activité transcriptionnelle à l'échelle du génome. Ces profils ont été étudiés dans les régions du TSS et révèlent que l'augmentation de la marque est associée au niveau de transcription : plus un gène est exprimé, plus la marque est abondante. Cependant, l'enrichissement plafonne rapidement à partir d'une expression moyenne. Selon les marques, les profils diffèrent : certaines marques sont présentes uniquement dans le TSS alors que d'autres se maintiennent en amont et en aval du TSS, avec une abondance

décroissante. Il est intéressant de noter que chez *S. mansoni*, les marques H3K9me3 et H3K27me3, actuellement associées à l'hétérochromatine, sont présentes dans des régions transcrites et que l'abondance de ces marques augmente avec le niveau de transcription.

Les analyses de transcriptomiques et épigénomiques étant toujours en cours, nos résultats préliminaires devront être validés ultérieurement par une méthode ciblée de PCR quantitative et devront être approfondis au regard des relations qui existent entre phénotype, transcriptome et épigénome. Il nous est néanmoins possible de donner d'ores et déjà des éléments de réponses aux différentes questions posées préalablement :

- (i) Suite à un stress environnemental, y-a-t'il apparition de modifications épigénétiques de nature chromatinienne ?

Ce qui est le cas, puisque le parasite qui s'est développé au sein d'un hôte intermédiaire allopatrique, présente des profils chromatinien différents. Ces résultats sont basés sur l'analyse de quatre marques d'histones : H3K9ac, H3K4me3, H3K9me3 et H3K27me3, nous permettant d'avoir des profils types.

- (ii) Le changement environnemental peut-il entraîner des modifications d'ordre transcriptomique ?

Effectivement, nous avons identifié 223 transcrits différentiellement exprimés dans la population issue de la condition allopatrique sur les trois générations.

- (iii) Ces modifications, tant épigénétiques que transcriptomiques sont-elles ciblées et/ou aléatoires dans le génome?

Les analyses sont encore en cours, mais nous avons tout de même observé pour les données de transcriptomique, que certains transcrits présentent une expression différentielle dans la condition allopatrique, sur les trois générations successives (~1%). Toutefois, d'autres transcrits semblent être influencés différemment entre les trois générations (~10%). On peut dès lors déduire que les modifications d'ordre transcriptomique sont essentiellement aléatoires même si certaines sont ciblées. Il est envisageable que certains *loci* soient plus sensibles aux modifications épigénétiques.

En ce qui concerne les données épigénétiques, seule la troisième génération est en cours d'analyse. Nous avons néanmoins préparé pour le séquençage, les échantillons pour la

première et deuxième génération afin d'observer les changements épigénétiques sur trois générations successives.

(iv) Existe-il une corrélation directe entre les modifications épigénétiques et les modifications des profils de transcription ?

Lors des analyses, nous avons établi que 18% des transcrits différentiellement exprimés dans la condition allopatrique présentent des profils de marques d'histones différents de la condition sympatrique. De plus, nous avons pu mettre en avant des profils types pour chaque marque dans la région du TSS. L'abondance des quatre marques dépend du niveau d'expression. Les modifications épigénétiques et de transcription sont corrélées.

(v) Ces modifications moléculaires sont-elles corrélées à des changements phénotypiques visibles ?

Dans l'évolution expérimentale mise en œuvre, nous avons pu constater que l'environnement hôte intermédiaire peut influencer différents traits phénotypiques. Une étude plus approfondie des données de transcriptomiques et épigénétiques sera nécessaire pour établir un éventuel lien entre les changements moléculaires et les changements phénotypiques.

(vi) Les changements engendrés tant phénotypiques que moléculaires sont-ils hérités par les générations futures ?

Comme précisé précédemment, les analyses épigénétiques doivent être menées sur les deux premières générations afin de comparer avec les données des trois générations. S'agissant des données de transcriptomiques, certains transcrits semblent être influencés à la première génération et ce changement perdurerait jusqu'à la troisième.

Au vu des observations faites lors de la deuxième énigme, une troisième en a découlé : (3) « Différenciation sexuelle en l'absence de gène W-spécifique, dans un système ZW ». Cette énigme ne permet pas de répondre à la question de l'influence de l'environnement sur le parasite, mais nous avons décidé d'investir cette question car c'est un phénomène adaptatif primordial pour la survie de l'espèce *S. mansoni*. En effet, d'une part ce parasite a sélectionné le mode de reproduction sexuée quand il s'est adapté aux animaux à sang chaud, ce qui

pourrait lui conférer des avantages en terme de diversité génétique, suite à l'élimination de mutations délétères.

D'autre part, le dimorphisme mâle/femelle permet d'avoir un mâle musculeux, pour faciliter la migration à contre courant du flux sanguin, et une femelle plus fine pouvant se déplacer dans les veinules (Read & Nee 1990), qui ne sont pas accessibles aux mâles, pour émettre les œufs là où ils ont une chance de pénétrer à travers le tissu intestinal sans être lavés par le système circulatoire.

Enfin, l'analyse approfondie du génome et épigénome (dans l'énigme 2) de cette espèce nous ont apporté certains éléments de réponse.

Nous avons mis en évidence, la présence d'un grand nombre de séquences répétitives, l'identification de régions spécifiques au chromosome Z ou au chromosome W et l'absence de gène W-spécifique. Nous avons cependant observé que la région spécifique du chromosome W est composée de plusieurs familles de séquences répétées, dont certaines sont transcrites uniquement aux stades larvaires, où aucun dimorphisme sexuel n'apparaît. Ces séquences répétées présentent une euchromatisation au stade miracidia mais ce caractère tend à diminuer tout au long du cycle.

De par cette étude, nous avons pu proposer un modèle hypothétique de différenciation sexuelle entre les mâles et les femelles, basée sur la transcription de séquences répétées femelles spécifiques induisant une hétérochromatisation de la région W-spécifique.

Il serait intéressant maintenant d'approfondir le rôle de ces séquences répétées et de comprendre comment l'hétérochromatisation de la région W-spécifique peut agir sur les régions situées à proximité.

Des études récentes au laboratoire ont mis en évidence la présence de 4 gènes potentiellement impliqués dans la différenciation et/ ou le déterminisme sexuel, les gènes de la famille des DMRTs. Dans le système ZW, les gènes DMRTs peuvent jouer en rôle dans le déterminisme sexuel. En effet, chez le xénope, deux gènes DM ont été identifiés : le DMRT1 sur le chromosome Z et son paralogue, le DM-W, sur le chromosome W. Pendant les stades précoces de cette espèce, le DM-W agirait comme compétiteur sur le même site de liaison à l'ADN que le DMRT1 et, chez les individus ZW, cette inhibition empêcherait le DMRT1 d'engager une différenciation mâle (Yoshimoto et *al.* 2008 ; Yoshimoto et *al.* 2010). Chez *Gallus gallus*, le poulet, il a été récemment montré par Smith et *al.* que la diminution de

quantité du transcrit DMRT1 chez un embryon ZZ (par RNA interférence) engage la voie d'un développement femelle (Smith *et al.* 2009). Ainsi, le gène DMRT1, présent uniquement sur le chromosome Z, est le déterminant mâle de cette espèce par un phénomène «dose dépendant».

Chez *S. mansoni*, grâce à la carte de liaison (Criscione *et al.* 2009), à notre séquençage massif et après validation par PCRq, les différents gènes DMRT ont pu être localisés sur les chromosomes Z et W. Parmi ces 4 gènes, le gène DMS1 est près d'une région Z-spécifique et très certainement dans une région hétérochromatique du chromosome W. Il peut donc être influencé par l'hétérochromatisation de la région W-spécifique, ce qui mettrait en avant une action en *cis*, de la région composée de séquences répétitives. Toutes les études focalisées sur le déterminisme et la différenciation sexuelle, ainsi que l'étude des gènes DMRT, dans notre espèce modèle, sont en cours d'analyse, dans le cadre de la thèse de Marion Picard.

# **CHAPITRE 6**

## **Bibliographie**

- Abdulla M.H., Ruelas D.S., Wolff B., Snedecor J., Lim K.C., Xu F., Renslo A.R., Williams J., McKerrow J.H. & Caffrey C.R. (2009). Drug discovery for schistosomiasis: hit and lead compounds identified in a library of known drugs by medium-throughput phenotypic screening. *Plos Neglected Tropical Diseases*, 3, e478.
- Abrusan G., Grundmann N., DeMester L. & Makalowski W. (2009). TEclass, a tool for automated classification of unknown eukaryotic transposable elements. *Bioinformatics*, 25, 1329-1330.
- Adema C.M., Hertel L.A., Miller R.D. & Loker E.S. (1997). A family of fibrinogen-related proteins that precipitates parasite-derived molecules is produced by an invertebrate after infection. *Proceedings of the National Academy of Sciences*, 94, 8691.
- Allis C.D., Jenuwein T. & Reinberg D. (2007). *Epigenetics*.
- Altschul S.F., Gish W., Miller W., Myers E.W. & Lipman D.J. (1990). Basic local alignment search tool. *Journal of molecular biology*, 215, 403-410.
- Aragon A.D., Imani R.A., Blackburn V.R. & Cunningham C. (2008). Microarray based analysis of temperature and oxidative stress induced messenger RNA in *Schistosoma mansoni*. *Molecular and Biochemical Parasitology*, 162, 134-141.
- Archer S. (1985). The chemotherapy of schistosomiasis. *Annual review of pharmacology and toxicology*, 25, 485-508.
- Archer S., El-Hamouly W., Seyed-Mozaffari A., Butler R.H., Pica-Mattoccia L. & Cioli D. (1990). Mode of action of the schistosomicide hycanthon: site of DNA alkylation. *Molecular and Biochemical Parasitology*, 43, 89-95.
- Basch P.F. (1990). Why do schistosomes have separate sexes? . *Parasitology Today*, 6, 160-3.
- Bech N., Beltran S., Portela J., Rognon A., Allienne J.F., Boissier J. & Théron A. (2010). Follow-up of the genetic diversity and snail infectivity of a *Schistosoma mansoni* strain from field to laboratory. *Infection, Genetics and Evolution*, 10, 1039-1045.
- Beltran S., Galinier R., Allienne J.F. & Boissier J. (2008). Cheap, rapid and efficient DNA extraction method to perform multilocus microsatellite genotyping on all *Schistosoma mansoni* stages. *Memorias do Instituto Oswaldo Cruz*, 103, 501-503.
- Beltran S. & Boissier J. (2009). Male-biased sex ratio: why and what consequences for the genus *Schistosoma*? *Trends in Parasitology*, 26, 63-69.
- Benson G. (1999). Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Research*, 27, 573-580.
- Berger S.L. (2007). The complex language of chromatin regulation during transcription. *Nature*, 447, 407-412.
- Berriman M., Haas B.J., LoVerde P.T., Wilson R.A., Dillon G.P., Cerqueira G.C., Mashiyama S.T., Al-Lazikani B., Andrade L.F., Ashton P.D., Aslett M.A., Bartholomeu D.C., Blandin G., Caffrey C.R., Coghlan A., Coulson R., Day T.A., Delcher A., DeMarco R., Djikeng A., Eyre T., Gamble J.A., Ghedin E., Gu Y., Hertz-Fowler C., Hirai H., Hirai Y., Houston R., Ivens A., Johnston D.A., Lacerda D., Macedo C.D., McVeigh P., Ning Z.M., Oliveira G., Overington J.P., Parkhill J., Perteira M., Pierce R.J., Protasio A.V., Quail M.A., Rajandream M.A., Rogers J., Sajid M., Salzberg S.L., Stanke M., Tivey A.R., White O., Williams D.L., Wortman J., Wu W.J., Zamanian M., Zerlotini A., Fraser-Liggett C.M., Barrell B.G. & El-Sayed N.M. (2009). The genome of the blood fluke *Schistosoma mansoni*. *Nature*, 460, 352-U65.
- Bird A. (2007). Perceptions of epigenetics. *Nature*, 447, 396-398.
- Blair D.M. (1958). Lucanthone Hydrochloride. *Bulletin Organisation mondiale Santé*, 18, 989-1010.
- Blair L., Webster J.P. & Barker G.C. (2001). Isolation and characterization of polymorphic microsatellite markers in *Schistosoma mansoni* from Africa. *Molecular Ecology Notes*, 1, 93-95.

- Boissier J., Morand S. & Mone H. (1999). A review of performance and pathogenicity of male and female *Schistosoma mansoni* during the life-cycle. *Parasitology*, 119, 447-454.
- Boissier J. & Moné H. (2000). Experimental observations on the sex ratio of adult *Schistosoma mansoni*, with comments on the natural male bias. *Parasitology*, 121.
- Boissier J. & Moné H. (2001). Relationship between worm burden and male proportion in *Schistosoma mansoni* experimentally infected rodents and primates. A meta-analytical approach. *International Journal for Parasitology*, 31, 1597-1599.
- Boissier J., Chlichlia K., Digon Y., Ruppel A. & Mone H. (2003a). Preliminary study on sex-related inflammatory reactions in mice infected with *Schistosoma mansoni*. *Parasitology Research*, 91, 144-150.
- Boissier J., Rivera E.R. & Moné H. (2003b). Altered behavior of the snail *Biomphalaria glabrata* as a result of infection with *Schistosoma mansoni*. *The Journal of Parasitology*, 89, 429-433.
- Boissier J., Coslédan F., Robert A. & Meunier B. (2009a). Evaluation of the in vitro activity of trioxaquines against *Schistosoma mansoni*. *Antimicrob Agents Chemother*, 53, 4903-4906.
- Boissier J., Coslédan F., Robert A. & Meunier B. (2009b). In vitro activities of trioxaquines against *Schistosoma mansoni*. *Antimicrobial agents and chemotherapy*, 53, 4903-4906.
- Bossdorf O., Richards C.L. & Pigliucci M. (2008). Epigenetics for ecologists. *Ecology Letters*, 11, 106-115.
- Caby S. & Pierce R.J. (2009). Quantitative chromatin immunoprecipitation (Q-ChIP) applied to *Schistosoma mansoni*. *Molecular and Biochemical Parasitology*, 166, 77-80.
- Campbell N. & Reece J. (2007). *Biologie*.
- Champagne F.A. (2008). Epigenetic mechanisms and the transgenerational effects of maternal care. *Frontiers in neuroendocrinology*, 29, 386-397.
- Chitsulo L., Loverde P. & Engels D. (2004). Schistosomiasis. *Nature*, 2, 12-13.
- Cioli D. & Pica-Mattoccia L. (1984). Genetic analysis of hycanthon resistance in *Schistosoma mansoni*. *The American Journal of Tropical Medicine and Hygiene*, 33, 80-88.
- Cioli D., Pica-Mattoccia L., Rosenberg S. & Archer S. (1985). Evidence for the mode of antischistosomal action of hycanthon. *Life sciences*, 37, 161-167.
- Cioli D., Pica-Mattoccia L. & Moroni R. (1992). *Schistosoma mansoni*: Hycanthon/Oxamniquine resistance is controlled by a single autosomal recessive gene. *Experimental Parasitology* 75, 425-432.
- Cioli D., Pica-Mattoccia L. & Archer S. (1993). Drug resistance in schistosomes. *Parasitology Today*, 9, 162-166.
- Cioli D. (1998). Chemotherapy of schistosomiasis: an update. *Parasitology Today* 14, 418-422.
- Cock P.J.A., Fields C.J., Goto N., Heuer M.L. & Rice P.M. (2010). The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. *Nucleic Acids Research*, 38, 1767-1771.
- Cosseau C., Azzi A., Smith K., Freitag M., Mitta G. & Grunau C. (2009). Native chromatin immunoprecipitation (N-ChIP) and ChIP-Seq of *Schistosoma mansoni*: critical experimental parameters. *Molecular and Biochemical Parasitology*, 166, 70-76.
- Cosseau C., Azzi A., Rognon A., Boissier J., Gourbière S., Roger E., Mitta G. & Grunau C. (2010). Epigenetic and phenotypic variability in populations of *Schistosoma mansoni* – a possible kick-off for adaptive host/parasite evolution. *Oikos*, 119, 669-678.
- Criscione C.D., Valentim C.L.L., Hirai H., LoVerde P.T. & Anderson T.J.C. (2009). Genomic linkage map of the human blood fluke *Schistosoma mansoni*. *Genome Biology*, 10.

- Croce D., Porazzi E., Foglia E., Restelli U., Sinuon M., Socheat D. & Montresor A. (2010). Cost-effectiveness of a successful schistosomiasis control programme in Cambodia (1995–2006). *Acta Tropica*, 113, 279-284.
- Cuénot L. (1925). *L'adaptation*.
- Curtis J., Sorensen R.E., Page L.K. & Minchella D.J. (2001). Microsatellite loci in the human blood fluke *Schistosoma mansoni* and their utility for other schistosome species. *Molecular Ecology Notes*, 1, 143-145.
- Darwin C. (1859). *De l'origine des espèces*.
- Dean A.C.R. & Hinshelwood C. (1963). Integration of cell reactions. *Nature*, 199.
- Delbrück M. (1949). In the discussion of T.M. Sonneborn and G.H. Beale: Influence des gènes, des plasmagènes et du milieu dans le déterminisme des caractères antigéniques chez *Paramecium Aurelia* (variété 4). In: *Colloques internationaux du Centre National de la Recherche Scientifique* (ed. génétique Ubddc), pp. 33-34.
- DeMarco R., Kowaltowski A.T., Machado A.A., Soares M.B., Gargioni C., Kawano T., Rodrigues V., Madeira A.M.B.N., Wilson R.A., Menck C.F.M., Setubal J.C., Dias-Neto E., Leite L.C.C. & Verjovski-Almeida S. (2004). Saci-1, -2, and -3 and Perere, four novel retrotransposons with high transcriptional activities from the human parasite *Schistosoma mansoni*. *Journal of Virology*, 78, 2967-2978.
- DeMarco R., Mathieson W., Manuel S.J., Dillon G.P., Curwen R.S., Ashton P.D., Ivens A.C., Berriman M., Verjovski-Almeida S. & Wilson R.A. (2010). Protein variation in blood-dwelling schistosome worms generated by differential splicing of micro-exon gene transcripts. *Genome Research*, 20, 1112-1121.
- Desprès L. & Maurice S. (1995). The evolution of dimorphism and separate sexes in schistosomes. *Proceedings of the Royal Society*, 262, 175-180.
- Dixon S.E., Stilger K.L., Elias E.V., Naguleswaran A. & Sullivan Jr W.J. (2010). A decade of epigenetic research in *Toxoplasma gondii*. *Molecular and Biochemical Parasitology*, 173, 1-9.
- Dong Y., Chollet J., Vargas M., Mansour N.R., Bickle Q., Alnouti Y., Huang J., Keiser J. & Vennerstrom J.L. (2010). Praziquantel analogs with activity against juvenile *Schistosoma mansoni*. *Bioorganic & Medicinal Chemistry Letters*, 20, 2481-2484.
- Dubois F., Caby S., Oger F., Cosseau C., Capron M., Grunau C., Dissous C. & Pierce R.J. (2009). Histone deacetylase inhibitors induce apoptosis, histone hyperacetylation and up-regulation of gene transcription in *Schistosoma mansoni*. *Molecular and Biochemical Parasitology*, 168, 7-15.
- Durand P., Sire C. & Théron A. (2000). Isolation of microsatellite markers in the digenetic trematode *Schistosoma mansoni* from Guadeloupe island. *Molecular Ecology*, 9, 997-998.
- Duvaux-Miret O., Stefano G.B., Smith E.M., Dissous C. & Capron A. (1992). Immunosuppression in the definitive and intermediate hosts of the human parasite *Schistosoma mansoni* by release of immunoreactive neuropeptides. *Proceedings of the National Academy of Sciences*, 89, 778.
- Eissa M.M., El-Azzouni M.Z., Amer E.I. & Baddour N.M. (2011). Miltefosine, a promising novel agent for schistosomiasis mansoni. *International Journal for Parasitology*, 41, 235-242.
- Fallon P.G. & Doenhoff M.J. (1994). Drug-resistant schistosomiasis: resistance to praziquantel and oxamniquine induced in *Schistosoma mansoni* in mice is drug specific. *The American Journal of Tropical Medicine and Hygiene*, 51, 83.
- Felsenstein J. (1974). The evolutionary advantage of recombination. *Genetics*, 78, 737-756.
- Feng X., Grossman R. & Stein L. (2011). PeakRanger: A cloud-enabled peak caller for ChIP-seq data. *BMC Bioinformatics*, 12, 139.

- Fenwick A., Savioli L., Engels D., Robert Bergquist N. & Todd M.H. (2003). Drugs for the control of parasitic diseases: current status and development in schistosomiasis. *Trends in Parasitology*, 19, 509-515.
- Fitzpatrick J.M., Johansen M.V., Johnston D.A., Dunne D.W. & Hoffmann K.F. (2004). Gender-associated gene expression in two related strains of *Schistosoma japonicum*. *Molecular and Biochemical Parasitology*, 136, 191-209.
- Fitzpatrick J.M., Johnston D.A., Williams G.W., Williams D.J., Freeman T.C., Dunne D.W. & Hoffmann K.F. (2005). An oligonucleotide microarray for transcriptome analysis of *Schistosoma mansoni* and its application/use to investigate gender-associated gene expression. *Molecular & Biochemical Parasitology*, 141, 1-13.
- Fitzpatrick J.M., Peak E., Perally S., Chalmers I.W., Barrett J., Yoshino T.P., Ivens A.C. & Hoffmann K.F. (2009). Anti-schistosomal intervention targets identified by lifecycle transcriptomic analyses. *Plos Neglected Tropical Diseases*, 3.
- Fraga M.F., Ballestar E., Paz M.F., Ropero S., Setien F., Ballestar M.L., Heine-Suner D., Cigudosa J.C., Urioste M. & Benitez J. (2005). Epigenetic differences arise during the lifetime of monozygotic twins. *Proceedings of the National Academy of Sciences of the United States of America*, 102, 10604.
- Franklin T.B. & Mansuy I.M. (2010). Epigenetic inheritance in mammals: Evidence for the impact of adverse environmental effects. *neurobiology of Disease*, 39, 61-65.
- Fukagawa T., Nogami M., Yoshikawa M., Ikeno M., Okazaki T., Takami Y., Nakayama T. & Oshimura M. (2004). Dicer is essential for formation of the heterochromatin structure in vertebrate cells. *Nature Cell Biology*, 6, 784-791.
- Fusco G. & Minelli A. (2010). Phenotypic plasticity in the development and evolution: facts and concepts. *Philosophical Transactions of The Royal Society B*, 365, 547-556.
- Gottsching D.E., Felsenfeld G., Allis C.D., Jenuwein T., Reinberg D., Grunstein M., Gasser S.M., Elgin S.C.R., Reuter G., R.C. A., Selker E.U., Meyer E., Chalker D.L., Mrtienssen R., Moazed D., Matzke M., Scheid O.M., Kouzarides T., Berger S.L., Grossniklaus U., Paro R., Henikoff S., Smith M.M., Karpen G.H., Hawley R.S., Strome S., Kelly W.G., Lucchesi J.C., Kuroda M.I., Brockdorff N., Turner B.M., Li E., Bird A., Barlow D.P., Bartolomei M.S., Surani M.A., Reik W., Busslinger M., Tarakhovsky A., Jaenisch R., Gurdon J., Zoghbi H.Y., Beaudet A.L., Baylin S.B. & Jones P.A. (2005). *Epigenetics*. Cold Spring Harbor Laboratory Press edn. Inglis John.
- Grant-Dawton R.T. & Dickinson H.G. (2005). Epigenetics and its implications for plant biology. *Annals of Botany*, 96, 1143-1164.
- Grevelding C.G. (1999). Genomic instability in *Schistosoma mansoni*. *Molecular and Biochemical Parasitology*, 101, 207-216.
- Griffiths R. & Holland P.W. (1990a). A novel avian W chromosome DNA repeat sequence in the lesser black-backed gull (*Larus fuscus*). *Chromosoma*, 99, 243-250.
- Griffiths R. & Holland P.W.H. (1990b). A novel avian W chromosome DNA repeat sequence in the lesser black-backed gull (*Larus fuscus*). *Chromosoma*, 99, 243-250.
- Grossman A.I., McKenzie R. & Cain G.D. (1980). Sex heterochromatin in *Schistosoma mansoni*. *The Journal of Parasitology*, 66, 368-370.
- Grossman A.I., Short R.B. & Cain G.D. (1981). Karyotype evolution and sex chromosome differentiation in schistosomes (Trematoda, Schistosomatidae). *Chromosoma*, 84, 413-430.
- Grunau C. & Boissier J. (2010). No evidence for lateral gene transfer between salmonids and schistosomes. *Nature Genetics*, 42, 918-919.
- Gryseels B., Mbaye A., De Vlas S.J., Stelma F.F., Guisse F., Van Lieshout L., Faye D., Diop M., Ly A. & Tchuem-Tchuente L.A. (2001). Are poor responses to praziquantel for

- the treatment of *Schistosoma mansoni* infections in Senegal due to resistance? An overview of the evidence. *Tropical Medicine & International Health*, 6, 864-873.
- Gryseels B., Polman K., Clerinx J. & Kestens L. (2006). Human schistosomiasis. *Lancet*, 368, 1106-1118.
- Hardy I. (2002). *Sex ratios: concepts and research methods*. Cambridge Univ Pr.
- Hertel L.A., Adema C.M. & Loker E.S. (2005). Differential expression of FREP genes in two strains of *Biomphalaria glabrata* following exposure to the digenetic trematodes *Schistosoma mansoni* and *Echinostoma paraensei*. *Developmental & Comparative Immunology*, 29, 295-303.
- Hinshelwood C.N. (1946). *The chemical kinetics of bacterial cell*.
- Hinshelwood C. (1953). Adaptation in micro-organisms and its relation to evolution. *Symposia of the Society for Experimental Biology*, 7, 31-42.
- Hirai H. & LoVerde P.T. (1995). FISH techniques for constructing physical maps on schistosome chromosomes. *Parasitology Today*, 11, 310-314.
- Hirai H., Hirai Y. & LoVerde P.T. (2012). Evolution of sex chromosomes ZW of *Schistosoma mansoni* inferred from chromosome paint and BAC mapping analyses. *Parasitology International*.
- Hower V., Evans S.N. & Pachter L. (2011). Shape-based peak identification for CHIP-Seq. *BMC Bioinformatics*, 12, 15.
- Ismail M., Botros S., Metwally A., William S., Farghally A., Tao L.F., Day T.A. & Bennett J.L. (1999). Resistance to praziquantel: direct evidence from *Schistosoma mansoni* isolated from Egyptian villagers. *The American Journal of Tropical Medicine and Hygiene*, 60, 932-935.
- Jablonka E., Oborny B., Molnar I., Kisdi E., Hofbauer J. & Czaran T. (1995). The adaptive advantage of phenotypic memory in changing environments. *Philosophical Transactions: Biological Sciences*, 133-141.
- Jablonka E. & Lamb M.J. (1999). *Epigenetic inheritance and evolution*.
- Jansma W.B., Hulbert P.B. & Bueding E. (1974). Induction of hycanthon resistance in *Schistosoma mansoni*. *Federation Proceedings* 22, 374-404.
- Jansma W.B., Rogers S.H., Liu C.L. & Bueding E. (1977). Experimentally produced resistance of *Schistosoma mansoni* to hycanthon. *The American Journal of Tropical Medicine and Hygiene*, 26, 926.
- Jenuwein T. & Allis C.D. (2001). Translating the histone code. *Science's STKE*, 293, 1074.
- Jordan P. & Randall K. (1962). Comparison of Lucanthon hydrochloride and TWSb in the treatment of *Schistosoma haematobium* infection in Tanganyika. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 56, 136-142.
- Kanellopoulou C., Muljo S.A., Kung A.L., Ganesan S., Drapkin R., Jenuwein T., Livingston D.M. & Rajewsky K. (2005). Dicer-deficient mouse embryonic stem cells are defective in differentiation and centromeric silencing. *Genes and Development*, 19, 489-501.
- Keiser J., Chollet J., Xiao S.H., Mei J.Y., Jiao P.Y., Utzinger J. & Tanner M. (2009). Mefloquine-aminoalcohol with promising antischistosomal properties in mice. *Plos Neglected Tropical Diseases*, 3, e350.
- Kejnovsky E., Hobza R., Cermak T., Kubat Z. & Vyskot B. (2009). The role of repetitive DNA in structure and evolution of sex chromosomes in plants. *Heredity*, 102, 533-541.
- Kohany O., Gentles A., Hankus L. & Jurka J. (2006). Annotation, submission and screening of repetitive elements in Repbase: RepbaseSubmitter and Censor. *BMC Bioinformatics*, 7, 474.

- Kornberg R.D. & Thomas J.O. (1974). Chromatin structure - Oligomers of histones. *Science*, 184, 865-868.
- Krautz-Peterson G. & Skelly P.J. (2008). *Schistosoma mansoni*: the dicer gene and its expression. *Experimental Parasitology*, 118, 122-128.
- Kurz A., Lampel S., Nickolenko J.E., Bradl J., Benner A., Zirbel R.M., Cremer T. & Lichter P. (1996). Active and inactive genes localize preferentially in the periphery of chromosome territories. *The Journal of Cell Biology*, 135, 1195-1205.
- Kvarnemo C. & Ahnesjö I. (1996). The dynamics of operational sex ratios and competition for mates. *Trends in Ecology & Evolution*, 11, 404-408.
- Ladrière O. & Roberty S. (in press). Shift to heterotrophy during coral bleaching by increasing cnidocytes density: possible key of coral survival? *Coral Reefs*.
- Lam M. (2009). Nature's tough guys turned cosmonaut. *Epoch Times UK Staff*.
- Langmead B., Trapnell C., Pop M. & Salzberg S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol*, 10, R25.
- Laurent S. (2006). Trioxanes antiparasitaires: mécanisme d'action et nouvelles molécules biologiquement actives. In: *Chimie Biologie Santé*. Université Paul Sabatier Toulouse.
- Laurent S.A.L., Boissier J., Coslédan F., Gornitzka H., Robert A. & Meunier B. (2008). Synthesis of "trioxaquantel"® derivatives as potential new antischistosomal drugs. *European journal of organic chemistry*, 2008, 895-913.
- Lederberg J. & Lederberg E.M. (1952). Replica plating and indirect selection of bacterial mutants. *Journal of Bacteriology*, 63, 399-406.
- Lee J.H., Tate C.M., You J.S. & Skalnik D.G. (2007). Identification and characterization of the human Set1B histone H3-Lys4 methyltransferase complex. *Journal of Biological Chemistry*, 282, 13419-13428.
- Lee J.S., Smith E. & Shilatifard A. (2010). The language of histone crosstalk. *Cell*, 142, 682-685.
- Lepesant J.M.J., Boissier J., Climent D., Cosseau C. & Grunau C. (in prep.). Experimental observations of life traits during the life cycle of *Schistosoma mansoni* infecting sympatric or allopatric intermediate host.
- Lepesant J.M.J., Grunau C. & Cosseau C. (2011). Towards an understanding of the epigenetics of schistosomes: a comparative epigenomic study. *Memorias do Instituto Oswaldo Cruz*, 106, 823-830.
- Lepesant J.M.J., Cosseau C., Boissier J., Freitag M., Portela J., Climent D., Perrin C., Zerlotini A. & Grunau C. (2012a). Chromatin structure changes around satellite repeats on the *Schistosoma mansoni* female sex chromosome suggest a possible mechanism for sex chromosome emergence. *Genome Biology*, 13, R14.
- Lepesant J.M.J., Roquis D., Emans R., Cosseau C., Arancibia N., Mitta G. & Grunau C. (2012b). Combination of de-novo assembly of massive sequencing reads with classical repeat prediction improves identification of repetitive sequences in *Schistosoma mansoni*. *Experimental Parasitology*, 130, 470-474.
- Li R., Yu C., Li Y., Lam T.W., Yiu S.M., Kristiansen K. & Wang J. (2009). SOAP2: an improved ultrafast tool for short read alignment. *Bioinformatics*, 25, 1966.
- Liberatos J.D. & Short R.B. (1983). Identification of sex of schistosome larval stages. *Journal of Parasitology*, 69, 1084-1089.
- Liu L., Li Y. & Tollefsbol T.O. (2008). Gene-environment interactions and epigenetic basis of human diseases. *Current issues in molecular biology*, 10, 25.
- Loker E.S. & Brant S.V. (2006). Diversification, dioecy and dimorphism in schistosomes. *Trends in Parasitology*, 22, 521-528.

- Luo R., Xue X., Wang Z., Sun J., Zou Y. & Pan W. (2009). Analysis and characterization of the genes encoding the Dicer and Argonaute proteins of *Schistosoma japonicum*. *Parasites and Vectors*, 3, 1-8.
- Luong D. (2009). <http://cnx.org/content/m26565/latest/>
- Luria S.E. & Delbrück M. (1943). Mutations of bacteria from virus sensitivity to virus resistance. *Genetics*, 28, 491.
- Martin M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. journal*, 17, 10-12.
- Maynard Smith J. (1990). Models of a Dual Inheritance System. *Journal of Theoretical Biology*, 143, 41-53.
- Mayr E. (1963). *Animal species and evolution*. Harvard University Press.
- Mayr E. (1982). *The growth of biological thought: diversity, evolution and inheritance*.
- Mc Guigan K. & Sgro C.M. (2009). Evolutionary consequences of cryptic genetic variation. *Trends in Ecology & Evolution*, 24, 305-311.
- Melcher M., Schmid M., Aagaard L., Selenko P., Laible G. & Jenuwein T. (2000). Structure-function analysis of SUV39H1 reveals a dominant role in heterochromatin organization, chromosome segregation, and mitotic progression. *Molecular and cellular biology*, 20, 3728-3741.
- Merrick C.J. & Duraisingh M.T. (2010). Epigenetics in Plasmodium: what do we really know? *Eukaryotic Cell*, 9, 1150-1158.
- Mitta G., Adema C.M., Gourbal B., Loker E.S. & Théron A. (2012). Compatibility polymorphism in snail/schistosome interactions: From field to theory to molecular mechanisms. *Developmental & Comparative Immunology*.
- Moné H. (1991). Influence of non-target molluscs on the growth of *Biomphalaria glabrata* infected with *Schistosoma mansoni*: Correlation between growth and cercarial production. *Journal of Molluscan Studies*, 57, 1-10.
- Moné H. (1997). Change in schistosome sex ratio under the influence of a biotic environmental-related factor. *Journal of Parasitology*, 83, 220-223.
- Moné Y., Gourbal B., Duval D., Du Pasquier L., Kieffer-Jaquinod S. & Mitta G. (2010). A large repertoire of parasite epitopes matched by a large repertoire of host immune receptors in an invertebrate host/parasite model. *Plos Neglected Tropical Diseases*, 4, e813.
- Morand S., Pointier J.P., Borel G. & Théron A. (1993). Pairing probability of schistosomes related to their distribution among the host population. *Ecology*, 74, 2444-2449.
- Morgan J.A.T., Dejong R.J., Snyder S.D., Mkoji G.M. & Loker E.S. (2001). *Schistosoma mansoni* and *Biomphalaria*: past history and future trends. *Parasitology*, 123, 211-228.
- Morgan J.A.T., Dejong R.J., Adeoye G.O., Ansa E., Barbosa C.S., Bremond P., Cesari I.M., Charbonnel N., Correa L.R. & Coulibaly G. (2005). Origin and diversification of the human parasite *Schistosoma mansoni*. *Molecular Ecology*, 14, 3889-3902.
- Mouahid A., Bouhaddioui N., Jana M., Combes C. & Moné H. (1992). Effect of different mollusc associations on target mollusc growth and parasite cercarial production in the triple system: *Bulinus wrighti*-*Schistosoma bovis* and *Melanopsis praemorsa*. *Journal of Molluscan Studies*, 58, 349-355.
- Nakayama J., Rice J.C., Strahl B.D., Allis C.D. & Grewal S.I.S. (2001). Role of histone H3 lysine 9 methylation in epigenetic control of heterochromatin assembly. *Science's STKE*, 292, 110.
- Oger F., Dubois F., Caby S., Nofól C., Cornette J., Bertin B., Capron M. & Pierce R.J. (2008). The class I histone deacetylases of the platyhelminth parasite *Schistosoma mansoni*. *Biochemical and biophysical research communications*, 377, 1079-1084.

- Ohno S., Wolf U. & Atkin N.B. (1968). Evolution from fish to mammals by gene duplication. *Hereditas*, 59, 169-187.
- Pal C. (1998). Plasticity, memory and the adaptive landscape of the genotype. *Proceedings of the Royal Society B-Biological Sciences*, 265, 1319-1323.
- Pal C. & Miklos I. (1999). Epigenetic inheritance, genetic assimilation and speciation. *Journal of Theoretical Biology*, 200, 19-37.
- Patz J.A., Graczyk T.K., Geller N. & Vittor A.Y. (2000). Effects of environmental change on emerging parasitic diseases. *International Journal for Parasitology*, 30, 1395-1405.
- Pekowska A., Benoukraf T., Zacarias- Cabeza J., Belhocine M., Koch F., Holota H., Imbert J., Andrau J.-C., Ferrier P. & Spicuglia S. (2011). H3K4 tri-methylation provides an epigenetic signature of active enhancers. *EMBO Journal*, 30, 4198-4210.
- Pembrey M.E. (2002). Time to take epigenetic inheritance seriously. *European Journal of Human Genetics*, 10, 669-671.
- Peters A.H.F.M., Kubicek S., Mechtler K., O'Sullivan R.J., Derijck A.A.H.A., Perez-Burgos L., Kohlmaier A., Opravil S., Tachibana M. & Shinkai Y. (2003). Partitioning and plasticity of repressive histone methylation states in mammalian chromatin. *Molecular Cell*, 12, 1577-1589.
- Perrin C., Lepesant J.M.J., Roger E., Duval D., Thuillier V., Alliene J.F., Grunau C., Mitta G. & Cosseau C. (under revision). Promoter regions of duplicated *Schistosoma mansoni* gene crucial for infections success undergo rapid epigenetic modifications in varying environments while remaining genetically stable.
- Pica-Mattoccia L., Cioli D. & Archer S. (1988). Binding of tritiated hycanthon and hycanthon N-methylcarbamate to macromolecules of drug-sensitive and drug-resistant schistosomes. *Molecular and Biochemical Parasitology*, 31, 87-96.
- Portela J., Grunau C., Cosseau C., Beltran S., Dantec C., Parrinello H. & Boissier J. (2010). Whole-genome in-silico subtractive hybridization (WISH) - using massive sequencing for the identification of unique and repetitive sex-specific sequences: the example of *Schistosoma mansoni*. *BMC Genomics* 11.
- Portela J., Boissier J., Gourbal B., Pradines V., Collière V., Coslédan F., Meunier B. & Robert A. (2012). Antischistosomal activity of trioxaquines: in vivo efficacy and mechanism of action on *Schistosoma mansoni*. *Plos Neglected Tropical Diseases*, 6, e1474.
- Pradines V., Portela J., Boissier J., Coslédan F., Meunier B. & Robert A. (2011). Trioxaquine PA1259 alkylates heme in the blood-feeding parasite *Schistosoma mansoni*. *Antimicrobial agents and chemotherapy*, 55, 2403-2405.
- Protasio A.V., Tsai I.J., Babbage A., Nichol S., Hunt M., Aslett M.A., De Silva N., Velarde G.S., Anderson T.J.C. & Clark R.C. (2012). A systematically improved high quality genome and transcriptome of the human blood fluke *Schistosoma mansoni*. *Plos Neglected Tropical Diseases*, 6, e1455.
- Rai G., Sayed A.A., Lea W.A., Luecke H.F., Chakrapani H., Prast-Nielsen S., Jadhav A., Leister W., Shen M. & Inglese J. (2009). Structure mechanism insights and the role of nitric oxide donation guide the development of oxadiazole-2-oxides as therapeutic agents against schistosomiasis. *Journal of medicinal chemistry*, 52, 6474-6483.
- Rea S., Eisenhaber F., O'Carroll D., Strahl B.D., Sun Z.W., Schmid M., Opravil S., Mechtler K., Ponting C.P. & Allis C.D. (2000). Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature*, 406, 593-599.
- Read A.F. & Nee S. (1990). Male schistosomes: More than just muscle? . *Parasitology Today*, 6.
- Reinhart B.J. & Bartel D.P. (2002). Small RNAs correspond to centromere heterochromatic repeats. *Science*, 297, 1831.

- Rice J.C. & Allis C.D. (2001). Histone methylation versus histone acetylation: new insights into epigenetic regulation. *Current Opinion in Cell Biology*, 13, 263-273.
- Rice J.C., Briggs S.D., Ueberheide B., Barber C.M., Shabanowitz J., Hunt D.F., Shinkai Y. & Allis C.D. (2003). Histone methyltransferases direct different degrees of methylation to define distinct chromatin domains. *Molecular Cell*, 12, 1591-1598.
- Richards E.J. (2006). Inherited epigenetic variation-revisiting soft inheritance. *Nature Reviews Genetics*, 7, 395-401.
- Rodin S.N., Parkhomchuk D.V. & Riggs A.D. (2005). Epigenetic changes and repositioning determine the evolutionary fate of duplicated genes. *Biochemistry (Moscow)*, 70, 559-567.
- Rodrigues N.B., Coura Filho P., de Souza C.P., Jannoti Passos L.K., Dias-Neto E. & Romanha A.J. (2002). Populational structure of *Schistosoma mansoni* assessed by DNA microsatellites. *International Journal for Parasitology*, 32, 843-851.
- Rodrigues N.B., Silva M.R., Pucci M.M., Minchella D.J., Sorensen R., Loverde P.T., Romanha A.J. & Oliveira G. (2007). Microsatellite-enriched genomic libraries as a source of polymorphic loci for *Schistosoma mansoni*. *Molecular Ecology Notes*, 7, 263-265.
- Roger E., Gourbal B., Grunau C., Pierce R.J., Galinier R. & Mitta G. (2008a). Expression analysis of highly polymorphic mucin proteins (*Sm* PoMuc) from the parasite *Schistosoma mansoni*. *Molecular and Biochemical Parasitology*, 157, 217-227.
- Roger E., Grunau C., Pierce R.J., Hirai H., Gourbal B., Galinier R., Emans R., Cesari I.M., Cosseau C. & Mitta G. (2008b). Controlled chaos of polymorphic mucins in a Metazoan parasite (*Schistosoma mansoni*) interacting with Its invertebrate host (*Biomphalaria glabrata*). *Plos Neglected Tropical Diseases*, 2.
- Roger E., Mitta G., Moné Y., Bouchut A., Rognon A., Grunau C., Boissier J., Théron A. & Gourbal B. (2008c). Molecular determinants of compatibility polymorphism in the *Biomphalaria glabrata/Schistosoma mansoni* model: New candidates identified by a global comparative proteomics approach. *Molecular and Biochemical Parasitology*, 157, 205-216.
- Rogers S.H. & Bueding E. (1971). Hycanthone resistance: development in *Schistosoma mansoni*. *Science*, 172, 1057.
- Rosi D., Peruzzotti G., Dennis E.W., Berberian D.A., Freele H., Tullar B.F. & Archer S. (1967). Hycanthone, a New Active Metabolite of Lucanthone. *Journal of medicinal chemistry*, 10, 867-876.
- Sayed A.A., Simeonov A., Thomas C.J., Inglese J., Austin C.P. & Williams D.L. (2008). Identification of oxadiazoles as new drug leads for the control of schistosomiasis. *Nature Medicine*, 14, 407-412.
- Shapiro J.A. & Von Sternberg R. (2005). Why Repetitive DNA is essential to genome function? *Biological reviews of the Cambridge Philosophical Society*, 80, 227-250.
- Shelford V.E. (1913). Law of toleration. *Animal communities in temperate America (Eds.)*, 302-303.
- Smit A.F.A., Hubley R. & Green P. (1996). *RepeatMasker Open-3.0*.
- Smith C.A., Roeszler K.N., Ohnesorg T., Cummins D.M., Farlie P.G., Doran T.J. & Sinclair A.H. (2009). The avian Z-linked gene DMRT1 is required for male sex determination in the chicken. *Nature*, 461, 267-271.
- Spyrou C., Stark R., Lynch A. & Tavaré S. (2009). BayesPeak: Bayesian analysis of ChIP-seq data. *BMC Bioinformatics*, 10, 299.
- Standen O.D. (1953). The relationship of the cercariae of *Schistosoma mansoni* into the skin and lymphatics of the mouse. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 47, 292-298.

- Stein L.D., Mungall C., Shu S.Q., Caudy M., Mangone M., Day A., Nickerson E., Stajich J.E., Harris T.W. & Arva A. (2002). The generic genome browser: a building block for a model organism system database. *Genome Research*, 12, 1599-1610.
- Strahl B.D. & Allis C.D. (2000). The language of covalent histone modifications. *Nature*, 403, 41.
- Théron A. (1980). Mise en évidence de races chronobiologiques de *Schistosoma mansoni*, agent de la bilharziose, à partir des cinétiques cercariennes. *C.R. Academie des Sciences de Paris*.
- Théron A., Pages J.R. & Rognon A. (1997). *Schistosoma mansoni*: Distribution patterns of miracidia among *Biomphalaria glabrata* snail as related to host susceptibility and sporocyst regulatory processes. *Experimental Parasitology*, 85, 1-9.
- Théron A. & Coustau C. (2005). Are *Biomphalaria* snails resistant to *Schistosoma mansoni*? *Journal of Helminthology*, 79, 187-191.
- Thomas F., Lefèvre T. & Raymond M. (2010). *Biologie évolutive*.
- Trapnell C., Roberts A., Goff L., Pertea G., Kim D., Kelley D.R., Pimentel H., Salzberg S.L., Rinn J.L. & Pachter L. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature Protocols*, 7, 562-578.
- Ugarkovic D. (2005). Functional elements residing within satellite DNAs. *EMBO Reports*, 6, 1035-1039.
- Utzinger J. & Keiser J. (2004). Schistosomiasis and soil-transmitted helminthiasis: common drugs for treatment and control. *Expert opinion on pharmacotherapy*, 5, 263-285.
- Utzinger J., Tanner M. & Keiser J. (2010). ACTs for schistosomiasis: do they act? *The Lancet Infectious Diseases*, 10, 579-581.
- Van Valen L. (1973). A new evolutionary law. *Evolutionary theory*, 1, 1-30.
- Verdel A., Songtao J., Scott G., Tomoyasu S., Steven G., Grewal S.I.S. & Moazed D. (2004). RNAi-mediated targeting of heterochromatin by the RITS complex. *Science*, 303, 672-676.
- Verstrepen K.J. & Fink G.R. (2009). Genetic and epigenetic mechanisms underlying cell-surface variability in protozoa and fungi. *Annual review of genetics*, 43, 1-24.
- Volpe T.A., Kidner C., Hall I.M., Teng G., Grewal S.I.S. & Martienssen R.A. (2002). Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. *Science*, 297, 1833.
- Wang Q.P., Kawahara T. & Horn D. (2010). Histone deacetylases play distinct roles in telomeric VSG expression site silencing in African trypanosomes. *Molecular Microbiology*, 77, 1237-1245.
- Weaver I.C.G., Cervoni N., Champagne F.A., D'Alessio A.C., Sharma S., Seckl J.R., Dymov S., Szyf M. & Meaney M.J. (2004). Epigenetic programming by maternal behavior. *Nature Neuroscience*, 7, 847-854.
- WHO (2011). Schistosomiasis, countries or areas at risk, 2011. In: *World Health Organization* (ed. WHO).
- Wright S. (1945). Genes as physiological agents. General considerations. *American Naturalist*, 74, 289-303.
- Wuhrer M., Koeleman C.A.M., Fitzpatrick J.M., Hoffmann K.F., Deelder A.M. & Hokke C.H. (2006). Gender-specific expression of complex-type N-glycans in schistosomes. *Glycobiology*, 16, 991-1006.
- Xiao S.H., Keiser J., Chollet J., Utzinger J., Dong Y., Endriss Y., Vennerstrom J.L. & Tanner M. (2007). In vitro and in vivo activities of synthetic trioxolanes against major human schistosome species. *Antimicrobial agents and chemotherapy*, 51, 1440-1445.

- Yoshida M., Horinouchi S. & Beppu T. (1995). Trichostatin A and trapoxin: novel chemical probes for the role of histone acetylation in chromatin structure and function. *BioEssays*, 17, 423-430.
- Yoshimoto S., Okada E., Umemoto H., Tamura K., Uno Y., Nishida-Umehara C., Matsuda Y., Takamatsu N., Shiba T. & Ito M. (2008). A W-linked DM-domain gene, DM-W, participates in primary ovary development in *Xenopus laevis*. *Proceedings of the National Academy of Sciences*, 105, 2469.
- Yoshimoto S., Ikeda N., Izutsu Y., Shiba T., Takamatsu N. & Ito M. (2010). Opposite roles of DMRT1 and its W-linked paralogue, DM-W, in sexual dimorphism of *Xenopus laevis*: implications of a ZZ/ZW-type sex-determining system. *Development*, 137, 2519-2526.
- Zavodna M., Sandland G.J. & Minchella D.J. (2008). Effects of intermediate host genetic background on parasite transmission dynamics: a case using *Schistosoma mansoni*. *Experimental Parasitology*, 120, 57-61.
- Zerbino D.R. & Birney E. (2008). Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Research*, 18, 821.

# **CHAPITRE 7**

## **Annexes**

## ANNEXE 1

Publication: **Lepesant J.M.J.**, Boissier J., Climent D., Cosseau C. & Grunau C. (in prep.).  
Experimental observations of life traits during the life cycle of *Schistosoma mansoni*  
infecting sympatric or allopatric intermediate host.

Research article

Title :

**Effects of sympatric vs allopatric intermediate host on *Schistosoma mansoni* life history traits : highlights on sex-ratio**

Julie M.J. Lepsant<sup>1,2</sup>, Jérôme Boissier<sup>1,2</sup>, Déborah Climent<sup>1,2</sup>, Céline Cosseau<sup>1,2</sup>, Christoph Grunau<sup>1,2</sup>

1- Université de Perpignan Via Domitia, Perpignan, F-66860, France

2- CNRS, UMR 5244, Ecologie et Evolution des Interactions (2EI), Perpignan, F-66860, France

**\*Corresponding author:**

Julie M.J. Lepsant

Université de Perpignan Via Domitia, UMR 5244 CNRS Écologie et Evolution des Interactions (2EI), 52 Avenue Paul Alduy, 66860 Perpignan Cedex, France

<http://2ei.univ-perp.fr>

Email: [julie.lepsant@free.fr](mailto:julie.lepsant@free.fr)

Tel: +33 4 68 66 21 80

Fax: +33 4 68 66 22 81

**Key words:** *Schistosoma mansoni*, allopatric, sympatric, life history traits, plasticity, sex-ratio

**Abstract**

For parasites that require multiple hosts to complete their development, the interplay of the host may impact parasite transmission and establishment in different hosts. In this study we work on the parasite *Schistosoma mansoni* which need two different hosts to complete his life cycle, an intermediate host snail *Biomphalaria glabrata* and a final host human or rodent. Even if the parasite is specific of one species of mollusc, he can, with lower success, infect different strain of *Biomphalaria glabrata* which mean different geographical isolates.

We performed experimental infections with a clonal population, on neutral microsatellites marks, of *S. mansoni* utilizing two strains of *B. glabrata*, one considered as the sympatric host and one as allopatric (two geographic different isolates). We assessed prevalences on infections in the snail and the cercarial infectivity, female ratio, immunopathology in the final host and genetic analyses (microsatellites) on individuals larvae at each generations. We performed the experiment, for the first time, on three succesives generations.

Even if the parasite population is clonal on neutral marks, *S. mansoni* keeps the ability of generating plasticity and variability for different life history traits in direct correlation with the persistence of the parasite when confront to an unusual environment, in this study the intermediate host.

## Introduction

For parasites that require a number of hosts to complete their development (complex life cycle), each host species can represent a potential filter that can modulate infection and establishment of the parasite (Zavodna *et al.* 2008). As proposed in Red Queen hypothesis and according to local co-evolutionary interactions, parasites are assumed to be more adapted to the most common local host genotypes (i.e. ‘sympatric association’) (Lively & Dybdahl 2000). Host-parasite sympatric associations would thus result in higher parasite fitness than in allopatric associations. Parasite fitness is mainly measured as compatibility parameter, which reflects both the ability of a parasite to develop and the incapacity of the host to resist to the parasite infection.

Schistosomes are parasitic plathyhelminthes responsible of the intestinal schistosomiasis. Schistosomiasis is an important parasitic human disease ranking second only to malaria in terms of parasite-induced human morbidity and mortality, with more than 200 million infected people (King 2010). The schistosome life cycle is characterized by passage through two obligatory hosts: the intermediate fresh-water snail *Biomphalaria glabrata* (or other *Biomphalaria* species, depending on the geographical location) for the asexual multiplication, and Human or rodents as final host for the sexual reproduction at the adult stage. Eggs are excreted with the host feces and free-swimming larvae (miracidia) are released when eggs become into contact with the water. Eggs of the parasite also accumulate into the host liver and cause the symptoms of the disease. Miracidia infect the freshwater mollusc host and transform into primary and secondary sporocysts. Finally, a third larval stage, the cercariae, capable of infecting the Human host, is released in the water.

The impact of allopatric *vs* sympatric associations on host-parasite compatibility has been extensively studied in schistosome-snail interactions. Some authors have observed differential infectivities when the parasite is confronted to either sympatric or allopatric snails (Woolhouse 1989; Webster & Woolhouse 1999; Sandland *et al.* 2009; Jones-Nelson *et al.* 2011). There are geographic patterns of compatibility between schistosomes and snails that suggest that there is a

tendency for sympatric combinations to be more compatible than allopatric combinations (Sandland *et al.* 2009). This pattern is in agreement with the patterns reported for other parasite-host systems (Sandland *et al.* 2009). However, reverse pattern has also been reported (Morand *et al.* 1996). Variations in *B. glabrata* susceptibility and *S. mansoni* infectivity from various geographic areas were reported by Files and Cram (Files & Cram 1949), who suggested that these variations were genetically based. Susceptibility is now known to vary between geographic areas, between individuals in the same population, and in the same snail at different ages (Richards 1976). Genetic variations in snail susceptibility and parasite infectivity result in a variety of host-parasite relations (Sullivan & Richards 1981).

Even if compatibility parameter is a prerequisite for parasite development success is not alone implicated in parasite transmission dynamic. Indeed, it has been shown that the mollusc host genetic background also influences both the parasite infectivity for the vertebrate host (Zavodna *et al.* 2008) and the adult sex-ratio (Zavodna *et al.* 2008). These last authors have shown that the worm sex-ratio recovered in experimentally infected mice is male-biased when the parasite pass through a sympatric mollusc host and female biased when the parasite pass through an allopatric mollusc host (Zavodna *et al.* 2008). However, because the sex-ratio was only measured at adult stage is not possible to detect the origin of such bias. It could originate from emitted cercariae (i.e. inside the mollusc host) or from developing adult worms (i.e. inside the vertebrate host).

In this study we used a single parasite population originated from Brazil on two distinct mollusc host populations from distinct geographic areas (Brazil and Guadeloupe). We looked at life history traits (prevalence, infectivity, host immunopathology and sex-ratio) and at population genetic structure using 14 microsatellites markers. This experiment is, for the first time, maintained on 3 successive generations.

Our results show that *Schistosoma mansoni* can infect two strains of *Biomphalaria glabrata* but with consequences on parasite life history traits. We show a female biased sex-ratio when the parasite is passed through the allopatric host (1.8 times more than in sympatric association). This

bias is maintained during three successive generations. The infectivity and the immunopathology stay unchanged for the two conditions and during the three generations, but prevalence are lower in the allopatric association than in sympatric association. Genetic analyses show very few diversity on the 14 neutral markers used suggesting that despite a low genetic diversity the parasite keep the capacity to generate plasticity and variability on different life history traits.

## Material and Methods

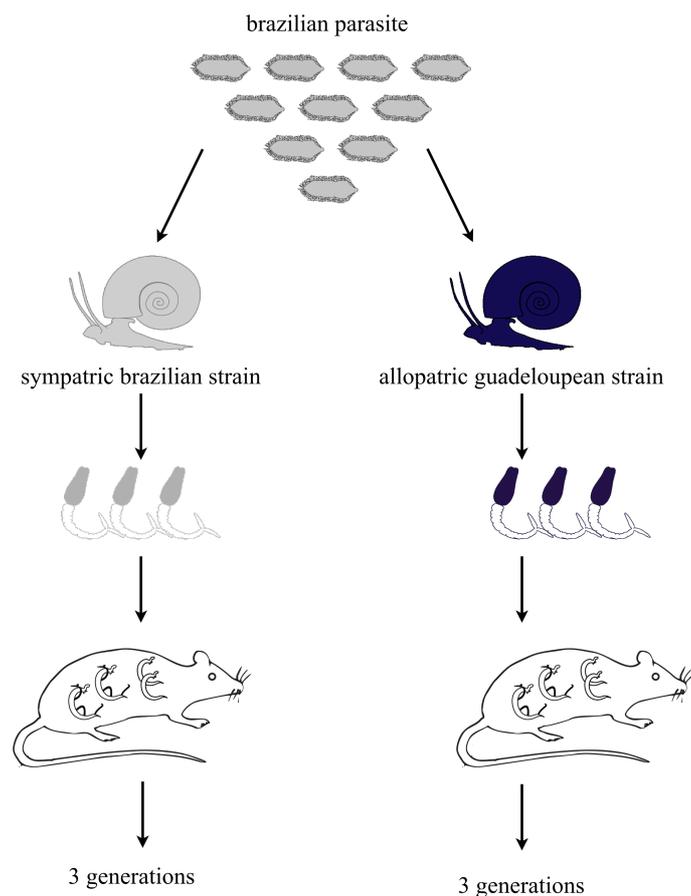
### *Parasite and host strains*

Two *Biomphalaria glabrata* snail populations were infected using the same *Schistosoma mansoni* (*SmBre*) parasite strain. One group of snails was selected from our laboratory population of a *B. glabrata* brazilian strain (*BgBRE*). The second group consisted of another laboratory population of a *B. glabrata* guadeloupan strain (*BgGUA*).

The brazilian snail is the sympatric host of the brazilian *Schistosoma mansoni* parasite (*SmBre*), whereas the *B. glabrata* guadeloupean strain (*BgGUA*) is used as an allopatric host (Figure 1).

### *Experimental procedure (figure) and parameters measured*

A brazilian population of *Schistosoma mansoni* were used to infect for three successive generations either a sympatric snail population (brazilian strain) or an allopatric snail population (guadeloupean strain). The vertebrate host (Swiss OF1 mouse) is the same for the two lineages. For each generation we have measured (i) the mollusc prevalence (the number of infected molluscs divided by the total number of exposed molluscs), (ii) the cercarial infectivity (the number of developing worms divided by the number of cercariae proposed to mouse infection), (iii) the host immunopathology assessed by the relative spleen weight (mouse spleen weight divided by body weight), (iv) the sex ratio at each parasite stage: miracidia, cercariae and adult and (v) the genetic population structure (using 14 microsatellite markers) at miracidia and cercarial stages.



*Figure 1: Experimental design. A brazilian population of Schistosoma mansoni were used to infect for three successive generations either a sympatric snail population (brazilian strain) or an allopatric snail population (guadeloupean strain). The vertebrate host (Swiss OF1 mouse) is the same for the two lineages.*

### Snails infections and maintenance

Eggs were axenically recovered from 60-day infected hamster livers and brazilian miracidia (*SmBre*) were hatched from eggs in 5 ml of spring water during 2-3 hours under light.

For infection, *B. glabrata* of each strain (*BgBRE* and *BgGUA*) of 4-5 mm diameter were individually exposed to ten *SmBre* miracidia in 5 ml of springwater. The snails were then individually isolated and maintained in round, clear plastic containers for 24 hr. On the next day, the exposed *B. glabrata* were separated in two groups: the sympatric association (*SmBre/BgBRE*) and the allopatric association (*SmBre/BgGUA*). Snails were fed with fresh lettuce ad libitum and the water was maintained at 25°C and changed weekly. Photoperiod during the entire experiment was equilibrated to 12L:12D (Boissier *et al.* 2003). The prevalence was calculated, 35 days post-

infection, by dividing the number of snails which presented sporocysts by the total number of snails.

### Mice infections

A group of 10 mice were individually exposed to *S. mansoni* cercariae emitted from, either the pooled infected snails of the sympatric association (SmBre/BgBRE) or the pooled infected snails of the allopatric association (SmBre/BgGUA). The mice were anesthetized with a mixture of Rompun (20 mg/ml; Bayer) 0.50 ml and Imalgène (100 mg/ml; Rhône Mérieux) 1.00 ml in 8.5 ml of autoclaved NaCl 8.5 (‰) by injection of 0,1 ml/10g of body weight. Mice abdomens were shaved and exposed to 160 randomly selected cercariae for 1 hour (Boissier & Moné 2000)

Mice were killed by a lethal intraperitoneal infection of sodium pentobarbital. Adults worms were recovered by retrograde perfusions of the hepatic portal system with citrate (7.5%) saline (8.5%) solution administrated through the left ventricle (Duvall & DeWitt 1967). Worms trapped in the liver or mesenteric system were collected after excising these organs. All worms were sexed and counted.

### ***Sex proportion and genotyping***

Before DNA extraction, individual miracidia or cercariae were individually vacuum-dried for 15 min in a Speedvac evaporator. Next, 20 µl of NaOH (250 mM) was added to each tube. After a 15 min incubation period at 25°C, the tubes were heated at 99°C for 2 min. Then, 10 µl HCl (250 mM), 5 µl of Tris-HCl (500 mM) and 5 µl Triton X-100 (2%) were added and a second heat shock at 99°C for 2 min was performed. The products were stored at -20°C.

### Sex proportion

The PCR amplifications were performed with female specific markers (scaff\_002739) and housekeeping gene (rhodopsin 3) in multiplex on 24 to 32 cercariae and miracidia for the sympatric and the allopatric association on the 3 generations.

Oligonucleotide primers were previously diluted in a final volume of 100 µl of water with concentrations of 0.1 µM for each female specific oligonucleotide primer and 0.2 µM for each housekeeping gene oligonucleotide primer (Table 1).

The PCR reactions were carried out in a total volume of 10 µl containing 5 µl of 2x QIAGEN (#206143) Multiplex PCR Master Mix (providing a final concentration of 3 mM MgCl<sub>2</sub>), 1 µl mix of primers previously diluted (providing a final concentration of 0.1 µM for each female specific oligonucleotide primer and 0.2 µM for each housekeeping gene oligonucleotide primer), 1 µl of diluted DNA (10<sup>-1</sup>) and DNase-free water q.s.p. 10 µl.

The PCR programme consisted in an initial denaturation phase at 95°C for 15 min, followed by 40 cycles at 94°C for 30 s, 57 °C for 90 s, 72°C for 60 s, and a final extension at 60°C for 30 min.

The PCR products were separated by electrophoresis through a 2% TBE agarose gel. When two bands appear it is a female and when there is only one band it is a male.

*Table 1: Primers used for calcul of sex proportion*

Targets	Forward sequences	Reverse sequences	Expected product sizes
Rhodopsin	GACGGCCACACTAAAG	AGTAAAATGGTCACTGCTAT	191 bp
Scaff_002739	TGTTTCGAATTTCACTTAC	CATTCACAGTTTGGCGAACA	391 bp

### *Genotyping*

All samples were genotyped on 14 microsatellite markers (Table 2). The relevant DNA fragments were amplified using Polymerase Chain Reaction (PCR). To maximise efficiency and minimize costs, these PCRs were performed in three multiplexes (see Table below) using the QIAGEN multiplex kit. These multiplexes were carried out according to the manufacturer's standard microsatellite amplification protocol in the final volume of 10 µl and with 57°C for annealing temperature. The PCR programme consisted in an initial denaturation phase at 95°C for 15 min,

followed by 40 cycles at 94°C for 30 s, 57 °C for 90 s, 72°C for 60 s, and a final extension at 60°C for 30 min.

PCR products were diluted in Sample Loading Solution (Beckman Coulter) with red-labeled size standard (CEQTM DNA size standard kit, 400, Beckman Coulter, Fullerton, CA, USA) and electrophoresed on an automatic sequencer (CEQTM 8000, Beckman Coulter). Genotypes were determined using the fragment analyzer package from Beckman Coulter. (Bech *et al.* 2010)

*Table 2: Microsatellite markers used in this study*

Microsatellites	Accession number	Forward sequences	Reverse sequences	Multiplex number
SMC1	AF325694	TGACGAGGTTGACCATAATTCTAC	AACACAGATAAGAGCGTCATGG	
SMD011	AF325698	TGTTTAAAGTCGTCGGTGCTG	CTACGCTAAA CTGGCAGGGT	
Snbr16	LO4480	TGTGACTTTGATGCCACTGA	GGCCTGATACAATTCTCCGA	1
R95529	R95529	GTGATTGGGGTGATAAAG	CATGTTTCTTCAGTGTC	
SMD57	AF202967	GTGATTGGGGTGATAAAG	GCAGTAATCCGAAAGATTAG	
SMDA28	AF325695	CATGATCTTAGCTCAGAGAGCC	AGCCAGTATAGCGTTGATCATC	
SmBr1	L81235	GAGTATACGGCTTCTTGGA	CGGAACGACAAGAAAATCAT	
SMD28	AF202966	CATCACCATCAATCACTC	TATTCACAGTAGTAGGCG	2
SCMSMOXII	M85305	TTCTACAATAATACCATCAAC	TTTTTCTCACTCATATACAC	
SMS7-1	AF330105	TCCTCCTCTATTTTCTCTTTG	ATTACGATTGCACAGCTACTTTTG	
Snbr10	DQ448293	CATGATCTTAGCTCAGAGAGC	GTACATTTTATGTCAGTTAGCC	
Snbr13	DQ137790	GTCACAGATACCTGACGAGCTG	ACTCCCAGCAATTTGTCC	3
SMD25	AF202965	GATTCCCAAGATTAATGCC	GCCATTAGATAATGTACGTG	
L46951	L46951	CAAACATATACATTGAATACAG	TGAATTGATGAATGATTGAAG	

## **Analyses**

Statistical analyses were performed using SPSS 18.0 (IBM). Fisher exact tests were used to compare prevalences. The cercarial infectivity and relative spleen weight were compared using non-parametric Kruskal-Wallis or Mann-Whitney tests. Population genetic analyses (Hardy Weinberg equilibrium, Fst, genetic diversity and allelic richness) were calculated using Fstat 2.9.

## **Ethical note**

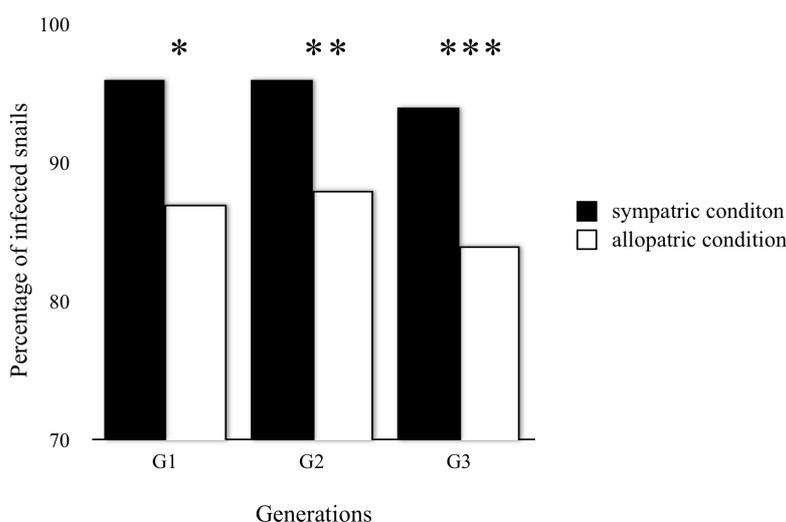
The French Ministère de l'Agriculture et de la Pêche and French Ministère de l'Education Nationale de la Recherche et de la Technologie provided permit A 66040 to our laboratory for experiments on animals and certificate for animal experimentation (authorization 007083, decree

87–848) for the experimenters. Housing, breeding and animal care followed the national ethical requirements.

## Results

### The prevalence changes and is lower when the parasite is confronted to an other strain of intermediate host (allopatric snail)

The comparison of prevalences between the sympatric (*SmBre/BgBRE*) and the allopatric association (*SmBre/BgGUA*) is presented in Figure 2. These results show that there are significative differences between the sympatric and the allopatric association with a higher prevalence in the sympatric than in the allopatric condition. These differences are found for the three generations (G1 :  $p = 0.04$ ; G2:  $p = 0.0176$ ; G3:  $p = 0.0021$ ). Within each association whatever sympatric or allopatric, the prevalence stay unchanged between the three generations ( $p > 0.05$ ).

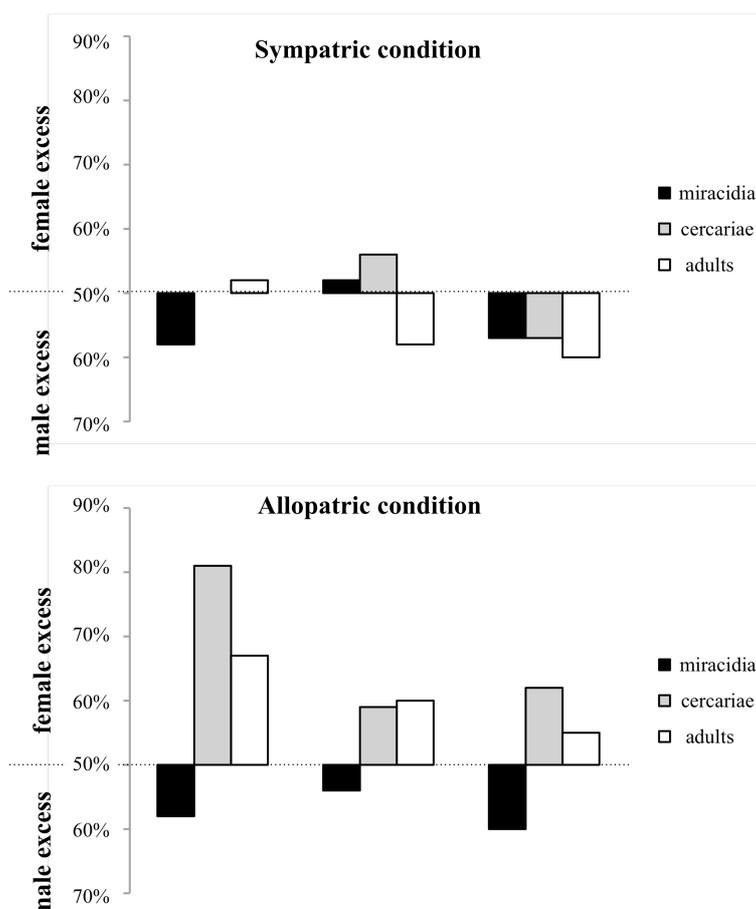


*Figure 2: Percentage of infected snails for the control group and the experimental group for 3 successive generations (G1, G2, G3). Sympatric association : brazilian Schistosoma mansoni on albinos brazilian Biomphalaria glabrata (black); Allopatric association : brazilian Schistosoma mansoni on non albinos guadeloupean Biomphalaria glabrata (white).; \*:  $p < 0.05$ , \*\*:  $p < 0.01$ , \*\*\*:  $p < 0.001$*

### Dramatic changes in sex-ratio when the parasite develops in an allopatric intermediate host

The initial miracidia sex ratio is a little bit male biased (58% male, 42% female). Molluscs whatever sympatric or allopatric were infected using this male biased miracidia population. When the miracidia infect the sympatric mollusc the emitted cercariae have a sex-ratio around the

equilibrium. The sex ratio become male biased when penetrating in the definitive host (adult sex ratio). This pattern is the same for the three generations. When the miracidia infect the allopatric mollusc the emitted cercariae have a strong female biased sex-ratio (20% male, 80% female). This bias is also found at lower level at the adult stage (35% male, 65% female). This pattern is the same for the three generations.



*Figure 3: Histogramm of the percentage of females of miracidia (black), cercariae (grey) and adults (white), for 3 generations and for the control group (first graphic) and experimental group (second graphic).*

*No differences appear in infectivity of cercariae, inflammatory response of the definitive host and genetic diversity*

Population genetic structure has been measured at miracidia and cercarial stages using 14 microsatellites markers. No statistical difference has been observed between all  $F_{st}$  values whatever the host-parasite combination (allopatric *SmBre/BgBRE* or sympatric *SmBre/BgGUA*) or the generation considered (mean value of  $F_{st}$  : 1.1%). Heterozygosity and allelic richness were very

low between 1.94% and 8.1%, and between sympatric and allopatric associations, for heterozygosity and allelic richness, respectively. For these two last parameters, no statistical difference has been detected whatever the host-parasite combination or the generation considered ( $p>0.05$ ). The cercarial infectivity and the inflammatory reaction measured as the relative spleen weight were the same whatever the host parasite system or the generation considered ( $p>0.05$ ).

## Discussion

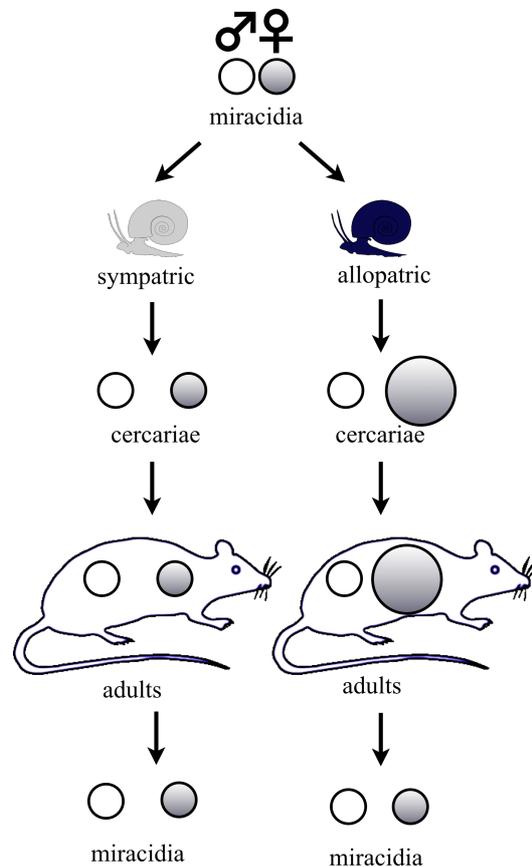
Our results show that when *S. mansoni* parasite is confronted to an allopatric mollusc host some life history parameters stay unchanged (cercarial infectivity and vertebrate host immunopathology) but other are modified. We have shown that in allopatric condition the prevalence is reduced and the sex ratio become female biased, compare to sympatric condition.

Prevalences show a significant decrease when the parasite is confronted to an allopatric intermediate host. The lower prevalence in the allopatric association is the same during 3 successive generations. *Schistosoma mansoni* is assumed to infest specific snail strain depending on their geographic area (Files & Cram 1949; Morand *et al.* 1996). This geographic influence leads to a geographic compatibility polymorphism with each parasite strain infecting a specific snail strain. In the *B. glabrata/S. mansoni* interaction a compatibility polymorphism has been shown (Richards & Shade 1987; Théron *et al.* 1997; Sandland *et al.* 2007; Roger *et al.* 2008; Sandland *et al.* 2009) and our results on prevalences match to studies on compatibility between host and parasite. Some authors show that in the snail hosts and regardless of snail line, outcrossed parasites exhibited signs of greater exploitation of host resources relative to inbred parasites. Snails infected by outcrossed parasites grew less than uninfected snails, as well as less than infected inbred snails in the novel line. Together with the fact that outcrossed parasites produced more cercariae than inbred parasites, this could be evidence of a more successful diversion of host resources by outcrossed parasites. Furthermore, increased mortality of outcrossed-infected snails is also consistent with increased exploitation and exhaustion of host energetic reserves (Woolhouse 1989; De Kock 1992; Webster & Woolhouse 1999; Sandland *et al.* 2009).

Our experiments also show a dramatic change in sex-ratio when the parasite develops in an allopatric intermediate host compare to sympatric intermediate host. Indeed, there is a 1.8 female enrichment in cercarial population and subsequently in the adult worm population in the allopatric

association, whereas the sex ratio stays male biased in the sympatric association. Schistosome sex ratio has puzzled parasitologists since several years and factors determining its variations are still unclear (see Beltran & Boissier 2009 for syntheses). *S. mansoni* sex ratio is known to be male-biased in natural vertebrate host infrapopulations (Boissier & Moné 2001). The origin of such bias has been studied using both experimental and analytical studies (see Beltran & Boissier 2009 for synthesis). It is now established that the bias originates inside the vertebrate host during the parasite development from cercariae to adult worm - the thin female being less able to resist to the blood flow than the muscly male. However, in the present study we show that the sex ratio changes before cercarial penetration and the adult sex ratio we have obtained is a consequence of sex ratio distortion inside the mollusc host. A modification in environmental conditions of the mollusc-parasite system is known to induce such sex ratio distortion. It has been shown that the presence of a non-vector mollusc (*Marisa cornuarietis*) in the vector mollusc environment (*B. glabrata*) induce a male biased sex ratio distortion. In this study, the author shows that the presence of the non-vector molluscs stimulate the growth of shistosome-vector mollusc and thus would increase the space available for the parasite (Moné 1997). In this better environment (i.e. bigger mollusc) male parasite seems to benefits more than female parasite. At the opposite female parasite seems to have selective advantageous in stressing environment. In host-parasite sympatric associations the sex ratio is male-biased and become female biased in allopatric environment (present study and Zavodna et al. 2008). Allopatric environment could be perceived as a stressing environment for the parasite. The fact that the prevalence is lower in allopatric associations compare to sympatric ones agrees with this last hypothesis. A difference in male vs female life history traits has even been proposed (Boissier et al. 1999). For instance, these last authors, using a meta-analytical approach, showed that female sporocysts produce more cercariae than male sporocysts. Our present study agrees with the fact that male and female schistosomes may be different in their life history performances and strategies to exploit their hosts (Boissier et al. 1999). In non stressing condition, male parasite is more efficient than female, the opposite is found in a stressing environment. In order to improve this hypothesis

more experiment need to be performed using other environmental conditions whatever stressed (starved mollusc for instance) or not (well nourished mollusc).



*Figure 4: Summarize of results on sex ratio when the parasite is confronted to a sympatric or an allopatric intermediate host. The female enrichment is found in the cercarial and adult population when the parasite is developed in an allopatric snail. This bias is not found after sexual reproduction.*

Finally genetic analyses show that our parasite population is genetically homogeneous

All the experimental evolution allows to show that depending on the intermediate host, parasite's life history traits could be modified, even if the population is genetically homogeneous. The parasite keeps the ability to generate plasticity in the new environment. Some traits are plastic and others not but this allow to increase the variability spectrum and so the ability to a rapid response ton changing environment.

**Literature Cited:**

- Bech N., Beltran S., Portela J., Rognon A., Allienne J.F., Boissier J. & Théron A. (2010). Follow-up of the genetic diversity and snail infectivity of a *Schistosoma mansoni* strain from field to laboratory. *Infection, Genetics and Evolution*, 10, 1039-1045.
- Beltran S. & Boissier J. (2009). Male-biased sex ratio: why and what consequences for the genus *Schistosoma*? *Trends in Parasitology*, 26, 63-69.
- Boissier J. & Moné H. (2000). Experimental observations on the sex ratio of adult *Schistosoma mansoni*, with comments on the natural male bias. *Parasitology*, 121.
- Boissier J. & Moné H. (2001). Relationship between worm burden and male proportion in *Schistosoma mansoni* experimentally infected rodents and primates. A meta-analytical approach. *International Journal for Parasitology*, 31, 1597-1599.
- Boissier J., Rivera E.R. & Moné H. (2003). Altered behavior of the snail *Biomphalaria glabrata* as a result of infection with *Schistosoma mansoni*. *Journal of Parasitology*, 89, 429-433.
- De Kock K.N. (1992). The effect of exposure to selected numbers of *Schistosoma mansoni* miracidia on survival of cohorts of *Biomphalaria pfeifferi*. *Annals of tropical medicine and parasitology*, 86, 557-562.
- Duvall R.H. & DeWitt W.B. (1967). An improved perfusion technique for recovering adults schistosomes from laboratory animals. *American Journal of Tropical Medicine and Hygiene*, 16, 483-486.
- Files V.S. & Cram E.B. (1949). A study on the comparative susceptibility of snail vectors to strains of *Schistosoma mansoni*. *The Journal of Parasitology*, 35, 555-560.
- Jones-Nelson O., Thiele E.A. & Minchella D.J. (2011). Transmission dynamics of two strains of *Schistosoma mansoni* utilizing novel intermediate and definitive hosts. *Parasitology Research*, 1-13.
- King C.H. (2010). Parasites and poverty: the case of schistosomiasis. *Acta Tropica*, 113, 95-104.
- Lively C.M. & Dybdahl M.F. (2000). Parasite adaptation to locally common host genotypes. *Nature*, 405, 679-681.
- Moné H. (1991). Influence of non-target molluscs on the growth of *Biomphalaria glabrata* infected with *Schistosoma mansoni*: Correlation between growth and cercarial production. *Journal of Molluscan Studies*, 57, 1-10.
- Moné H. (1997). Change in schistosome sex ratio under the influence of a biotic environmental-related factor. *Journal of Parasitology*, 83, 220-223.
- Morand S., Manning S.D. & Woolhouse M.E.J. (1996). Parasite-host coevolution and geographic patterns of parasite infectivity and host susceptibility. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 263, 119.
- Mouahid A., Bouhaddioui N., Jana M., Combes C. & Moné H. (1992). Effect of different mollusc associations on target mollusc growth and parasite cercarial production in the triple system: *Bulinus wrighti*-*Schistosoma bovis* and *Melanopsis praemorsa*. *Journal of Molluscan Studies*, 58, 349-355.
- Richards C.S. (1976). Variations in infectivity for *Biomphalaria glabrata* in strains of *Schistosoma mansoni* from the same geographical area. *Bulletin of the World Health Organization*, 54, 706.
- Richards C.S. & Shade P.C. (1987). The genetic variation of compatibility in *Biomphalaria glabrata* and *Schistosoma mansoni*. *The Journal of Parasitology*, 1146-1151.
- Roger E., Mitta G., Moné Y., Bouchut A., Rognon A., Grunau C., Boissier J., Théron A. & Gourbal B. (2008). Molecular determinants of compatibility polymorphism in the *Biomphalaria glabrata*/*Schistosoma mansoni* model: New candidates identified by a global comparative proteomics approach. *Molecular and Biochemical Parasitology*, 157, 205-216.

- Sandland G.J., Foster A.V., Zavodna M. & Minchella D.J. (2007). Interplay between host genetic variation and parasite transmission in the *Biomphalaria glabrata*-*Schistosoma mansoni* system. *Parasitology Research*, 101, 1083-1089.
- Sandland G.J., Wethington A.R., Foster A.V. & Minchella D.J. (2009). Effects of host outcrossing on the interaction between an aquatic snail and its locally adapted parasite. *Parasitology Research*, 105, 555-561.
- Sullivan J.T. & Richards C.S. (1981). *Schistosoma mansoni*, NIH-Sm-PR-2 strain, in susceptible and nonsusceptible stocks of *Biomphalaria glabrata*: Comparative histology. *The Journal of Parasitology*, 702-708.
- Théron A., Pages J.R. & Rognon A. (1997). *Schistosoma mansoni*: Distribution patterns of miracidia among *Biomphalaria glabrata* snail as related to host susceptibility and sporocyst regulatory processes. *Experimental Parasitology*, 85, 1-9.
- Webster J.P. & Woolhouse M.E.J. (1999). Cost of resistance: relationship between reduced fertility and increased resistance in a snail-schistosome host-parasite system. *Proceedings of the Royal Society of London*, 266, 391-396.
- Woolhouse M.E.J. (1989). On the interpretation of age-prevalence curves for schistosome infections of host snails. *Parasitology*, 99, 47-56.
- Zavodna M., Sandland G.J. & Minchella D.J. (2008). Effects of intermediate host genetic background on parasite transmission dynamics: a case using *Schistosoma mansoni*. *Experimental Parasitology*, 120, 57-61.

## ANNEXE 2

Publication: **Lepesant J.M.J.**, Cosseau C., Boissier J., Freitag M., Portela J., Climent D., Perrin C., Zerlotini A. & Grunau C. (2012). Chromatin structure changes around satellite repeats on the *Schistosoma mansoni* female sex chromosome suggest a possible mechanism for sex chromosome emergence. *Genome Biology*, 13, R14. (IF 9.036)

## RESEARCH

## Open Access

# Chromatin structural changes around satellite repeats on the female sex chromosome in *Schistosoma mansoni* and their possible role in sex chromosome emergence

Julie MJ Lepesant<sup>1\*</sup>, Céline Cosseau<sup>1</sup>, Jérôme Boissier<sup>1</sup>, Michael Freitag<sup>2</sup>, Julien Portela<sup>1</sup>, Déborah Climent<sup>1</sup>, Cécile Perrin<sup>1</sup>, Adhemar Zerlotini<sup>3</sup> and Christoph Grunau<sup>1</sup>

## Abstract

**Background:** In the leuphotochozoan parasitic platyhelminth *Schistosoma mansoni*, male individuals are homogametic (ZZ) whereas females are heterogametic (ZW). To elucidate the mechanisms that led to the emergence of sex chromosomes, we compared the genomic sequence and the chromatin structure of male and female individuals. As for many eukaryotes, the lower estimate for the repeat content is 40%, with an unknown proportion of domesticated repeats. We used massive sequencing to *de novo* assemble all repeats, and identify unambiguously Z-specific, W-specific and pseudoautosomal regions of the *S. mansoni* sex chromosomes.

**Results:** We show that 70 to 90% of *S. mansoni* W and Z are pseudoautosomal. No female-specific gene could be identified. Instead, the W-specific region is composed almost entirely of 36 satellite repeat families, of which 33 were previously unknown. Transcription and chromatin status of female-specific repeats are stage-specific: for those repeats that are transcribed, transcription is restricted to the larval stages lacking sexual dimorphism. In contrast, in the sexually dimorphic adult stage of the life cycle, no transcription occurs. In addition, the euchromatic character of histone modifications around the W-specific repeats decreases during the life cycle. Recombination repression occurs in this region even if homologous sequences are present on both the Z and W chromosomes.

**Conclusion:** Our study provides for the first time evidence for the hypothesis that, at least in organisms with a ZW type of sex chromosomes, repeat-induced chromatin structure changes could indeed be the initial event in sex chromosome emergence.

## Background

The origin and evolution of sexuality is one of the most fascinating topics in evolutionary biology. Sex can be determined by several mechanisms, such as environmental stimuli (environmental sex determination) or genetic differences between males and females (genetic sex determination). Genetic sex determination is mainly based on the acquisition of sex chromosomes, a more stable strategy than environmental determinism, especially when the environment becomes variable. The principle steps leading to the emergence and evolution of sex chromosomes

have been proposed by Charlesworth *et al.* [1] and Rice [2]. In this model, the emergence of a locus with female fertility and male sterility and another locus with male fertility and female sterility led to the establishment of a small sex-determining region on ordinary autosomes in hermaphrodite ancestors. These so-called proto-sex chromosomes are hardly distinguishable. To prevent the production of infertile individuals, recombination of these loci becomes restricted [3,4]. This crucial step is intensively debated and two mechanisms of action have been proposed: (i) structural changes by translocation or inversion (reviewed in [5]); or (ii) chromatin status changes involving heterochromatinization of the heterosexual chromosome [4,6-9]. Heterochromatinization of the sex-determining region has been shown in species with

<sup>1</sup>Université de Perpignan Via Domitia, CNRS, UMR 5244 Ecologie et Evolution des Interactions (ZEI), 52 Avenue Paul Alduy, 66860 Perpignan Cedex, France Full list of author information is available at the end of the article

primitive or nascent sex chromosomes, such as in papaya or tilapia (reviewed in [10]). The suppression of recombination between the heterochromosome and its homologue would trigger gradual degradation of the heterochromosome (Y in XY systems, or W in WZ systems) because genes that are not essential for males (in XY systems) or females (in WZ systems) show accelerated rates of mutation and deletion. Consequently, the heterochromosome becomes progressively gene-poor (for example, [11]) and in the extreme case the degradation process can lead to the complete loss of the heterochromosome (for example, [12]).

We decided to investigate the role of chromatin structural changes in sex chromosome emergence by using a basal metazoan species harboring a ZW system, the acoelomate *Schistosoma mansoni*. Schistosomes are parasitic plathyhelminthes that are responsible for schistosomiasis (bilharziosis), an important parasitic human disease ranking second only to malaria in terms of parasite-induced human morbidity and mortality [13]. *S. mansoni*'s life cycle is characterized by passage through two obligatory hosts: the fresh-water snail *Biomphalaria glabrata* (or other *Biomphalaria* species, dependent on the geographical location), for the asexual stage; and human or rodents for the sexual adult stage. The sex of the parasite is determined in the eggs (syngamic determination). Eggs are excreted with the host feces and free-swimming larvae (miracidia) are released when the eggs come into contact with water. These miracidiae infect the freshwater mollusk host and transform into primary and secondary sporocysts. Finally, a third larval stage, the cercariae, capable of infecting the vertebrate host, is released into the water. Once in the human or rodent host, morphological differences between female and male adults develop, and these then mate and produce eggs. In the larval stages, schistosome males and females are genetically different but morphologically identical; the sexual dimorphism (that is, the phenotypic expression of sex differentiation) is restricted to the adult stage. All stages are experimentally accessible, which allows the study of chromatin structural modifications for all stages of the life cycle.

Analysis of metaphase spreads indicates that sex is determined in schistosomes by sex chromosomes, with female being the heterogametic sex (ZW) and male the homogametic sex (ZZ) [14]. In some schistosoma species, there is a clear size difference between W and Z, while in other species, such as *S. mansoni*, discrimination is solely based on chromatin structure [15]. This makes *S. mansoni* a model of choice to study the involvement of chromatin structural changes in sex determination of a model harboring a ZW system. In addition, and in contrast with most other plathyhelminth species, schistosomes are gonochoric [16]. This

suggests that, in general, being hermaphrodite is an advantage in this phylum, probably through minimizing the risk that is associated with finding a mate inside the host [17]. In Schistosomatidae, the acquisition of separated sexes was concomitant with the invasion of warm-blooded animals [16]. This could be explained by the benefit that genetic diversity provides against the sophisticated immune system of warm-blooded vertebrate hosts and/or by the specialization of each gender for a limited set of 'domestic tasks' [16,18,19]. This particular feature of schistosomes in the plathyhelminth phylum provides the opportunity to study sex chromosome emergence.

The genome of *S. mansoni* was sequenced and initially only partially assembled (version 3.1 with 19,022 scaffolds) [20]. During the preparation of this manuscript, an improved version with assembly at the chromosome level became available (version 5.2 with 882 scaffolds) [21], and Criscione *et al.* [22] constructed a linkage map for 210 version 3.1 scaffolds using microsatellite markers. They identified eight linkage groups corresponding to the seven autosomes and one sex chromosome [22], indicating that the sex chromosomes recombine. Nevertheless, Criscione *et al.* discovered a small region of roughly 18 Mb on the sex chromosome that shows recombination repression. Several open questions remain to be answered. First, it is not clear what are the genetic differences between W and Z chromosomes of *S. mansoni*, or in other words, what are the W- and what are the Z-specific sequences. Second, the mechanism of recombination repression between *S. mansoni* sex chromosomes is not clear. As outlined above, either inversion events or heterochromatization [7,9,23] have been proposed for other species. The specific objectives of the present study were to determine what the sex-specific DNA sequences of *S. mansoni* are, and how heterochromatization of the W chromosome might be initiated. We present here evidence that *S. mansoni* sex chromosomes contain large pseudoautosomal regions. Outside these regions, Z-specific sequences are composed of unique sequences and interspersed repeats. W-specific sequences are almost entirely composed of satellite-type repeats located in the heterochromatic region of the W chromosome. While no female-specific gene could be identified, many of the female repeats are transcribed in the larval stages of the parasite but never in the adults. This loss of transcriptional activity and the development into adults is accompanied by chromatin structural changes around the W-specific repeats. We develop a model in which female-specific repeats are expressed to induce a change in chromatin structure of the W chromosome specifically in the sexual part of the life cycle, leading to functional heterogametism.

## Results

### The *S. mansoni* sex chromosomes Z and W share large pseudoautosomal regions

We had previously sequenced genomic DNA of female and male *S. mansoni* individuals of the DFO strain using Illumina sequencing (National Center for Biotechnology Information Sequence Read Archive (NCBI SRA) submission number SRA012151). We aligned the 8,600,198 sequences from the male samples and the 9,355,380 sequences from the female samples to the 19,022 known scaffolds of the *S. mansoni* genome assembly using SOAP. We then calculated for each scaffold the ratio between sequences that match with the scaffold in question ('hit') for the male and the female DNA. The rationale behind this approach was that, in males (ZZ), Z-specific scaffolds should show two times higher hit counts than in females (WZ). We searched for scaffolds with at least 10 hits per 1 kb in the female and the male genome, at least 10 kb in length, and a male/female hit-count ratio  $\geq 1.68$ . Using these parameters we identified 15 scaffolds spanning 6,436,718 bp (roughly 10% of the estimated size of the sex chromosomes [22]). We consider these scaffolds (Smp\_scaff000398, Smp\_scaff018906, Smp\_scaff000301, Smp\_scaff001995, Smp\_scaff000218, Smp\_scaff000465, Smp\_scaff000514, Smp\_scaff000425, Smp\_scaff001883, Smp\_scaff001948, Smp\_scaff000059, Smp\_scaff000044, Smp\_scaff000576, Smp\_scaff000019, Smp\_scaff018900) to be specific for the Z chromosome. We confirmed these *in silico* results for representative regions in a subset of 13 arbitrarily chosen scaffolds (5 Z-specific and 8 pseudoautosomal) by quantitative PCR (qPCR; Table 1). With the exception of one scaffold (Smp\_scaff000120), qPCR confirmed next generation sequencing hit-count ratios. When the working draft of the fully assembled sequence W/Z chromosome became available [21], we repeated the SOAP alignment. In this new assembly, Smp\_scaff000019 was placed on chromosome 2. We showed before [24] that at least 105 scaffolds (436,269 bp) were specific for the W chromosome in females (male DNA did not align to these scaffolds). In conclusion, in genome assembly version 3.1 more than 90% of the non-repetitive part of the Z chromosome and the W chromosome are identical (pseudoautosomal). In version 5.2, the pseudoautosomal region spans 70% of the assembled W/Z chromosome.

### The Z-specific region of the Z chromosome is composed of unique sequences and interspersed repeats

The region that is covered by the 15 Z-specific scaffolds contains 205 putative genes (according to the gene predictions in SchistoDB). For 118 genes, a function could be predicted based on sequence similarities (Additional file 1). Among those there are at least four genes that code for proteins that are predicted to be involved in

**Table 1 Comparison of the ratio of relative amounts of genomic DNA in male and female adults of *S. mansoni***

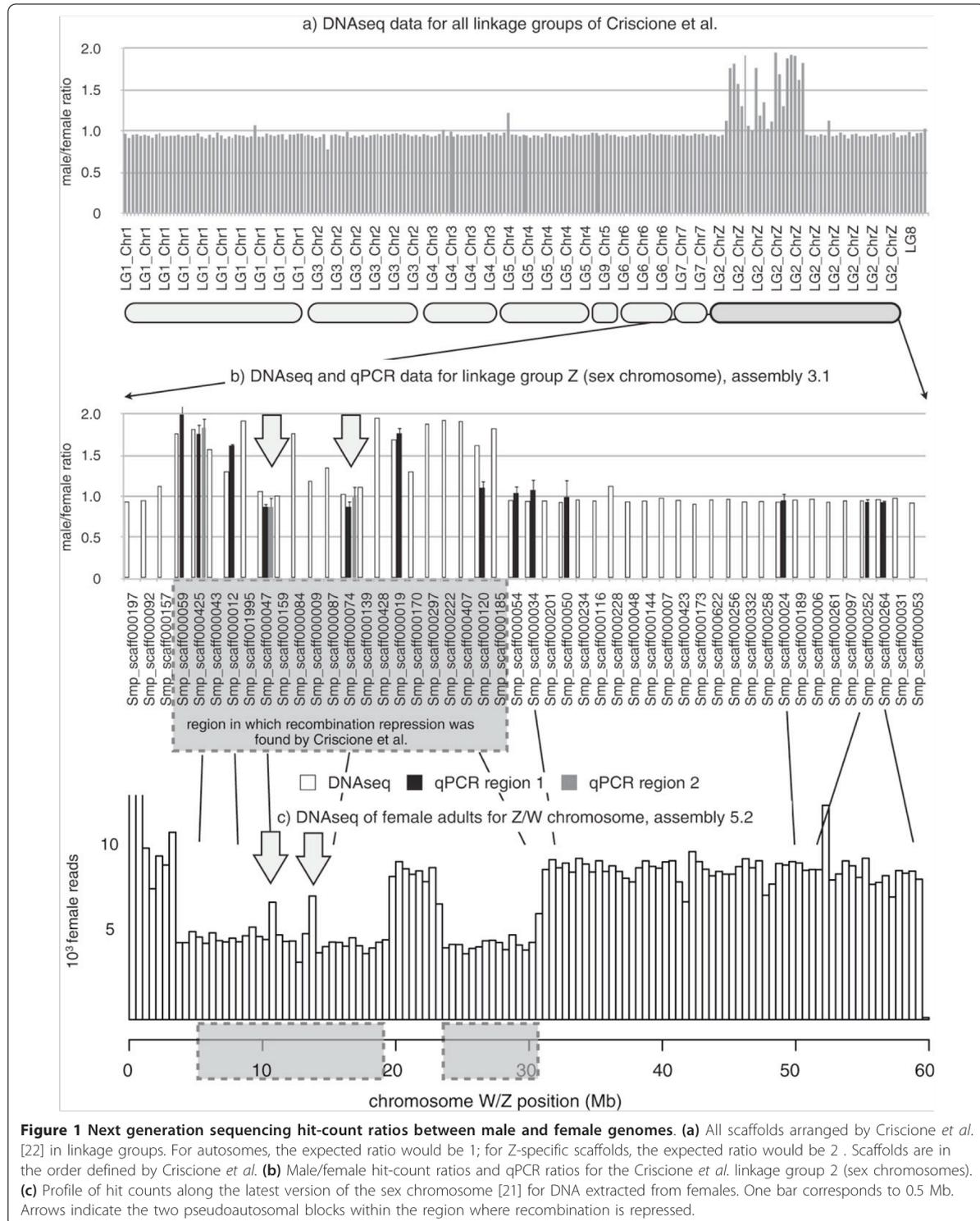
Scaffold	Male (ZZ)/female (WZ)	
	NGS hit-count ratio	qPCR ratio
Smp_scaff000059	1.76	2.00 $\pm$ 0.15
Smp_scaff000425	1.82	1.76 $\pm$ 0.11 (region 1)
Smp_scaff000425	1.82	1.83 $\pm$ 0.11 (region 2)
Smp_scaff000012	1.3	1.61 $\pm$ 0.03
Smp_scaff000047	1.06	0.87 $\pm$ 0.04 (region 1)
Smp_scaff000047	1.06	0.87 $\pm$ 0.12 (region 2)
Smp_scaff000074	1.02	0.87 $\pm$ 0.06 (region 1)
Smp_scaff000074	1.02	0.99 $\pm$ 0.13 (region 2)
Smp_scaff000019	1.69	1.76 $\pm$ 0.07
Smp_scaff000120	1.61	1.10 $\pm$ 0.08
Smp_scaff000054	0.95	1.04 $\pm$ 0.09
Smp_scaff000034	0.94	1.07 $\pm$ 0.13
Smp_scaff000050	0.93	0.99 $\pm$ 0.21
Smp_scaff000024	0.93	0.95 $\pm$ 0.09
Smp_scaff0000252	0.95	0.93 $\pm$ 0.03
Smp_scaff000264	0.96	0.93 $\pm$ 0.02

Relative amounts of genomic DNA in male and female adults of *S. mansoni* were measured by next generation sequencing (NGS) hit counts or qPCR in 13 scaffolds (3 scaffolds were sampled in the 2 different regions, 'region 1' and 'region 2').

spermatogenesis or for which paralogous genes show testis-specific expression. Nevertheless, for the moment it cannot be concluded that these genes are involved in sex differentiation and further analysis is necessary to clarify the role of these genes. Interspersed repeats were also observed in this genomic region but none of them are Z-specific. The Z-specific region in assembly 3.1 is 6.5 Mb in size. In assembly version 5.2 it spans about 18 Mb and, according to [21], contains 782 genes.

### A region on the sex chromosomes with repressed recombination contains Z-specific sequences but also pseudoautosomal sequences

Having identified pseudoautosomal scaffolds and Z- and W-specific sequences, we searched for the location of these sequences on the chromosomes. For the Z-specific scaffolds we explored an existing linkage map for the sex chromosomes [22]. The results are represented in Figure 1. All mapped Z-specific scaffolds are located in a region of the Z-chromosome for which repression of recombination was described. However, this region also contains pseudoautosomal scaffolds with a hit-count ratio of around 1. Consequently, recombination repression in this region is not due only to absence of sister chromatid sequences. This result was confirmed with assembly version 5.2. In this assembly, a block of sequences originally identified as linkage group 8 [22] was inserted at position 20 to 25 Mb. Consequently, this region recombines, but two smaller regions at 12 to 15



Mb are homologous on the Z and W chromosome but recombination repression occurs (Figure 1).

#### The female-specific region of the W chromosome is composed of repetitive sequences

As mentioned above, in an earlier publication we had shown that at least 105 scaffolds (436,269 bp) are specific for the W chromosome in females. We had also indications that a large part of female-specific sequences are composed of repetitive sequences because they matched to known repeats in a repeat database. Nevertheless, 15 to 19% of the massive sequencing data did not correspond to any of the known scaffolds and repeats [24]. These results suggested that they might relate, at least in part, to unknown repetitive sequences. We therefore *de novo* assembled all massive sequencing reads that did not match with unique sequences in the *S. mansoni* genome. SOAP was used to remove *in silico* all female and male reads that correspond to unique sequences, and velvet in combination with a commercial long read assembler was used to assemble the remaining sequences into 8,594 individual repeat contigs (minimum length 80 bp; maximum length 2,169 bp; average length 168 bp). The minimum length corresponds to the used velvet parameter. We then applied our earlier described whole-genome *in silico* subtractive hybridization (WISH) approach [24] to identify female-specific repeats. Thirty-three new repeat sequences were identified to be specific for the female W chromosome, giving a total of 36 W-specific repeats (combined literature data and our data). Several *in silico* methods were used to classify the repeats and their specificity was confirmed by PCR on male and female individuals (abundant in females, very weak signal or absence of amplification in males). The results are summarized in Table 2. Three repeats were already known, 33 repeats are new. The size of the consensus sequence for each assembled repeat was confirmed by PCR on female and male individuals (Additional files 2 and 3). EST data and RT-PCR show that at least eight repeats are transcribed. For a subset, copy number was estimated by qPCR and is moderate (100 to 400 copies), with the exception of SMA1phafem-1 (several thousand copies, confirming earlier estimations [25]). The copy number was estimated using quantitative DNA with a unique W-specific region on scaffold Smp\_scaff018821 as reference (positions 2,194 to 2,312).

We used SchistoDB to identify genes that could be located within the region that is spanned by the repeats. Eight putative genes were identified in the vicinity of the repeats (not more than 5 kb away). Manual inspection of all loci showed that female next generation sequencing hits can be found for four putative genes, and male hits

are absent (Smp\_186230, Smp\_190410, Smp\_117150, Smp\_117160). However, three genes (Smp\_190410, Smp\_117150, and Smp\_117160) are identical and the predicted coding regions are small (243 bp for Smp\_190410, 327 bp for Smp\_186230). No significant similarity to known proteins could be found with blastx. Blast against the genome shows that these putative genes are not unique and it remains to be answered whether these sequences are actually transcribed and code proteins.

#### Female-specific repeats are arranged as large satellite type blocks in the heterochromatic region of chromosome W

To identify the localization of the most abundant female-specific repeats, W1, W3-8 and W13, we used fluorescent *in situ* hybridization (FISH) on late secondary sporocyst metaphases (Figure 2). All studied repeats are (i) arranged as large satellite blocks and (ii) localized in the heterochromatic region of the W chromosome (darker propidium iodide staining), either in the pericentromeric region or on the euchromatin/heterochromatin boundary of the long arm. None of the tested repeats was found on the short arm of chromosome W. Repeats W6 and W7 are specific for the pericentromeric region of the q-arm, and W1 and W4 are located on the frontier of the heterochromatic region. W1 was already known [26] and we confirm the earlier FISH results that localized it to the distal part of the heterochromatic region of Wq [27]. Hirai *et al.* [27] described a euchromatic gap region (eg3) in the vicinity of the W1 chromosome. We did not see this gap, which might be due to the lower resolution of our equipment or differences between the used *S. mansoni* strains. W1 shows genetic instability and in some cases was also found in males [28]. The reason for this could be the close proximity to euchromatin and one might expect such a behavior also for W4. W3, W5 and W8 can be found in both the pericentromeric and the frontier region. W13 is localized roughly in the middle of the heterochromatic part of the q-arm. Results are summarized in Table 2. Copy number estimates (Table 2) correspond to what was found in the literature: 500 to 1,000 copies per genome for W1 [26] and 20,000 to 200,000 for SMA1phafem-1 [25].

#### Several of the female-specific repeats are transcribed in larvae but not in adults

EST data suggested that some of the repeats could be transcribed and transcription of W1 and SMA1phafem-1 was described for cercaria [29]. We extracted RNA from different life cycle stages and quantified the transcription level for repeats W3, W4, W5 and SMA1phafem-1. For repeats W3 we did not find significant transcription above background; however, repeats W4, W5 and SMA1phafem-1 are transcribed in the larval stages. No transcripts could be detected in adult couples or

**Table 2 W-chromosome-specific repeats of *S. mansoni***

GenBank accession number	Name	Length (bp)	Percentage female hits	Repeat family <sup>a</sup>	Transcription evidence	FISH localization	Copy number estimate <sup>b</sup>	Reference
U12442	SMA1phafem-1	338	99.86	SMA1alpha retroposon	RT-PCR	p-arm	60,000 - 70,000	[61,62]
J04665	W1	482	100	Retro		Middle of q-arm at frontier between heterochromatin and euchromatin as satellite, middle of q-arm	500	[26,27]
U10109	W2	715	100				400	[62]
HQ880214	W3	786	100	LTR, highly similar to W2, highly similar to R = 407	No transcription (RT-PCR)	As satellite in the middle of q-arm at frontier between heterochromatin and euchromatin or also in the pericentromeric region	200	
HQ880209	W4	1132	99.52	Highly similar to R = 879	RT-PCR	Same location as W1	800	
HQ880217	W5	1129	99.27	LTR, similar to Perere-2, identical to R = 564	EST and RT-PCR	Either at the frontier of heterochromatin and euchromatin of the q-arm or in the pericentromeric region, or at both locations		
HQ880215	W6	310	99.88	Retro	EST	In the pericentromeric region		
HQ880210	W7	1000	100	DNA transposon, 97% identical to GenBank accession number XP_002570219 (hypothetical protein Smp_186230)		In the pericentromeric region		
HQ880218	W8	266	99.97	Tandem repeat (previously described as TR266), DNA transposon		Either at the frontier of heterochromatin and euchromatin of the q-arm or in the pericentromeric region		
HQ880211	W9	803	100	LINE2, similar to Sjr2 retrotransposon	EST			
HQ880212	W10	682	100	LTR				
HQ880213	W11	376	100	LINE, similar to R = 170				
HQ880216	W12	264	100	Retro, 97 to 100% identical to several hypothetical S.m. proteins				
HQ880219	W13	258	100	Retro		In the middle of the heterochromatic part of the q-arm as satellite		
HQ880220	W14	209	100	DNA transposon, similar to R = 170				
HQ880221	W15	185	96.62	DNA transposon				
HQ880222	W16	164	100	Similar to R = 116				
HQ880223	W17	160	100	LTR				
HQ880224	W18	160	100	LTR	EST			
HQ880225	W19	139	100	Retro, 100% identical to GenBank accession number XP_002569391 (hypothetical protein Smp_181820)	EST			
HQ880226	W20	138	100					
HQ880227	W21	138	99.84	DNA transposon, similar to R = 116				
HQ880228	W22	132	100	DNA transposon				
HQ880229	W23	125	100	DNA transposon				

**Table 2 W-chromosome-specific repeats of *S. mansoni* (Continued)**

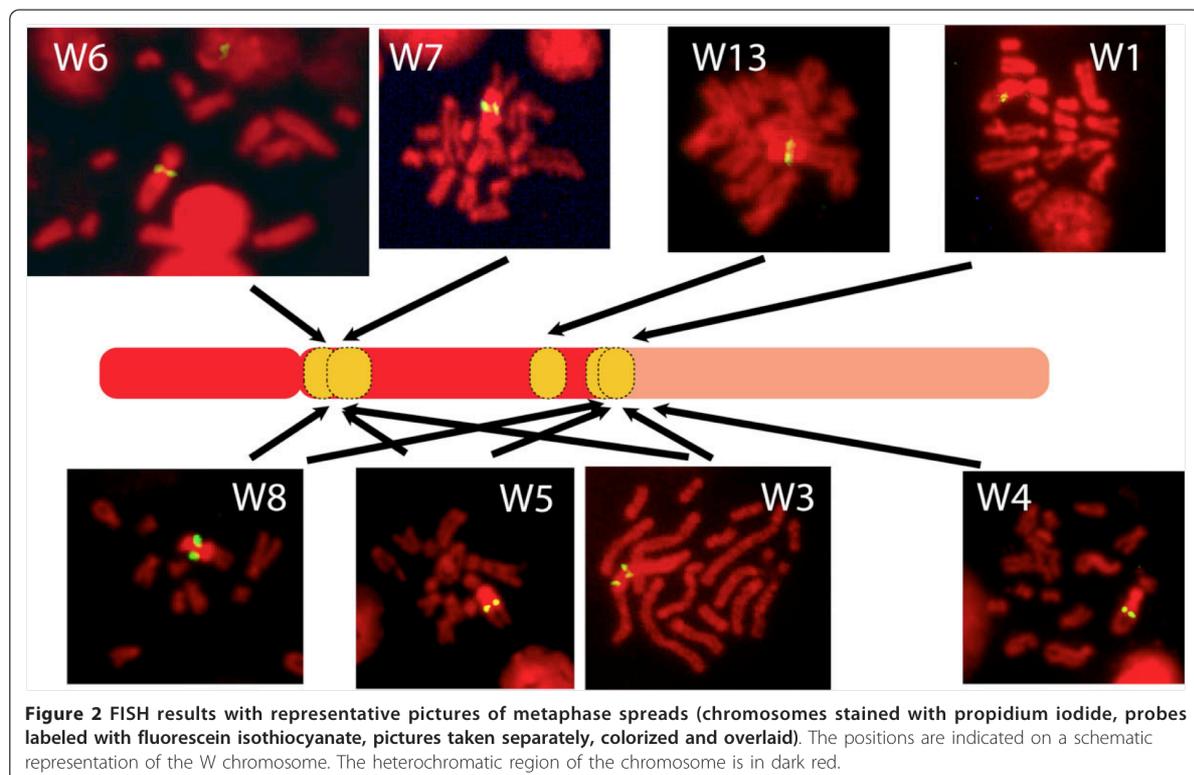
HQ880230	W24	115	99.56	Retro, similar to Sh122 repeat and R = 31	
HQ880231	W25	112	99.65	LTR	
HQ880232	W26	111	100	DNA transposon	
HQ880233	W27	110	100	DNA transposon, similar to R = 133 and Sh microsatellite C2	
HQ880234	W28	108	100	Similar to Sh microsatellite C140	
HQ880235	W29	97	100	Similar to Sb Sbov20 repeat	
HQ880236	W30	96	99.05	LTR	
HQ880237	W31	92	100	LTR	
HQ880238	W32	89	96.88	DNA transposon	
HQ880239	W33	86	99.86	LTR	
HQ880240	W34	82	100	DNA transposon	
HQ880241	W35	80	100	Retro	EST

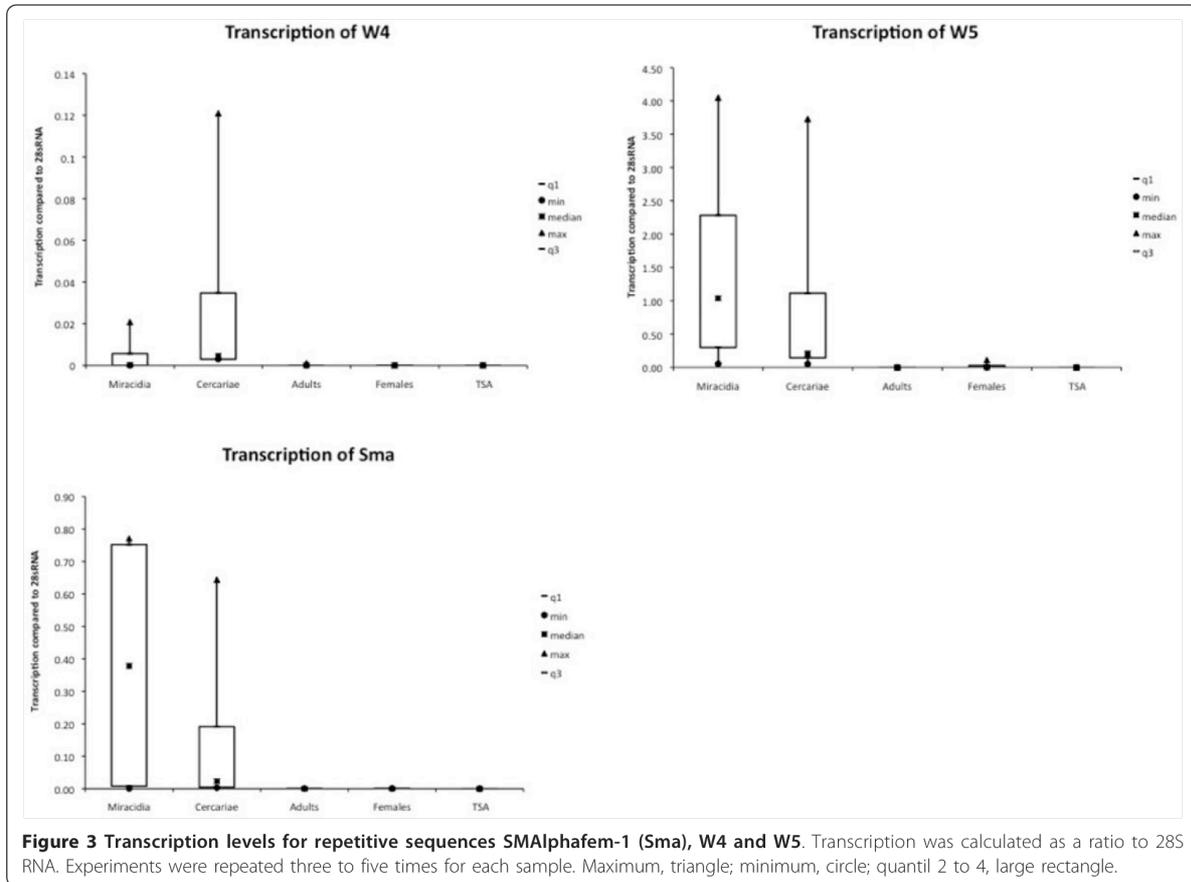
Because of the WZ-type chromosome set of females, these repeats are female-specific. <sup>a</sup>Combined Censor, Blast, TeClass results. <sup>b</sup>qPCR based. LINE, long interspersed element; LTR, long terminal repeat.

immature females (Figure 3). At the genomic DNA level, we observed  $\leq 20\%$  differences in repeat copy numbers (measured by qPCR) between different biological samples, but we did not observe a decrease in copy number, that is, shrinking of repeats, during the life cycle. Absence of transcription in adults is not, therefore, due to absence of repeats in the genome.

**The chromatin structure around the female-specific repeats changes during the life cycle**

Repeat transcription has been linked to chromatin structural changes [30]. We therefore analyzed histone isoforms that could potentially be associated with the female-specific repeats. Chromatin immunoprecipitation followed by massive sequencing (ChIP-Seq) was used to analyze the



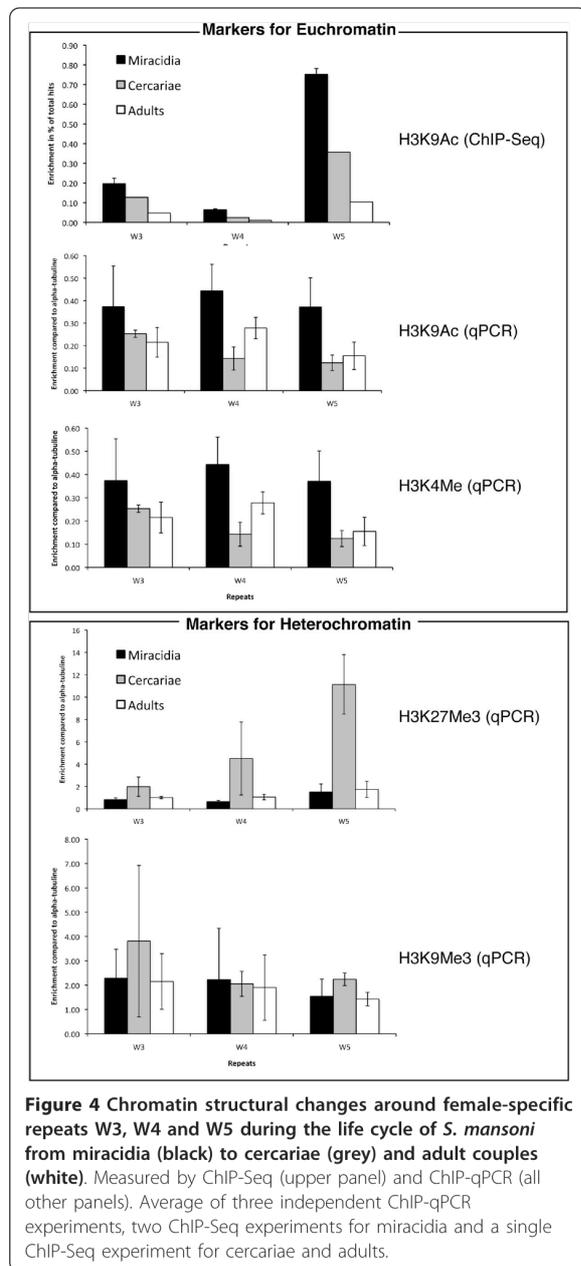


abundance of acetylated histone H3K9 (H3K9Ac) around the repeats in miracidia, cercariae and adult couples. All 36 female-specific repeats show a characteristic gradual decrease in H3K9 acetylation level from the larval stages to the adult stages. Among the total of 8,594 repeats in the genome, only 1,113 repeats show such a gradual decrease in H3K9 acetylation. The probability that such a pattern could be observed by chance for all 36 W-specific repeats is negligible (the individual term binomial distribution probability is  $1.1^{-32}$ ). To verify the ChIP-Seq data by ChIP combined with qPCR, we focused on two transcribed repeats (W4 and W5) and one non-transcribed repeat (W3). We used antibodies against H3K9Ac, tri-methylated H3K4 (H3K4Me3) that are characteristic for actively transcribed euchromatin, and the heterochromatin markers tri-methylated H3K9 (H3K9Me3), and tri-methylated H3K27 (H3K27Me3). A region in the body of the alpha-tubulin gene was used as reference for calculating the relative amount of immunoprecipitated DNA. The results are shown in Figure 4. Both euchromatic markers (H3K9Ac and H3K4Me3) are enriched at the repeats in the miracidia stages where transcription was observed. In contrast, there

are much fewer euchromatic markers around the repeats in adults. In the qPCR-based experiments, cercariae occupy an intermediate position. Based on the combined ChIP-seq and ChIP-qPCR data, we conclude a clear decrease in H3K9 acetylation from miracidia to cercaria and adults. Also, the abundance of the second euchromatic marker, methylation of H3K4, decreases from miracidia to cercaria and remains constant during the development into adults. The heterochromatic markers H3K9Me3 and H3K27Me3 are abundant in cercaria but low in miracidia and adults. In summary, around the female-specific repeats we observed three distinct types of chromatin structure in the three different life-cycle stages: in miracidia the repeats are clearly euchromatic, in cercaria a large proportion is heterochromatic, and in adults we can find a peculiar chromatin structure without classical euchromatic or heterochromatic markers, but associated with transcriptional silence.

#### Histone deacetylase inhibition does not induce transcription of W-specific repeats in adults

We tested whether the observed changes in chromatin structure are a result or the cause of the changes in



transcription. If hypoacetylation of histones were the cause of transcriptional inactivation, then inactivation of histone deacetylase would relieve repression. On the other hand, if transcription of repeats is the origin of chromatin structural changes, inhibition treatment should not lead to detectable changes in transcription because each transcriptional increase would reinforce deacetylation and counteract the inhibition. We treated adult parasites with trichostatin A (TSA), an inhibitor of

histone deacetylases at increasing concentrations *in vitro*. After 2 hours of treatment with  $\geq 20 \mu\text{M}$  TSA, mobility changes were observed (worms first straightened up and ceased moving, and convulsive movements were observed at higher concentrations and longer incubation times (Additional file 4)). We then measured the transcription levels for repeats W4, W5 and Sm-alpha-female at  $20 \mu\text{M}$  TSA and for 4 hours. In none of the cases was transcription activated. In contrast, an increase of transcription of retrotransposons Perere3 and Saci7, used as control, was observed (by 45 and 23%, respectively). The lactate dehydrogenase test shows no difference in cytotoxicity between TSA-treated and mock-treated worms.

## Discussion

Despite tremendous advancements in the past, the elements that are responsible for the establishment of sex chromosomes remain still enigmatic. According to Müller's ratchet model, sexual reproduction evolved because deleterious mutations could be eliminated by recombination between the parental autosomes [31]. To maintain isolation of two different sexes, recombination must, however, be repressed (at least partially) between the sex chromosomes. Zones in which recombination is repressed between sex chromosomes were meanwhile identified in many species. Accumulation of repeats on the heterogametic sex chromosome was also found in many examples, although their role is unknown and many authors still consider them as junk DNA. The view that repetitive DNA is non-functional was challenged by the discovery of transcription from repeats on autosomes and the production of small RNA that could be related to heterochromatinization events [32]. The presence of large heterochromatic blocks is also a common feature of sex chromosomes. So far, these observations were made in isolation from each other, and generally in different species, which makes the construction of a hypothetical model difficult. Here we present for the first time a comprehensive analysis of sequence composition, gene and repeat content, chromatin structure and repeat transcription of the sex-specific chromosome regions of the Z and W chromosomes of our biological model *S. mansoni*. Recombination repression has been described before in this region of interest [22]. Our data, in relation to previous reports, allows the current models for the suite of events that led to sex chromosome differentiation in *S. mansoni* to be refined and could represent a general model for this process in species with genetic sex determination of the Z/W type.

## Z- and W-specific sequences

Criscione *et al.* [22] identified a region of 20 scaffolds in which recombination repression was observed and

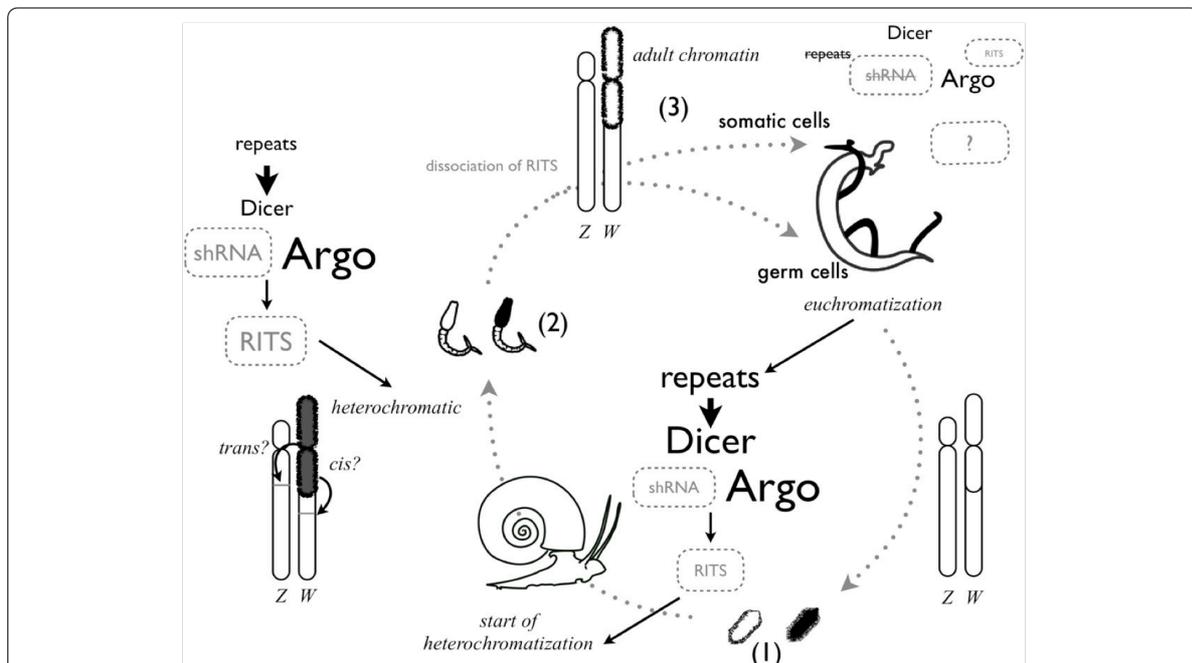
suggested that these are Z-specific sequences. We indeed found a male/female sequence reads hit and/or qPCR ratio of  $\geq 1.5$  for 13 of these scaffolds, indicating an overrepresentation in the male genome. However, seven scaffolds in this region showed no disequilibrium of hit counts and/or qPCR between males and females (male/female hit ratio  $\leq 1.4$ ), that is, the sequences are not specific to the Z chromosome (Figure 1). In other words, recombination is repressed but the homologous sequences on the sister chromosomes are still present. We find at least two blocks of sequences that are shared between the Z and W chromosome located in the large region with recombination repression. This result was confirmed with the most recent version of the genome assembly. We see three possible conclusions that can be drawn from our results. Either the Z/W sequence blocks are inverted, and additionally or alternatively the sequences are heterochromatic, thus preventing recombination. It is also possible that the scaffolds in the original assembly of the *S. mansoni* genome were chimeric. Indeed, of the 48 scaffolds originally found in linkage group Z/W [22], 4 are on other chromosomes in the 5.2 assembly. It will be difficult to formally exclude the possibility that our results are due to misassembly.

We did not find any paralogues to sex determination genes among the predicted genes on the Z-specific scaffolds. The specific region of the W chromosome is largely composed of large satellite blocks of at least 36 different W-specific repeats. These repeats are abundant on the W chromosome but our PCR analysis on different male individuals indicates that these sequences can also sometimes be found on other chromosomes. The strength of the PCR signal suggests, however, that they are present in very low copy number there. Analysis of the genomic sequence shows that they can occur intermingled with other repeats on autosomal scaffolds as individual sequences or as small blocks of up to five repeats in tandem. Our understanding of these results is that these repeats exist as large satellite blocks on the W chromosome but can occasionally be transferred to autosomes by a so far unknown mechanism. Such a behavior was described for W1 [28] and could depend on the chromatin structure around the repeats and/or flanking regions. Several of these W-specific repeats are transcribed in the miracidia and cercariae stages but never in the adults.

#### Role of W-specific repeats

In most species that possess sex chromosomes of the Y or W type it was found that (i) repetitive sequences accumulate on these chromosomes, (ii) large regions are heterochromatic and (iii) these chromosomes deteriorate or are completely absent in the extreme case. We show that the W chromosome in *S. mansoni* is no exception

to this rule. What is unknown, however, is the suite of events in the evolution of sex chromosomes and the role of the different elements in sex determination. We believe our present study sheds some light on this matter. Heterochromatization of the W chromosome in schistosomes has been known for a long time and has been even used as a marker for sex identification in morphologically indistinguishable cercariae [14,15,33]. Based on cytogenetic analysis, some authors argued that heterochromatization of the W starts in miracidia [14]. Since it is impossible to determine chromosome banding in miracidia and then reuse the larvae for infection and production of adults, these results are difficult to verify. Our results clearly show that the repeats that are located in the W heterochromatic region carry a euchromatic signature in miracidia and lose their euchromatic character progressively during the development into adults. This process is accompanied by a decrease of transcription until complete silencing of the repeats in the sexually mature adult stage. During the miracidia to cercaria transition - that is, precisely when sexual dimorphism starts to develop - the repeats heterochromatize. Sex-specific repeats are found in many species [34-36]. In some cases transcription has been described and it was suspected that these repeats play a role in the sex determination process [37,38]. The transcription of repetitive elements of the satellite type in *S. mansoni* is particularly interesting in the light of the recent discovery of stage-dependent expression of the elements that constitute the RNA interference (RNAi) pathway in schistosomes [39,40]. In many organisms RNAi and chromatin structural changes are linked [32,41-43] and it is tempting to speculate that transcription of W-specific repeats is actually the origin of chromatin compaction on the W chromosome during the life cycle. A hypothetical scheme is shown in Figure 5. In our model, reset of the repeat chromatin structure occurs during early embryogenesis (formation of miracidia). In the miracidia, repeats are euchromatic and several of them are transcribed. Transcripts are processed through a pathway that has similarity to RNAi and a hypothetical repeat-induced silencing complex is formed that induces the formation of heterochromatin around the repeats. At this stage, miracidia have infected the mollusk host and develop via sporocyst stages into cercaria. In cercaria, most of the repeats are heterochromatic and not transcribed. We hypothesize that the heterochromatization extends beyond the repeat frontiers and that nearby loci are silenced. If a sex determination locus is found among these loci, the heterochromatization would lead to a dose effect that could be the origin of the formation of the female adult phenotype. Once the task of silencing this locus in *cis* (or *trans*) is accomplished, repeats are not anymore transcribed and the chromatin



**Figure 5 Hypothetical model of the relationship between W-specific repeat transcription and heterochromatin formation.** In black, experimentally confirmed situation; in grey, hypothetical elements. Letter size corresponds to relative strength of the phenomenon. (1) Miracidia do not show sex dimorphism. The W chromosome is euchromatic and repeats are transcribed. Large amounts of Dicer and Argonaute proteins are present [39,40,58]. Dicer could produce small heterochromatic RNA (shRNA) that could bind to Argonautes and could build a RITS (RNA-induced initiation of transcriptional gene silencing) complex similar to those in yeast [42] that initiates heterochromatization around the repeat region of the W chromosome. (2) After infection of the snail host, cercariae are produced. Also, these larvae do not show sex dimorphism. Repeats are still transcribed but transcription of Dicer has decreased [39]. Accumulation of RITS complex around the repeat region progressively leads the loss of its euchromatic character. Finally (3), in dimorphic adults, the repeat region becomes transcriptionally inactive, and heterochromatin is maybe locked by HP1 or related chromatin proteins such as KDM2A [59], or a second histone modification mark is missing [60]. Dicer and Argonaute are less abundant. RITS is no longer necessary. Sexual reproduction occurs during this stage. Recombination is repressed by non-permissive chromatin. Late in germ cell production or during embryogenesis, erasure of chromatin marks occurs (epigenetic reset). The cycle restarts with a euchromatic W chromosome.

structure of the pericentromeric W chromosome is fixed into an unknown but transcriptionally silent configuration. We can only speculate about the proteins that are involved since our data indicate that neither the euchromatic markers H3K9Ac and H3K4Me3 nor the heterochromatic markers H3K9Me3 and H3K27Me3 are abundant. This model is supported by our finding that *in vitro* treatment of adults does not lead to detectable transcription from the W-specific repeats while autosomal retrotransposons can be activated.

One could argue that the function of repeat-induced silencing is purely defensive and down-regulates retrotransposon expression in general. Such a mechanism was described as the repeat-associated small interfering RNA (rasiRNA)-mediated pathway [44] in *Drosophila* ovary cells and is believed to protect the (female) germ line from transposable elements. If this were the case for *S. mansoni*, transcription should be observed in the ovary. Our data do not support this view.

### Conclusions

Most authors agree that suppression of recombination is an initial event in sex chromosome emergence, although it is not clear by what mechanism it is caused. Chromosome rearrangements (for example, inversions) or the action of modifier genes have been proposed (reviewed, for example, in [45]). Other authors see conformation differences (chromatin structural changes, differences in replication timing) as the origin for recombination inhibition [3,5]. Accumulation of repeats is a general feature of Y/W-type chromosomes. Some consider it an important feature with unknown function [36], while others see repeat accumulation as the result of recombination suppression [1] or solely as a genome defense mechanism [7], placing it late in the suite of events that characterize evolution of sex chromosomes.

With the present work we contribute two new elements that allow us to exclude some of the current hypotheses and to refine others. First, we show that the

presence of satellite repeats on the W chromosome does not lead in all life cycle stages to heterochromatinization. Consequently, it is not their presence itself that induces the heterochromatin formation. We show that all W-specific repeats are euchromatic in the miracidia stage. Our ChIP-Seq data tell us that this is not a general feature of autosomal and pseudoautosomal repeats, but specific for the W-specific satellites. Second, we demonstrate that the euchromatinization occurs concomitantly with transcription and that transcription always precedes heterochromatinization.

Based on these findings, we propose two not necessarily exclusive scenarios for the emergence of sex chromosomes. In the first model, transcription of non-coding RNA from repetitive DNA elements was the initial event in sex chromosome evolution of schistosomes. Non-coding RNA would have induced heterochromatinization and suppression of recombination. Both favored expansion of repeats and organization in large blocks (satellites). Satellite expansion would have reinforced the system and led finally to the beginning of genetic changes in the W chromosome. The very basal phylogenetic position of leuphotrochozoans such as *S. mansoni* permits a general model for the main stages of sex chromosome evolution to be proposed: the establishment of a sex-determining region, recruitment of repeats for production of non-coding RNA, RNA-directed heterochromatinization and repeat expansion, local suppression of recombination, and shrinkage of the chromosome by deletion.

In the second model, a small mutation and/or local heterochromatinization could have been the initial event, leading to recombination repression in the first place. Repetitive DNA accumulated subsequently. During germ cell formation or during early embryogenesis euchromatinization occurs. Cytogenetic evidence in other species in which the female is the heterogametic sex shows that the W chromosome is often condensed in somatic cells, and becomes euchromatic in early oocytes (reviewed in [46]). This releases transcription repression and repeats are transcribed, leading subsequently to heterochromatinization. Our preliminary data suggest that chromatin structural changes do not occur in *trans* - that is, not on the Z chromosome but on the adjacent regions of the W chromosome (not shown).

We cannot formally exclude that sex determination is based on a specific protein-coding gene that is absent or present on the W chromosome. But we show that the most pronounced difference in transcription between ZZ and ZW individuals is at the level of 'non-coding' RNA. We therefore favor the hypothesis that sex differentiation in *S. mansoni* is based on developmental stage-dependent tagging of the W chromosome by non-coding RNA and a chromatin marking system. Our

model predicts that chromatin structural changes influence transcription of one or several genes in the close vicinity of the core heterochromatic region and that transcriptional activation or inactivation of these leads to morphological and/or physiological changes that are the bases for development of the male and female phenotypes in the adult stage.

## Materials and methods

### Parasite culture and drug treatment

Eggs were asexually recovered from 60-day infected hamster livers and miracidia were hatched from eggs in 5 ml of spring water over 2 to 3 hours under light. Miracidia were concentrated by sedimentation on ice for 15 minutes. Cercariae were recovered from infected snails (4 weeks post-infection) and collected by pipetting. They were then concentrated by cold centrifugation (4°C) at 1,200 rpm for 5 minutes and the supernatant was removed. Eight-week-old adult worms were recovered by portal perfusion of hamsters with 0.8% (w/v) NaCl and 0.8% (w/v) trisodium citrate [47]. If necessary, miracidia, cercariae and adults were kept at -80°C.

For infection with a single sex, *B. glabrata* snails 4 to 5 mm in diameter were individually exposed to a single miracidium in 5 ml of springwater. The snails were then each isolated and maintained in round, clear plastic containers for 24 hours and kept all together for 5 weeks. Snails were fed fresh lettuce *ad libitum* and the water was maintained at 25°C and changed weekly. The photoperiod during the entire experiment was equilibrated to 12 hours light:12 hours dark [48].

Adults were recovered by portal perfusion of hamsters. Ten individuals were kept in 250 µl RPMI medium (Invitrogen-Gibco, Carlsbad, USA) and treated with an ethanol solution of the histone deacetylase inhibitor TSA (Invitrogen) at different final concentrations (2 µM, 20 µM, 50 µM, 100 µM and 200 µM). To the untreated control, a corresponding volume of ethanol was added. The cytotoxic effect of the drug was measured using the Roche Cytotoxicity Detection Kit (Roche no. 04744926001), which is based on the measurement of lactate dehydrogenase activity released from dead and lysed cells into the supernatant [49]. Behavior was observed every hour until 6.5 hours and after 21 hours of treatment. Individuals were filmed with a conventional numerical camera adapted to a stereomicroscope after 5, 6.5 and 21 hours of treatment.

### Sequencing of genomic DNA, alignment, and assembly of repeats

Solexa sequencing was performed at the sequencing facilities of GenomiX Montpellier (France) on a Genome Analyzer II (Illumina) by single end sequencing (36 bp) according to the manufacturer's protocol. The software

SOAP is usually employed to map unique sequences and reject repetitive sequences. We took advantage of this algorithm and used SOAP 2.17 [50], evoking the *-u* and *-r 0* options to split the sequence reads into those corresponding to unique or repetitive sequences. The resulting fasta files of unmapped reads (*-u*) was assembled with velvet using a coverage cutoff of 4 and a minimum contig length of 80 bp. For a second assembly round Sequencher v4.5 was used with minimum match 93%, minimum overlap 60 bp.

#### ***In silico* analysis**

Velvet-assembled repeats were then used for the whole-genome *in silico* subtractive hybridization (WISH) procedure [24]. This method compares different massive sequencing datasets with a reference genome and identifies sequences that are under-represented in one data set. Censor [51], Teclass [52] and blast [53] were used for repeat annotation.

For identification of genes in the vicinity of W-specific repeats, all repeat sequences were compared to the genome using blast searches of the SchistoDB database [54] and genes 5 kb upstream and downstream of regions containing these repeats were manually analyzed.

#### **Confirmation of sex-specific sequences by PCR**

PCRs were carried out in a final volume of 25  $\mu$ l containing 0.2  $\mu$ mol of each oligonucleotide primer (Additional file 5), 0.2 mmol of each dNTP (Promega), 0.625 U of GoTaq polymerase (Promega) used with the recommended buffer and completed to the final volume with DNase-free water. The PCR program consisted of an initial denaturation phase at 95°C for 5 minutes followed by 20 cycles at 95°C for 30 s, 60°C for 90 s, 72°C for 30 s and a final extension at 72°C for 5 minutes. The PCR products were separated by electrophoresis through a 2% TBE agarose gel.

#### **FISH on *S. mansoni* metaphases**

Metaphase spreads were prepared essentially as described by Hirai and LoVerde [55]. Sporocysts were obtained by dissection of two to three snails, each infected with five miracidia, at 28 to 29 days post-infection. Probes for repetitive DNA were prepared by cloning PCR products (for primers see Additional file 5) on genomic DNA as template into pCR2.1-TOPO (Invitrogen #K4510-20). Clones were sequenced to confirm the repeat assembly, labeled with the BioPrime DNA labeling system (Invitrogen #18094-011) and hybridized as described before [55]. Chromosomes were counterstained with propidium iodide and observed under an epifluorescence microscope (AKIOSKOP 2, Zeiss) equipped with a Leica DC 300 FX digital camera.

Between 7 and 34 female metaphases were studied for each repeat.

#### **RNA extraction, cDNA synthesis and qPCR**

Total RNA was purified from three independent preparations of larvae and adults. For the larval stages, RNA was extracted from 10,000 miracidia and 10,000 cercariae using 500  $\mu$ l Trizol (Invitrogen). Fifty adult couples were solubilized in 500  $\mu$ l Trizol with a MagNA Lyser and Green beads (Roche). RNA was treated with DNase I (Invitrogen) for 15 minutes at 37°C, followed by inhibition of the enzyme for 10 minutes at 65°C. PCR of 28s rDNA was used to test for genomic DNA contaminations. The DNase I treatment was repeated as many times as necessary to eliminate contaminations with genomic DNA. RNA was purified with the QIAGEN RNeasy kit. First strand cDNA was synthesized using 10  $\mu$ l of the total RNA preparation, in a final volume of 20  $\mu$ l (10 mM dNTPs, 0.1 M DTT, 40 U RNase out, 0.15  $\mu$ M random primers) with 200 U of SuperScript II RT (Invitrogen). After reverse transcription, the cDNAs were purified with the PCR clean-up system (Promega) and eluted into 40  $\mu$ l 10 mM Tris/Cl (pH 7.5). Real-time PCR analyses were performed using the LightCycler 2.0 system (Roche Applied Science) and LightCycler Fast-start DNA Master SYBR Green I kit (Roche Applied Science).

qPCR amplification was done with 2.5  $\mu$ l of cDNA in a final volume of 10  $\mu$ l (3 mM MgCl<sub>2</sub>, 0.5  $\mu$ M of each primer, 1  $\mu$ l of master mix). Primers were designed with the LightCycler Probe design software or the primer3plus web based interface [56]. The following protocol was used: denaturation, 95°C 10 minutes; amplification and quantification (40 cycles), 95°C for 10 s, 60°C for 5 s, 72°C for 16 s; melting curve, 60 to 95°C with a heating rate of 0.1 C/s and continuous fluorescence measurement, and a cooling step to 40°C. For each reaction, the crossing point (Ct) was determined using the 'fit point method' of the LightCycler Software 3.3. PCR reactions were done in duplicates and the mean value of Ct was calculated. 28s rRNA was used as an internal control and the amplification of a unique band was verified by electrophoresis through 2% TBE agarose gels for each qPCR product. Primer sequences and expected PCR product size are listed in Additional file 5. For all qPCR, efficiency was at least 1.89.

#### **Chromatin status analysis by ChIP and qPCR**

Native ChIP and ChIP-Seq were performed as described before [57]. In brief, antibodies against histone isoforms (Table 3) were used to precipitate chromatin in sporocysts, cercaria and adults. The resulting DNA was analyzed either by ChIP-Seq or qPCR. ChIP-Seq data are available at the NCBI SRA under accessions SRX088545, SRX088544, SRX088543 and SRX087825. For ChIP-Seq

**Table 3 Antibodies used for native ChIP (N-ChIP)**

Antibody	Host	Product	Lot	Saturating quantity used for N-ChIP <sup>a</sup>
H3K9ac	Rabbit	Upstate, 07-352	DAM16924924	8 µl
H3K4me3	Rabbit	Upstate, 04-745	NG1680351	4 µl
H3K9me3	Rabbit	Abcam, Ab8898	733951	4 µl
H3K27me3	Rabbit	Diagenode, pAb-069-050	A29900242	8 µl

<sup>a</sup>Saturating quantities for H3K9ac and H3K9me3 antibodies were previously determined [56]. Saturating quantities for H3K4me3 and H3K27me3 were determined in this study by a titration experiment (data not shown).

analysis, a repeat pseudogenome was constructed in which each identified repeat sequence occurred only once. Then SOAP2 [50] was used to align roughly 100,000 36-bp reads for miracidia of two strains (GH2 and BRE), cercaria and adult couples (both GH2) to this pseudogenome. Hit counts for each repeat were normalized by the total number of aligned reads and compared for the different stages.

### Additional material

**Additional file 1: List of male-specific scaffolds with putative genes.**

**Additional file 2: Video of adult schistosomes treated with TSA at 100 µM.** Individuals were filmed with a conventional numerical camera adapted to a stereomicroscope after 5 hours of treatment.

**Additional file 3: Video of adult mock-treated schistosomes.** Individuals were filmed with a conventional numerical camera adapted to a stereomicroscope after 5 hours of treatment.

**Additional file 4: Photographs of ethidium bromide stained PCR products after migration through 2% agarose gels.** PCR amplification was used to confirm size and sex-specificity of assembled W-specific repeats. Genomic DNA of two female (F1, F2) and two male individuals (M1, M2) was used as template.

**Additional file 5: Primers used in this study.**

### Abbreviations

bp: base pair; ChIP: chromatin immunoprecipitation; ChIP-qPCR: chromatin immunoprecipitation followed by quantitative PCR; ChIP-Seq: chromatin immunoprecipitation followed by massively parallel sequencing; EST: expressed sequence tag; FISH: fluorescence *in situ* hybridization; H3K27Me3: histone H3 tri-methylated on lysine 27; H3K4Me3: histone H3 tri-methylated on lysine 4; H3K9: histone H3 lysine 9; H3K9Ac: histone H3 acetylated on lysine 9; H3K9Me3: histone H3 tri-methylated on lysine 9; NCBI SRA: Sequence Read Archive at the National Center for Biotechnology Information; PCR: polymerase chain reaction; qPCR: quantitative PCR; RNAi: RNA interference; TSA: trichostatin A.

### Acknowledgements

The authors are grateful to the Plant Genome and Development Laboratory (UMR5096) of the University of Perpignan for access to their fluorescence microscope. Anne Rognon, Bernard Dejean and Kristina Smith provided important support. The work received financial support from the CNRS (PostDoc fellowship to CC) and the programs 'Schistophepigen' and 'Monogamix' from the French National Agency for Research (ANR).

### Author details

<sup>1</sup>Université de Perpignan Via Domitia, CNRS, UMR 5244 Ecologie et Evolution des Interactions (2EI), 52 Avenue Paul Alduy, 66860 Perpignan Cedex, France. <sup>2</sup>Department of Biochemistry and Biophysics, ALS 2011, Oregon State University, Corvallis, OR 97331-7305, USA. <sup>3</sup>CEBio - Centro de Excelência em Bioinformática, Rua Araguari, 741/301 - Barro Preto - BH/MG - CEP 30190-110, Brazil.

### Authors' contributions

JMUL did most of the experimental work and wrote the manuscript, CC conducted ChIP experiments, JB performed TSA treatment and edited the manuscript, MF performed ChIP-Seq, JP and DC did PCR and qPCR confirmation of W- and Z-specific sequences, CP contributed to data analysis, AZN did part of the bioinformatics work, and CG designed the experiment, performed FISH experiments, edited the manuscript and analyzed the massive sequencing data.

### Competing interests

The authors declare that they have no competing interests.

Received: 23 November 2011 Revised: 13 February 2012  
Accepted: 29 February 2012 Published: 29 February 2012

### References

- Charlesworth D, Charlesworth B, Marais G: Steps in the evolution of heteromorphic sex chromosomes. *Heredity* 2005, **95**:118-128.
- Rice WR: Sexually antagonistic male adaptation triggered by experimental arrest of female evolution. *Nature* 1996, **381**:232-234.
- Jablonka E: The evolution of the peculiarities of mammalian sex chromosomes: an epigenetic view. *BioEssays* 2004, **26**:1327-1332.
- Nicolas M, Marais G, Hykelova V, Janousek B, Laporte V, Vyskot B, Mouchiroud D, Negrutiu I, Charlesworth D, Moneger F: A gradual process of recombination restriction in the evolutionary history of the sex chromosomes in dioecious plants. *PLoS Biol* 2005, **3**:e4.
- Jablonka E, Lamb MJ: The evolution of heteromorphic sex chromosomes. *Biol Rev Camb Philos Soc* 1990, **65**:249-276.
- Zhang YE, Vibranovski MD, Landback P, Marais GA, Long M: Chromosomal redistribution of male-biased genes in mammalian evolution with two bursts of gene gain on the X chromosome. *PLoS Biol* 2010, **8**:e1000494.
- Steinemann S, Steinemann M: Y chromosomes: born to be destroyed. *BioEssays* 2005, **27**:1076-1083.
- Gorelick R: Evolution of dioecy and sex chromosomes via methylation driving Muller's ratchet. *Biol J Linn Soc* 2003, **80**:353-368.
- Griffin DK, Harvey SC, Campos-Ramos R, Ayling LJ, Bromage NR, Masabanda JS, Penman DJ: Early origins of the X and Y chromosomes: lessons from tilapia. *Cytogenet Genome Res* 2002, **99**:157-163.
- Fraser JA, Heitman J: Chromosomal sex-determining regions in animals, plants and fungi. *Curr Opin Genet Dev* 2005, **15**:645-651.
- Handley LJ, Ceplitis H, Ellegren H: Evolutionary strata on the chicken Z chromosome: implications for sex chromosome evolution. *Genetics* 2004, **167**:367-376.
- Just W, Baumstark A, Suss A, Graphodatsky A, Rens W, Schafer N, Bakloushinskaya I, Hameister H, Vogel W: *Ellobius lutescens*: sex determination and sex chromosome. *Sex Dev* 2007, **1**:211-221.
- King CH: Parasites and poverty: the case of schistosomiasis. *Acta Trop* 2010, **113**:95-104.
- Liberatos JD, Short RB: Identification of sex of schistosome larval stages. *J Parasitol* 1983, **69**:1084-1089.
- Grossman AI, Short RB, Cain GD: Karyotype evolution and sex chromosome differentiation in Schistosomes (Trematoda, Schistosomatidae). *Chromosoma* 1981, **84**:413-430.
- Loker ES, Brant SV: Diversification, dioecy and dimorphism in schistosomes. *Trends Parasitol* 2006, **22**:521-528.
- Despres L, Maurice S: The evolution of dimorphism and separate sexes in schistosomes. *Proc R Soc Lond B* 1995, **262**:175-180.
- Basch PF: Why do schistosomes have separate sexes? *Parasitol Today* 1990, **6**:160-163.

19. Read AF, Nee S: **Male schistosomes: more than just muscle?**. *Parasitol Today* 1990, **6**:297, author reply 297.
20. Berriman M, Haas BJ, LoVerde PT, Wilson RA, Dillon GP, Cerqueira GC, Mashiyama ST, Al-Lazikani B, Andrade LF, Ashton PD, Aslett MA, Bartholomeu DC, Blandin G, Caffrey CR, Coghlan A, Coulson R, Day TA, Delcher A, DeMarco R, Djikeng A, Eyre T, Gamble JA, Ghedin E, Gu Y, Hertz-Fowler C, Hirai H, Hirai Y, Houston R, Ivens A, Johnston DA, et al: **The genome of the blood fluke *Schistosoma mansoni***. *Nature* 2009, **460**:352-358.
21. Protasio AV, Tsai IJ, Babbage A, Nichol S, Hunt M, Aslett MA, De Silva N, Velarde GS, Anderson TJ, Clark RC, Davidson C, Dillon GP, Holroyd NE, Loverde PT, Lloyd C, McQuillan J, Oliveira G, Otto TD, Parker-Manuel SJ, Quail MA, Wilson RA, Zerlotini A, Dunne DW, Berriman M: **A systematically improved high quality genome and transcriptome of the human blood fluke *Schistosoma mansoni***. *PLoS Negl Trop Dis* 2012, **6**:e1455.
22. Criscione CD, Valentim CL, Hirai H, LoVerde PT, Anderson TJ: **Genomic linkage map of the human blood fluke *Schistosoma mansoni***. *Genome Biol* 2009, **10**:R71.
23. Zhang W, Wang X, Yu Q, Ming R, Jiang J: **DNA methylation and heterochromatinization in the male-specific region of the primitive Y chromosome of papaya**. *Genome Res* 2008, **18**:1938-1943.
24. Portela J, Grunau C, Cosseau C, Beltran S, Dantec C, Parrinello H, Boissier J: **Whole-genome in-silico subtractive hybridization (WISH) - using massive sequencing for the identification of unique and repetitive sex-specific sequences: the example of *Schistosoma mansoni***. *BMC Genomics* 2010, **11**:387.
25. DeMarco R, Kowaltowski AT, Machado AA, Soares MB, Gargioni C, Kawano T, Rodrigues V, Madeira AM, Wilson RA, Menck CF, Setubal JC, Dias-Neto E, Leite LC, Verjovsky-Almeida S: **Saci-1, -2, and -3 and Perere, four novel retrotransposons with high transcriptional activities from the human parasite *Schistosoma mansoni***. *J Virol* 2004, **78**:2967-2978.
26. Webster P, Mansour TE, Bieber D: **Isolation of a female-specific, highly repeated *Schistosoma mansoni* DNA probe and its use in an assay of cercarial sex**. *Mol Biochem Parasitol* 1989, **36**:217-222.
27. Hirai H, Tanaka M, LoVerde PT: ***Schistosoma mansoni*: chromosomal localization of female-specific genes and a female-specific DNA element**. *Exp Parasitol* 1993, **76**:175-181.
28. Greveling CG: **Genomic instability in *Schistosoma mansoni***. *Mol Biochem Parasitol* 1999, **101**:207-216.
29. Fitzpatrick JM, Protasio AV, McArdle AJ, Williams GA, Johnston DA, Hoffmann KF: **Use of genomic DNA as an indirect reference for identifying gender-associated transcripts in morphologically identical, but chromosomally distinct, *Schistosoma mansoni* cercariae**. *PLoS Negl Trop Dis* 2008, **2**:e323.
30. Reinhart BJ, Bartel DP: **Small RNAs correspond to centromere heterochromatic repeats**. *Science* 2002, **297**:1831.
31. Felsenstein J: **The evolutionary advantage of recombination**. *Genetics* 1974, **78**:737-756.
32. Fukagawa T, Nogami M, Yoshikawa M, Ikeno M, Okazaki T, Takami Y, Nakayama T, Oshimura M: **Dicer is essential for formation of the heterochromatin structure in vertebrate cells**. *Nat Cell Biol* 2004, **6**:784-791.
33. Grossman AI, McKenzie R, Cain GD: **Sex heterochromatin in *Schistosoma mansoni***. *J Parasitol* 1980, **66**:368-370.
34. Tone M, Nakano N, Takao E, Narisawa S, Mizuno S: **Demonstration of W chromosome-specific repetitive DNA sequences in the domestic fowl, *Gallus g. domesticus***. *Chromosoma* 1982, **86**:551-569.
35. Griffiths R, Holland PW: **A novel avian W chromosome DNA repeat sequence in the lesser black-backed gull (*Larus fuscus*)**. *Chromosoma* 1990, **99**:243-250.
36. Kejnovsky E, Hobza R, Cermak T, Kubat Z, Vyskot B: **The role of repetitive DNA in structure and evolution of sex chromosomes in plants**. *Heredity* 2009, **102**:533-541.
37. Shapiro JA, von Sternberg R: **Why repetitive DNA is essential to genome function**. *Biol Rev Camb Philos Soc* 2005, **80**:227-250.
38. Ugarkovic D: **Functional elements residing within satellite DNAs**. *EMBO Rep* 2005, **6**:1035-1039.
39. Krautz-Peterson G, Skelly PJ: ***Schistosoma mansoni*: the dicer gene and its expression**. *Exp Parasitol* 2008, **118**:122-128.
40. Luo R, Xue X, Wang Z, Sun J, Zou Y, Pan W: **Analysis and characterization of the genes encoding the Dicer and Argonaute proteins of *Schistosoma japonicum***. *Parasit Vectors* 2010, **3**:90.
41. Volpe TA, Kidner C, Hall IM, Teng G, Grewal SI, Martienssen RA: **Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi**. *Science* 2002, **297**:1833-1837.
42. Verdel A, Jia S, Gerber S, Sugiyama T, Gygi S, Grewal SI, Moazed D: **RNAi-mediated targeting of heterochromatin by the RITS complex**. *Science* 2004, **303**:672-676.
43. Kanellopoulou C, Muljo SA, Kung AL, Ganesan S, Drapkin R, Jenuwein T, Livingston DM, Rajewsky K: **Dicer-deficient mouse embryonic stem cells are defective in differentiation and centromeric silencing**. *Genes Dev* 2005, **19**:489-501.
44. Pelisson A, Sarot E, Payen-Groschene G, Bucheton A: **A novel repeat-associated small interfering RNA-mediated silencing pathway downregulates complementary sense gypsy transcripts in somatic cells of the *Drosophila* ovary**. *J Virol* 2007, **81**:1951-1960.
45. Bergero R, Charlesworth D: **The evolution of restricted recombination in sex chromosomes**. *Trends Ecol Evol* 2009, **24**:94-102.
46. Jablonka E, Lamb MJ: **Meiotic pairing constraints and the activity of sex chromosomes**. *J Theor Biol* 1988, **133**:23-36.
47. Theron A, Pages JR, Rognon A: ***Schistosoma mansoni*: distribution patterns of miracidia among *Biomphalaria glabrata* snail as related to host susceptibility and sporocyst regulatory processes**. *Exp Parasitol* 1997, **85**:1-9.
48. Boissier J, Rivera ER, Mone H: **Altered behavior of the snail *Biomphalaria glabrata* as a result of infection with *Schistosoma mansoni***. *J Parasitol* 2003, **89**:429-433.
49. Cosseau C, Azzi A, Rognon A, Boissier J, Gourbière S, Roger E, Mitta G, Grunau C: **Epigenetic and phenotypic variability in populations of *Schistosoma mansoni*-a possible kick-off for adaptive host/parasite evolution**. *Oikos* 2010, **119**:669-678.
50. Li R, Yu C, Li Y, Lam TW, Yiu SM, Kristiansen K, Wang J: **SOAP2: an improved ultrafast tool for short read alignment**. *Bioinformatics* 2009, **25**:1966-1967.
51. Censor. [<http://www.girinst.org/censor/>].
52. Abrusan G, Grundmann N, DeMester L, Makalowski W: **TEclass - a tool for automated classification of unknown eukaryotic transposable elements**. *Bioinformatics* 2009, **25**:1329-1330.
53. BLAST. [[http://blast.ncbi.nlm.nih.gov/Blast.cgi?CMD=Web&PAGE\\_TYPE=BlastHome](http://blast.ncbi.nlm.nih.gov/Blast.cgi?CMD=Web&PAGE_TYPE=BlastHome)].
54. Zerlotini A, Heiges M, Wang H, Moraes RL, Dominini AJ, Ruiz JC, Kissinger JC, Oliveira G: **SchistoDB: a *Schistosoma mansoni* genome resource**. *Nucleic Acids Res* 2009, **37**:D579-582.
55. Hirai H, LoVerde PT: **FISH techniques for constructing physical maps on schistosome chromosomes**. *Parasitol Today* 1995, **11**:310-314.
56. Primer3. [<http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi>].
57. Cosseau C, Azzi A, Smith K, Freitag M, Mitta G, Grunau C: **Native chromatin immunoprecipitation (N-ChIP) and ChIP-Seq of *Schistosoma mansoni*: critical experimental parameters**. *Mol Biochem Parasitol* 2009, **166**:70-76.
58. Chen J, Yang Y, Guo S, Peng J, Liu Z, Li J, Lin J, Cheng G: **Molecular cloning and expression profiles of Argonaute proteins in *Schistosoma japonicum***. *Parasitol Res* 2010, **107**:889-899.
59. Frescas D, Guardavaccaro D, Kuchay SM, Kato H, Poleshko A, Basrur V, Elenitoba-Johnson KS, Katz RA, Pagano M: **KDM2A represses transcription of centromeric satellite repeats and maintains the heterochromatic state**. *Cell Cycle* 2008, **7**:3539-3547.
60. Brunmeir R, Lagger S, Simboeck E, Sawicka A, Egger G, Hagemkruys A, Zhang Y, Matthias P, Miller WJ, Seiser C: **Epigenetic regulation of a murine retrotransposon by a dual histone modification mark**. *PLoS Genet* 2010, **6**:e1000927.
61. Hirai H, Spotila LD, LoVerde PT: ***Schistosoma mansoni*: chromosomal localization of DNA repeat elements by in situ hybridization using biotinylated DNA probes**. *Exp Parasitol* 1989, **69**:175-188.
62. Drew AC, Brindley PJ: **Female-specific sequences isolated from *Schistosoma mansoni* by representational difference analysis**. *Mol Biochem Parasitol* 1995, **71**:173-181.

doi:10.1186/gb-2012-13-2-r14

Cite this article as: Lepesant et al.: Chromatin structural changes around satellite repeats on the female sex chromosome in *Schistosoma mansoni* and their possible role in sex chromosome emergence. *Genome Biology* 2012 **13**:R14.

### ANNEXE 3

Publication: **Lepesant J.M.J.**, Roquis D., Emans R., Cosseau C., Arancibia N., Mitta G. & Grunau C. (2012). Combination of de-novo assembly of massive sequencing reads with classical repeat prediction improves identification of repetitive sequences in *Schistosoma mansoni*. (IF 2.122)



Contents lists available at SciVerse ScienceDirect

# Experimental Parasitology

journal homepage: [www.elsevier.com/locate/yexpr](http://www.elsevier.com/locate/yexpr)

## Research Brief

### Combination of *de novo* assembly of massive sequencing reads with classical repeat prediction improves identification of repetitive sequences in *Schistosoma mansoni*

Julie M.J. Lepesant, David Roquis, Rémi Emans, Céline Cosseau, Nathalie Arancibia, Guillaume Mitta, Christoph Grunau \*

Université de Perpignan Via Domitia, Perpignan F-66860, France  
CNRS, UMR 5244, Ecologie et Evolution des Interactions (2EI), Perpignan F-66860, France

#### ARTICLE INFO

##### Article history:

Received 24 November 2011  
Received in revised form 7 February 2012  
Accepted 8 February 2012  
Available online 21 February 2012

##### Keywords:

*Schistosoma mansoni*  
Repetitive sequences  
Massive sequencing  
*de novo* Assembly

#### ABSTRACT

The genome of the parasitic platyhelminth *Schistosoma mansoni* is composed of approximately 40% of repetitive sequences of which roughly 20% correspond to transposable elements. When the genome sequence became available, conventional repeat prediction programs were used to find these repeats, but only a fraction could be identified. To exhaustively characterize the repeats we applied a new massive sequencing based strategy: we re-sequenced the genome by next generation sequencing, aligned the sequencing reads to the genome and assembled all multiple-hit reads into contigs corresponding to the repetitive part of the genome. We present here, for the first time, this *de novo* repeat assembly strategy and we confirm that such assembly is feasible. We identified and annotated 4,143 new repeats in the *S. mansoni* genome. At least one third of the repeats are transcribed. This strategy allowed us also to identify 14 new microsatellite markers, which can be used for pedigree studies. Annotations and the combined (previously known and new) 5,420 repeat sequences (corresponding to 47% of the genome) are available for download (<http://methdb.univ-perp.fr/downloads/>).

© 2012 Elsevier Inc. All rights reserved.

Despite their abundance, repetitive sequences of the genome are often considered as “junk”, “selfish”, or “parasitic” DNA that is tolerated by the genome but has no biological or evolutionary functions. This view is about to change. In 2005, Shapiro and von Sternberg discussed the importance of the repetitive sequences for the establishment of the frontiers between heterochromatin and euchromatin, and their influence on homologous and nonhomologous recombination (Shapiro and von Sternberg, 2005). Depending on their position, repetitive sequences can play a part in activation or repression of gene transcription (Goodier and Kazazian, 2008). Some repeats have important structural functions such as telomeric repeats or the long satellite blocks that make up the centromeres of mammals and insects (Kejnovsky et al., 2009). In some cases, transcription of repeats and subsequent processing into small RNA was described. These transcripts are involved in heterochromatinization (small heterochromatin inducing RNA – shiRNA) (Reinhart and Bartel, 2002). Transposable elements, constituting a substantial share of the repetitive DNA, are known to have an impact on the genome evolution. Some were even selected

to play a precise role in the cell (“domesticated repeats”) (Shapiro and von Sternberg, 2005). Taken together, repetitive elements can no longer be considered as a side-aspect of the genome and deserve a deeper investigation.

*Schistosoma mansoni* is a parasitic platyhelminth responsible for intestinal schistosomiasis. This parasitic human disease ranks second only to malaria in terms of parasite-induced human morbidity and mortality, with more than 200 million infected people. The economic burden caused by the disease is tremendous as, for example, people disabled by the disease have limited job performances and are less likely to contribute to the local development. It was estimated that schistosomiasis burden represents 25–50 million disability-adjusted life-years (DALY) (King, 2010). The life cycle of the parasite is characterized by passage through two obligatory hosts: a fresh-water snail (*Biomphalaria* species, depending on the geographical location) as intermediate host, and humans or rodents as the final host. Miracidia infect the snail and transform into primary and secondary sporocysts, from which cercariae, capable of infecting the human host, are released into the water. Based on RepeatScout data, the genome of *S. mansoni* was thought to contain approximately 40% repetitive sequences (Berriman et al., 2009), of which roughly 20% correspond to transposable elements (Simpson et al., 1982). Over the last 30 years, roughly a dozen repetitive sequences have been identified by classical molecular

\* Corresponding author. Address: Université de Perpignan Via Domitia, UMR 5244 CNRS Ecologie et Evolution des Interactions (2EI), 52 Avenue Paul Alduy, 66860 Perpignan Cedex, France. Fax: +33 468662281.

E-mail addresses: [grunau@methdb.net](mailto:grunau@methdb.net), [christoph.grunau@univ-perp.fr](mailto:christoph.grunau@univ-perp.fr) (C. Grunau).

biology methods e.g. (Copeland et al., 2003, 2006). When the genome sequence became available, conventional repeat prediction programs were used to identify additional repetitive sequences. These 1,225 repeat sequences are available from the J. Craig Venter Insitute Institute ([ftp://ftp.tigr.org/pub/data/Eukaryotic\\_Projects/s\\_mansoni/preliminary\\_annotation/homology\\_evidence/sma1-repeats.gz](ftp://ftp.tigr.org/pub/data/Eukaryotic_Projects/s_mansoni/preliminary_annotation/homology_evidence/sma1-repeats.gz)) and their naming convention suggests that RepeatScout was used for prediction. Fifty-five repeats were available in GenBank. At this point, when we had re-sequenced the genome by massively sequencing (next-generation sequencing, NGS) our mapping results suggested that a large number of additional repeats must exist in the *S. mansoni* genome. We reasoned that by combining alignment information to identify reads that correspond to multiple locations on the genome and short-read assembly it should be possible to identify all repeats in a genome without *a priori* information.

A Brazilian strain (Bre) and a Guadeloupean strain (GH2), maintained respectively in their sympatric *B. glabrata* strain, were used in this study. Miracidia and eight-week adult worms were recovered as described before (Theron et al., 1997) and kept at  $-80^{\circ}\text{C}$ . The French Ministère de l'Agriculture et de la Pêche and French Ministère de l'Education Nationale de la Recherche et de la Technologie provided permit A 66040 to our laboratory for experiments on animals and certificate for animal experimentation (authorization 007083, decree 87–848) for the experimenters. Housing, breeding and animal care followed the national ethical requirements. Genomic DNA was extracted from 10 adult couples using the phenol–chloroform protocol. For total RNA purification, three independent preparations of each larvae and adults were used. For the larval stages, RNA was extracted from 10,000 miracidia using 500  $\mu\text{l}$  Trizol (Invitrogen™). Ten adult couples were solubilized in 500  $\mu\text{l}$  Trizol with a MagNA Lysor and Green beads (Roche). RNA was treated with DNA-free (Ambion #cat: AM1907) for 45 min at  $37^{\circ}\text{C}$ , followed by inactivation of the enzyme using the inactivation reagent. PCR of 28S rDNA was used to test for genomic DNA contaminations. First strand cDNA was synthesized using 20 ng of the total RNA preparation, in a final volume of 20  $\mu\text{l}$  with 200 U of RevertAid (Fermentas, #cat: G2101). To assemble the repeat genome, we used a total of 38,004,342 36-bp single-end reads generated on a Genome Analyzer II (Illumina) according to the manufacturer's protocol, at the MGX and Oregon State University sequencing facilities. Sequences are available at the NCBI sequence read archive (study accession numbers SRA012151.6 and SRA043796.1). Reads were aligned to the reference genome v.3.1 with SOAP2/SOAPaligner (Li et al., 2009) evoking the  $-r 0$  (not repeats) and  $-u$  (write unmapped reads into a file) options. The rationale behind this approach was that in this case, SOAP would only align reads with a single occurrence in the genome. All other reads correspond either to unknown sequences or occurring more than once, i.e. are repetitive. The 12,535,613 unmapped sequence reads (33% of total) were then assembled using Velvet 0.7.01 (Zerbino, 2010) with the  $-\text{cov\_cutoff } 4$   $\text{min\_contig\_length } 100$  options resulting in 8,608 contigs. A long read assembler (Sequencher version 4.5 (Gene Codes)  $\text{min\_match} = 93\%$ ,  $\text{min\_overlap} = 60$  bp) was used to produce finally 8,594 contigs. Each repeat was assembled individually and therefore the assemblies may be composed of two or more distinct, but very similar repeats. First pass annotation of the 8,594 presumed repeat contigs was done with Blast (Altschul et al., 1990), Censor/Repbase (Kohany et al., 2006), TEclass (Abrusan et al., 2009) and Tandem Repeats Finder (Benson, 1999). Blast2GO v2.4.8 (Conesa et al., 2005) was used to carry out various types of BLAST searches (conditions in Supplementary Table 1 and results in Table 1). CENSOR (<http://www.girinst.org/censor/>) (Kohany et al., 2006) and the Repbase Update (<http://www.girinst.org/repbase/>) (Jurka et al., 2005) were applied to find sequences sharing similarity to known repeats. Parameters were: *Sequence source*: all; *Forced translated search*: no; *Search for identity*: no; *Mask simple*

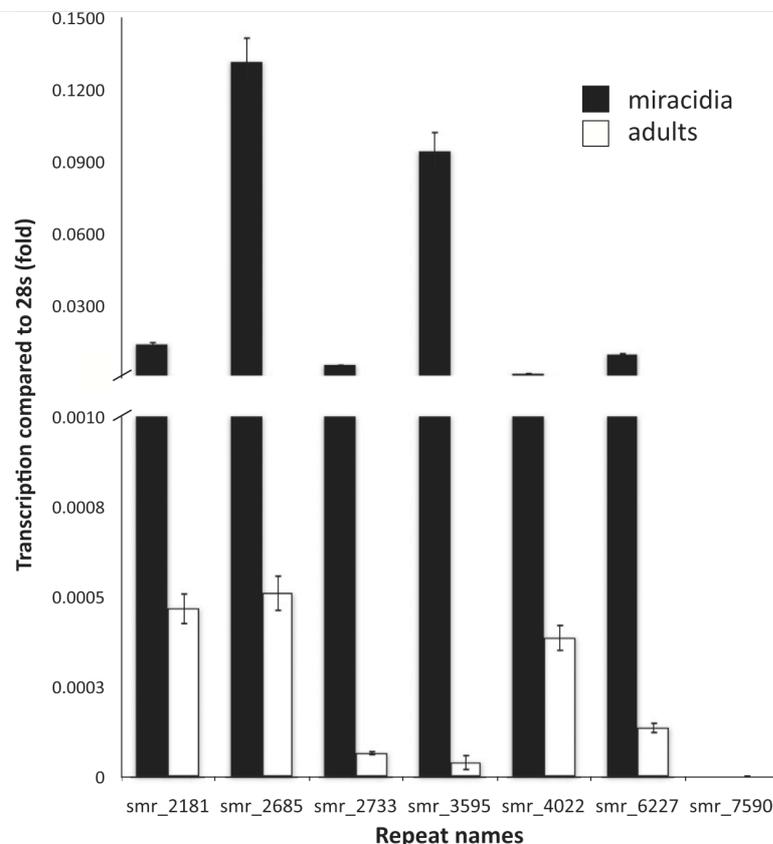
*repeats*: yes; *Mask pseudogenes*: yes. Results were evaluated according to the 80/80/80 principle (80% of identity on 80% of the sequence spanning a minimum of 80 bp) (Wicker et al., 2007). The web application TEclass (<http://www.compgen.uni-muenster.de/tools/teclass/>) (Abrusan et al., 2009) was used to predict potential transposable elements in a sequence with default parameters. Finally we used Tandem Repeats Finder (TRfinder, <http://tandem.bu.edu/trf/trf.html>) (Benson, 1999) to identify tandem repeats. Parameters were optimized for highest sensitivity and specificity (respectively 81% and 97%) using *in silico* generated training sequences: *Minimum alignment score*: 30; *alignment parameters*: 2–7–7. TRfinder was also used to find sequences with a period of 2, 3 or 4 bp that could serve as new microsatellite markers. Candidates were verified to have only one occurrence in the genome, to be polymorphic (by comparison with trace files used for the genome assembly) and it was checked if they were located in a gene.

For confirmation of *in silico* results, PCRs were carried out in a final volume of 25  $\mu\text{l}$  containing 0.2  $\mu\text{mol}$  of each oligonucleotide primer (Supplementary Table 2), 0.2 mmol of each dNTP (Promega), 1.25 U of GoTaq polymerase (Promega, #cat: M3175) used with the recommended buffer and completed to the final volume with DNase-free water ( $95^{\circ}\text{C}$  for 10 min followed by 35 cycles at  $95^{\circ}\text{C}$  for 30 s,  $53^{\circ}\text{C}$  for 30 s,  $72^{\circ}\text{C}$  for 1 min and a final extension at  $72^{\circ}\text{C}$  for 10 min). The PCR products were separated by electrophoresis through a 2% TBE agarose gel. Real-time quantitative PCR analyses were performed using the LightCycler 2.0 system (Roche Applied Science) and LightCycler Fast-start DNA Master SYBR Green I kit (Roche Applied Science) with 2.5  $\mu\text{l}$  of cDNA in a final volume of 10  $\mu\text{l}$  (3 mM  $\text{MgCl}_2$ , 0.5  $\mu\text{M}$  of each primer, 5  $\mu\text{l}$  of master mix). The primers were designed with the LightCycler Probe design software or the Primer3Plus web based interface (<http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi>). The following protocol was used:  $95^{\circ}\text{C}$  for 10 min; 40 cycles:  $95^{\circ}\text{C}$  for 10 s,  $60^{\circ}\text{C}$  for 5 s,  $72^{\circ}\text{C}$  for 16 s; melting curve,  $60$ – $95^{\circ}\text{C}$  with a heating rate of  $0.1^{\circ}\text{C/s}$  and continuous fluorescence measurement, and a cooling step to  $40^{\circ}\text{C}$ . For each reaction, the crossing point (Ct) was determined with the “second derivative method” of the LightCycler Software 3.3. PCR reactions were done in quadruplicate and the mean value of Ct was calculated. 28S rRNA was used as an internal control and the amplification of a unique band was verified by electrophoresis through 2% agarose gels for each qPCR product. Primer sequences and expected PCR product sizes are listed in Supplementary Tables 2 and 3. For all qPCR, efficiency was at least 1.95.

Our assembly of SOAP-sorted massive sequencing reads delivered 8594 contigs. Contigs smr\_2181, smr\_2685, smr\_2733, smr\_3000, smr\_3595, smr\_3826, smr\_4022, and smr\_6227 – that we tested by PCR – showed a band at the predicted size indicating that assembly is correct in most cases. For two contigs (smr\_3000 and smr\_3826) a supplementary band was present, with a molecular weight twice as high as the major band, suggesting repetition in that these fragments correspond to an amplicon of 2 tandem repeats in the genome. Among the 8594, we clearly identified 6,531 (76%) as repeats via *in silico* analysis using the following criteria: two or more occurrences in the reference genome and no Blast annotation (against the nr database from NCBI) related to a known gene or protein (with the exception of proteins typical of transposable elements, such as transposase, reverse transcriptase or GAG polyprotein), using an e-value cut-off of  $1.0\text{E}^{-30}$ . Using the information obtained from the Blast done on the nr database, we identified 306 contigs related to genes or gene families, which include 40 mitochondrial genes. A Blast search against the reference genome of *S. mansoni*, which allowed us to count the number of occurrences of each repeat, showed that 1,332 sequences (other than the ones identified as genes) were unique, and 230 were absent from the reference genome. All Blast conditions are listed in

**Supplementary Table 1.** At this point of our analysis, a total of 10 sequences are suspected to be possible contamination, as they are neither present in the reference genome, nor in the trace files, but match perfectly sequences from other organisms. Four sequences correspond to rodents and six to small freshwater organisms such as *Hydra magnipapillata*. It is likely that the sequences are due to contaminations from host tissue and spring-water used for parasite culture. Out of the 6,531 repeated sequences, TRfinder detected 516 containing tandem repeats. Two thousand four hundred and forty-five repeated sequences matched fully or partially to already known repeats, *i.e.* those present in the RepeatScout generated database of repeat predictions. In conclusion, assembly of repeats is possible by our new approach and so far, 4,143 unknown repeats have been identified. Based on the combined results of Blast and Censor we classified the 6,531 repeats into 9 groups, and into 5 with TEClass. This categorization is based on the hierarchy of classes of repeats suggested by the Censor results (list of classes and subclasses available on the GIRI website, <http://www.girinst.org/censor/help.html>). The majority of repeat annotations belongs to class I and class II (retro) transposons (Table 1b). For 1,921 of the 6,531 repeats (29%) we found ESTs through Blast searches (Supplementary Table 1), indicating that a large part is transcribed. Blast searches in stage specific ESTs revealed a homogeneous distribution among the life cycle stages (data not shown). For seven repeats with 0–52 EST hits, we verified transcription in miracidia and adults. Six (smr\_2181, smr\_2685, smr\_2733, smr\_3595, smr\_4022 and smr\_6227) showed low, but significant above background transcription, while for a single repeat (smr\_7590) no transcription was found (Fig. 1). There was no

correlation between the number of EST hits and transcription level measured by qPCR. We used TRfinder, Blast searches against the trace file database and Blast searches against the *S. mansoni* genome followed by visual inspection to identify 14 new polymorphic microsatellite markers. Five of them are located in predicted gene coding regions and the remaining nine are probably neutral markers (Supplementary Table 4). To generate a combined database we used the abovementioned 1,225 predicted repeats. We then analyzed the 55 repeats available in GenBank and identified three duplicates: Sm\_SR-AB2 (AF025674.1), Sm\_SR1\_pol\_3 (U66331.1) and Sm\_Sinbad\_iS4-T (AY965073.1). For Sm\_salmonid (AY834402.1) we have shown previously (Grunau and Boissier, 2010) that it is not present in *S. mansoni*. Consequently, we removed these 4 repeats from the database, and added the remaining 51 repeats sequences available in GenBank and a tandem repeat earlier identified in our laboratory (TandemRepeat\_266) to the combined database (1,225 predicted + 55 GenBank – 3 duplicates – 1 wrong + 1 previously identified = 1277). In total, our new *Schistosoma mansoni* repeat database contains 5,420 sequences. The workflow and a summary of results are shown in Fig. 2. This repeat database was used for the annotation of assembly 3.1 of the *S. mansoni* genome. We employed RepeatMasker (Smit et al., 1996) evoking the -cutoff 250 -norna options. The repeatmasking was done with the original Sanger/TIGR library and with our combined library. Repeatmasker identified repeats in 47.40% of the genome. Fragmentation and overlapping repeatmasking were considerably lower with our new library resulting in annotation of 623,983 repeats compared to 881,451 obtained with the previous Sanger/TIGR data. Examples that show the improved quality of annotation



**Fig. 1.** Transcription level of arbitrarily chosen repeats in miracidia and adults of *S. mansoni* measured by RT-qPCR (4 replicates). Transcription is expressed in fold of 28s rRNA that served as reference.

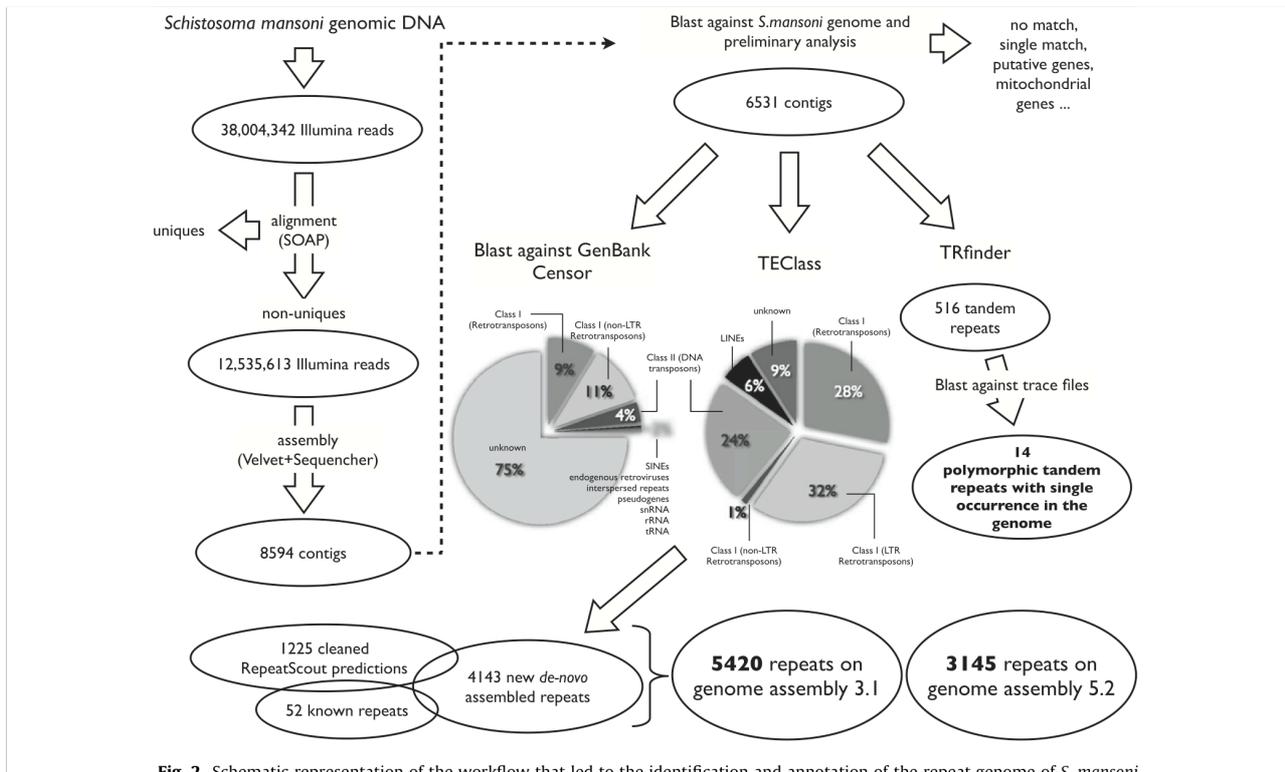


Fig. 2. Schematic representation of the workflow that led to the identification and annotation of the repeat genome of *S. mansoni*.

are shown in Supplementary Figure 1. During the preparation of the manuscript, version 5.2 of the *S. mansoni* genome became available in which scaffold redundancy was reduced (Protasio et al., 2012). We reasoned that some of our new repeats might be absent from the new assembly and repeated the blast search with an *e*-value of  $1.0E^{-30}$ . We removed all repeats that did not match to the genome 5.2 or that matched only once resulting in

3145 different repeat sequences. Repeatmasking showed that 47.73% is repetitive, *i.e.* essentially the same value as for the 3.1 assembly. Repeat annotations were converted into GFF and loaded on an in-house genome browser for visualization. The repeats are available as fasta files and annotations as GFF files (<http://meth-db.univ-perp.fr/downloads/>) (See Table 1a).

We describe here – to our knowledge for the first time – a *de novo* assembly strategy for repetitive sequences making use of a filter option in short read alignment programs such as SOAP. Short reads that match to more than one region of the reference genome can be handled in two different ways by these programs. Either, they are randomly positioned to the multiple matching loci, or they are not at all aligned. We made use of this latter option and employed the alignment program as a filter for repetitive sequences. A potential caveat of this method is that unique sequences that for some reasons are not present in the reference genome (or not aligned to it) will also initially be included in the sequence bin that will serve as the input for the repeat assembly process. These sequences must be eliminated after assembly by alignment to the genome. The rationale behind this approach is that true repeats will be detected – by their very nature – more than once in the genome. All other sequences must be removed. This strategy allowed us to identify 4,143 new repeats in the *S. mansoni* genome in addition to the 1,225 previously predicted and the 52 experimentally found repeat sequences available in GenBank or through our own work. The strategy can naturally also be used for other species for which a reference sequence and massive sequencing reads are available. We then annotated the repeats with conventional annotation programs in order to classify them. It is interesting to note that a substantial proportion of the repeats (at least 30% but probably more) is transcribed. It is tempting to speculate that this transcription has a biological function as observed in other species. Further work will be necessary to explore this option in schistosomes. Finally, our approach allowed us to identify 14

Table 1a  
Summary of combined Blast and Censor results.

Repeat class	Number	%
Class I (Retrotransposons)	565	8.6
Class I (non-LTR Retrotransposons)	720	11
Class II (DNA transposons)	239	3.7
SINES	31	0.5
Endogenous retroviruses	18	0.3
Interspersed repeats	21	0.3
Pseudogenes, snRNA, rRNA, tRNA	38	0.6
Unknown	4899	75
Total	6531	

Table 1b  
Summary of TEClass results.

Repeat class	Number	%
Class I (Retrotransposons)	1841	28
Class I (LTR Retrotransposons)	2067	32
Class I (non-LTR Retrotransposons)	70	1
Class II (DNA transposons)	1563	24
LINEs	398	6
Unknown	592	9
Total	6531	

new polymorphic microsatellite markers, of which 9 are probably neutral markers. Microsatellite markers are popular tools in many fields of biology such as pedigree studies or epidemiology. While being of great interest, they are notoriously difficult to find, and during the last decade only 43 were discovered (Supplementary Table 4). Our new markers could bring this number to 57, providing a larger choice that will be useful in situations where the so far used markers are not sufficiently polymorphic.

### Acknowledgments

The authors are grateful to Michael Freitag (Oregon State University) for library construction and sequencing of data SRA043796.1, and to the staff of the MGX-Montpellier GenomiX (MGX) sequencing center for generation of data SRA012151.6. Patricia Ruy (SchistoDB, Centro de Pesquisas Rene Rachou, Belo Horizonte, Brazil) provided support for Blast against *S.mansoni* EST databases. This work received funding from the French National Agency for Research (ANR), project ANR-2010-BLAN-1720-01 (EPI-GEVOL). The funding source had no involvement in the study design and publication decision. The authors declare to have no conflict of interest.

### Appendix A. Supplementary data

Supplementary data associated with this article can be found in the online version, at [doi:10.1016/j.exppara.2012.02.010](https://doi.org/10.1016/j.exppara.2012.02.010).

### References

- Abrusan, G., Grundmann, N., DeMester, L., Makalowski, W., 2009. TEclass—a tool for automated classification of unknown eukaryotic transposable elements. *Bioinformatics* 25, 1329–1330.
- Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. *Journal of Molecular Biology* 215, 403–410.
- Benson, G., 1999. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Research* 27, 573–580.
- Berriman, M. et al., 2009. The genome of the blood fluke *Schistosoma mansoni*. *Nature* 460, 352–358.
- Conesa, A., Gotz, S., Garcia-Gomez, J.M., Terol, J., Talon, M., Robles, M., 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21, 3674–3676.
- Copeland, C.S., Brindley, P.J., Heyers, O., Michael, S.F., Johnston, D.A., Williams, D.L., Ivens, A.C., Kalinna, B.H., 2003. Boudicca, a retrovirus-like long terminal repeat retrotransposon from the genome of the human blood fluke *Schistosoma mansoni*. *Journal of Virology* 77, 6153–6166.
- Copeland, C.S., Lewis, F.A., Brindley, P.J., 2006. Identification of the Boudicca and Sinbad retrotransposons in the genome of the human blood fluke *Schistosoma haematobium*. *Memorias do Instituto Oswaldo Cruz* 101, 565–571.
- Goodier, J.L., Kazazian Jr, H.H., 2008. Retrotransposons revisited: the restraint and rehabilitation of parasites. *Cell* 135, 23–35.
- Grunau, C., Boissier, J., 2010. No evidence for lateral gene transfer between salmonids and schistosomes. *Nature Genetics* 42, 918–919.
- Jurka, J., Kapitonov, V.V., Pavlicek, A., Klonowski, P., Kohany, O., Walichewicz, J., 2005. Repbase update, a database of eukaryotic repetitive elements. *Cytogenetic and Genome Research* 110, 462–467.
- Kejnovsky, E., Hobza, R., Cermak, T., Kubat, Z., Vyskot, B., 2009. The role of repetitive DNA in structure and evolution of sex chromosomes in plants. *Heredity* 102, 533–541.
- King, C.H., 2010. Parasites and poverty: the case of schistosomiasis. *Acta Tropica* 113, 95–104.
- Kohany, O., Gentles, A.J., Hankus, L., Jurka, J., 2006. Annotation, submission and screening of repetitive elements in Repbase: RepbaseSubmitter and Censor. *BMC Bioinformatics* 7, 474.
- Li, R., Yu, C., Li, Y., Lam, T.W., Yiu, S.M., Kristiansen, K., Wang, J., 2009. SOAP2: an improved ultrafast tool for short read alignment. *Bioinformatics* 25, 1966–1967.
- Protasio, A.V. et al., 2012. A systematically improved high quality genome and transcriptome of the human blood fluke *Schistosoma mansoni*. *PLoS Neglected Tropical Diseases* 6, e1455.
- Reinhart, B.J., Bartel, D.P., 2002. Small RNAs correspond to centromere heterochromatic repeats. *Science* 297, 1831.
- Shapiro, J.A., von Sternberg, R., 2005. Why repetitive DNA is essential to genome function. *Biological Reviews of the Cambridge Philosophical Society* 80, 227–250.
- Simpson, A.J., Sher, A., McCutchan, T.F., 1982. The genome of *Schistosoma mansoni*: isolation of DNA, its size, bases and repetitive sequences. *Molecular and Biochemical Parasitology* 6, 125–137.
- Smit, A.F.A., Hubley, R., Green, P., 1996. RepeatMasker Open-3.0.
- Theron, A., Pages, J.R., Rognon, A., 1997. *Schistosoma mansoni*: distribution patterns of miracidia among *Biomphalaria glabrata* snail as related to host susceptibility and sporocyst regulatory processes. *Experimental Parasitology* 85, 1–9.
- Wicker, T. et al., 2007. A unified classification system for eukaryotic transposable elements. *Nature Reviews. Genetics* 8, 973–982.
- Zerbino, D.R., 2010. Using the Velvet de novo assembler for short-read sequencing technologies. *Curr Protoc Bioinformatics Chapter* 11, Unit 11.5.

## ANNEXE 4

Publication : **Lepesant J.M.J.**, Grunau C. & Cosseau C. (2011). Towards an understanding of the epigenetics of schistosomes: a comparative epigenomic study. *Memorias do Instituto Oswaldo Cruz*, 106, 823-830. (IF 2.058)

## Towards an understanding of the epigenetics of schistosomes: a comparative epigenomic study

Julie Mireille and Joé Lepesant, Christoph Grunau, Céline Cosseau/+

Laboratoire Ecologie et Evolution des Interactions, Université de Perpignan Via Domitia,  
52 Avenue Paul Alduy 66860, UMR 5244 CNRS Perpignan, France

*As in perhaps all eukaryotes, schistosomes use a supplementary information transmitting system, the epigenetic inheritance system, to shape genetic information and to produce different phenotypes. In contrast to other important parasites, the study of epigenetic phenomena in schistosomes is still in its infancy. Nevertheless, we are beginning to grasp what goes on behind the epigenetic scene in this parasite. We have developed techniques of native chromatin immunoprecipitation (N-ChIP) and associated the necessary bioinformatics tools that allow us to run genome-wide comparative chromatin studies on *Schistosoma mansoni* at different stages of its life cycle, on different strains and on different sexes. We present here an application of such an approach to study the genetic and epigenetic basis for a phenotypic trait, the compatibility of *S. mansoni* with its invertebrate host *Biomphalaria glabrata*. We have applied the ChIP procedure to two strains that are either compatible or incompatible with their intermediate host. The precipitated DNA was sequenced and aligned to a reference genome and this information was used to determine regions in which both strands differ in their genomic sequence and/or chromatin structure. This procedure allowed us to identify candidate genes that display either genetic or epigenetic difference between the two strains.*

Key words: epigenetic - ChIP - genetic - adaptation - *Schistosoma mansoni* - compatibility polymorphism

In eukaryotes, genotypes and epigenotypes are manifested in a complex nuclear structure called chromatin. This nucleoprotein structure contains histone and non-histone proteins that interact with the genomic DNA. Chromatin exists either as a relaxed structure that is permissive to gene expression, called euchromatin, or as a condensed structure that is typically silent, called heterochromatin (Luijsterburg et al. 2008). The genotype refers to the DNA sequence of an organism. “Epigenotype” refers to the sum of chemical modifications of DNA (e.g., DNA methylation), DNA-associated proteins (e.g., histones) that can affect chromatin structure and other mediators of epigenetic information such as non-coding RNA and gene location within the nucleus (Dillon 2008, Umlauf et al. 2008). Posttranslational modifications of histones (e.g., methylation, acetylation and phosphorylation) are widely studied and have been shown to play a role in chromatin compaction and control of gene transcription (Lee et al. 2010). Orthologues of all histones apart from MacroH2A are predicted in the current genome assembly of *Schistosoma mansoni* and our laboratory and others have confirmed the presence of histone H3 and H4 and the isoforms H3K4Me3, H3S10P (unpublished observations), H3K4Me2, H3K9Ac, H3K9Me3, H3K27Me1 and H4Ac (Souza et al. 2002, Caby & Pierce 2009, Cosseau

et al. 2009). An analysis of data available in SchistoDB (Zerlotini et al. 2009) indicates that genes for all canonical histone modifying enzymes can be found.

A major advancement in the study of histone modifications in *Schistosoma* is the chromatin immunoprecipitation (ChIP) assay. ChIP allows the analysis of protein-DNA interaction in vivo. The N-terminal ends of *Schistosoma* histones H3 and H4 are conserved with other model species and, therefore, commercial antibodies can be used to immunoprecipitate modified histones and identify DNA to which they are bound. Cross-linked and native ChIP (N-ChIP) have been successfully applied to schistosomes (Caby & Pierce 2009, Cosseau et al. 2009). Massive sequencing analysis of the immunoprecipitated DNA (ChIP-Seq) and alignment against the *S. mansoni* genome has afforded a whole genome approach to visualise enrichment of specific epigenetic marks at specific *Schistosoma* loci. ChIP-Seq analysis can be performed using different antibodies that recognise different epigenetic marks responsible for euchromatic or heterochromatic structures. This technique allows the comparison of chromatin structure between life cycle stages (Cosseau et al. 2009), male or female adults (unpublished observations) or different strains. This comparative epigenomic approach has highlighted differences in chromatin structure that may impact the physiology of *S. mansoni*, such as development, sexual dimorphism and adaptation to environmental change.

In this study we have used such a comparative approach on two strains of parasites that are either compatible (C) or incompatible (IC) with a mollusc originally isolated in Brazil.

Compatibility polymorphism is a phenomenon that is observed in the interaction between *Biomphalaria glabrata* and *S. mansoni*, where some snail/schistosome

Financial support: Schistophepigen (ANR-07-BLAN-0119-02) of the French National Agency for Research, CNRS through a post-doctoral fellowship

+ Corresponding author: celine.cosseau@univ-perp.fr

Received 18 April 2011

Accepted 30 May 2011

combinations are compatible and others are not (Theron & Coustau 2005). In the C interaction, the parasite manages to infect the host whereas in the IC interaction, the majority of parasites fail to infect the host. It has been hypothesised that the success or the failure of *B. glabrata*/*S. mansoni* infection relies on the matched or mismatched status of the host and parasite phenotypes. Genetic polymorphism and/or epigenetic distinction between the two strains may be responsible for this matched or mismatched status. Some molecular determinants of the compatibility polymorphism have already been described (Roger et al. 2008a, c).

In this study, we have mapped histone H3 acetylation on lysine 9 (H3K9Ac), which is a mark generally detected in active promoters (Lee et al. 2010), in both the C and IC strains of miracidia. We used an *in silico* comparison approach to identify genomic regions in which H3K9Ac enrichment differs between both strains. This method, termed ChIPDIFF, combined with quantitative polymerase chain reaction (qPCR) allowed us to highlight loci that display either genetic polymorphism or differential acetylation between the two strains.

#### MATERIALS AND METHODS

**Parasite strains and cell culture** - A C strain (Brazilian strain) and an IC strain (Guadelupian strain) toward the same mollusc (*B. glabrata* isolated from Brazil) were used in this study. Each strain was maintained in its sympatric *B. glabrata* strain and in hamsters (*Mesocricetus auratus*) as described previously (Dissous et al. 1981, Theron et al. 1997). Eggs were asexually recovered from 50-day infected hamster livers and miracidia were hatched from eggs as previously described (Theron et al. 1997). Miracidia were concentrated by sedimentation on ice for 15 min and stored at -80°C until ChIP experiments.

**Chromatin status analysis by ChIP** - N-ChIP was performed as previously described (Cosseau et al. 2009) ([methdb.univ-perp.fr/cgrunau/methods/native\\_chip\\_sm.html](http://methdb.univ-perp.fr/cgrunau/methods/native_chip_sm.html)). Briefly, chromatin was extracted from both C and IC miracidia and fragmented with micrococcal nuclease. An antibody against modified histones was used to precipitate the fragmented chromatin (Table I). Immunoprecipitated DNA was purified using phenol-chloroform concentrated by precipitation and finally suspended in 80 µL Tris-ethylenediamine tetraacetic acid 10 mM.

**Generation of Illumina/Solexa Libraries/processing IG data** - Generation of Illumina/Solexa Libraries was performed as previously described (Cosseau et al. 2009). The immunoprecipitated DNA was sequenced at the Oregon State University Centre for Genome Research and Biocomputing core facility using an Illumina IG analyzer. Sequenced DNA was mapped to the *S. mansoni* reference genome (<ftp://ftp.sanger.ac.uk/pub/pathogens/Schistosoma/mansoni/genome/Assembly-v3.1/>) with SOAP2 software (Li et al. 2009) to generate a list of reads and their genomic locations. FindPeaks (Fejes et al. 2008) and CASHX (Fahlgren et al. 2009) were used for peak calling and generation of histograms in a genome browser (GBrowse) (Stein et al. 2002) as previously described (Cosseau et al. 2009).

**In silico comparative analysis** - To identify regions of statistically significant differences in H3K9Ac occurrence, we used the ChIPDiff software (Xu et al. 2008). The SOAP2 output was reformatted into the ChIPDiff input format using a custom Perl script. ChIPDiff parameters were as follows: maxIterationNum 500, minP 0.60, maxTrainingSeqNum 10000, minFoldChange 1.5, minRegionDist 500. The resulting output file was used to guide visual inspection on GBrowse. Regions of interest were selected if putative genes were in proximity.

**qPCR analysis** - The primers for qPCR were designed with the primer3 plus web-based interface ([bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi](http://bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi)). The genomic sequences from the reference genome (<ftp://ftp.sanger.ac.uk/pub/pathogens/Schistosoma/mansoni/genome/Assembly-v3.1/>) were used as a template. Primer sequences, scaffold position, expected product size and efficiency are listed in Table II. qPCR amplification was performed using 2.5 µL immunoprecipitated DNA in a final volume of 10 µL on a LightCycler® 480 Real-Time Instrument (3 µL H<sub>2</sub>O, 0.5 µM of each primer, 5 µL of master mix). The following protocol was used: denaturation, 95°C 10 min, amplification and quantification (repeat 40 times), 95°C for 10 s, 60°C for 10 s and 72°C for 20 s, melting curve, 65-97°C with a heating rate of 0.11°C/s and continuous fluorescence measurement. For each reaction, the crossing point cycle threshold (Ct) was determined using the “second derivative” method of the LightCycler® 480 Software release 1.5. Reactions were performed in duplicate and the mean Ct was calculated. DNA recovered in the immunoprecipitated fraction was quantified by calculat-

TABLE I  
Antibodies used for native chromatin immunoprecipitation (N-ChIP)

Antibody	Host	Product	Lot	Saturating quantity used for N-ChIP (µL)
H3K9Ac	Rabbit	Upstate, 07352	DAM1394804 for ChIP-seq DAM1576933 for ChIP-qPCR	8
H3K9Met3	Rabbit	Abcam, Ab8898	733951	4
H3K4Met3	Rabbit	Upstate, 04745	JBC1355340	4

qPCR: quantitative polymerase chain reaction.

TABLE II  
Primers used for quantitative polymerase chain reaction (qPCR) in this study

Target	Forward (5'-3')	Reverse (5'-3')	Efficiency (IC strain)	Efficiency (C strain)	Amplicon length (bp)
Smp_173370	AGGGGGTCCCTCCCTCGAAACA Smp_scaff000539:15007	CGCGCAAGTAGGCAGATAGCCA Smp_scaff000539:15143	1.95	ND	136
Smp_171100	CCTGGAATGGAGAGCTCGTATGACA Smp_scaff000421:12602	GGAGTGTGGATTGAACGGCTCG Smp_scaff000421:12724	1.91	ND	122
Smp_159800	CCGCCTGTGAACACCACTCAGC Smp_scaff000234:2330	TGCTGGACATTTAGTTACAGTTGCGA Smp_scaff000234:2352	2	ND	122
Smp_159810	GCTTAAGTGACACTGCACAGCGA Smp_scaff000234:63585	TGGTACTCTGGATAAGCTACCGTGC Smp_scaff000234:63731	1.94	ND	146
Smp_159820	AGGCAGCCAGTCACTTGAAAGC Smp_scaff000234:74614	CGAACAGGTTCCGATGAGGACGA Smp_scaff000234:74748	2	ND	134
Smp_159830	AGCAATTGTTGTGACCAATCGACA Smp_scaff000234:161592	TCAACTCTCCCAAGGCCGTTCA Smp_scaff000234:161723	1.92	ND	131
Smp_159840	AGCTTCGACGGCTTGTGGC Smp_scaff000234:217554	GCCCATGCACCAGAAGAGGCA Smp_scaff000234:217688	1.92	ND	134
Smp_scaff000131	TGCGCTTCCCTCATCGGCCTT Smp_scaff000131:58261	TCGTGTTCATAGGTGTGTAGTGTGGT Smp_scaff000131:58367	2	ND	106
Smp_scaff000154.2	CCGGGCGGTGACCGACCTAT Smp_scaff000154:384067	TGCACCTGCGAAAGGCTCGACA Smp_scaff000154:384167	2	2	100
Smp_scaff000154.3	TTCCACGGGATGGGGTCGCT Smp_scaff000154:532853	GGGCCCCATAATTAACCTCGCCTGT Smp_scaff000154:532960	1.94	1.95	107
Smp_scaff000154.5	ACGGTGGCTTGAGGGATCAAGAA Smp_scaff000154:519335	ACAGTACGGCGGTTCCCTGGC Smp_scaff000154:519465	ND	1.96	130

the region targeted is indicated by a gene name or a scaffold name when no hypothetical open reading frame has been annotated. Primer sequence is given from 5'-3' DNA strand. The 5' position on the scaffold is indicated. The qPCR efficiency has been determined when possible on the incompatible (IC) and compatible (C) strains. ND: not determined.

ing the percent input recovery (%IR) normalised with the %IR obtained at a reference locus ( $\alpha$ Tub or 28S) as previously described (Cosseau et al. 2009). Correct melting curves were checked using the Tm calling method of the LightCycler<sup>®</sup> 480 Software release 1.5. The amplification of a unique band was verified by 2% agarose gel electrophoresis for each qPCR product.

## RESULTS

*Identification of candidate genetic or epigenetic differences between isolates by ChIP-Seq* - Our N-ChIP procedure was applied to miracidia from the C and IC strains. We then performed genome-wide sequencing of the immunoprecipitated DNA using an Illumina/Solexa 1G Genome Analyzer and aligned it to the reference genome. There were matches in 3,131,677 and 3,367,822 reads for the C and IC strain, respectively, corresponding to roughly 120 Mb.

ChIPDiff was used to identify differences in aligned sequence between the two strains. Our parameters instructed the software to screen the immunoprecipitated aligned sequences in 1 kb blocks and highlight those displaying more than a 1.5 fold change in H3K9 acetylation. The C, IC and reference strains are three different *S. mansoni* isolates. If the genetic polymorphism between these strains is greater than the mismatches allowed for alignment (> 2 in 36 bp), then a difference in sequence alignment may occur at some loci. In this case, the ChIPDiff method allows for identification of genetic polymorphism between the C and IC strains. If there are differences in copy number between these strains (duplicated sequences), then an apparent difference in enrichment of sequences may be observed. If there are no genetic differences and no copy number effects, then the differences detected by ChIPDiff are due to differences in epigenetic marks. It should be pointed out that epigenetic differences may accumulate with genetic differences. Conversely, an epigenetic difference may be compensated by a genetic difference. In these cases, the ChIPDiff method will not be able to distinguish between these scenarios.

Using the ChIPDiff method, we identified 321 differences between the two strains. There were 128 “1 kb blocks” that were enriched for sequences from the IC strain and the remaining 193 “1 kb blocks” were enriched for sequences from the C strain (Fig. 1). We discarded the individual “1 kb blocks” and only further considered regions that contained clusters of “1 kb blocks” spanning several kb regions. We selected nine regions, some of which contain hypothetical annotated genes. Results are listed in Table III. Of particular interest, we identified a 300 kb region on Smp\_scaff000234 with more aligned sequences in the IC strains compared with the C strains. This region contains four genes encoding for Egg Secreted Proteins (ESP15 family) recently classified as MEG2 family genes [micro exon genes (MEG)] (DeMarco et al. 2010). Another region enriched in IC sequences contains a gene highly expressed in eggs and miracidia that encodes a metalloprotease from the neprelysin-2 family. Most of the region with more aligned sequences in the C strains contains genes that encode for hypothetical proteins.

*Confirmation of genetic and epigenetic difference between the C and IC strains by qPCR* - To determine whether the observed difference in aligned sequences was due to epigenetic or genetic reasons, we further performed qPCR experiments. We selected 11 loci for which ChIPDiff had identified a difference in aligned sequences between the C and IC strains. These loci may or may not be located in a predicted open reading frame. Primers were designed to target these loci and the %IR was calculated for each position. Results are presented in Table IV. Six primer sets amplified the ChIP products from the IC strain with good efficiency, but did not amplify anything from the C strains. Conversely, one primer set amplified the ChIP products from the C strain, but no amplification could be obtained from the IC strain. For these positions, we hypothesised that the targets are located in regions where genetic polymorphism between the two strains occurs, which is the case for the region containing the MEG2 family genes. For two regions, qPCR did not allow for confirmation of an observed difference by ChIPDiff with no apparent genetic polymorphism. However, a copy number effect was detected for these two regions. Finally, we confirmed a difference in H3K9Ac enrichment in the IC strains compared with the C strain in the Smp\_171100 and Smp\_scaff000154.3 regions, which indicates that these two loci are more euchromatic in the IC strains compared with the C strain. For the Smp\_171100 locus, a heterochromatic epigenetic mark [H3 trimethylated on lysine 9 (H3K9Me3)] was further tested by ChIP-qPCR and it was confirmed that this region is more euchromatic in the IC strains compared with the C strain (Fig. 2).

## DISCUSSION

Throughout its life cycle, *S. mansoni* encounters and must adapt to different types of environments. Interaction with host partners is an example of adaptation to such a changing environment. The compatibility polymorphism described for the *S. mansoni/B. glabrata* interaction is an illustration of such an adaptation displayed by the

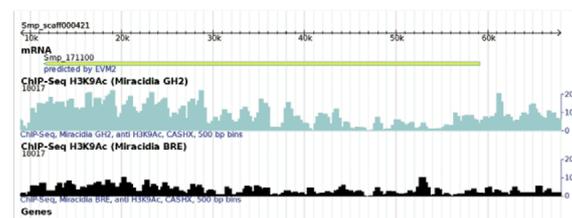


Fig. 1: Smp\_scaff000421:9000..68000. Immunoprecipitated sequences using an antibody that recognizes H3 acetylated on lysine 9 (H3K9Ac) was aligned against *Schistosoma mansoni* reference genome. Aligned sequences can be visualized using an inhouse *S. mansoni* genome browser. The figure represents the data from Smp\_scaff000421, position 9000..68000. mRNA represents a predicted gene (Smp\_171100) in this region. Chromatin immunoprecipitation (ChIP)-Seq H3K9Ac (miracidia GH2 and BRE) represents the number of hits in a 500 bp (CASHX data) window obtained for the incompatible (IC) and compatible (C) strains after ChIP using an antibody against H3 acetylated on lysine 9.

parasite toward its invertebrate host (Theron & Coustau 2005). Molecular candidates for this compatibility polymorphism have been previously described (Roger et al. 2008a, c). In this work, we further raised the question of the implications of genetic vs. epigenetic information for the compatibility polymorphism phenotype. We immunoprecipitated chromatin from two strains, C and IC, being compatible and incompatible, respectively, with the mollusk *B. glabrata*. We used an antibody against histone H3 acetylated on lysine 9, which is a mark found in euchromatin, and we compared the ChIP-Seq profile of these two strains by ChIPDiff. This method allowed us to identify at least nine regions that displayed quantitative differences in aligned sequences against the reference genome, thus highlighting either genetic or epigenetic differences between the two strains. To confirm the genetic and/or epigenetic differences, we used a qPCR approach. We assumed that no PCR amplification in one strain (with a positive result in the other strain) means that there is genetic polymorphism between these two strains. We identified seven loci with a genetic difference, two loci with a difference in copy number and two other loci with an epigenetic difference.

Among the genetic candidates, we identified a 300 kb region, located on Smp\_scaff000234, that contains four copies of genes annotated as MEG. These MEG-type genes have only been described in *S. mansoni* (DeMarco et al. 2010). They are composed of 18 different family members and are a molecular system for creating protein variation through alternate splicing of short exons. The MEG-2 member is detected in our study as a genetic candidate for the compatibility polymorphism phenotype. It is expressed in eggs and miracidia and has been shown to be glycosylated and secreted in eggs. Its potential role in the context of immune evasion strategy has been previously discussed (DeMarco et al. 2010). It is of particular interest that MEG-2 shares many common features with the *SmPoMuc* genes, previously described as major molecular determinants for the compatibility polymorphism (Roger et al. 2008b). *SmPoMuc* genes encode highly polymorphic glycosylated proteins that are secreted in miracidia and their cDNA variants seem to be generated by the “pick and mix” strategy as described for the MEG genes (Berriman et al. 2009). *SmPoMuc* genes display MEG-like characteristics and their potential role as polymorphic antigenic variants in

TABLE III

Regions containing a difference in the number of aligned sequences between the compatible (C) and incompatible (IC) strains

Scaffold	Positions	Predicted open reading frame	Annotation	Expression
IC > C				
Smp_scaff000234	0-310 kb	Smp_159800	MEG2 family (ESP15)	Egg
		Smp_159810	MEG2 family (ESP15)	Egg
		Smp_159820	MEG2 family (ESP15)	Egg
		Smp_159830	MEG2 family (ESP15)	Egg
		Smp_159840	Putative dynein heavy chain	Egg, schistosomula, adults
Smp_scaff000421	10-70 kb	Smp_171100	Zinc metalloprotease	Egg, miracidia, female
Smp_scaff000496	0-2.3 kb and 10-15 kb	None		
C > IC				
Smp_scaff000131	0-70 kb	Smp_149750	Hypothetical protein	Egg, germband, cercaria
		Smp_051690	Hypothetical protein	No EST in database
		Smp_149760	Hypothetical protein	No EST in database
Smp_scaff000490	8-16 kb	None		
Smp_scaff000538	0-24 kb	Smp_093470	Hypothetical protein	No EST in database
Smp_scaff000539	0-35 kb	Smp_173370	M13 unassigned peptidase	No EST in database
Smp_scaff019017	10-40 kb	Smp_121530	Hypothetical protein	Germband, adult
Alternative IC > C/C > IC				
Smp_scaff000154	449-500 kb	Smp_152600	Paired box protein pax-6	Egg, miracidium
		Smp_152610	Hypothetical protein	No EST in database
		Smp_152620	Hook homolog	Egg, germband, schistosomula
		Smp_152630	Hypothetical protein	No EST in database

nine regions that display a difference between the C and IC strains in aligned sequences have been selected. Positions of the concerned region on the selected scaffold are indicated. Information on putative genes contained in these regions, their annotations and their corresponding expressed sequence tag (EST) are given (available from the SchistoDB: [schistodb.net/schistodb20/](http://schistodb.net/schistodb20/)). MEG: micro exon genes.

the context of an arm race hypothesis has been discussed (Moné et al. 2010). These features make the MEG-2 genes promising candidates for a second determinant of the compatibility polymorphism phenotype.

Among the candidates with an epigenetic difference, we identified a 60 kb region on Smp\_scaff000421 that contains a gene (Smp\_171100) annotated as a metalloprotease of the neutral endopeptidase (NEP) family. The role of *S. mansoni* NEP for the immuno-modulation of *B. glabrata* has been previously described (Duvaux-Miret et al. 1992). This gene is expressed in eggs and miracidia and we showed in this study that it is located in a region that is more euchromatic in the IC strain compared with the C strain. Therefore, this gene constitutes an excellent candidate to study the epigenetic bases of the compatibility polymorphism phenotype.

While the role of epigenetics for adaptation, development and immune evasion strategy has been largely described in other parasite models (Dixon et al. 2010, Merrick & Duraisingh 2010, Wang et al. 2010), epigenetic studies in *S. mansoni* are still in their infancy. Development of tools to study the role of histone modifications and the recent release of the *S. mansoni* genome (Berriman et al. 2009) have allowed for performing genome-wide epigenomic studies on this organism. In this study, we applied such tools to study adaptation to invertebrate hosts. This relatively straightforward approach that consists of a combination of low-coverage ChIP-Seq, bioinformatics analysis and confirmation by qPCR has

enabled highlighting of candidates for the compatibility polymorphism phenotypes that we believe are relevant for immune evasion strategies.

*Perspectives* - In addition to the study of the histone modifications presented in this paper, it will be of particular interest to broaden the study of schistosome's chromatin structure to other epigenetic events. Of particular interest is the study of non-coding RNA. We recently identified female specific satellite DNA se-

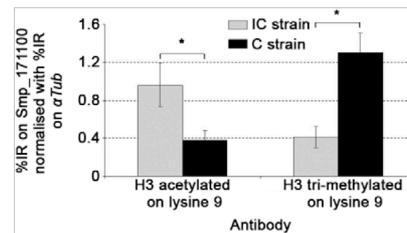


Fig. 2: percent input recovery (%IR) of gDNA in the compatible (C) and incompatible (IC) strains after chromatin immunoprecipitation (ChIP). ChIP was applied on C and IC miracidia gDNA using two antibodies that recognize H3 acetylated on lysine 9 and H3 tri-methylated on lysine 9. Quantitative polymerase chain reaction was used to quantify the %IR on the target locus (Smp\_171100) and reference genes ( $\alpha$ Tub) as indicated in the method section. Results represent the mean of three biological repeats of the %IR of Smp\_171100 normalised with  $\alpha$ Tub. Asterisk means a t-test result below 0.05 between BRE and GH2 data.

TABLE IV  
Comparison of chromatin immunoprecipitation-quantitative polymerase chain reaction (ChIP-qPCR) and sequencing analysis of the immunoprecipitated DNA (ChIP-Seq) results

Target	Number of loci	PCR on IC	PCR on C	Differences between	Differences between	Comparison of $\Delta$ Ct(IC) and $\Delta$ Ct(C)	Genetic or epigenetic?
				C and IC strains using ChIP-qPCR	C and IC strains using ChIP-Seq		
Smp173370	2	+	+	IC > C	C > IC	$\Delta$ Ct(IC) > $\Delta$ Ct(C)	No confirmation
Smp171100	2	+	+	IC > C	IC > C	$\Delta$ Ct(IC) = $\Delta$ Ct(C)	Epigenetic
Smp159800	2	+	-	ND	IC > C	ND	Genetic
Smp159810	4	+	-	ND	IC > C	ND	Genetic
Smp159820	2	+	-	ND	IC > C	ND	Genetic
Smp159830	1	+	-	ND	IC > C	ND	Genetic
Smp159840	1	+	-	ND	IC > C	ND	Genetic
Scaff000131	1	+	+	IC = C	C > IC	$\Delta$ Ct(IC) > $\Delta$ Ct(C)	No confirmation
Scaff000154.2	1	+	-	ND	IC > C	ND	Genetic
Scaff000154.3	1	+	+	IC > C	IC > C	$\Delta$ Ct(IC) = $\Delta$ Ct(C)	Epigenetic
Scaff000154.5	1	-	+	ND	C > IC	ND	Genetic

the region targeted is indicated by a gene name or a scaffold name when no hypothetical open reading frame has been annotated. The primer sequences for qPCR amplification and accurate positions for these targets are given in Table II. The number of loci targeted by the primer sets is given in the column "loci". PCR amplification of the ChIP products is indicated for compatible (C) and incompatible (IC) strains. We assume that no amplification in one strain (but working on the other strain) means that genetic polymorphism occurs between these two strains and prevent the primer set to work correctly. Differences of immunoprecipitated DNA between the C and IC strains using H3 acetylated on lysine 9 are indicated for a detection using qPCR or Solexa sequencing. The  $\Delta$ Ct represents the Cycle threshold (Ct) (target) - Ct ( $\alpha$ Tub),  $\Delta$ Ct in the IC and C strains are compared in order to determine if the target copy number in the C and IC strains are identical (we assume that  $\alpha$ Tub copy number is the same in the C and IC strains). ND: not determined.

quences for which we have hypothesised a role in the sexual dimorphism observed at the adult stages (unpublished observations). MicroRNAs candidates have also been reported in the genome of *S. mansoni* (Simões et al. 2011) and in *Schistosoma japonicum* (Huang et al. 2009). The identification of regulatory non-coding RNA molecules in *Schistosoma* opens new avenues that will likely broaden the epigenetic understandings of schistosomes. Although DNA methylation has not been detected in schistosomes (Fantappiè et al. 2001), a bioinformatic survey of the sequenced *S. mansoni* genome revealed the presence of a DNA methyl transferase (DNMT2). We anticipate that low levels of DNA methylation may occur in schistosomes that likely mediate fine-tuning of the development of this parasite at specific stages of its life cycle. This was previously described for other “DNMT2-only” organisms (Phalke et al. 2009). It will certainly be necessary to increase the study of these other epigenetic events to decipher the impact of epigenetics for the schistosomes physiology.

#### ACKNOWLEDGEMENTS

To Guillaume Mitta, for many discussions during the preparation of this work, and to A Azzi, Bernard Dejean and Anne Rognon, for valuable technical support.

#### REFERENCES

- Berriman M, Haas BJ, LoVerde PT, Wilson RA, Dillon GP, Cerqueira GC, Mashiyama ST, Al-Lazikani B, Andrade LF, Ashton PD, Aslett MA, Bartholomeu DC, Blandin G, Caffrey CR, Coghlan A, Coulson R, Day TA, Delcher A, DeMarco R, Djikeng A, Eyre T, Gamble JA, Ghedin E, Gu Y, Hertz-Fowler C, Hirai H, Hirai Y, Houston R, Ivens A, Johnston DA, Lacerda D, Macedo CD, McVeigh P, Ning Z, Oliveira G, Overington JP, Parkhill J, Pertea M, Pierce RJ, Protasio AV, Quail MA, Rajandream MA, Rogers J, Sajid M, Salzberg SL, Stanke M, Tivey AR, White O, Williams DL, Wortman J, Wu W, Zamanian M, Zerlotini A, Fraser-Liggett CM, Barrell BG, El-Sayed NM 2009. The genome of the blood fluke *Schistosoma mansoni*. *Nature* 460: 352-358.
- Caby S, Pierce RJ 2009. Quantitative chromatin immunoprecipitation (Q-ChIP) applied to *Schistosoma mansoni*. *Mol Biochem Parasitol* 166: 77-80.
- Cosseau C, Azzi A, Smith K, Freitag M, Mitta G, Grunau C 2009. Native chromatin immunoprecipitation (N-ChIP) and ChIP-Seq of *Schistosoma mansoni*: critical experimental parameters. *Mol Biochem Parasitol* 166: 70-76.
- DeMarco R, Mathieson W, Manuel SJ, Dillon GP, Curwen RS, Ashton PD, Ivens AC, Berriman M, Verjovski-Almeida S, Wilson RA 2010. Protein variation in blood-dwelling schistosome worms generated by differential splicing of micro-exon gene transcripts. *Genome Res* 20: 1112-1121.
- Dillon N 2008. The impact of gene location in the nucleus on transcriptional regulation. *Dev Cell* 15: 182-186.
- Dissous C, Dissous C, Capron A 1981. Isolation and characterization of surface antigens from *Schistosoma mansoni* schistosomula. *Mol Biochem Parasitol* 3: 215-225.
- Dixon SE, Stilger KL, Elias EV, Naguleswaran A, Sullivan WJ Jr 2010. A decade of epigenetic research in *Toxoplasma gondii*. *Mol Biochem Parasitol* 173: 1-9.
- Duvaux-Miret O, Stefano GB, Smith EM, Dissous C, Capron A 1992. Immunosuppression in the definitive and intermediate hosts of the human parasite *Schistosoma mansoni* by release of immunoreactive neuropeptides. *Proc Natl Acad Sci USA* 89: 778-781.
- Fahlgren N, Sullivan CM, Kasschau KD, Chapman EJ, Cumbie JS, Montgomery TA, Gilbert SD, Dasenko M, Backman TW, Givan SA, Carrington JC 2009. Computational and analytical framework for small RNA profiling by high-throughput sequencing. *RNA* 15: 992-1002.
- Fantappiè MR, Gimba ER, Rumjanek FD 2001. Lack of DNA methylation in *Schistosoma mansoni*. *Exp Parasitol* 98: 162-166.
- Fejes AP, Robertson G, Bilenky M, Varhol R, Bainbridge M, Jones SJ 2008. FindPeaks 3.1: a tool for identifying areas of enrichment from massively parallel short-read sequencing technology. *Bioinformatics* 24: 1729-1730.
- Huang J, Hao P, Chen H, Hu W, Yan Q, Liu F, Han ZG 2009. Genome-wide identification of *Schistosoma japonicum* microRNAs using a deep-sequencing approach. *PLoS ONE* 4: e8206.
- Lee JS, Smith E, Shilatifard A 2010. The language of histone cross-talk. *Cell* 142: 682-685.
- Li R, Yu C, Li Y, Lam TW, Yiu SM, Kristiansen K, Wang J 2009. SOAP2: an improved ultrafast tool for short read alignment. *Bioinformatics* 25: 1966-1967.
- Luijsterburg MS, White MF, van Driel R, Dame RT 2008. The major architects of chromatin: architectural proteins in bacteria, archaea and eukaryotes. *Crit Rev Biochem Mol Biol* 43: 393-418.
- Merrick CJ, Duraisingh MT 2010. Epigenetics in *Plasmodium*: what do we really know? *Eukaryot Cell* 9: 1150-1158.
- Moné Y, Gourbal B, Duval D, Du Pasquier L, Kieffer-Jaquinod S, Mitta G 2010. A large repertoire of parasite epitopes matched by a large repertoire of host immune receptors in an invertebrate host/parasite model. *PLoS Negl Trop Dis* 4: e813.
- Phalke S, Nickel O, Walluscheck D, Hortig F, Onorati MC, Reuter G 2009. Retrotransposon silencing and telomere integrity in somatic cells of *Drosophila* depends on the cytosine-5 methyltransferase DNMT2. *Nat Genet* 41: 696-702.
- Roger E, Gourbal B, Grunau C, Pierce RJ, Galinier R, Mitta G 2008a. Expression analysis of highly polymorphic mucin proteins (Sm-PoMuc) from the parasite *Schistosoma mansoni*. *Mol Biochem Parasitol* 157: 217-227.
- Roger E, Grunau C, Pierce RJ, Hirai H, Gourbal B, Galinier R, Emans R, Cesari IM, Cosseau C, Mitta G 2008b. Controlled chaos of polymorphic mucins in a metazoan parasite (*Schistosoma mansoni*) interacting with its invertebrate host (*Biomphalaria glabrata*). *PLoS Negl Trop Dis* 2: e330.
- Roger E, Mitta G, Moné Y, Bouchut A, Rognon A, Grunau C, Boissier J, Théron A, Gourbal BE 2008c. Molecular determinants of compatibility polymorphism in the *Biomphalaria glabrata*/*Schistosoma mansoni* model: new candidates identified by a global comparative proteomics approach. *Mol Biochem Parasitol* 157: 205-216.
- Simões MC, Lee J, Djikeng A, Cerqueira GC, Zerlotini A, da Silva-Pereira RA, Dalby AR, Loverde P, El-Sayed NM, Oliveira G 2011. Identification of *Schistosoma mansoni* microRNAs. *BMC Genomics* 19: 12-47.
- Souza PP, Santos DN, Pena SDJ, Franco GR 2002. Cloning and molecular characterization of the *Schistosoma mansoni* genes RbAp48 and histone H4. *Mem Inst Oswaldo Cruz* 97 (Suppl. 1): 77-84.
- Stein LD, Mungall C, Shu S, Caudy M, Mangone M, Day A, Nickerson E, Stajich JE, Harris TW, Arva A, Lewis S 2002. The generic genome browser: a building block for a model organism system database. *Genome Res* 12: 1599-1610.

830 Study of the compatibility polymorphism • Julie Mireille and Joé Lepasant et al.

- Theron A, Coustau C 2005. Are *Biomphalaria* snails resistant to *Schistosoma mansoni*? *J Helminthol* 79: 187-191.
- Theron A, Pages JR, Rognon A 1997. *Schistosoma mansoni*: distribution patterns of miracidia among *Biomphalaria glabrata* snail as related to host susceptibility and sporocyst regulatory processes. *Exp Parasitol* 85: 1-9.
- Umlauf D, Fraser P, Nagano T 2008. The role of long non-coding RNAs in chromatin structure and gene regulation: variations on a theme. *Biol Chem* 389: 323-331.
- Wang QP, Kawahara T, Horn D 2010. Histone deacetylases play distinct roles in telomeric VSG expression site silencing in African trypanosomes. *Mol Microbiol* 77: 1237-1245.
- Xu H, Wei CL, Lin F, Sung WK 2008. An HMM approach to genome-wide identification of differential histone modification sites from ChIP-seq data. *Bioinformatics* 24: 2344-2349.
- Zerlotini A, Heiges M, Wang H, Moraes RL, Daminini AJ, Ruiz JC, Kissinger JC, Oliveira G 2009. SchistoDB: a *Schistosoma mansoni* genome resource. *Nucleic Acids Res* 37: D579-782.

## ANNEXE 5

Publication : Perrin C., **Lepesant J.M.J.**, Roger E., Duval D., Thuillier V., Alliene J.F., Grunau C., Mitta G. & Cosseau C. (under revision). Promoter regions of duplicated *Schistosoma mansoni* gene crucial for infections success undergo rapid epigenetic modifications in varying environments while remaining genetically stable.

Research article

Title:

Promoter regions of duplicated *Schistosoma mansoni* genes crucial for infection success undergo rapid epigenetic modifications in varying environments while remaining genetically stable

Perrin C<sup>1,2</sup>, Lepesant JM<sup>1,2</sup>, Roger E<sup>3</sup>, Duval D<sup>1,2</sup>, Fneich S<sup>1,2</sup>, Thuillier V<sup>1,2</sup>, Alliene JF<sup>1,2</sup>, Mitta G<sup>1,2</sup>, Grunau C<sup>1,2</sup> and Cosseau C<sup>1,2</sup>.

1- Université de Perpignan Via Domitia, Perpignan, F-66860, France

2- CNRS, UMR 5244, Ecologie et Evolution des Interactions (2EI), Perpignan, F-66860, France

3- Center for Infection and Immunity of Lille, Inserm U1019, CNRS UMR 8204, Institut Pasteur de Lille, University Lille Nord de France, Lille, F-59019, France.

**\*Corresponding author:**

Céline Cosseau

Université de Perpignan Via Domitia, UMR 5244 CNRS Ecologie et Evolution des Interactions (2EI), 52 Avenue Paul Alduy, 66860 Perpignan Cedex, France

<http://2ei.univ-perp.fr>

Email: [celine.cosseau@univ-perp.fr](mailto:celine.cosseau@univ-perp.fr)

Tel: +33 4 68 66 21 80

Fax: +33 4 68 66 22 81

**Key words:** host-parasite evolution, epigenetic inheritance system, *Schistosoma mansoni*, *SmPoMuc*,

**Running head:** High epigenetic variability in key genes of a parasite

**Abbreviations:** qPCR stands for quantitative Polymerase Chain Reaction, ChIP stands for Chromatin ImmunoPrecipitation, FREPs stands for Fibrinogen RElated Proteins, C stands for Compatible, IC stands for Incompatible, SD stands for Standard Deviation.

**Abstract**

The digenetic trematode *Schistosoma mansoni* is a human parasite that uses the mollusc *Biomphalaria glabrata* as intermediate host. Specific *S. mansoni* strains can infect efficiently only certain *B. glabrata* strains (compatible strains C) while others are incompatible (IC). Differences in transcription of polymorphic mucins (*SmPoMucs*) between strains of *S. mansoni* are one of the principle determinants for compatibility with *B. glabrata*. *SmPoMucs* are polymorphic glycoproteins that interact with *B. glabrata* Fibrinogen Related Proteins (FREPs). FREPS are involved in the immune response of the mollusc to the parasite. *SmPoMucs* are encoded by a multi-gene family. Importantly, the high degree of protein variability is derived from a limited number of *SmPoMuc* genes. The *SmPoMuc* coding sequences of C and IC strains are very similar. To generate high variability and differences between strains based on a relatively low number of *SmPoMuc* genes, *S. mansoni* has evolved a complex cascade of mechanisms, a “controlled chaos”, acting at the transcriptional, translational and post-translational level. In the present study, we investigated the bases of the control of *SmPoMuc* expression that evolved to evade *B. glabrata* diversified antigen recognition molecules in the promoter of *SmPoMuc* genes. We compared two strains of different geographic origin and either compatible (C) or incompatible (IC) with a reference snail host. We reveal that although sequence differences are observed between active promoter regions of *SmPoMuc* genes, the sequences of the promoters are not diverse and are conserved between IC and C strains, suggesting that genetics alone cannot explain the evolution of compatibility polymorphism. In contrast, promoters carry epigenetic marks that are significantly different between the C and IC strains. Moreover, we show that these modifications of the structure of the chromatin of the parasite lead to enhanced transcription of *SmPoMuc* in the IC strain compared to the C strain and correlate with the presence of additional combinations of *SmPoMucs* transcripts only observed in the IC phenotype. These

results indicate that epigenetic marks change even if the genetic information remains stable, and suggest that epigenetic changes may be important for the early steps in the evolution of an adaptive trait *in S. mansoni*.

## Introduction

The interaction of hosts and parasites is one of the best-studied examples of evolution in a changing environment. Their reciprocal antagonistic co-evolution can be illustrated by an arms race in which host and parasite develop mechanisms to circumvent counter-measures developed by their opponents. Under certain conditions, parasite virulence and host defence can be in equilibrium leading to a phenomenon called compatibility. Compatibility occurs in a host-parasite system when the parasite species is capable of infection and transmission through the host species. The phenomenon that some parasite strains are compatible with certain host strains but not with others (and *vice versa*) is called compatibility polymorphism. Such phenomenon was described in the platyhelminth *Schistosoma mansoni* and its intermediate host, the mollusc *Biomphalaria glabrata* (Theron and Coustau 2005). *S. mansoni* is a human parasite whose life cycle is characterised by the passage through two obligatory sequential hosts: the fresh-water snail *B. glabrata* (or dependent on the geographical location other *Biomphalaria* species) for asexual replication, and human or rodents as hosts for sexual reproduction. The molecular mechanisms underlying compatibility polymorphism between *S. mansoni* and *B. glabrata* were recently investigated by comparing the proteoms of two *S. mansoni* laboratory strains: one strain that is compatible (the C strain) and one that is incompatible (the IC strain) with the same reference *B. glabrata* strain from Brazil. The study identified *S. mansoni* Polymorphic Mucins (*SmPoMucs*) as key markers for compatibility (see Mitta et al. 2011 for a recent review). *SmPoMucs* glycoproteins have a mucin-like structure with an N-terminal domain containing a variable number of tandem repeats (VNTR) (Roger et al. 2008a). *SmPoMuc* proteins are highly polymorphic (Roger et al. 2008a) and interact with the Fibrinogen Related Proteins (FREPs) of the mollusc (Moné et al. 2010). FREPS are diversified antigen recognition molecules playing a central role in the secondary immune response to digenetic trematodes (Adema et al. 1997, Hanington 2010, Zhang et al. 2008).

The extraordinary level of *SmPoMucs* polymorphism is generated by a complex cascade of mechanisms, a “controlled chaos”, acting at the transcriptional, translational and post-translational level (Roger et al. 2008b). *SmPoMucs* are encoded by a multigene family with at least 10 members that are organised in 4 clusters on the genome. They recombine frequently and generate new alleles (Roger et al. 2008b). Each individual miracidia (the larva that infects the mollusc) expresses only a specific subsets of *SmPoMuc* genes (Roger et al. 2008b). The mechanisms controlling this expression polymorphism of *SmPoMucs* remain unclear. The recent finding that Trichostatin A, a modifier of chromatin structure, influences *SmPoMuc* transcription patterns (Cosseau et al. 2010) suggests that epigenetic mechanisms participate in transcription control.

Epigenetic information is information on the status of gene activity that is heritable, for which changes are reversible and that is not based on the DNA sequence. The scientific debate about the reason of the evolution of an epigenetic inheritance system (EIS) in most organisms is intense. Others and we have suggested that EIS provides a basis for modifications in the reaction norms that do not require changes of genotypes (Pal and Miklos 1999, Cosseau et al. 2010), resulting in increased phenotypic plasticity on the individual level or increased phenotypic variability on the population level. We have shown that this is true for *SmPoMuc* expression in populations of *S. mansoni* miracidia (Cosseau et al. 2010). If EIS influences on the capacity to generate different phenotypes, both the better adapted phenotype and the capacity to generate this phenotype will be selected for and carried into the next generation. This transgenerational quality of an EIS was described as “transgression potential” (Johannes et al. 2011) and allows for rapid adaptive evolution (Roux et al. 2011). It was hypothesized that in changing environments, in particular parasite-host interactions with fluctuating hosts, EIS will be important (Lachmann and Jablonka, 1996). Theoretical models predict that in

these systems, epigenetic changes, *e.g.* chromatin structure marks, occur before genetic changes (Pal and Miklos 1999).

To obtain experimental evidences for these theoretical considerations, we investigated in the present study the genetic and epigenetic changes that occurred during the evolution of the phenotypic compatibility polymorphism in two *S. mansoni* strains. We focused on the sequence of the promoters of active *SmPoMuc* genes and investigated whether there exist differences in the promoter sequences between *S. mansoni* compatible and incompatible strains. Our study revealed that IC and C strains display very little genetic variability within strain, and limited nucleotide differences between promoter sequences of the two strains, but show strong chromatin structure differences. Our findings indicate that, in the *SmPoMuc* promotor regions, epigenetic changes occurred while the nucleotide sequences remained stable, and that EIS can control a heritable adaptive trait.

## **Materials and Methods**

### ***Culture of Schistosoma mansoni***

A compatible strain (C) (Brazilian strain), an incompatible *S. mansoni* strain (IC) (Guadeloupean strain) and a reference mollusc strain (*B. glabrata* BRE isolated from Brazil) were used in this study. For initial breeding, each strain was maintained in its sympatric (compatible) *B. glabrata* strain, and in hamsters (*Mesocricetus auratus*) as described previously (Theron et al. 1997). Adult worms and miracidia were obtained as described in Roger et al. (2008b). The French Ministère de l'Agriculture et de la Pêche and French Ministère de l'Education Nationale de la Recherche et de la Technologie provided permit A 66040 to our laboratory for experiments on animals and certificate for animal experimentation (authorization 007083, decree 87-848) for the experimenters. Housing, breeding and animal care followed the national ethical requirements.

### ***PCR screening for promoters of SmPoMuc genes, cloning and sequencing***

We searched for sequences of promoter regions of *SmPoMuc* genes in the genomic database of the *S. mansoni* NMRI strain (assembly version 3.1) using BLAST searches. Contigs matching to *SmPoMuc* genes were assembled with the Sequencher software (Gene Codes Corporation) to recover the sequences of the promoter regions of the genes. From the blast search and manual assemblage of relevant contigs, scaffolds of promoter regions were constructed for the different *SmPoMuc* genes in groups 1 - 4. Primers were designed on these contigs to amplify the promoter regions of the different *SmPoMuc* genes in the C and IC strains of *S. mansoni*. The DNA templates to generate PCR products were either genomic DNA (C and IC strains), BACs library (NMRI strain) or a phage library (IC strain). Genomic DNA was extracted from adult worms as described in Roger et al. (2008b, c). The production of the phage library is described below. Promoter regions were amplified using the Advantage

2 PCR Enzyme System (Clontech) (see Supplementary file 1 for primer sequences, amplified fragment lengths and sources of DNA). PCR products were either cloned into pCR-XL-TOPO (TOPO TA Cloning kit for sequencing, Invitrogen) and plasmid DNA was purified using the Wizard Plus SV Miniprep DNA purification system (Promega), or sequenced directly. We sent PCR amplificons or plasmids containing the promoter regions to GATC (GATC Biotech, Germany) for cycle sequencing on both directions and performed primer walking up to 2.0 kb upstream of the transcription start sites (TSS) of *SmPoMuc* genes (for primer sequences see Supplementary file 2). We checked trace data and aligned nucleotide sequences manually using the BioEdit software.

### ***Production and screening of a phage lambda library of IC genomic DNA***

The presence of multiple copies of some *SmPoMuc* genes prevented sometimes the amplification of a single copy and assembly of a gene with its corresponding promoter. To address this problem, we constructed a phage library of the IC strain using the Lambda Fix II vector system from Stratagene. The expected size of inserts was 15 to 23 kb corresponding to the size range of *SmPoMuc* genes (10 – 30 kb). Details of the construction of the phage library and screening are available at [http://methdb.univ-perp.fr/cgrunau/methods/lambda\\_library.html](http://methdb.univ-perp.fr/cgrunau/methods/lambda_library.html). Genome coverage of the library was four fold. The library was screened for *SmPoMuc* genes using probe UR1, a highly conserved intronic sequence spanning the region between two repeat units of the *SmPoMuc* genes (Roger et al. 2008b). The probe was labeled with the DIG High Prime DNA Labeling and Detection Starter Kit II using Random primed DNA labeling with digoxigenin-dUTP, alkali-labile and chemiluminescence with CSPD (Roche). Screening was performed according to the manufacturers instruction. Secondary and tertiary screening rounds were performed with the same probe to isolate individual phage clones. Phages that scored positive for *SmPoMuc*

repeat units were screened by PCR using a combination of diagnostic primers for each group of *SmPoMuc* genes (Supplementary file 2) with the Advantage 2 PCR Enzyme System (Clontech). Selected phages were subsequently purified and used as template to PCR amplify *SmPoMuc* group 3.1(r1-r2) as described in the section “PCR screening for promoters of *SmPoMuc* genes, cloning and sequencing”.

### ***Sequence variation of promoter regions of SmPoMuc genes between S. mansoni IC and C strains***

#### *Sequence annotation and promoter prediction*

The 5'UTR and ORF were previously characterised using 5'RACE-PCR experiments (Roger et al. 2008a). The core promoter including a TATA box and the TSS was predicted using Neural Network Promoter Prediction Tool ([http://www.fruitfly.org/seq\\_tools/promoter.html](http://www.fruitfly.org/seq_tools/promoter.html)) (Reese 2001). We identified repetitive elements in the promoter region sequences using the CENSOR software (Kohany et al. 2006). We searched for duplications, recombinations and gene conversions using dot plots among sequences and the programs RDP3 (Martin et al. 2010). *SmPoMuc* promoter sequences were annotated using CLC Sequence Viewer v6.5.1 (CLC Bio 2011). We colour-coded paralogous sequence blocks, portions of repetitive elements, duplications and recombination to visualise the evolution of paralogous and orthologous *SmPoMuc* promoter sequences. The number of substitutions per site for pairwise comparisons and searched for conserved regions was calculated with DnaSPv4.50.3 (Rozas et al. 2003).

#### *Phylogenetic analysis*

We performed Bayesian phylogenetic analyses using MrBayes 3.2.0 (Ronquist et al., 2012). We sampled across the substitution model space in the Bayesian Markov Chain Monte Carlo (MCMC) itself (Huelsenbeck et al., 2004). The model selected was the HKY model.

Insertion/deletion (indel) events were coded as binary characters (presence/absence) and included as a separate binary data partition in the analysis (Ronquist et al., 2012). We ran the MCMC for 120,000 generations, trees being sampled every 100 generations. It allowed the final average standard deviations of split frequencies to reach below 0.01 and the potential scale reduction factors (PSRF) for all parameters to be close to 1, indicating that the runs had converged onto the stationary distribution. The first 1,000 trees were discarded as burn-in to compute the consensus tree. We repeated the analyses three times to ensure the posterior probabilities were stable. Trees were rooted with a sequence of the promoter sequence of the *SmPoMuc* pseudogene group 4.

### ***Sequence variation and gene diversity***

We used DnaSP to characterise promoter sequence variation within and between groups of *SmPoMuc* promoter sequences as the number of polymorphic sites, number of mutations between strains, net number of substitutions per site between strains and between groups of *SmPoMuc* promoter sequences.

### ***Sequence variation of the promoter region of a single copy gene, SmFTZ-F1, between S. mansoni IC and C strains***

We amplified and sequenced the promoter region of the *SmFTZ-F1* gene (De Mendonca et al. 2002) in 1 and 2 individuals of *S. mansoni* strains IC and C, respectively, from genomic DNA with primers *Smftzf1-F* (5'-ATGAGATGTTTCTGAGCAATGGC-3') and *Smftzf1-R* (5'-TCTTCTCGTAGCTGAATCTGACC-3') using the Advantage 2 PCR Enzyme System (Clontech). PCR amplicons were then sequenced and analysed for sequence variation and gene diversity as described above.

## ***Heterologous expression of promoter regions of SmPoMuc genes***

### *Cell culture and transfection*

HeLa cells were maintained in Dulbecco's modified Eagle's medium (DMEM) and 10% fetal calf serum (FCS) containing an antibiotic/antimycotic mixture (penicillin 100 units/ml, streptomycin 0.1 mg/ml, amphotericin B 0.25 µg/ml; Sigma) at 37°C. Transfections were performed on Lab-Tek™ chamber slides (0.8 cm<sup>2</sup>/wells) with 250 ng of DNA using jetPRIME™ according to the manufacturer's instruction (Polyplus transfection). Briefly, 20,000 cells were seeded per well in 350 µl of cell growth medium 24 h prior to transfection. Two hundred and fifty ng of plasmid DNA diluted into 25 µl jetPRIME™ buffer were incubated with 1 µl jetPRIME™ transfectant for 10 min at room temperature. The transfection mix was added directly to the cells. After 72 h, we washed HeLa cells with PBS and fixed in -20°C methanol for 5 min. Cells were washed twice with PBS and counterstained with DAPI (100 µg/ml) for 10 sec and mounted with fluorescent mounting medium (Dako). Fluorescence was observed with a Zeiss Axioskop2 (Zeiss) using a camera Leica DC350FX coupled to imaging software (Leica FW4000).

### *SmPoMuc group 3.1(r1-r2) promoter construction*

To determine a minimal functional promoter, we cloned roughly 1 kb of the SmPoMuc group 3.1(r1-r2) promoter directly upstream of the EGFP cDNA of a promoterless vector, pEGFP1 (Clontech). This sequence was amplified by PCR using primers containing *SacI* and *BamHI* restriction sites (5' primer CCCTTTgagctcATGGAGCAGTGAGTTA and 3' primer TATggatccTGGCTTTATATATATTTTAC). The PCR product was gel-purified (Wizard® SV gel and Clean-Up system, Qiagen), digested with both restriction enzymes and cloned into a *SacI* and *BamHI* digested pEGFP-1 reporter vector with T4 DNA ligase (New England Biolabs). The construct was verified by sequencing both DNA strands. Plasmids pEGFP-1

and pCMV-EGFP driving EGFP expression, under the control of the CMV-promoter, were used as negative and positive controls in the transfection assay.

***Sequence variation of promoter regions of SmPoMuc group 1 gene between S. mansoni IC and C strains at the population level***

A 3.3 kb region of the *SmPoMuc* group 1 gene promoter region was amplified using primers SmpomucpromGP3.1.f2 and BR2 (Supplementary file 1) in individuals of each of *S. mansoni* IC and C strain. The PCR products span from 1.8 kb upstream of the TSS to the first repeat unit of the *SmPoMuc* gene and covers the promoter region. 1.4 Kb of the promoter region was sequenced for 20 and 18 individuals of the IC and C strains, respectively, by primer walking (Supplementary file 2). We used Arlequin 3.1 to characterise *SmPoMuc* group 1 promoter diversity within the two strains as the expected unbiased gene diversity, the nucleotide diversity, corrected for sample size and incorporating nucleotide information (Nei 1987). We tested for sequence variation between the two strains using population comparisons and differentiation in Arlequin 3.1. Estimations incorporated Tamura-Nei distances between sequences and allele frequencies (Nei's  $\Phi$ -estimator of  $F_{ST}$ ). The significance of genetic differentiation was tested by permuting 2,000 times the alleles among all samples. We also estimated inbreeding coefficient in each strain using  $f$  and genetic differentiation between the two strains using  $F_{ST}$  estimator  $\theta$  (Weir and Cockerham 1984, incorporating allele frequencies only). Inbreeding coefficients and genetic differentiation for departure from the null hypothesis ( $f = 0$ ,  $\theta = 0$ ) were tested using 2,000 permutations in GENETIX 4.05 (Belkhir et al. 2006).

***Allelic variation of 14 microsatellite loci between S. mansoni IC and C strains at the population level***

Nineteen individuals of each of the IC and C strains were genotyped using 14 microsatellite loci (Bech et al. 2010). We estimated genetic diversity of microsatellite loci as mean number of alleles per locus ( $A$ ), observed and expected unbiased heterozygosities ( $H_O$  and  $\hat{H}$  respectively) under the assumption of Hardy–Weinberg equilibrium (Nei 1987). We estimated inbreeding coefficient  $f$  in each strain, genetic differentiation between the two strains  $R_{ST}$  estimator (Rousset 1996, Slatkin 1995) and  $F_{ST}$  estimator  $\theta$  as above.

### ***Trichostatin-A treatment, mRNA extraction, cDNA synthesis and transcription analysis***

Trichostatin-A (TSA) (invivoGen met-tsa-5) was dissolved in ethanol to 20 mM and added to the 1000 IC or C miracidia pool at 20  $\mu$ M and 200  $\mu$ M during 4h. We had shown previously the effect of TSA at these concentrations on development, morphology, mobility and gene expression without any cytotoxicity for the larvae (Cosseau et al. 2010, Azzi et al. 2009). To the untreated control, an equal volume of ethanol was added (mock treatment). After 4 h, metamorphosis arrest was observed for larvae treated with TSA at 200  $\mu$ M as expected for a positive effect with this drug (Azzi et al. 2009). Miracidia were then spun down at 12,000g during 5 min and suspended in 100  $\mu$ l of lysis buffer (Dynabeads mRNA DIRECT Micro kit, DYNAL BIOTECH) in Rnase-free tubes and stored at -80°C. Messenger RNAs were extracted using the Dynabeads mRNA isolation Kit according to the manufacturer's instructions. mRNA poly-A residues were eluted from the surface of the paramagnetic beads by a final denaturing step of 10\_min at 75°C in 20 $\mu$ l of Tris-HCl 10mM. cDNA synthesis was carried out using 10  $\mu$ l of mRNA in a final volume of 20  $\mu$ l according to manufacturer instructions (0.5 mM dNTPs, 0.01mM DTT, 1X first strand buffer, 2U RNase out, 10U SuperScript II RT (Invitrogen) during 50 min at 42°C. After reverse transcription, the cDNAs were purified with the PCR clean-up system (Promega) and eluted into 100  $\mu$ l 10 mM Tris/Cl (ph 7.5).

Specific primers for qPCR from groups 1, 2 and 3.1 (r1-r2) were designed based on sequence alignment performed on cDNA variant representative of each group (Supplementary file 2). Their specificity was tested using as template a plasmid in which a cDNA variant of group 1, 2 or 3.1 (r1-r2) was cloned. Group 4 genes contain a STOP codon in exon 8 of the gDNA sequence and its cDNA has never been detected. Therefore, transcripts of the group 4 genes were not targeted in this study. qPCR amplifications were performed as described below. Results were normalised with the  $\alpha$ Tub gene. The  $2^{\Delta Ct}$  value was calculated. Statistical test were performed on at least 3 different biological samples.

### ***Chromatin status of SmPoMuc promoters by ChIP-qPCR***

Native chromatin immunoprecipitation was performed as described before (Cosseau et al. 2009). In brief, antibodies against histone isoforms were used to precipitate chromatin in miracidia from IC and C strains (Supplementary file 3). DNA was extracted from the precipitated complex and analysed by qPCR using specific primers of *SmPoMuc* group 1, 3.1 and 3.1(r1-r2). Primers targeting specifically these genes were designed based on sequence alignment of *SmPoMuc* promoter sequences (Supplementary file 2). We tested their specificity using as template plasmids with promoters of group 1, 3.1 or 3.1 (r1-r2). It was not possible to design primer sets that would hybridize specifically to the promoter sequences of the other groups or subgroups because conservation in the sequences resulted in cross-amplification between these groups. The amount of target DNA recovered in the immunoprecipitated fraction was quantified by calculating the percent input recovery (% IR) normalised with the percent input recovery obtained with a reference locus ( $\alpha$ Tub) as previously described (Cosseau et al. 2008).

**qPCR analysis**

qPCR amplifications were performed with 2.5µl of immunoprecipitated DNA or cDNA in a final volume of 10µl on a LightCycler® 480 II real Time instrument (1.5µl H<sub>2</sub>O, 0.5µM of each primer, 5µl of master mix). The following protocol was used: denaturation, 95°C 10 minutes; amplification and quantification (40 times), 95°C for 10 seconds, 60°C for 10 seconds, 72°C for 20 seconds; melting curve, 65-97°C with a heating rate of 0.11 C/s and continuous fluorescence measurement, and a cooling step to 40°C. For each reaction, the crossing point (Ct) was determined using the “2nd derivative” method of the LightCycler® 480 Software release 1.5. PCR reactions were performed in duplicates and the mean value of Ct was calculated. Correct melting curves were checked using the T<sub>m</sub> calling method of the LightCycler® 480 Software release 1.5. The amplification of a unique band was verified by electrophoresis separation through a 2% agarose gel for each qPCR product.

**Results*****Transcription of SmPoMuc genes is different in IC and C strains of S. mansoni***

Transcription of *SmPoMuc* genes in groups 1, 2 and 3.1(r1-r2) was compared between miracidia of the IC and C strains. *SmPoMuc* genes group 1, 2 and 3.1(r1-r2) are approximately 2.9, 3.5 and 25.8 more transcribed in the IC than in the C strain (fig.1). The 3.1 subgroup containing intermingled r1-r2 repeats is highly transcribed in the IC strain but was nearly not detectable in the C strain. This result is consistent with a previous study on individuals of the IC and C strains, which showed that variants containing the r1-r2 combinations are only expressed in the IC strain (Roger et al. 2008a, b).

***The SmPoMuc minimal promoter region is located within 1,000 bp upstream of the TSS***

To investigate the mechanisms underlying differences of transcription between *SmPoMuc* groups and subgroups, we characterized the minimal promoter region of the *SmPoMuc* genes. We sequenced 1.04 - 2.00 kb upstream of the transcriptional start site (TSS) for 4 groups of *SmPoMuc* and produced PCR products spanning 1 kb of the region. Plasmids containing these sequences upstream a reporter gene (EGFP) were transfected into HeLa cells and fluorescence was observed under a microscope (fig. 2). These experiments showed that a 1 kb sequence is sufficient to drive the heterologous expression of the reporter gene and contains the minimal promoter sufficient for transcription.

***Sequence variations of promoter regions of SmPoMuc genes between S. mansoni IC and C strains are small***

As a first approach to investigate a putative genetic basis for this difference in transcription level between strains, we investigated the paralogous and orthologous relationships between the four groups of *SmPoMuc* gene promoters and between the two *Schistosoma mansoni* IC and C strains using a phylogenetic analysis, reciprocal Blast dot-plots and comparison of repetitive elements, duplication, recombination events and gene conversions (fig. 3). We annotated the sequences and visualised them by colour-coding of blocks with less than 95% identity (fig. 3). A recombination event was detected using BootScan (Salminen et al. 1995; Martin et al. 2005), Maximum Chi Square (Maynard Smith 1992; Posada and Crandall, 2001) and Sister Scanning method (Gibbs et al. 2000) methods in RDP3 and the recombination break points were estimated (fig. 3). We observed in both strains one duplication in group 3.1(r1-r2) promoters resulting in an insertion, several insertions/deletions (indels) including one large deletion in group 3.1 promoters and a recombination event from group 2 to group 1 promoter. High similarity to a repeat DNA element was detected in group 2 promoter,

however it constituted only a small fragment of the complete repeat – 61 bp out of 385 bp of the DIVER2 LTR (*Drosophila*).

The estimated divergence time between the two *S. mansoni* strains IC and C is about 400 years (Morgan et al. 2005) and the promoter sequences between the two strains are highly conserved (0.000-0.004 net substitutions per site, Table 1). The number of fixed differences between the two strains varied between 0 in the promoter region of *SmPoMuc* group 2 genes, to 3 in group 3.1, 4 in groups 1 and 8 in group 3.1(r1-r2) (Table 1). No substitution was observed in the TATA signal and TSS regions of the promoters between the two strains. *SmPoMuc* promoter sequences were divided into four paralogous sequence groups and sequence differences between strains (orthologous relationships) within groups were much lesser than the differences observed between groups of the *SmPoMuc* gene family - net substitutions per site varied from 0.000-0.004 within groups of promoter sequences between strains compared to 0.024-0.041 between promoter groups (Table 1). The number of *SmPoMuc* promoter sequence differences between strains was equal or slightly higher than the number of sequence differences for the promoter of the single copy gene *SmFTZ-F1*, which shows no difference between strains (Table 1). Six of 14 microsatellite loci also did not show any sequence difference between the two strains (one unique allele). The two strains share the molecular evolution and phylogeny of the promoter region of the four groups of the *SmPoMuc* gene family (fig. 3) – indels, recombination and duplication events. These findings indicate that the divergence between groups of the *SmPoMuc* gene family from a common gene ancestor is ancient and largely predates the time of separation between the two *S. mansoni* strains.

***Low level of promoter nucleotide diversity within the IC and C strains***

At this step of the study we hypothesized that *SmPoMuc* expression differences in C and IC strains could be due to nucleotide differences in the promoter regions of the genes. The sequencing of 1.4 Kb of *SmPoMuc* group 1 promoter region for 20 and 18 individuals of the IC and C strains, respectively, revealed a very low number of alleles and genotypes (Table 2) – one genotype in the IC strain and 3 genotypes in the C strain. In the C strain, sequence variation was minimal, with the three alleles differing by only one base pair from each other, resulting in insignificant nucleotide diversity (Table 2). All individuals were homozygotes. The IC strain allele of the *SmPoMuc* promoter group 1 differed from the three C strain alleles by four to five base pairs, a sequence divergence of 0.29 to 0.36 %. In summary, nucleotide sequence differences between the two strains are surprisingly small.

***SmPoMuc group 1 promoter population sequence difference between IC and C strains is not higher than average genome-wide difference***

Promoter diversity within strain and divergence between strains of *SmPoMuc* group 1 genes were similar to those of 14 microsatellite loci that may be used to reflect genome-wide diversity and divergence (Bech et al. 2010). Promoter diversity of *SmPoMuc* group 1 was 0.00 (one allele) in the IC strain compared to 0.22 (3 alleles) in the C strain (Table 2), while expected heterozygosity was 0.000 (one allele) for both strains for 14 microsatellite loci (Data not shown). All individuals were homozygotes. Six out of 14 microsatellite loci showed no divergence between the two strains. At eight microsatellite loci, the IC strain alleles differed from the C strain alleles by one to eleven microsatellite repeats. The promoter region of the single copy *SmFTZ-F1* gene displayed a unique sequence common to the two strains. We estimated extremely high and significant genetic differentiation between the two strains for both *SmPoMuc*s group 1 promoter sequences and microsatellite loci using  $\theta$ ,  $\Phi_{ST}$  and  $R_{ST}$

estimators (Table 3). However, we detected almost no heterozygotes and high significant inbreeding coefficients  $f$  in both strains and for both *SmPoMuc* group 1 promoter sequences and the microsatellite loci (Table 3). Therefore those high values of divergence are likely the result of high clonality in the two strains as discussed previously in Bech et al. (2010). Nonetheless, the distribution of alleles matched the pattern of differentiation as we detected fixed alleles that were different in the two strains. We reasoned that the small genetic differences in the promoter region are simply a by-product of clonality and not the reason for expression differences. We therefore explored an alternative hypothesis, *i.e.* that the expression differences are due to dissimilarity in the epigenetic information.

#### ***HDAC inhibitors have an effect on SmPoMuc transcription***

As the difference in *SmPoMuc* transcription phenotype cannot be easily explained by genetic differences in the promoter region, we investigated the putative implication of epigenetic mechanisms. As a previous study showed that histone modifications are clearly involved in *S. mansoni* epigenetic mechanisms (Cosseau et al. 2010, Dubois et al. 2009), we tried to influence the epigenotype and phenotype (*SmPoMuc* expression pattern) of *S. mansoni* using trichostatin-A (TSA) that is a specific and reversible inhibitor of class I and II Histone Deacetylase (HDAC). Treatment with this drug prevents histone deacetylation and is expected to increase the overall acetylation of histones and therefore gene expression (Azzi et al. 2009). The influence of TSA treatment on the transcription of *SmPoMuc* genes (group 1, 2 and 3.1(r1-r2)) of both C and IC strains was tested in miracidia larvae exposed during 4 h to the drug. A Friedman non-parametric test was performed to test the significance of the TSA effect (Supplementary file 4). We observed a statistically significant increase in transcription of groups 1 and 2 after TSA treatment in the IC strain only ( $P$ -value = 0.05). This indicates that changes in histone acetylation correlate with increased expression for *SmPoMuc* group 1 and

2 in the IC strain and has no effect in the C strain. Control genes were also tested for their response to TSA in order to test that TSA effect was not pleiotropic. No effect of TSA was observed for these genes (GAPDH, Smp\_011030, Smp\_152710.1, Smp\_054160, Smp\_158110.1, GST.B, Glyoxalase, data not shown).

### ***Strain hybrids express both C and IC strain specific SmPoMucs***

Since the TSA treatment influences overall histone acetylation, it could not be excluded that the observed effect is an indirect one and that *SmPoMuc* expression control is posttranscriptional and/or posttranslational such as selective RNA or protein degradation. We reasoned that in the offspring of crosses between IC and C strain transcriptional control would produce an additive pattern of *SmPoMuc* proteins while control by selective degradation of gene products would produce a subtractive pattern. Western blots show that in miracida that are produced from crosses between the strains (either paternal C or maternal C) an additive pattern of the C and IC specific bands can be observed (fig. 4). This indicates that regulation operates on the transcriptional and not on post-transcriptional level and further supports the view that chromatin structure plays a role in the generation of specific *SmPoMuc* profiles for each strain.

### ***There are strong chromatin status differences in the SmPoMuc promoters between S. mansoni C and IC strains***

Since all experiences had delivered results in favour of a difference in chromatin structure of *SmPoMuc* locus between strains, we decided to investigate the chromatin status in these loci by ChIP experiments. Primers were designed upstream of the start codon in regions that displayed sequence degeneration in order to target specifically *SmPoMuc* group 1, 3.1 and 3.1(r1-r2). ChIP experiments were performed using an antibody that recognised Histone 3

(H3) acetylated on lysine 9 and H3 tri-methylated on lysine 4 which are euchromatic marks and an antibody that recognised H3 tri-methylated on Lysine 9, which is a heterochromatic mark. Immunoprecipitation with the antibody that targets H3 tri-methylated on lysine 4 (H3K4Me3) did not show any enrichments in the *SmPoMuc* region tested for both IC and C strains whereas controls ( $\alpha$ Tub and 28S) were positive (data not shown). The H3K4Me3 is probably very sharp and the high degree of sequence similarity between the gene groups did not allow for designing primers to produce PCR products that cover the entire region. Both *SmPoMuc* group 1 and 3.1(r1-r2) from the IC strain displayed a higher level of H3 acetylated on lysine 9 compared to the C strain (fig. 5). Consistently with this result, the C strain displayed a higher level of the heterochromatic mark for group 1 and 3.1 (r1-r2). In the IC strain, epigenetic marks showed differences among *SmPoMuc* groups 1, 3.1 and 3.1(r1-r2) (Supplementary file 5). Promoter of group 3.1(r1-r2) is the most acetylated and the less heterochromatic. This result is consistent with expression analysis after TSA treatment where no effect of TSA was observed for expression of group 3.1(r1-r2). This absence of TSA effect may be explained by the fact that acetylation on this promoter is already saturated and cannot be further increased.

## Discussion

The co-evolution of host-parasite interactions can be viewed as an arms race in which both the host and the parasite develop mechanisms to circumvent the weapons developed by their opponent. In this context, variability-generating processes are crucial for survival on both sides of the interaction (red queen hypothesis, Van Valen 1974). The mechanisms that are responsible for these (heritable) phenotypic variations are a current and fundamental question in evolutionary biology. Traditionally, random genetic changes are seen as sole source of phenotypic variation. But the picture is probably more complex: heritable adaptive phenotypic shifts could be partly controlled by epigenetic factors that were underrated until recently (Jablonka and Lamb 2005, Bossdorf et al. 2008). A high rate of heritable epigenetic changes would generate phenotypic variation, that in turn could allow a rapid response to selection pressures (Cosseau et al. 2010; Danchin et al. 2011). This could allow for a transient and efficient response to changes in the environment, and could subsequently be followed by stabilisation through genetic changes (Jablonka et al. 1998, Pigliucci et al. 2006). Epigenetic modifications affect the transcription status of a gene in a heritable way without changes in the DNA sequence (Umlauf et al. 2008, Dillon 2008, Lee et al. 2010). Epigenetic information can be based on a chromatin marking system. Chromatin exists either as a relaxed structure that is permissive to gene expression and is called euchromatin, or as a condensed structure that is typically silent and is called heterochromatin (Luijsterburg et al. 2008). Therefore, these different chromatin states alter gene expression and, ultimately, influence phenotypic outcomes without changes in DNA sequence. Evolutionary implications of epigenetic inheritance systems and its potential link to stress-induced phenotypic variation has been discussed in several models (Rapp and Wendel 2005; Grant-Downton and Dickinson 2006; Richards 2006; Bossdorf et al. 2008, Bossdorf et al. 2010, Boyko and Kovalchuk 2008; Jablonka and Raz 2009). While it is clear now that induced epigenetic modifications are

heritable (e.g. Verhoeven et al. 2010), there are very few reports that show that epigenetic events lead to modification of gene expression profiles, production of new phenotypes and adaptation to environment (Uchida et al. 2011).

In the present work, we addressed the question of the relative importance of genetic and epigenetic differences between two strains of *S. mansoni* that show clear differences in an ecological important adaptive trait: the capacity to infect their intermediate host. We had previously identified the *SmPoMuc* genes as key markers for host compatibility and had shown that nucleotide differences in the coding region could not explain differences in transcription. We focused therefore on the promoter regions in the present work. Our comparative survey of sequence variation in the different groups of *SmPoMuc* gene family from IC and C strains revealed a high conservation of the promoter sequences of *SmPoMuc* genes between the two strains. The molecular evolution of *SmPoMuc* promoters were uniform between all strains analysed, IC, C and NMRI. The sequence difference between the IC, C and NMRI strains within each group of *SmPoMuc* promoter was small, and the number of substitutions between the IC and C strains was equal or slightly higher than in the monomorphic single-copy gene *SmFTZ-F1* and consistent with sequence differences at 14 microsatellite loci. To assess whether substitutions between the two strains could have an effect on transcription, we searched for functional regions of the active promoters. None of the substitutions between the IC and C strains occurred in TATA signal or TSS regions. Other functional regions of the active promoter might be characterised by highly conserved sequences among groups of promoters. One to two substitutions between strains were observed at the exception of group 2 in otherwise conserved regions among *SmPoMuc* groups of promoters (fig. 3). The nucleotide differences between the two strains consisted of zero in group 2 to eight substitutions in group 3.1(r1-r2), resulting in net nucleotide substitutions per site similar or lower than the ones observed in presumably neutral *SmPoMuc* introns (Table

2). At the population level, our analysis of *SmPoMuc* group 1 promoters in the IC and C strains revealed very low allelic and nucleotide variability within strain and high allele frequency differences between the IC and C strains due to fixed substitutions. All individuals were homozygotes at *SmPoMuc* group 1 promoter, similarly to the genotypes at 14 microsatellite loci, suggesting that *S. mansoni* strains present genome-wide homozygosity. Both strains are characterised by high significant inbreeding coefficient, resulting from high clonality in the two strains (Bech et al. 2010), that may have arisen because of the bottleneck due to the strain maintenance in laboratory conditions. Despite the lack of diversity within strains, alleles fixed in each strain for *SmPoMuc* group 1 promoter and nine microsatellites were different, resulting in high genetic differentiation between the two strains as estimated by  $F_{ST}$ . It contrasted with the promoter of the single-copy gene *SmFTZ-F1* and six microsatellite loci, which displayed a unique sequence common to the two strains. Therefore, high inbreeding in each strain makes it difficult to estimate the level of variation between the two strains. However, nucleotide differences appeared small within and between strains and are unlikely to explain the level of cDNA variation observed within and between strains, in particular for *SmPoMuc* group 2 (Roger et al. 2008b).

In summary, our analysis of the genetic information shows that (i) both strains are genetically monomorph, including the *SmPoMuc* promoter regions, (ii) both strains are different in terms of alleles, i.e. they do not share the same alleles, but (iii) these alleles are not very different in the number of base substitutions. In the view of our results, we argue that *SmPoMuc* gene and promoter sequences and their regulatory functions are conserved between the IC and C strains. The most parsimonious explanation of the observed differences in alleles would be the fixation of random neutral nucleotide changes through multiple bottlenecks. Thus, genetic differences between sequences within each group of *SmPoMuc* promoters were unlikely to solely dictate high variation in *SmPoMuc* and compatibility polymorphism phenotypes.

We therefore further investigated the epigenetic basis for such phenotypes. TSA treatment was used to study the impact of overall acetylation status of histones on miracidia larvae where *SmPoMuc* is expressed. This drug is known to be a specific histone desacetylase (HDAC) inhibitor and has been previously shown to influence phenotypic traits in *S. mansoni* (Cosseau et al. 2010). A dose dependant effect of TSA is observed for *SmPoMuc* expression (all groups taken together) in the IC strain whereas no effect is observed in the C strain. This result suggests that the acetylation status of histones in the promoter sequences is differentially regulated between the IC and C strain. At least, HDAC play a more prominent role in regulating the acetylation level in the IC strain that allowed us to pinpoint a TSA effect in this strain. More specifically, a TSA effect is reported on groups 1 and 2 of the IC strain whereas no effect is observed for group 3.1 (r1-r2) for which acetylation is the strongest. This also suggests that a differential regulation by HDAC exists in-between the *SmPoMuc* groups in the same strain. Further support for regulation on transcriptional level comes from a crossing experiment in which strain hybrids were produced. Western blots show that in the hybrids the C-specific and the IC-specific *SmPoMucs* are expressed. One could hypothesize that production of *SmPoMuc* variants is due to post-transcriptional strain-specific regulation. In this scenario all genes would be expressed, but the gene products would be processed in a strain-specific form. In the hybrids, in which the hypothetical post-transcriptional regulation pathway for both strains is now present, we should see a diminution of non-IC and the non-C *SmPoMuc* forms. This is not the case. In summary, all lines of evidence point towards a chromatin-based regulation of *SmPoMuc* expression.

The chromatin configuration was further investigated by ChIP analysis using antibody that recognises heterochromatic and euchromatic marks. ChIP results clearly demonstrate that different epigenetic marks occur on the *SmPoMuc* promoter of group 1 and group 3.1 (r1-r2) between the IC and C strains likely resulting in a different chromatin configuration. On these

loci, chromatin is indeed more enriched in H3 acetylated on lysine 9 in the IC compared to the C strain and less enriched in the opposite mark H3 trimethylated on Lysine 9. Therefore, the local chromatin structures differ between the two strains for groups 1 and 3.1 (r1-r2) and are consistent with expression data as stronger acetylation correlates with enhanced expression.

Most current models of adaptive evolution are based on the Wright-Fisher model that assumes that random changes in the genetic information lead to changes in fitness towards or away from a fitness maximum in the fitness landscape. It is, however, now established that the phenotype is not only product of genetic processes, but expression of an ensemble that is composed of genetic and epigenetic components. Others and we have proposed that this additional system allows for rapid adaptive evolution without necessarily changing the genotype initially. A theoretical framework for this model was provided by Pal and Miklos (1999), and more recently by Klironomos, Berg and Collins (personal communication). Essentially, these authors propose that higher rate of random changes in epigenetic marks compared to genetic mutations transmitted from one generation to the next in a population generates increased phenotypic variations that can be selected for if the environment changes. In the numerical model of Klironomos, the phenotypes encoded by the low-fidelity (epigenetic) inheritance system reach the fitness maximum first, while the high-fidelity (genetic) system remains neutral until it catches up. For such a behavior as the term mutational genetic assimilation of a phenotype was proposed (Jablonka and Raz 2009). We have shown that in *S. mansoni* the augmentation of epigenetic variation leads indeed to an increase in phenotypic variation (Cosseau et al. 2010) providing evidence for the first assumption in the above-mentioned model. Here we show that two isolates of *S. mansoni* with distinct phenotypic traits, in particular their compatibility with a reference host, show low nucleotide differences in both coding sequence and promoters of *SmPoMuc* but high epigenetic differences in the promotor regions. Both parasite populations are in a situation where the

fitness value of genetically encoded phenotypes has not changed significantly, but epigenetic variation have produced phenotypic variants that are adapted to different environments (compatible hosts). Our experimental result goes in line with the Pal-Klironomos model. Our proposed scenario is as follows: in the 15th-16th century the ancestor strain of contemporary strains IC and C migrated via the slave trade from Africa to the West Indian Islands and the South American continent, respectively (Morgan et al. 2005). There, they had to adapt to a new intermediate host. The initial bottleneck resulting from the migration of only a limited number of parasites and the expected strong selective pressure acting on both genetic and epigenetic variants of the key-molecules for compatibility with the new snail hosts, *SmPoMucs*, may have reduced significantly genetic and epigenetic variation in the newly formed laboratory IC and C strains compared to the ancestral strain. Now, it is likely that epigenetic variation retained from the ancestral strain and the higher rate of occurrence of epigenetic changes in subsequent generations rather than the strain genetic variation enabled the parasite to adapt rapidly to their host and new environment. A conundrum with the "epigenetic mutation system first" hypothesis is that epigenetic information concerns the transcriptional activity of a gene but not its coding potential, in other words, a gene can be switched on and off by the surrounding chromatin but the resulting protein cannot be changed. Loss of function of genes can easily be imagined through an epigenetic mechanism, but for gain of function a complex inhibitor-based mechanism would be necessary. The classical Ohno hypothesis of gene duplications as way to provide material for evolution (Ohno et al. 1968) could deliver a solution. Rodin and Riggs have shown that duplicated genes have a tendency to be heterochromatic (Rodin et al. 2005). It is interesting to note that the *SmPoMuc* proteins essential for host compatibility are encoded by duplicated genes. Our analysis shows that the duplication events predate the IC/C separation and occurred in the strain's common ancestor, *i.e.* gene duplication was not a result of divergence of the two strains. We postulate

that *SmPoMuc* duplicated genes provide an additional system for phenotypic variation. Duplicated genes are randomly modulated in their relative transcriptional activity through chromatin structure changes as evidenced by our results and those of Cosseau et al. (2010), resulting in new combinations of *SmPoMuc* genes expressed and subsequent increased phenotypic variation. If the parasite encounters new intermediate hosts, the probability for the phenotypes to match is increased, thus allowing for adaptive evolution. Similar epigenetic-based variation generating systems, hence on a much smaller time-scale, are known from protozoan parasites such as *P. falciparum* or *T. brucei* (Verstrepen and Fink, 2009) but here to cope with the adaptive immune system of the final host. In contrast, the system described here for *S. mansoni* could be a general mode of evolution of metazoan.

A question that remains open and that we have not addressed in the present work concerns the origin of chromatin structure differences. One might argue that small nucleotide differences are sufficient to provoke modulation of histone modification. Such leverage effect of SNPs cannot be excluded but has so far not been observed in heavily studied models such as human, *Drosophila melanogaster* and *Arabidopsis thaliana*. It could also be that strain-specific loci exist that regulates chromatin structure of the *SmPoMuc* genes *in trans*. Our work does not exclude these possibilities but simply shows that in a gene family that codes for an adaptive phenotypic trait, epigenetic changes are more important than genetic changes. This finding provides support for theoretical models of adaptive evolution in which epimutations occur more rapidly than mutations.

**Supplementary material:**

Supplementary file 1. Origin of the sequences used for phylogenetic analysis of fig. 3.

Supplementary file 2. Primers used in this study

Supplementary file 3. Antibodies used in this study

Supplementary file 4. Expression of each *SmPoMuc* groups in C and IC strains and TSA effect.

Supplementary file 5. Immunoprecipitation of miracidia chromatin: Comparison of the chromatin state of the different group within a strain.

**Acknowledgements:**

The authors are indebted to Bernard Dejean and Anne Rognon for providing valuable technical support. The authors acknowledge Dr Jerome Boissier for his contribution to statistical analysis. The work received funding from the Program Schistophepigen ANR-07-BLAN-0119-02 of the French National Agency for Research and from the CNRS through a post-doctoral fellowship.

**Literature Cited:**

Adema CM, Hertel LA, Miller RD, Loker ES. 1997. A family of fibrinogen-related proteins that precipitates parasite-derived molecules is produced by an invertebrate after infection. *Proc Natl Acad Sci USA*. 94:8691-8696.

Azzi A, Cosseau C, Grunau C. 2009. *Schistosoma mansoni*: developmental arrest of miracidia treated with histone deacetylase inhibitors. *Exp Parasitol*. 121:288-291.

Bech N, Beltran S, Portela J, Rognon A, Allienne JF, Boissier J, Théron A. 2010. Follow-up of the genetic diversity and snail infectivity of a *Schistosoma mansoni* strain from field to laboratory. *Infect Genet Evol.* 10:1039-1045.

Belkhir K, Dawson KJ, Bonhomme F. 2006. A comparison of rarefaction and bayesian methods for predicting the allelic richness of future samples on the basis of currently available samples. *J Hered.* 97:483-492.

Bergman Y, Cedar H. 2010. Epigenetic control of recombination in the immune system. *Semin Immunol.* 22:323-329.

Bossdorf O, Arcuri D, Richards CL, Pigliucci M. 2010. Experimental alteration of DNA methylation affects the phenotypic plasticity of ecologically relevant traits in *Arabidopsis thaliana*. *Evol Ecol.* 24:541-553.

Bossdorf O, Richards CL, Pigliucci M. 2008. Epigenetics for ecologists. *Ecology Letters.* 11: 106-115.

Boyko A, Kovalchuk I. 2008. Epigenetic control of plant stress response. *Environ Mol Mutagen.* 49: 61-72.

Buscaglia CA, Campo VA, Frasch AC, Di Noia JM. 2006. *Trypanosoma cruzi* surface mucins: host-dependent coat diversity. *Nat Rev Microbiol.* 4:229-236.

Cosseau C, Azzi A, Rognon A, Boissier J, Gourbière S, Roger E, Mitta G, Grunau C. 2010. Epigenetic and phenotypic variability in populations of *Schistosoma mansoni* – a possible kick-off for adaptive host/parasite evolution. *Oikos*. 119: 669–678.

Cosseau C, Azzi A, Smith K, Freitag M, Mitta G, Grunau C. 2009. Native chromatin immunoprecipitation (N-ChIP) and ChIP-Seq of *Schistosoma mansoni*: Critical experimental parameters. *Mol Biochem Parasitol*. 166:70-76.

Danchin É, Charmantier A, Champagne FA, Mesoudi A, Pujol B, Blanchet S. 2011. Beyond DNA: integrating inclusive inheritance into an extended theory of evolution. *Nat Rev Genet*. 12:475-486.

De Mendonça RL, Bouton D, Bertin B, Escriva H, Noël C, Vanacker JM, Cornette J, Laudet V, Pierce RJ. 2002. A functionally conserved member of the FTZ-F1 nuclear receptor family from *Schistosoma mansoni*. *Eur J Biochem*. 269:5700-5711.

Dillon N. 2008. The impact of gene location in the nucleus on transcriptional regulation. *Dev Cell*. 15: 182-186.

Dubois F, Caby S, Oger F, Cosseau C, Capron M, Grunau C, Dissous C, Pierce RJ. 2009. Histone deacetylase inhibitors induce apoptosis, histone hyperacetylation and up-regulation of gene transcription in *Schistosoma mansoni*. *Mol Biochem Parasitol*. 168:7-15.

Felsenstein J. 2004. PHYLIP (Phylogeny Inference Package), version 3.6. Department of Genome Sciences, University of Washington, Seattle, Washington.

Gibbs MJ, Armstrong JS, Gibbs AJ. 2000. Sister-scanning: a Monte Carlo procedure for assessing signals in recombinant sequences. *Bioinformatics*. 16:573-582.

Grant-Downton RT and Dickinson HG. 2006. Epigenetics and its implications for plant biology 2: the “epigenetic epiphany”: epigenetics, evolution and beyond. *Ann. Bot.* 97:11-27.

Hanington PC, Forys MA, Dragoo JW, Zhang SM, Adema CM, Loker ES. 2010. Role for a somatically diversified lectin in resistance of an invertebrate to parasite infection. *Proc Natl Acad Sci U S A*. 107:21087-21092.

Jablonka E and Lamb MJ. 2005. *Evolution in Four Dimensions: Genetic, Epigenetic, Behavioral, and Symbolic Variation in the History of Life*. MIT Press, Cambridge.

Jablonka E, Lamb RM, Avital E. 1998. Lamarckian mechanisms in Darwinian evolution. *Trends Ecol Evol*. 13, 206-210.

Jablonka E, Raz G. 2009. Transgenerational epigenetic inheritance: prevalence, mechanisms, and implications for the study of heredity and evolution. *Q Rev of Biol*. 84: 131-176.

Johannes F, Colomé-Tatché M. 2011. Quantitative epigenetics through epigenomic perturbation of isogenic lines. *Genetics*. 188:215-227.

Kohany O, Gentles AJ, Hankus L, Jurka J. 2006. Annotation, submission and screening of repetitive elements in Repbase: RepbaseSubmitter and Censor. *BMC Bioinformatics*. 7:474.

Kulakova L, Singer SM, Conrad J, Nash TE. 2006. Epigenetic mechanisms are involved in the control of *Giardia lamblia* antigenic variation. *Mol Microbiol.* 61:1533-1542.

Lachmann M, Jablonka E. 1996. The inheritance of phenotypes: an adaptation to fluctuating environments. *J Theor Biol.* 181:1-9.

Lee JS, Smith E, Shilatifard A. 2010. The language of histone crosstalk. *Cell.* 142: 682-685.

Luijsterburg MS, White MF, van Driel R, Dame RT. 2008. The major architects of chromatin: architectural proteins in bacteria, archaea and eukaryotes. *Crit Rev Biochem Mol Biol.* 43: 393-418.

Martin DP, Lemey P, Lott M, Moulton V, Posada D, Lefevre P. 2010. RDP3: a flexible and fast computer program for analyzing recombination. *Bioinformatics.* 26:2462-2463.

Mitta G, Adema CM, Gourbal B, Loker ES, Theron A. 2011. Compatibility polymorphism in snail/schistosome interactions: From field to theory to molecular mechanisms. *Dev Comp Immunol.* in press.

Moné Y, Gourbal B, Duval D, Du Pasquier L, Kieffer-Jaquinod S, Mitta G. 2010. A large repertoire of parasite epitopes matched by a large repertoire of host immune receptors in an invertebrate host/parasite model. *PLoS Negl Trop Dis.* 4. pii: e813.

Morgan JA, Dejong RJ, Adeoye GO et al. (33 co-authors) 2005. Origin and diversification of the human parasite *Schistosoma mansoni*. *Mol Ecol*. 14:3889-3902.

Nei M. 1987. *Molecular Evolutionary Genetics*. Columbia University Press, New York.

Ohno S, Wolf U, Atkin NB. 1968. Evolution from fish to mammals by gene duplication. *Hereditas*. 59:169-187.

Osipovich O, Oltz EM. 2010. Regulation of antigen receptor gene assembly by genetic-epigenetic. *Semin Immunol*. 2010. 22:313-322.

Pal C, Miklos I: 1999. Epigenetic inheritance, genetic assimilation and speciation. *J Theor Biol*. 200:19-37.

Pigliucci M, Murren CJ, Schlichting CD. 2006. Phenotypic plasticity and evolution by genetic assimilation. *J Exp Biol*. 209:2362-2367.

Portela J, Grunau C, Cosseau C, Beltran S, Dantec C, Parrinello H, Boissier J. 2010. Whole-genome in-silico subtractive hybridization (WISH)--using massive sequencing for the identification of unique and repetitive sex-specific sequences: the example of *Schistosoma mansoni*. *BMC Genomics*. 11:387.

Prucca CG, Slavin I, Quiroga R, Elias EV, Rivero FD, Saura A, Carranza PG, Lujan HD. 2008. Antigenic variation in *Giardia lamblia* is regulated by RNA interference. *Nature*. 456:750-754.

Rapp RA, Wendel JF. 2005. Epigenetics and plant evolution. *New Phytologist*. 168:81–91.

Reese MG. 2001. Application of a time-delay neural network to promoter annotation in the *Drosophila melanogaster* genome. *Comput Chem*. 26:51-56.

Richards EJ. 2006. Inherited epigenetic variation – revisiting soft inheritance. *Nat. Rev. Genet*. 7:395-401.

Rodin SN, Parkhomchuk DV, Riggs AD. 2005. Epigenetic changes and repositioning determine the evolutionary fate of duplicated genes. *Biochemistry (Mosc)*. 70:559-567.

Roger E, Gourbal B, Grunau C, Pierce RJ, Galinier R, Mitta G. 2008a. Expression analysis of highly polymorphic mucin proteins (*SmPoMuc*) from the parasite *Schistosoma mansoni*. *Mol Biochem Parasitol*. 157:217-227.

Roger E, Grunau C, Pierce RJ, Hirai H, Gourbal B, Galinier R, Emans R, Cesari IM, Cosseau C, Mitta G. 2008b. Controlled chaos of polymorphic mucins in a metazoan parasite (*Schistosoma mansoni*) interacting with its invertebrate host (*Biomphalaria glabrata*). *PLoS Negl Trop Dis*. 2:e330.

Roger E, Mitta G, Moné Y, Bouchut A, Rognon A, Grunau C, Boissier J, Théron A, Gourbal BE. 2008c. Molecular determinants of compatibility polymorphism in the *Biomphalaria glabrata*/*Schistosoma mansoni* model: new candidates identified by a global comparative proteomics approach. *Mol Biochem Parasitol.* 157: 205-216.

Ronquist F, Teslenko M, van der Mark P, Ayres D, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. 2012. MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol.* 3:539-342.

Rousset F. 1996. Equilibrium values of measures of population subdivision for stepwise mutation processes. *Genetics.* 142:1357-1362.

Roux F, Colomé-Tatché M, Edelist C, Wardenaar R, Guerche P, Hospital F, Colot V, Jansen RC, Johannes F. 2011. Genome-wide epigenetic perturbation jump-starts patterns of heritable variation found in nature. *Genetics.* 188:1015-1017.

Rozas J, Sánchez-DelBarrio JC, Messeguer X. 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics.* 19:2496-2497.

Rudenko G. 2011. African trypanosomes: the genome and adaptations for immune evasion. *Essays Biochem.* 51:47-62.

Scherf A, Lopez-Rubio JJ, Riviere L. 2008. Antigenic variation in *Plasmodium falciparum*. *Annu Rev Microbiol.* 62:445-70.

Slatkin M. 1995. A measure of population subdivision based on microsatellite allele frequencies. *Genetics*. 139:457-462.

Tamura K, Dudley J, Nei M, Kumar S. 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol*. 24:1596-1599.

Theron A, Coustau C. 2005. Are Biomphalaria snails resistant to *Schistosoma mansoni*? *J Helminthol*. 79:187-191.

Theron A, Pages JR, Rognon A. 1997. *Schistosoma mansoni*: distribution patterns of miracidia among *Biomphalaria glabrata* snail as related to host susceptibility and sporocyst regulatory processes. *Exp Parasitol*. 85:1-9.

Uchida S, Hara K, Kobayashi A, Otsuki K, Yamagata H, Hobara T, Suzuki T, Miyata N, Watanabe Y. 2011. Epigenetic status of Gdnf in the ventral striatum determines susceptibility and adaptation to daily stressful events. *Neuron*. 69:359-72.

Umlauf D, Fraser P, Nagano T. 2008. The role of long non-coding RNAs in chromatin structure and gene regulation: variations on a theme. *Biol Chem*. 389:323-331.

Van Valen, L. 1974. Molecular evolution as predicted by natural selection. *J. Mol. Evol*. 3:89-101.

Verhoeven KJ, Jansen JJ, van Dijk PJ, Biere A. 2010. Stress-induced DNA methylation changes and their heritability in asexual dandelions. *New Phytol.* 185:1108-18.

Verstrepen KJ, Fink GR. 2009. Genetic and epigenetic mechanisms underlying cell-surface variability in protozoa and fungi. *Annu Rev Genet.* 43:1-24.

Waddington CH. 1957. *The Strategy of the Genes: A Discussion of Some Aspects of Theoretical Biology.* London (UK): Allen & Unwin.

Zhang SM, Zeng Y, Loker ES. 2008. Expression profiling and binding properties of fibrinogen-related proteins (FREPs), plasma proteins from the schistosome snail host *Biomphalaria glabrata*. *Innate Immun.* 14:175-89.

Table 1. Sequence variation between *Schistosoma mansoni* IC and C strains and between groups of *SmPoMuc* promoter sequences.

Sequences	Seq. length (Kb)	<i>N</i>	Nb polym. sites	Nb mutat.	Nb fixed diff. IC vs. C	<i>D<sub>a</sub></i> IC vs. C	<i>D<sub>a</sub></i> group 1 vs.	<i>D<sub>a</sub></i> group 2 vs.	<i>D<sub>a</sub></i> group 3.1 vs.	<i>D<sub>a</sub></i> group 3.1(r1-r2) vs.
<i>SmPoMuc</i> promoters										
Group 1	1.80	2	4	4	4	0.002	-	0.021	0.031	0.035
Group 2	1.74	3	0	0	0	0.000	-	-	0.040	0.026
Group 3.1	1.04	3	4	4	3	0.003	-	-	-	0.040
Group 3.1(r1-r2)	2.00	2	8	8	8	0.004	-	-	-	-
<i>SmPoMuc</i> introns										
Group 1	7.14	2	12	16	16	0.002	-	0.068	0.031	
Group 2	7.72	2	32	32	32	0.004	-	-	0.068	
Group 3.1	7.11	2	43	43	43	0.006	-	-	-	
<i>SmFTZ-F1</i> promoter										
<i>SmFTZ-F1</i>	1.4	3	0	0	0	0.000				

Seq. Length, sequence length;  $N$ , number of sequences; Nb polym. sites and Nb mutat., number of polymorphic sites and mutations, respectively; Nb fixed diff. IC vs. C, number of nucleotide differences between IC and C strains' promoter sequences;  $D_a$  IC vs. C, number of net nucleotide substitutions per site between IC and C strains' promoter sequences;  $D_a$  group X vs., number of net nucleotide substitutions per site between groups of promoter sequences.

Table 2. Sequence diversity of the promoter of *SmPoMuc* group 1 genes for *Schistosoma mansoni* IC and C strains.

	<i>N</i>	<i>A</i>	<i>S</i>	Gen. Div. $\pm$ SD	Nuc. Div. $\pm$ SD	Tajima's <i>D</i>	<i>P</i>
IC	20	1	0	0.00 $\pm$ 0.00	0.000 $\pm$ 0.000	0.000	1.000
C	18	3	2	0.22 $\pm$ 0.12	0.000 $\pm$ 0.000	-1.508	0.048

*N*, number of *S. mansoni* samples; *A*, number of alleles; *S*, number of substitutions; Gen. Div.  $\pm$  SD and Nuc. div.  $\pm$  SD, expected unbiased gene diversities and nucleotide diversities, respectively. Tajima's *D*, *P*, Tajima's *D* and test of significance of *D*, respectively.

Sequence length: 1.4kb.

Table 3. Genetic differentiation between *Schistosoma mansoni* IC and C strains estimated by  $\Phi_{ST}$  and  $\theta$ , and inbreeding coefficients *f* within strains.

	$\Phi_{ST} / R_{ST}$	$\theta$	<i>f</i> in IC	<i>f</i> in C
<i>SmPoMuc</i> promoter group 1	0.974 <sup>S</sup>	0.898 <sup>S</sup>	N/A	1.00 <sup>S</sup>
Microsatellites	six loci N/A (one allele) eight loci 1.000 <sup>S</sup>	six loci N/A (one allele) eight loci 1.000 <sup>S</sup>	fourteen loci N/A	N/A

<sup>S</sup>: significant departure from 0 at *P* = 0.05.

**List of figures:**

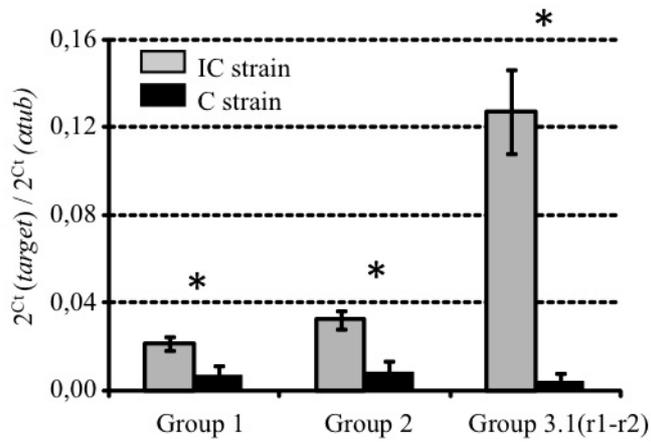


Figure 1: *SmPoMuc* expression of each group in C and IC strains

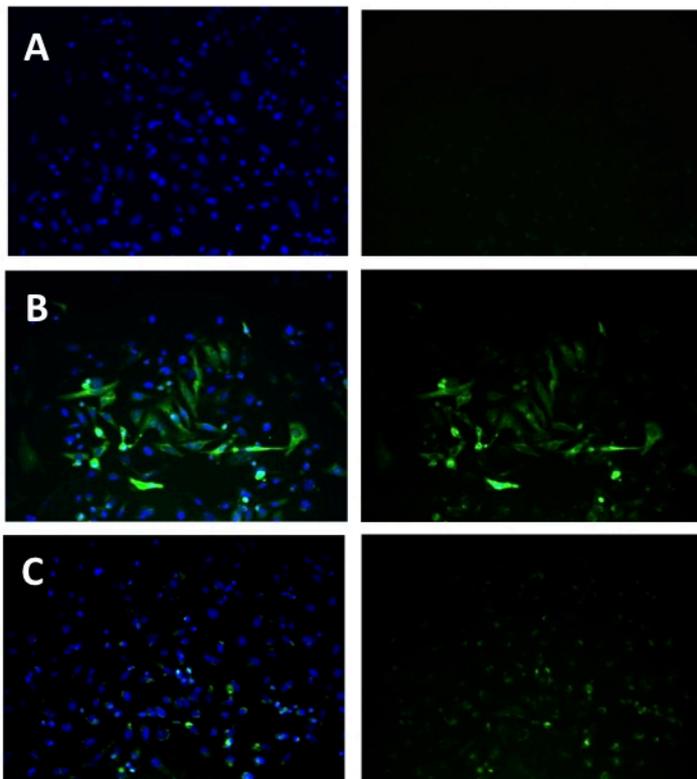


Figure 2: Expression of EGFP under control of the *SmPoMuc* group 3.1(r1-r2) promoter in HeLa cells

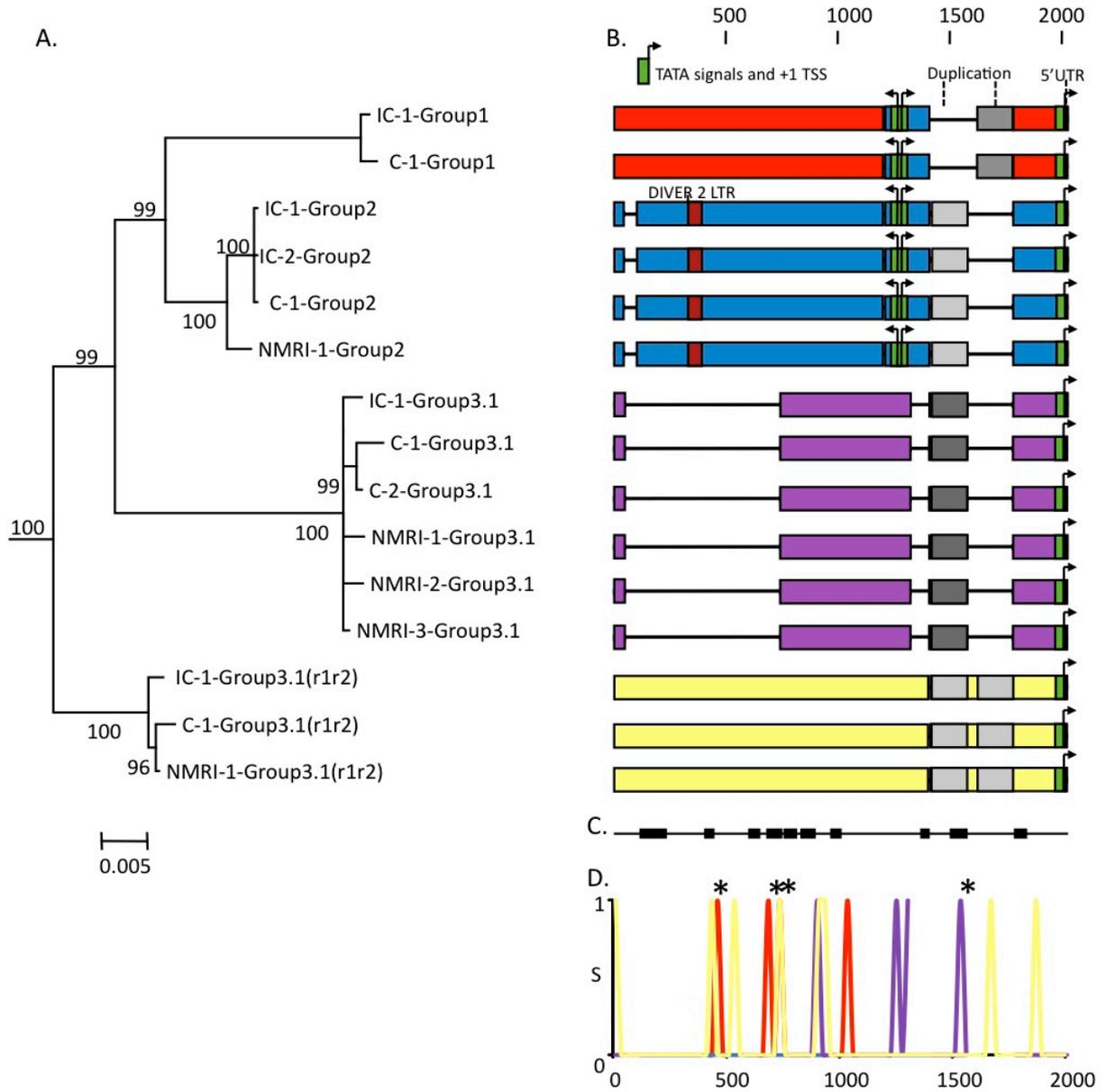


Figure 3: Promoters of *SmPoMuc* group 1, 2, 3.1 and 3.1(r1-r2), paralogous and orthologous relationships among groups between *Schistosoma mansoni* IC and C strains.

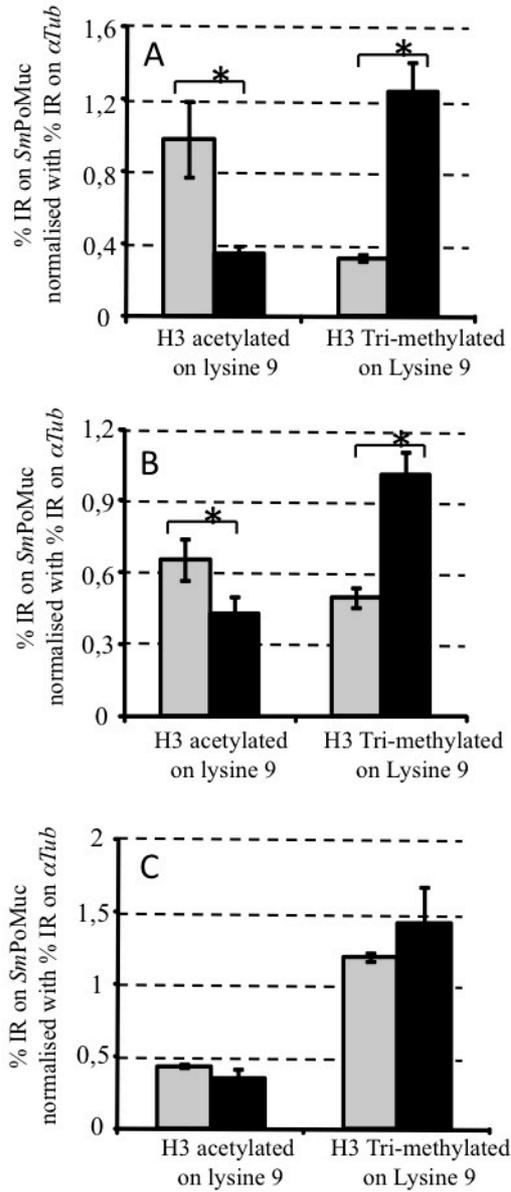


Figure 4: Immunoprecipitation of miracidia chromatin: Comparison of the chromatin state of each group between C and IC strains.

**Legends:**

Figure 1: mRNA were extracted from miracidia pool from the IC and C strain and qPCR were performed with primers targeting *SmPoMuc* group 1, group 2, group 3.1 (r1-r2). Results represents the mean value of 3 biological repeats, \* indicates a p value below 0.05.

Figure 2: HeLa cells were transiently transfected with (A) promoterless pEGFP1 vector backbone, (B) pCMV-EGFP or (C) pSmPoMuc-EGFP. Cell nuclei were labelled with DAPI. Visualization of the fluorescence separately (right panel) or overlaid (left panel) is presented. Magnification is x400.

Figure 3:

(A.) Bayesian analysis of phylogenetic relationship among *SmPoMuc* promoter sequences with scores as posterior probability values above 70 indicated an associated nodes. (B.) Schematic diagram of aligned *SmPoMuc* promoter sequences corresponding to sequences in panel A. Numbers show the nucleotide position in relation to the transcriptional start site in the alignment. We annotated the sequences by colour-coding blocks of less than 95% identity: Group 1 (red), Group 2 (blue), Group 3.1 (purple) and Group 3.1(r1-r2) (yellow). The 5'UTR were characterised and are represented in orange. TATA signals, here in green, and Transcription Starting Sites (+1 TSS) were predicted using Neural Network Promoter Prediction Tool. Deletions are represented by black lines. A recombination event was detected from Group 2 to Group 1 promoter sequences (in blue in Group 1 sequences). One duplication event resulted in an insertion in Group 3.1(r1-r2) (in greys). Traces of a retrotransposon insertion (DIVER2 – LTR) is present in Group 2 (in burgundy). (C.) Black blocks are the conserved regions among groups of *SmPoMuc* promoters, ignoring the variation within

groups of promoters (nucleotide positions as above). (D.) Number of substitutions per site among promoter sequences between strains within *SmPoMuc* groups (S on Y-axis, colour-codes described above), along the sequence alignment (nucleotide positions on X-axis as above). \* are the substitutions between the two strains positioned within region conserved among groups of promoters. The number of substitutions among *SmPoMuc* promoters between the two strains varied from 0 in Group 2 to 8 in Group 3.1(r1-r2). No substitution was observed in TATA signals and TSS sites between the two strains. The sequences have GenBank accession number JQ615951 to JQ615965 (See supplementary file 1 for details).

Figure 4: ChIP experiments were performed on chromatin isolated from miracidia from both the IC (grey bars) and C strain (black bars) with antibodies against H3 acetylated on lysine 9 and H3 tri-methylated on lysine 9. Immunoprecipitated chromatin was analysed by qPCR using primers that hybridize with specific sequences of *SmPoMuc* group 3.1(r1-r2) (panel A), group 1 (panel B) and group 3.1 (panel C). Results represent the percentage input recovery (%IR) normalised with %IR of a reference gene ( $\alpha$ Tub). Results are the average of 3 biological samples. \* indicates a p value < 0,05 of a student ttest.

## ANNEXE 6

Curriculum vitae

**Lepesant Julie M.J.**

27 rue du Stade  
66690 Saint André  
FRANCE

Phone: + 33 (0)6 74 13 64 28

Email : [julie.lepesant@free.fr](mailto:julie.lepesant@free.fr)  
: [julie.lepesant@univ-perp.fr](mailto:julie.lepesant@univ-perp.fr)

[http://2ei.univ-perp.fr/?page\\_id=960](http://2ei.univ-perp.fr/?page_id=960)

Born December, 22<sup>th</sup>, 1986

## Curriculum vitae

### FORMATION

---

- 2009-2012    **PhD Student (present)**  
Group: Environmental and Evolutionary Epigenetics of *Schistosoma mansoni*  
at the UMR 5244 CNRS-UPVD (Perpignan)
- 2007-2009    **Master Integrative Biology and Physiology (with honor)**  
**Major: Comparative Biology and Physiology**  
University Pierre and Marie Curie (Paris IV)
- 2006-2007    **Third year of Organisms Biology Licence (Undergraduate)**  
**Major: Aquatic Ecology**  
University of Montpellier (UM II)
- 2004-2006    **First and second year of Biology Licence (Undergraduate)**  
University Joseph Fourier (Grenoble I)
- 2004          **Scientific High School Diploma, Major: Biology, European section (teaching in French and German)**  
Highschool Champollion of Grenoble

### RESEARCH EXPERIENCES

---

- 2009-2012    **PhD Student**
- **Laboratory:** Ecology and Evolution of Interactions UMR 5244 CNRS-UPVD (Perpignan)
  - **Group:** Environmental and Evolutionary Epigenetics of *Schistosoma mansoni*
  - **Subject:** The role of epigenetics in the generation of phenotypic variability in *Schistosoma mansoni* and its impact on infection success. Supervisors Mr. Guillaume Mitta (PhD) and Mr. Christoph Grunau (PhD).

- **Techniques:** Dissections, DNA and RNA extractions, QPCR, RT-PCR, Chromatin Immunoprecipitation, Massively parallel sequencing
- 2009 (6 months)      Research training in Arago Laboratory in Banyuls sur Mer (Master 2)
- **Laboratory:** Factors of the Mediums and Adaptive Mechanisms
  - **Subject:** Study of the circadian rhythms and the gonadotropines in the seabass *Dicentrarchus labrax*. Supervisor Mr Jack Falcon (PhD)
  - **Techniques:** Dissections of fishes brains, proteins extraction and quantification, enzymatic activity, QPCR, cloning, In Situ Hybridization
- 2008 (3 months)      Research training in Ecolag Laboratory in Montpellier II (Master 1)
- **Laboratory:** Immunity Reactions, Aquaculture and Environment
  - **Subject:** Contribution under investigation of the immune reaction in oyster *Crassostrea gigas*. Supervisor Ms Viviane Boulo (PhD).
  - **Techniques:** RNA extractions, RT-PCR, QPCR, Agarose gels, Northern Blot
- 2007 (1 month)      Voluntary Research training in Ecolag Laboratory in Montpellier II (Master 1)
- **Laboratory:** Ecophysiological Adaptation and Ontogenesis
  - **Subject:** Study of the physiological adaptations of the seabass *Dicentrarchus labrax* to variations of salinity, Supervisor Ms Patricia Cucchi-Mouillot (PhD).
  - **Technics:** Fishery, immunohisto/cytochemistry, histology, Western Blot

## TEACHING EXPERIENCES

---

2009-2012      **Practical classes** of Animal Physiology (Licence), Animal cellular Physiology (Licence), Informatic certificate (Licence), Bibliography studies (Licence), Molecular Biology (Master) and supervision of licence and master student for training in laboratories

## PUBLICATIONS LIST

---

2012

**Julie M.J. Lepasant**, David Roquis, Rémi Emans, Céline Cosseau, Nathalie Arancibia, Guillaume Mitta, Christoph Grunau. De-novo assembly of massive sequencing reads into the repeat genome of *Schistosoma mansoni* and first-pass annotation. *Experimental Parasitology*. {IF 2.122}

**Julie M.J. Lepasant**, Céline Cosseau, Jérôme Boissier, Michael Freitag, Julien Portela, Déborah Climent, Cécile Perrin, Adhemar Zerlotini Neto, Christoph Grunau. The role of heterochromatization in sex chromosome emergence: lessons from the human parasite *Schistosoma mansoni*. *Genome Biology*. {IF 9.036}.

Cécile Perrin, **Julie M.J. Lepsant**, Emmanuel Roger, David Duval, Virginie Thuillier Jean-François Alliene, Christoph Grunau, Guillaume Mitta, Céline Cosseau. Promoter regions of duplicated *Schistosoma mansoni* genes that are crucial for infection success undergo rapid epigenetic modifications in different environments but remain genetically stable (under revision).

**Julie M.J. Lepsant**, Jérôme Boissier, Déborah Climent, Anne Rognon, Céline Cosseau, Christoph Grunau. Experimental observations of life traits during the life cycle of *Schistosoma mansoni* infecting sympatric or allopatric intermediate host (in prep.)

## 2011

**Julie M.J. Lepsant**, Céline Cosseau, Christoph Grunau. Towards an understanding of the epigenetics of schistosomes: A comparative epigenomic study. *Memorias do Instituto Oswaldo Cruz* 106(7): 823-30 {IF 2.097}

## COMMUNICATIONS

### ORAL PRESENTATIONS

## 2012

**Lepsant J.M.J.**, Cosseau C., Boissier J., Freitag M., Portela J., Climent D., Perrin C., Zerlotini A., Mitta G. & Grunau C. (2012) Possible role of epigenetic mechanisms in sex chromosome emergence in *Schistosoma mansoni*, a human parasite. EMOP XI (session IV K.E. Mott symposium on schistosomiasis and food borne trematodiasis / A. Schistosomiasis), Cluj Napoca, Roumania, July 24-29 2012

**Lepsant J.M.J.**, Cosseau C., Roquis D., Boissier J., Théron A., Rognon A., Mitta G. & Grunau C. (2012) A role for epigenetics in the adaptive evolution of a human parasite, *Schistosoma mansoni*, in response to environmental stress. EMOP XI (session IV K.E. Mott symposium on schistosomiasis and food borne trematodiasis / A. Schistosomiasis), Cluj Napoca, Roumania, July 24-29 2012

Grunau C., Cosseau C., **Lepsant J.M.J.**, Fneich S., Roquis D. & Mitta G. (2012) The épigénome of *Schistosoma mansoni* and its plastic response to environmental stress. 13rd International Symposium of Schistosomiasis. Belo Horizonte, Brazil, September 16-19 2012

## 2010

**Lepsant J.M.J.**, Cosseau C., Perrin C., Mitta G. & Grunau C. (2010) Rôle de l'épigénétique dans l'évolution adaptative. Congrès Ecologie 2010 (session : Intégration de l'interaction Génotype x Environnement dans l'étude des capacités d'adaptation), Montpellier, France, September 2-4 2010

**Lepesant J.M.J.**, Cosseau C., Perrin C., Mitta G. & Grunau C. (2010) A role for epigenetics in adaptative evolution. 13e congrès Young Researchers in Life Science (session: Host-pathogen interactions), Paris, France, June 7-9 2010

## POSTERS

### 2012

Fneich S., **Lepesant J.M.J.**, Cosseau C., Reichelt M., Dheilly N., Dejan B. & Grunau C. (2012) DNA méthylation in *Biomphalaria glabrata* the intermediate host of a human parasite *Schistosoma mansoni* : rôle of epigenetics in the coévolution of host/parasite interactions. SMBE (Society for Molecular Biology & Evolution), Dublin, Ireland, June 23-26 2012

Roquis D., **Lepesant J.M.J.**, Clément J., Cosseau C., Emans R., Mitta G. & Grunau C. (2012) Making next generation sequencing a routine tool for epigenome, genome and transcriptome studies : challenges and solutions. 13rd International Symposium of Schistosomiasis. Belo Horizonte, Brazil, September 16-19 2012

### 2011

**Lepesant J.M.J.**, Cosseau C., Boissier J., Rognon A. & Grunau C. (2011) Changes in the epigenome of a human parasite during experimental adaptive evolution: antihelminth drug. ESF-EMBO Symposium : Epigenetics in Context: From Ecology to Evolution, Sant Feliu de Guixols, Spain, September 18-23 2011

Herrero M.J., Isorna E., Besseau L., Fuentes M., **Lepesant J.M.J.**, Sauzet S., Delgado M.J. & Falcon J. (2011) Daily variations of clock genes : testing endogenous rhythms and influence of exogenous melatonin in pituitary of sea bass (*Dicentrarchus labrax*). 8th congress of iberic association of compared endocrinology, Madrid, Spain, September 5-7 2011

### 2010

**Lepesant J.M.J.**, Cosseau C., Portela J., Boissier J., Dejean B. & Grunau C. (2010) Chromatin status and transcription patterns of several new classes of female specific repeats of *Schistosoma mansoni*. ICOPA XII, Melbourne, Australia, August 15-20 2010

### 2009

Grunau C., Boissier J., Cosseau C., **Lepesant J.M.J.**, Portela J., Théron A., Rognon A., Dejean B., Arancibia N. & Mitta G. (2009) What makes a girl a girl ? ASTMH, Washington, United States, November 2009

## SCIENTIFIC VULGARISATION

**2012            Organization of a regional congress for PhD students in Perpignan.**

**2009            Joining author of the teaching case “Darwin, l’évolution c’est tout bête”.**

Case financed by CNRS, composed of 5 posters and 4 workshops (heredity, DNA, natural selection, tree of life), realized within the framework of a partnership between UMR CNRS 5096 and 5244 and association “Les Petits Débrouillars”. Inaugurated and animated at the time of the village of Sciences 2009 (November 23-23) at the convention center of Perpignan then diffused via the association in the school structures.

## **INFORMATICS**

---

- MAC et PC : Microsoft Word, Excel, Power Point, Keynote, Internet, free software R, GraphPad, logiciel LightCycler

## **OTHERS**

---

- French (maternal language)
  - English (read, spoken and written)

## **REFEREES**

---

- Dr. Grunau Christoph, PhD supervisor
  - [christoph.grunau@univ-perp.fr](mailto:christoph.grunau@univ-perp.fr)
  - +33 4 68 66 21 80
- Prof. Mitta Guillaume, Unit director
  - [mitta@univ-perp.fr](mailto:mitta@univ-perp.fr)
  - +33 4 68 66 21 85



## Résumé

La bilharziose est une maladie parasitaire humaine qui se trouve au second rang mondial, après la malaria, en termes de morbidité et de mortalité. Plus de 200 millions de personnes sont infestées dans 74 pays et plus de 200 000 personnes décèdent par an (Chitsulo *et al.* 2004; Gryseels *et al.* 2006). *Schistosoma mansoni*, ici étudié, est responsable de la bilharziose intestinale.

Ce parasite possède un cycle de vie complexe, passant obligatoirement à travers deux hôtes : un hôte intermédiaire mollusque d'eau douce et un hôte définitif humain ou murin. Mes travaux de thèse concernent l'influence de l'environnement sur l'épigénome de *Schistosoma mansoni*. L'information épigénétique est une source supplémentaire pour la création de variants phénotypiques. Nous proposons que certains de ces variants puissent être hérités et sélectionnés, et ainsi conférer un avantage au parasite.

Pour investir cette hypothèse nous avons décidé de modifier l'environnement du parasite de façon biotique et abiotique, lorsqu'il se développe au sein de l'hôte définitif ou lorsqu'il est au contact de l'hôte intermédiaire. En effet, *Schistosoma mansoni* est un endoparasite donc chaque hôte, dans lequel il va se maintenir, représente son environnement direct. Pour cela nous avons utilisé une drogue anti-helminthe et un hôte intermédiaire allopatrique.

La réponse du parasite aux différents changements d'environnements a été analysée de la population (traits d'histoire de vie) à la molécule (génome, épigénome, transcriptome). Les nouvelles techniques de biologie moléculaire et de séquençage massif nous ont permis de réaliser des études globales.

Lors de ces travaux de thèse, les différentes évolutions expérimentales ont mené à deux énigmes : (1) « **Adaptation rapide de *Schistosoma mansoni* à une drogue anti-helminthes** » et (2) « **Adaptation rapide de *Schistosoma mansoni* à l'Amérique du Sud** ». Les résultats obtenus pour cette dernière ont amené une troisième énigme : (3) « **Différenciation sexuelle en l'absence de gène W-spécifique dans un système ZW** ».

