Frequency and mitotic heritability of epimutations in Schistosoma mansoni
David Roquis, Anne Rognon, Cristian Chapparo, Jérôme Boissier, Nathalie Arancibia, Céline Cosseau, Hughes Parrinello, Christoph Grunau

To cite this version:

HAL Id: hal-01311089
https://hal-sde.archives-ouvertes.fr/hal-01311089
Submitted on 1 Apr 2019

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.
Frequency and mitotic heritability of epimutations in Schistosoma mansoni

David Roquis1,2, Anne Rognon1,2, Cristian Chapparo, Jerome Boissier1,2, Nathalie Arancibia1,2, Celine Cosseau1,2, Hugues Parrinello3 & Christoph Grunau1,2

1 Université de Perpignan Via Domitia, Perpignan, F-66860, France,
2 CNRS, UMR 5244, Interactions Hôtes-Pathogènes-Environnements (IHPE), Perpignan, F-66860, France,
3 CNRS, GenomiX IBiSA, Montpellier, France,

Abstract

Schistosoma mansoni is a parasitic platyhelminth responsible for intestinal bilharzia. It has a complex life cycle, infecting a freshwater snail of the Biomphalaria genus, and then a mammalian host. S. mansoni adapts rapidly to new (allopatric) strains of its intermediate host. To study the importance of epimutations in this process we infected sympatric and allopatric mollusk strains with parasite clones. ChIP-Seq was done on four histone modifications (H3K4me3, H3K27me3, H3K27ac and H4K20me1) in parallel with genome-wide DNA resequencing (i) on parasite larvae shed by the infected snails, and (ii) on adult worms that had developed from the larvae. No change in single nucleotide polymorphisms (SNP) and no mobilization of transposable elements (TE) was observed, but 58 – 105 copy number variations (CNV) within the parasite clones in different mollusks were detected. We also observed that the allopatric environment induces three types of chromatin structure changes: (i) host-induced changes on larvae epigenomes in 51 regions of the genome that are independent of the parasites’ genetic background, (ii) spontaneous changes (not related to experimental condition or genotype of the parasite) at 64 locations, and (iii) 64 chromatin structure differences dependent on the parasite genotype. Up to 45 % of the spontaneous, but none of the host-induced chromatin structure changes were transmitted to adults. In our model, the environment induces epigenetic changes at specific loci but only spontaneous epimutations are mitotically heritable and have therefore the potential to contribute to transgenerational inheritance. We also show that CNV are the only source of genetic variation and occur at the same order of magnitude as epimutations.
Introduction

Adaptive evolution relies on the generation of heritable phenotypic variations. There is now an intense scientific discussion about the source of these variants (Laland et al. 2014). During the last couple of years, it has become clear that there are considerable non-genetic sources of heritable phenotypes (Jablonka et al. 1992; Jablonka & Lamb 2008; Danchin et al. 2011). In particular epigenetic modifications, that we define here tentatively as heritable changes in chromatin structure that can have an impact on gene expression, are in the center of a lively scientific debate (Kovalchuk 2011). Particularities of epimutations are (I) their metastability, i.e. changes in chromatin structure can be reversible and (II) their relations with the environment. In the genetic inheritance system, the environment acts solely as selective agent, but in the epigenetic system it could play a dual role, acting as an inducer of epimutations (Guerrero-Bosagna & Skinner 2011; Skinner et al. 2015) as well as applying selection on the resulting phenotypes. The question whether epimutations can generate new phenotypes is solved (Peaston & Whitelaw 2006; Nilsson & Skinner 2015). There are now many examples for phenotypes that are entirely or partially encoded by epialleles presenting changes in DNA methylation or histone modifications (Duncan et al. 2014). The role of the environment, however, is far from being comprehended. There is empirical support that epimutations are generated spontaneously, independently of environment but at a higher frequency than genetic mutations (van der Graaf et al. 2015). It was also argued that environmental stress could provoke adaptive and heritable changes through targeted épimutations. Others concluded that stress leads to an increase of the epimutation rate, thus representing an ephemeral, but inducible version of the genetic inheritance system, i.e. a way to induce reversible but possibly inheritable phenotypic changes (Jiang et al. 2014). The objective of the present study was to determine the frequency and nature of histone-based epimutations during a change of environment, and to measure their mitotic heritability. To address this question we used a parasite-host system. In these systems reciprocal selective forces, such as infectivity for the parasite and resistance for the host, are strong and adaptive evolution is expected to be fast (Thrall et al. 2007; Tellier et al. 2014).

The plathelminth Schistosoma mansoni is the causing agent of intestinal schistosomiasis, a human disease affecting more than 67 million people worldwide (Oliveira et al. 2004; King 2010). Its complex life cycle (Supplementary Figure 1) starts with eggs hatching in fresh water. A free-swimming larvae (miracidium) is released and will enter the tegument of the intermediate mollusk host of the Biomphalaria genus. Miracidia quickly transform into primary sporocysts, multiply asexually and differentiate into secondary sporocysts. They generate multiple cercariae, another free-swimming larvae, which leave the mollusk, infect the definitive mammalian host (rodent or primate), and transform into schistosomula. They follow a complex maturation and migration process in the mammalian body to reach the adult worm stage. Male and female worms form couples in the mesenteric veins, where eggs are released, go through the intestine wall and are excreted in the feces. S. mansoni can infect a wide range of intermediate and definitive host species (Martins 1957; DeJong et al. 2001). Changes in the intermediate host strain and species has a reproducible phenotypic impact on the parasite, and more specifically on its life history traits (sex-ratio, growth, prevalence and intensity of the infection) (Dias et al. 1988; Lepesant et al. 2007).
2013; Theron et al. 2014), but the underlying mechanisms are still unknown. Compatibility polymorphism is classically used to explain infectivity between *S. mansoni* and *B. glabrata* (Roger et al. 2008; Mitta et al. 2012). However, recent work has uncovered that epigenetic modifications could also play a role in the adaptation to new host strains (Knight et al. 2013). Knowing this, we were interested to find out what would be the impact of different host strains on the genome and epigenome of *S. mansoni*. In order to disentangle the impact of epimutations and mutations, we decided to work with genetically identical parasites and to let them develop in different snails. Miracidia, coming from sexual reproduction, are not clonal, and although it is relatively easy to infect snails with sibling miracidia with low genetic diversity between them, there will still be genetic diversity. To overcome this problem, we used sporocyst grafts (Jourdane & Theron 1980; Jourdane 1984). This technique allows, after a mono-miracidium infection, to extract secondary sporocysts and graft them into new mollusks. Secondary sporocysts revert to primary sporocysts once grafted, and then undergo the normal development until cercariae emission (Jourdane & Theron 1980; Jourdane 1984).

Here, we used a parasite strain with known genetic background (*Sm*Bre from Brazil; Clément et al. 2013) to infect a sympatric *Biomphalaria glabrata* strain (*Bg*Bre from Brazil) and an allopatric one (*Bg*Gua from Guadeloupe). Interactions between *Sm*Bre and the two mollusk strains have been previously characterized in great detail and the differences in life traits are known (Lepesant et al. 2013; Theron et al. 2014). Prevalence in the sympatric interaction (*Sm*Bre on *Bg*Bre) is 50-100%, and 30-85% in the allopatric interaction (*Sm*Bre on *Bg*Gua), depending on the number of miracidia used for infection (Theron et al. 2014). This shift in the environment does not represent a strong stress but has still measurable effects on the life history traits of the parasite (Lepesant et al. 2013). We sequenced the genomes of cercariae and adult parasites to search for somatic mutations, copy number variations and mobilization of transposable elements. DNA methylation is only present at a very low level in *S. mansoni* (Geyer et al. 2011), which allowed us to focused on chromatin marks based on histone modifications. We used chromatin immunoprecipitation followed by sequencing (ChIP-Seq). ChIP-Seq was done on cercariae emitted after grafts, but also on adult worms obtained by infecting mice with the same cercariae. The objective was to see if chromatin structure changes, potentially induced by development in different mollusks, are transmitted to the next developmental stage of the parasite. We selected four histone modifications (H3K4me3, H3K27ac, H3K27me3, H4K20me1), associated with different biological functions. H3K4me3 form narrow peaks (rarely larger than 2 kb wide) usually associated with the transcription start site (TSS) and promoters of transcriptionally competent genes (Berger 2007; Gu & Fire 2009; Kharchenko et al. 2010; Zentner & Henikoff 2013). H3K27me3 plays an important role in sequential expression and regulation of developmental genes, and is generally associated with repression of transcription (Kouzarides 2007; Trojer & Reinberg 2007; Bannister & Kouzarides 2011). Acetylation of H3K27 serves to discriminate active enhancers (where the mark is present) from poised ones (where the mark is absent) (Tie et al. 2009; Creyghton et al. 2010; Liu et al. 2011; Shlyueva et al. 2014). It is described as a very punctate mark, similar to H3K4me3. The role of H4K20me1 is more controversial. Some authors report it being associated with distinct regions of silent chromatin (Sims et al. 2006; Vielle et al. 2012) in human and *C. elegans*, while others describe it as an essential mark for transcription activation (Wang et al. 2008; Li et al. 2011). It is
also involved in DNA repair and cell cycle (Jorgensen et al. 2013), and genome-wide increase of this mark impairs genome integrity (Schotta et al. 2008).

Our data show that changes in SNP and mobilization of TE do not contribute to genetic variation, but that CNV are relatively frequent. Roughly 10% of spontaneous CNV are mitotically heritable. In addition, we observed the occurrence of environment-directed (i.e. environment-dependent) and genotype-independent epimutations, as well as spontaneous epimutations (not related to experimental conditions or genotypes). However, while mitotic heritability of spontaneous epimutations can reach 45%, none of the environmentally targeted chromatin structure changes are transmitted to the adult worms. Our findings are therefore in favor of spontaneous epimutations as a possible source for heritable phenotypic variants that could be selected by the environment.

2. Materials and Method

The experimental procedure is schematically summarized in Figure 1.

Ethics statement

Housing, feeding and animal care followed the national ethical standards established in the writ of February 1st, 2013 (NOR: AGRG1238753A) setting the conditions for approval, planning and operation of establishments, breeders and suppliers of animals used for scientific purposes and controls. The French Ministère de l’Agriculture et de la Pêche and French Ministère de l’Éducation 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.

Origin of the parasites and their hosts

The Schistosoma mansoni strain SmBre was originally sampled in Recife, Brazil in the 1960’s. It was provided to our laboratory in 1975 by Pr. Y. Golvan (Faculté de Médecine de Paris – Saint Antoine) and has been since then maintained in its sympatric intermediate host strain BgBre of the mollusk Biomphalaria glabrata and Mus musculus (SWISS OF1) as definitive vertebrate host. BgBre (albino strain) is also from Recife and was also acquired in 1975. Albinism is of genetic origin and does not have an impact on the experimental design. The allopatric mollusk strain BgGua originates from the town of Dans Fond and arrived at our laboratory in 2005 (pers. comm. A. Théron). Mollusks were maintained as described in Boissier & Moné (2000).

Infection and sex determination

Thirty-two B. glabrata BgBre individuals were each infected with a single SmBre miracidium. Miracidia were freshly hatched from the liver of a single infected mouse. Miracidia hatching, as well as snails and mice infections, were performed as described in Boissier & Moné (2000). Mollusks were screened 30 days later for presence of parasites and cercariae shed by each positive B. glabrata were genotyped with sex markers (Portela et al. 2010). We chose to work only with female parasites, as it was previously shown that there is an increase in female cercariae production when SmBre develops in BgGua hosts (Lepesant et al. 2013), and also to avoid possible
bias caused by gender-specific epialleles. We selected two infected snails with a large number of sporocysts to be donors for sporocyst grafts. We named the female genotypes of the parasites in these two snails Fa and Fb.

**Sporocyst graft**

The grafts were done as detailed in Jourdane & Theron (1980) and Jourdane (1984). Large (10-12 mm) BgBre and BgGua mollusks were selected as receivers. They were anesthetized by incubating 4 hours in spring water mixed with sodium pentobarbital at a final concentration of 1.2 mg/mL. Donor mollusks had their shells carefully removed with tweezers and their digestive gland (where secondary sporocysts are located) was recovered. Using tweezers, explants of 1 mm³ containing 1-3 sporocysts were prepared and grafted within an hour. Explants were grafted in the cephalopedal sinus of receiver snails with a custom-made glass micro-needle attached to a 1 mL syringe. A small incision of the tegument above the genital pore was done to reach the cephalopedal sinus. Grafted snails were then kept in normal growing conditions. Fa sporocysts were grafted in 15 BgGua and 14 BgBre, and Fb sporocysts were transplanted in 17 BgGua and 15 BgBre.

**Parasite preparation**

Fifty days after transplantations, mollusks were screened under a stereomicroscope to look for secondary sporocysts. Uninfected mollusks were discarded. Cercariae were collected from infected snails 53 days after infection by gently pipetting from spring water, avoiding mucus and feces, and sedimented on ice. Water was removed and cercariae from each snail were stored separately at -80°C before being used for ChIP-Seq. As we were interested to see if epimutations induced by the development in different snail strains would be maintained in adults, we also infected mice with cercariae from some of the snails. For both Fa and Fb parasite genotype, we selected BgBre and BgGua snails emitting the highest amount of cercariae (> 5,000) to use for infection. For parasite genotype Fa, we had one BgBre and one BgGua, and for Fb we had two snails of each strain. For each snail, three mice were infected with 300 cercariae each, as described in Boissier & Moné (2000). Mice were sacrificed 42 days later by injection of sodium pentobarbital and female adult 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. Worms from each mouse were collected separately and stored at -80°C.

**Genomic DNA sequencing**

Cercariae emitted from a mollusk infected with a single miracidium are generally considered genetically identical, but evidences exist that some genetic diversity can occur (Bayne & Grevelding 2003). We could not exclude that sporocyst grafts and development in an allopatric host were source of stress, which could activate mobilization of TE (Wijayawardena et al. 2015). Knowing this, we performed DNA pool sequencing (Pool-Seq) on adult worms developed from cercariae emitted after the grafts. We chose to sequence DNA of adults rather than cercariae because current tools for the analysis of Pool-Seq data do not perform well on large numbers of potentially different haploid genomes. It is not feasible to obtain sufficient amount of DNA for sequencing from less than 1,000 cercariae while 10 adults are enough. We selected four pools of
12, 9, 11 and 17 Fb genotype female adult worms. The first two pools originated from two
different sympatric (BgBre) snails, while the last two are from a single allopatric (BgGua) snail, but
isolated from two different mice. DNA was extracted following Beltran et al. (2008) and
sequencing was done as paired-end 125 bp reads on an Illumina HiSeq 2500 at the Montpellier
GenomiX facility (http://mgx.cnrs.fr). Sequencing was designed to have a genomic coverage of at
least 3X per individual per sample.

Chromatin immunoprecipitation and sequencing (ChIP-Seq)

Native ChIP was performed following the protocol developed for Schistosoma mansoni by Cosseau
et al. (2009) (also available online at http://methdb.univ-perp.fr/cgrunau/methods/native_chip_sm.html) on cercariae and adult worms. We used at least 5,000 cercariae or 20 adult worms.

Immunoprecipitation was performed using the following antibodies: H3K4me3 (Millipore, cat# 04-745 lot# NG1680351, 4 µL per reaction), H3K27ac (Abcam, cat# ab4729 lot# GR-150-367-2, 4 µL per reaction), H3K7me3 (Diagenode, cat# C15410069 lot# A1821D/2, 8 µL per reaction), and H4K20me1 (Abcam, cat# ab9051 lot# GR15874-1, 4 µL per reaction). For each sample, we used a control without antibody to assess unspecific background (bound fraction) and input (unbound fraction). Inputs were used for normalization in all subsequent bioinformatics analyses. Antibodies were carefully tested for specificity as described in Cosseau & Grunau (2011) and saturating quantities were used. We had earlier shown that the abovementioned catalogue/lot numbers for anti-H3K4me3 and anti-H3K27me3 can be used for ChIP-Seq (Roquis et al. 2015). For H3K27ac and H4K20me1, we performed all the validation steps described by Cosseau & Grunau (2011), including a chromatin titration to assess specificity of the antibodies. Titrations, using increasing amount of antibodies over the same quantity of S. mansoni chromatin, proved that antibodies were indeed specific and able to immunoprecipitate all target chromatin using 4 µl of the antibody for 5,000 cercariae or 20 adult female worms (Supplementary Figure 2).

ChIP products were sequenced as single-end 50 bp reads on an Illumina HiSeq 2500 at the McGill University and Génome Québec Innovation Centre (http://gqinnovationcenter.com/index.aspx). Briefly, fragmented DNA was quantified using a 2100 Bioanalyzer (Agilent Technologies). Libraries were generated robotically with 10 ng of fragmented DNA (range 100-300 bp) using the Kapa HTP Library Preparation Kit (Kapa Biosystems) as per the manufacturer’s recommendations, except that adapters and PCR primers were diluted 100-fold. The size selection step was done after the PCR step and the number of PCR cycles was increased by six. Adapters and PCR primers were purchased from Integrated DNA Technologies whereas size selection has been performed on a Pippin Prep instrument (SAGE Biosciences Inc). Libraries were quantified using the Quant-iT PicoGreen dsDNA Assay Kit (Life Technologies) and the Kapa Illumina GA with Revised Primers SYBR Fast Universal kit (D-Mark). Average size fragment was determined using a LabChip GX (PerkinElmer) instrument. Cluster formation on the flow cell was performed using the cBot instrument (Illumina) with four indexed libraries per lane. Sequencing, in the form of 50-cycle single end reads, was performed on a HiSeq 2000/2500 (Illumina) running HCS software version 2.2.38. Demultiplexed FASTQ files were generated by allowing up to one mismatch in the index.

Quality control, alignment and peak calling

All data treatment was carried out under a local galaxy instance (Goecks et al. 2010)
Sequences were aligned to the *S. mansoni* reference genome (Protasio et al. 2012) with Bowtie2.1 (Langmead & Salzberg 2012) using parameters --end-to-end, --sensitive, --gbar 4 and filtered for unique match with samtools (Li et al. 2009) (samtools view -Sh -q quality value 40-42 -F 0x0004 - | grep -v XS:i). SAM alignment files were converted into the bed format with pyicos (Althammer et al. 2011) and sorted with sortBed -i of the bedtools suite (Quinlan & Hall 2010). PCR duplicates were removed using the remduplicates function from Pyicos (Althammer et al. 2011), allowing a maximum of 4 duplicates per position. Random sampling using a custom Perl script was performed on each library to generate an equal number of reads per library for peak calling. Although random sampling is not mandatory, we found in previous work that it reduces false positives when searching for histone modification differences between samples. We took 7 million reads for H3K4me3, 23 millions for H3K27ac and H3K27me3, and 19 millions for H4K20me1. One of the samples, Cercariae Fb Gua 8, had a higher amount of PCR duplicates than the other samples, and we could only select 14 million reads for H3K27ac and 10 millions for H4K20me3. Ranger of Peakranger v1.16 (Feng et al. 2011) with P-value cutoff 0.0001, FDR cutoff 0.01, Read extension length 200, Smoothing bandwidth 99 and Delta 1 was used for peak identification. Input samples were used as negative control (-c). Wiggle files were visualized under IGV (Thorvaldsdottir et al. 2013). Reads mapping to repetitive regions were aligned on the 3,145 repetitive consensus sequences obtained from the same strain of *S. mansoni* (*SmBre*) in a previous study (Lepesan et al. 2012) (available at http://methdb.univ-perp.fr/downloads/) using Bowtie2.1 (Langmead & Salzberg 2012) evoking parameters --end-to-end, --sensitive, --gbar 4. In order to obtain identical effective library sizes for each sample, the same number of aligned reads was randomly sampled for each of the ChIP experiments (H3K4me3 = 2.5 million, H3K27ac = 7 million, H3K27me3 = 7 million, H4K20me1 = 8 million). After the counts had been normalized, they were compared between groups (allopatric vs sympatric, with three replicates; genotyp type Fa vs Fb, with two and four replicates) and between individual samples using the DESeq package (Anders & Huber 2010) of the R software. This package uses the negative binomial distribution to calculate the P-value for each element and for each comparison between two samples and to find elements that present significant differences in their enrichment levels.

**Chromatin structure profiles**

The analysis of the different histone modification combinations on the *S. mansoni* genome was performed using EpiCSeg (Mammana & Chung 2015), using count normalization (normalizecounts) and binsize of 500bp (--binsize 500). We tested 4 to 16 different possible chromatin states and concluded that five states (--nstates 5) was the optimal number to avoid redundancy between states. Profiles were done on cercariae and adults, each time using the three datasets from the sympatric (BRE) condition as replicates for the bin count. In a similar way, we also compared sympatric (BRE) and allopatric (GUA) chromatin profiles for adults and cercariae. Annotation files used to generate chromatin profiles on genes, transcription start sites (TSS) and transcription end site (TES) were derived from an earlier established transcriptome (Roquis et al. 2015), by keeping a subset of genes longer than 5 kb (2,428 genes).
**Comparative epigenomic analysis**

For comparison of chromatin profiles of each mark, DiffBind v1.8.3 (Stark & Brown 2013) was used with FDR of 0.1. Comparisons were done for each histone modification between allopatric and sympatric conditions (3 replicates for each histone modification). Comparisons were also done for genotype Fa (two replicates) vs Fb (four replicates), and for each individual sample (no replicate). One library (cercariae Fb5 H3K4me3) was systematically excluded from these DiffBind analyses because it had more than 20% of PCR duplicates, indicating probably a too few starting material, thus reducing the number of replicates for this histone modification by one. Heatmaps were generated evoking dba.plot Heatmap (sample, contrast=1, correlations=FALSE) and PCA was done with dba.plotPCA(sample, contrast=1,th=.05). DiffBind (allopatric vs sympatric condition and genotype Fa vs Fb with minOverlap = 2) will test a peak if it is identified in at least two of the six total samples. It is the variance amongst all the replicates in each condition that drive the p-value calculated for the comparison. For sample-by-sample comparison, non-parametric bootstrapping or default value minOverlap = 1 was used for each sample. Both methods provided the same results. Differentially modified chromatin regions were saved as BED files for further analysis. Gene ontology enrichment analysis was done with BLAST2GO (Conesa et al. 2005). For analyses of significant differences in histone isoform enrichment in repetitive sequences, normalized aligned read counts were compared using the DESeq package of the R software (Anders & Huber 2010).

**Comparative genomic analysis**

For the detection of SNP frequency distortion, i.e. significant differences in allele frequency between sympatric and allopatric conditions, we used paired-end data for genotype Fb and single-end data corresponding to the ChiP-Seq input libraries for both Fa and Fb. All raw data were of excellent quality and no quality filter was necessary. Reads were aligned with Bowtie 2 (Langmead & Salzberg 2012) and default parameters. Unique reads were identified using the Bowtie 2 ‘XS:i’ flag and sorted with SamTools (Li et al. 2009). Alignment rates were between 54% and 59%, which is consistent to what has been previously achieved on S. mansoni (Roquis et al. 2015) and can be explained by the fact that we only align on unique sequences of the genome (47% of the genome is composed of repetitive elements (Lepesant et al. 2012). To homogenize effective library size, 56 million aligned reads were randomly subsampled for the paired-end libraries and used for further analysis. For single-end data, 24 million reads were subsampled. Pileup files were generated by SamTools invoking -B -C 0 -d 250 -q 0 -Q 13. For the comparison of allele frequencies we used PoPoolation2 (Kofler et al. 2011). Fisher’s Exact Test was chosen to estimate the significance of allele frequency differences in samples for which we had no replicates (Fa), and the Cochran-Mantel-Haenszel (CMH) test was used to detect consistent allele frequency changes with replicates (Fb). Pileup files were converted to sync file with parameter --min-qual 20. And Fisher and CMH tests were used invoking --min-count 2 --min-coverage 10 --max-coverage 200 --suppress-noninformative for single-end data and --min-count 12 for paired-end data. Copy-number variation (CVN) are another source of genetic variability and we used CNV-seq (Xie & Tammi 2009) to estimate the probability of CNV differences between the sympatric and the allopatric conditions. We used default parameters (log2 = 0.6, pvalue = 0.001, minw = 4) with a genome size of 364,541,798 bp. Results were converted into IGV or GWAS compatible formats and visualized under IGV. Differences in genomes Fa and Fb were detected using FreeBayes v0.9.20.
(Garrison & Marth 2012) using parameters --theta 0.001 --ploidy 2 -J -m 1 -q 0 -R 0 -Y 0 -e 1000 -F 0.2 -C 2 --min-alternate-qsum 0 -G 1 --min-coverage 20. Resulting VCF files were compared with vcf-isec of the VCF tools (Danecek et al. 2011). CNV were searched for as described above.

The search for transposable element (TE) insertions was done by aligning the reads of the 4 Fb paired-end pools using bwa-mem from the BWA software (Li & Durbin 2009) to the reference genome of *S. mansoni*. Detection of the insertions and excision of elements was done using the TEMP software (Zhuang et al. 2014) with the required repeat and annotation files, parameters -x 10 and -c 12 for the insertion detection, and default parameters for the absence (excision) detection. The principle of this analysis is to use differences in mapping of forward and reverse reads to detect TE insertion events. BED files were generated from the output of TEMP and the files were sorted. In order to eliminate constitutive differences between the sequenced genomes and the reference genome we subtracted the positions that were common to all four genomes and visually inspected the remaining regions.

**Comparative Transcriptomics**

Parasites were used to infect sympatric and allopatric snail hosts for three generations as described in Lepesant et al. (2013). For RNA extraction, cDNA generation and sequencing, and comparative analysis with the TopHat2 and CuffLinks we followed Roquis et al. (2015). Briefly, reads were aligned to the genome and exon-intron structures are established individually for each sample. Cuffmerge was used to generate a consensus transcription annotation file in GFF format. We then used htseq-count with the “union” option (Anders et al. 2015) to obtain read-counts for each of the 18,844 loci (“genes”) identified by Cufflinks/Cuffmerge. A single count (“pseudocount”) was added to each readcount to avoid division by 0 problems. DESeq (Anders & Huber 2010) with default parameters was used for differential gene expression analysis. A gene ontology enrichment analysis was done with BLAST2GO (Conesa et al. 2005) and prediction of protein interaction networks was done with STRING10 (Szklarczyk et al. 2015).

**Results**

*There are no significant differences in SNP or TE mobilization between SmBre in sympatric and allopatric condition*

SNP frequencies were evaluated using Fisher’s exact test (for Fa without biological replicates) and the CMH test for Fb (with replicates). Both tests are robust but the CMH-test showed the highest power to identify interesting loci (Kofler & Schlotterer 2014). Following others (Barsh et al. 2012, Panagiotou et al. 2012, Kofler & Schlotterer 2014, Broer et al. 2013) we decided to establish a significant threshold for p-values at p = 1x10^{-5}. Despite this relatively relaxed threshold, we detect no significant allele frequency distortion. We noted nevertheless that there are large blocks of 100 kb range in the genome (roughly one per chromosome) that show enrichment of non-significant AF differences between the sympatric and allopatric condition in all replicates, whatever genotype was used. We calculated nucleotide diversity between Fa and Fb using private SNP (i.e. with allele frequency 1). We found 1,570 SNP differences (nucleotide diversity 4.3x10^{-6} based on the total
genome length of 364,541,798 bp) and 72 CNV between cercariae Fa and Fb emitted from the sympatric host BgBre.

The analysis of transposable element insertion produced several thousand candidate regions which showed differences in the number of supporting reads or the detection of one of the two insertion borders with respect to the reference sequence. The elimination of the redundant positions and those pointing to opposite insertion borders showed that the detected positions were due to the difference between the sequenced genome and the reference genome. Once these differences were eliminated the analysis showed that no neoininsertions of elements could be detected.

There are significant differences in CNV between sympatric and allopatric SmBre

CNV-seq detected 58 - 105 CNV covering 0.1-0.2% of the genome in all comparisons (Table 1a & 1b, Supplementary File 1). Interestingly, some CNV covered or flanked the above-mentioned blocks of non-significant AF differences. To answer the question whether mitotically heritable CNV are systematically found in the allopatric condition, we compared these variants between Fa and Fb. CNV between sympatric Fa and Fb were excluded since they could represent simply differences to the reference genome. None of the regions sensu stricto fulfilled the condition to be a CNV in cercariae allopatric vs sympatric in both Fa and Fb, and CNV in adults allopatric vs sympatric in both Fa and Fb. Nevertheless, there is in one region (Schisto_mansoni.Chr_3:1,725,500-1,762,000) with a clustering of environment dependent CNV. Three single exon genes are predicted in the region (Smp_192030, Smp_101720 and Smp_201970) for which we did, however, not find transcripts in our RNA-Seq data. CNV between biological replicates in sympatric or allopatric condition do not show significant differences (Table 1a & 1b). CNV locations are provided in the supplementary material (Supplementary File 1).

Taken together, our data provide evidence that, upon exposure to the allopatric condition, point mutations and TE mobilization are below the detection limit. CNV occurs at appreciable frequency, but are not linked to a specific environmental condition (allopatric or sympatric). The frequency of CNV is not different between the sympatric and allopatric hosts. Twelve to fifteen percent of random CNV are maintained through development to adults. One region on chromosome 3 could be a CNV hot-spot activated upon exposure to the allopatric environment.

In two cases (Schisto_mansoni.Chr_2:717900-817000 and Schisto_mansoni.Chr_1:58049000-58059000), we found epimutations close to CNV, but not to SNP or transposable elements. After having established the degree and type of genetic mutations we attempted to characterize the epimutations.

The epigenome of S. mansoni comes in five main chromatin colors

We then performed ChIP-Seq experiments with the abovementioned antibodies for two genotypes (Fa and Fb) of cercariae that had developed in sympatric or allopatric snail hosts, and of the resulting female adults (a total of 60 libraries, details in Supplementary File 2). The analysis of the different histone modification combinations on the S. mansoni genome shows five main states
(also called “colors”; Filion et al. 2010), in both cercariae (Figure 2) and adults (Supplementary Figure 3). Both profiles are qualitatively similar. The most frequent histone mark combination (state 3 on Figure 2 and Supplementary Figure 3) is a combination of H3K27me3-H3K27ac-H4K20me1, which covers 60% of the genome in cercariae and only 46% in adults. This state has a low presence around the transcription start site of genes (TSS) but is strongly enriched at the transcription end site (TES). The second most frequent state (state 2) is characterized by an enrichment of H3K27ac and H3K20me1, covering 17% of the genome of cercariae and 24% of adults. It is enriched upstream of the TSS, decreases drastically in the first kb past the TSS before being enriched again in the gene body and decreases once more near the 3’ end of the gene. The absence of any of the four histone marks (state 5) covers 14% of the genome in both developmental stages of the parasite. It is not very common in gene bodies, but is particularly depleted around TSS and TES. State 4, characterized by a stronger presence of H3K27me3, combined with a lower amount of H3K27ac and H4K20me1 is very lowly represented around genes. Outside of genic regions, it can be found covering 5% of cercariae and 14% of adult genomes. The last combination of marks (state 1) presents a very strong enrichment in H3K4me3 located right after the TSS (3% of cercariae and 2.3% of adult genomes). We also compared chromatin state between sympatric and allopatric cercariae, as well as adults, but we did not observe any large-scale chromatin structure profile differences (data not shown). It should be noted that except for H3K27ac and H3K27me3, that are mutually exclusive, our approach can neither confirm nor exclude that histone modifications occur at the same histone or same nucleosome. All we can say it that there is co-occurrence at the same genomic site.

There are three types of chromatin changes upon exposure to allopatric hosts

For comparison of chromatin changes in the sympatric and allopatric conditions, one would ideally use these five chromatin colors. Unfortunately, no algorithm exists yet to do so. Instead, we performed what was feasible, i.e. a comparative analysis histone mark by histone mark for the four modifications using peak-calling (PeakRanger) and differential enrichment analysis (DiffBind).

For the cercaria we found in average 8,378 peaks for H3K4me3, 15,866 peaks for H3K27ac, 20,784 peaks for H3K27me3, and 11,175 peaks for H4K20me1. In adults there were 11,967 peaks for H3K4me3, 9,1110 peaks for H3K27ac, 22,435 peaks for H3K27me3, and 17,403 peaks for H4K20me1 (Supplementary File 3).

Heatmaps generated by DiffBind provide a general overview of the chromatin landscape. They clearly show chromatin differences in cercariae clones exposed to the sympatric and allopatric host, and the resulting female adults (Figure 3A and 3B). Principal component analysis, using genomic positions with differences on chromatin profiles, separated sympatric and allopatric samples extremely well, based on the number of differences in histone marks between samples (Supplementary Figure 4A and 4B). We observed in cercariae three different types of chromatin changes depending on condition and genotype. In principle, to classify epimutations, one would fit a statistical model to test for a genotype effect, an environmental effect, and possibly the interaction, on peak presence and shape. In such cases one could detect a significant effect of (for instance) environment even if not all replicates showed the pattern consistently. To our
knowledge, such statistical model has not been developed and tested. We use here a simpler
approach and classified epimutations into (1) host-induced epimutations, (2) spontaneous
epiputations, and (3) genotype dependent epimutations. For convenience, we will also use the
term “targeted epimutation” or “induced epimutations” for type 1 epimutation, and “random
epiputations” for type 2. In other words, at a given developmental stage (cercariae or adults),
type 1 epimutations must be present in all samples of the allopatric condition and absent in the
sympatric (or vice versa). Type 2 epimutations are present in at least one sample of a condition or
a genotype, but must not be present in all samples of this condition or genotype. Type 3
epiputations are present in all samples of one genotype (independently of the environmental
condition) and absent in all samples of the other genotype. A potential caveat of this classification
method lies in the relatively low number of replicates, leading potentially to an overestimation of
type 1 epimutations. In addition, type 2 epimutations might not be stochastic in the strict sense.
For instance, they might also be environment or genotype-specific but with incomplete
“penetrance” among replicate samples.

For type 1 epimutations, using a FDR 0.1 (Diffbind’s default), we identified 131 differences for
H4K20me2, 592 for H3K4me3, 346 for H3K27ac and 203 for H3K27me3, present in all replicates
between sympatric and allopatric cercariae. All were verified by visual inspection, in which we
looked for background changes that DiffBind had probably mistakenly identified as real
differences. Basically, if a trained human eye (considering intensity, shape, and strength of
neighboring peaks) cannot detect the differences spotted by Diffbind, we considered it as a false
positive (see Supplementary Figure 5 for examples). In general, validated epimutations
corresponded to a p-value ≤ 1x10^{-5} for H3K4me3 and ≤ 1x10^{-15} for the other marks. It reduced the
number of type 1 differences to a total of 51 genomic regions, among those all of them showed
changes for H3K4me3, 11 for H3K27ac, 2 for H3K27me3, and 3 for H4K20me1. Examples are
shown in Figure 4A and a detailed list of differences is available in Supplementary File 4.
Interestingly, epigenetic diversity is higher in samples from the allopatric condition (Figure 3A and
Supplementary Figure 4A). We found that 76.5% of induced epimutations are in transcribed
regions previously identified with RNA-Seq in cercariae (Roquis et al. 2015).

As outlined above we defined spontaneous epimutations (type 2) as a chromatin structure change
that occurs in one or more samples, but that is not consistently shared among replicates (Figure
4B). We detected 64 genomic locations where these random epimutations occurred, very often
with multiple changes in the same region: 60 changes in H3K4me3, 9 in H3K27ac, 6 in H3K27me3,
and 5 in H4K20me1. We identified that 79.7% of the differences were in transcribed regions
details available in Supplementary File 4). We detected these epimutations by comparing
samples one by one on Diffbind results (Supplementary File 5), and verifying them through the
same visual inspection as for type 1 epimutations. We noted that for the two replicates of
genotype Fb in the allopatric condition, we systematically found several hundreds of random
epimutation for the four histone isoforms. We believe that this is due to a technical issue leading
to higher background in these ChIP-Seq libraries, which is consistent with our visual inspection.
We also looked for genotype related epimutations (type 3), *i.e.* chromatin structure changes that are identical in both environmental conditions, but different when comparing cercariae of genotype Fa and Fb (Figure 4C). After visual inspection (same as type 1 & 2), we observed a total of 64 genomic locations with chromatin structure differences between the two genotypes, often with two or more marks changing: 39 differences for H4K20me2, 41 for H3K4me3, 38 for H3K27ac and 42 for H3K27me3. 73.4% of the changes between Fa and Fb are in transcribed genomic regions. BED files are available as Supplementary File 5.

**Spontaneous epimutations, but not host-induced epimutations in cercariae are inherited to adults**

By infecting mice with some of the cercariae collected from each sympatric and allopatric snail, we produced adult worms and investigated if the different types of epimutations we had spotted in cercariae were transmitted to the next developmental stage. None of the type 1 chromatin structure changes, shared by all replicates, where maintained in adults. However, Diffbind heatmaps (with the exception of H3K4me3) still show differences between adults developed from cercariae that have experienced either the sympatric or allopatric condition (Figure 3B), regardless of the genotype (Fa or Fb). With the exception of H3K27me3, PCA of adults separate adults by the environmental condition they experienced as cercariae (Supplementary Figure 4A). While host-induced epimutations were not inherited *sensu stricto* from cercariae to adults, the developmental trajectory was altered, resulting in different chromatin profiles between adults from allopatric or sympatric origin (Figure 3A and 3B; Supplementary Figure 4A and 4B).

Since we had replicates for genetically identical clones Fb, we investigated whether spontaneous epimutation (type 2) would be transmitted to the corresponding adults. Indeed, 45% of the sites with epimutations have the same chromatin structure in adults: 29 differences in H3K4me3, 3 in H3K27ac, 5 in H3K27me3 but none in H4K20me1 were heritable *sensu stricto*, *i.e.* they had occurred in the sporocysts and were maintained in cercariae and adults.

We found genotype-associated epimutations to be most mitotically heritable. Roughly 77% of the genomic locations with epimutations between Fa and Fb cercariae were found to be identical in adults. It corresponds to 31 changes in H3K4me3, 23 in H3K27ac, 24 in H3K27me3 and in 30 H4K20me1.

We then used the average number of peaks detected by PeakRanger (Supplementary File 3) to calculate frequency and heritability of histonic epimutations. Environmentally triggered epimutations (type 1) occur at a frequency of 4.2x10⁻³ (H3K4me3), 1.2x10⁻³ (H3K27ac), 8.9x10⁻⁵ (H3K27me3), and 1.7x10⁻⁴ (H4K20me1). Random epimutations (type 2) are observed in the same order of magnitude: 5x10⁻³ (H3K4me3), 1x10⁻³ (H3K27ac), 2.7x10⁻⁴ (H3K27me3), and 2.9x10⁻³ (H4K20me1). As mentioned above, heritability of type 1 epimutations is 0 (reversibility = 1). In contrast, heritability of type 2 epimutations reaches up to 0.83. In summary, frequency and heritability of histonic epimutations depends on the chemical nature of the modification and their type. Mean frequency is in the 10⁻³ order of magnitude.
Based on their locations, epimutations for the different histone marks were manually compiled into a single table for further analysis (Supplementary File 4).

**Negligible chromatin structure changes occur around repetitive elements**

Mapping our ChIP-Seq library over the 3,145 identified repetitive sequences of *S. mansoni* genome and analyzing the read counts between samples, we did not find any differences present in all replicates of an environmental condition in cercariae or adults. Only when we compared environmental conditions within a single genotype, or between genotypes in the same conditions, we detected changes in 21 repetitive elements (17 of which being transposable elements). Only a single histone mark was modified in these 21 repetitive elements. None of the histone marks were systematically changed and we found no heritability of the changes from cercariae to adults (Supplementary File 6).

**Chromatin structure changes are not associated with specific pathways or molecular/biological functions**

After having established that chromatin architecture changes upon exposure to a different environment and that this occurs primarily in regions of the genome that contain genes, we wondered whether a phenotype was associated with these epimutations. GO enrichment analysis was done for all the transcribed regions of the genome where chromatin structure changes were detected (host-induced, spontaneous and genotype-associated epimutations). No specific pathway or enrichment in molecular or biological functions was observed in any type of comparison. STRING analysis showed that Smp_016630 (Trimeric G-protein alpha o subunit), Smp_141980 (cAMP-specific 3,5-cyclic phosphodiesterase) and Smp_147260 (Ras-like) could potentially interact in a signaling pathway. All three genes have environmentally triggered complex epimutations (Supplementary File 4).

Next, we exposed miracidia for three generations to the allopatric host, extracted RNA from the resulting cercariae and compared their RNA profile to cercariae produced in sympatric conditions. We had earlier shown that transcription is stalled in cercariae (Roquis *et al.* 2015) and we expected only small differences in RNA profiles. We compared the RNA levels of genes at or close to the type 1 epimutations using the three generations as biological replicates (Supplementary File 7). Differences in RNA levels of epimutated genes were moderate with a few exceptions: Smp_130050 codes probably for a PDZ domain containing RING finger protein and its RNA is 3-5 fold overrepresented in allopatric cercariae. Smp_151290 codes for a putative multidrug and toxin extrusion protein 2 and its RNA is 2.5 times enriched in cercariae from allopatric hosts. Smp_015630 (Glutamate-gated chloride channel subunit) is threefold underrepresented in allopatric cercariae.

We also identified genes with strong differences *i.e.* adjusted p-value < 0.01, and/or 10 fold overrepresentation in one of the conditions (Supplementary File 8). For these genes we verified chromatin structure in cercariae and adults. None of the genes were in the regions that we had qualified as epimutated. However, it is noteworthy that 80% (50 of 62) of these genes do not
contain a canonical H3K4me3 peak at the TSS. This is two times higher than the genome average. It could therefore be that with the antibody combination we used we did not capture the histone mark structure of these genes. We also noticed subtle differences in the extent and position of chromatin states 1 to 5 (Supplementary Figure 6). We would have considered this as random noise but in some cases the chromatin state profile was faithfully inherited to the adults, suggesting that these small differences are stable and might have an impact on gene expression.

Discussion

In this experiment we set out to investigate the impact of the environment on the genome and epigenome of S. mansoni larvae. The environment could be (i) a selective agent, (ii) could induce specific mutations and/or epimutations, and (iii) could change mutation and/or epimutation rate, or (iv) might have no effect. The two S. mansoni genotypes we used in this study (Fa and Fb) come from the same inbred population. Nucleotide diversity is 0.0004% based on SNP, but they show relatively strong differences on the level of CNV (0.2% of the genome). We also found that they bear epigenetic differences. Here, we considered genotype-associated epimutations as chromatin structure changes present within a genotype in both allopatric and sympatric condition, but absent in the other genotype. We observed 64 of these differences (FDR ≤ 0.1, p≤1x10^-5 and after visual inspection). The majority (73.4%) of epimutations are in transcribed regions in cercariae, which is similar to what we observed with spontaneous and induced epimutations. In at least two striking cases that we investigated, genotype associated epimutations were linked to CNV. In the case of clone Fa, there is an important CNV difference between cercariae emitted from sympatric and allopatric snails, and could be a result of the environment, although without replicates for this genotype we cannot assess it without doubt. It has been previously observed (Skinner et al. 2015) that environmentally induced epimutations can lead to genetic mutation and CNV. We noticed that the CNV were clustered (Supplementary File 1). Centromere positions on S. mansoni genome are not mapped to the genome, and it is also possible that these CNV clusters are located in pericentromeric regions, which are hotspots of segmental duplications and prone to generate CNV (Makino et al. 2013). Apparition of these CNV could also possibly explain some of the genetic heterogeneity between S. mansoni clonal sporocysts observed before (Bayne & Grevelding 2003).

We then looked at environmentally triggered epigenetic changes, i.e. epimutations that occur in all the samples of a given mollusk condition. We found a total of 51 genomic regions, where at least one histone mark is different in all three cercariae samples emitted from sympatric and allopatric mollusks. Changes were often subtle (peaks of low intensities compared to neighbor peaks). As we used whole organisms for ChIP-Seq, there is the possibility that these changes only reflect a small subset of cells or tissues within cercariae. No gene pathway or gene ontology term enrichments were uncovered in genes spanning induced epimutations, and none of these epimutations were transmitted to adults. This type of ephemeral epigenetic change in response to an environmental stimulus is reminiscent of what was found for instance in smokers and former smokers where roughly a dozen CpG sites change their methylation status but revert progressively to normal when smoking was abandoned (Wan et al. 2012; Elliott et al. 2014).
No major chromatin remodeling was observed in the 3,145 known repetitive sequences of the *S. mansoni* genome, and although some of the epigenetic differences are located in transposable elements, we did not detect new insertion sites. Transposable elements movement is known to generate new genetic, epigenetic and phenotypic diversity in other works (Biemont 2010; Fedoroff 2012; Holoch & Moazed 2015), but they do not play a major role in this study.

Subsequently, we investigated the presence of spontaneous (type 2) epimutations through the genome of cercariae. We first detected type 2 epimutations through comparison of replicates of clone Fb. Sixty-four sites on the genome showed random epimutation, and 45% of them maintained their epimutated chromatin structure at the adult stage. Our data suggest that a change of the environment has an impact on epimutations since chromatin profiles of cercariae emitted from *Bg*Gua (and the resulting adults) present more differences between samples (as observed from the PCA on Supplementary Figure 4A and 4B) than chromatin profiles of cercariae and adults emitted from *Bg*Bre. We had shown before that *S. mansoni* worms exposed transiently to hycanthone, an antimelminth drug, possess 57 days later a characteristic chromatin profile different from unexposed worms (Roquis et al. 2013). At that time, we were not able to tell if the drug directly induced the different profiles, or if it acted as a selection agent, killing most of the parasites in the population, except those with a pre-existing distinct chromatin structure. In light of our new findings we are now in favor of the hypothesis that random epimutations had occurred prior to the drug exposure, and had produced in some individuals a resistant phenotype to the treatment.

The average frequency of histone modification based epimutations in our system is about $10^{-3}$. This is one order of magnitude higher than what was reported for DNA methylation based epimutations in CpG pairs (Schmitz et al. 2011; van der Graaf et al. 2015). In this issue, Kronholm & Collins (2015) underline the importance of stability of epimutations for adaptive evolution. We show that heritability (i.e. stability) depends strongly on the type of epimutation and that for most of the detected epimutations reversion is actually above the threshold of $10^{-1}$ that Kronholm and Collins define as the reversion rate that limits the adaptive effect of epimutations. In other words, our data support the view that in *S. mansoni* epimutations do not contribute to adaptive walks, unless they provide a strong fitness gain.

*Schistosoma mansoni* proves to be an interesting model for this kind of epigenetic studies, due to its ecological/healthcare importance, the possibility to generate clones, and different phenotypically distinct developmental stages allowing to observe mitotic inheritance of epimutations. However, one of the caveats of this organism is its small size, making it impossible at the moment to perform ChIP-Seq on specific tissues or individuals (we needed at least 5,000 cercariae for our experiments). It means that we can only observe epimutations that are present in most of the cells and most individuals in the pools. Tissue-specific epimutations, as well as rare epimutations within the clone population, cannot be detected. Conversely, this also means that the chromatin structure changes we observed concern the majority of cells and individuals in our pools, and hence are more likely to have a significant adaptive impact. We also keep in mind that our sample size (three for each environmental condition and developmental stage) is small, and
more replicates, from the same and new genotypes, need to be done in order to strengthen our findings.

**Conclusion**

In conclusion, our findings indicate that in *S. mansoni* sporocysts, there are two major sources of mitotically heritable variation: CNV and spontaneous epimutations. This means that at least in our biological model (i) epimutations occur at appreciably frequency and since they are overrepresented in coding regions they can have an impact on gene expression, and (ii) that CNV occur at the same order of magnitude as epimutations, potentially contributing as much as the latter to phenotypic variation. We could not fully answer the question whether or not environmental changes increase epimutations. That means that at least for *S. mansoni* theoretical models in which epimutations serve to produce phenotypic variants that explore the fitness landscape for invariant genotypes (e.g. Klironomos *et al.* 2013) are fully applicable but future models must include CNV and must take into account the high reversion rate of histone-based epimutations.

**Acknowledgement and data availability**

Authors are grateful to Rory Stark, Aaron Taudt, Maria Colomé-Tatche and Frank Johannes for helpful discussions. Data for ChIP-Seq, Pool-Seq and RNA-Seq are available as fastq files BioProject Accession PRJNA236156 at the NCBI SRA. Chromatin profiles and transcripts can be visualized at our GBrowse instance (Stein *et al.* 2002) ([http://genome.univ-perp.fr](http://genome.univ-perp.fr)). Funding was provided by the French National Research Agency ANR, project ANR-10-BLAN-1720 (EPIGEVOL) and the GenEpi program of the UMR5244/IHPE ([http://ihpe.univ-perp.fr/en/ihpe-transversal-epigenetics/](http://ihpe.univ-perp.fr/en/ihpe-transversal-epigenetics/)).

**Figure legends**

**Figure 1:** Experimental procedure for the sporocyst grafts. Brazilian mollusks (*BgBre*) were infected with a single miracidium of the sympatric strain of *S. mansoni* (*SmBre*). Thirty days later, secondary sporocysts were surgically removed from the infected mollusks and grafted in several sympatric (*BgBre*) and guadeloupean allopatric (*BgGua*) mollusks. Emitted cercariae were collected for chromatin immunoprecipitations followed by sequencing (ChIP-Seq) and infection of mice to obtain adult worms. Adult worms were recovered for ChIP-Seq and paired-end DNA sequencing. The experiment was performed twice, with two different genotypes of the parasite (*Fa* and *Fb*). In *Fb* genotype, two replicates were done in sympatric and allopatric conditions.

**Figure 2:** Five different chromatin states (“colors”) as identified by EpiCSeg in cercariae. (A) Composition in histone marks of the five chromatin states. Color code of the states is used in all other parts of this figure. (B) Proportion of the genome covered by each state in cercariae. Enrichment for each state around gene bodies (C), at transcription start site (TSS) (D) and transcription end site (TES) (E) of cercariae. Enrichment profiles were done on 2,428 genes longer than 5 kb. Results for adults are very similar and shown in Supplementary Figure 2.
**Figure 3:** Heatmaps for the four studied histone marks in cercariae (A) and adults (B). Only sites with statistically significant differences (epimutations) were used. Red bars highlight conserved changes. Red snail symbols represent sympatric host *B. glabrata* (B *Bre*), blue snail symbols stand for allopatric snail host *B. glabrata* (B *Gua*). Cercariae sample GuaFb5 for H3K4me3 was excluded from analysis (see methods).

**Figure 4:** Typical examples of epimutations. (A) Environmentally dependent epimutations (green frame) in H3K4me3 (blue profile) and H3K27ac (red profile): there are differences in all three samples Fa and two replicates Fb between sympatric and allopatric cercariae. (B) Random epimutations in H3K4me3, present in cercariae in three samples and transmitted to adult worms in corresponding three samples. (C) Genotype-associated epimutations in H3K4me3, H3K27ac, H3K27me3 and H4K20me1 between genotypes Fa and Fb in sympatric condition. The track Ref. transcriptome *S. mansoni* displays transcripts (exons as blocks and introns as arrows) deduced from RNA-Seq data in cercariae.

**Supplementary Figure 1:** Life cycle of the parasite *Schistoma mansoni*. The endoparasites infect two hosts, (1) a mollusk intermediate host (*Biomphalaria* genus) in which it multiplies asexually as sporocysts and then cercariae, and (2) a vertebrate definitive host (primate or rodent) in which adult worms reproduce sexually. Two types of free-swimming larvae, miracidia and cercariae, are the infective form for the mollusk and vertebrate hosts, respectively. Eggs laid by adult worm couples in the mesenteric vein of the vertebrate host can go through the intestine wall to be excreted in the feces, and hatch in contact with freshwater.

**Supplementary Figure 2:** ChIP titration, using the same amount of *S. mansoni* chromatin (equivalent to 20 adult worms) and an increasing amount of antibody. Percentage of input recovery of chromatin was measured by qPCR on two housekeeping genes, alpha tubulin and 28S, as described in (Cosseau *et al.* 2009). Titrations were done to measure the efficiency of immunoprecipitation for antibody H3K27ac Abcam cat# ab4729 lot# GR-150-367-2 (A and B) and H4K20me1 Abcam, cat# ab9051 lot# GR158874-1 (B and C).

**Supplementary Figure 3:** Five different chromatin states (“colors”) as identified by EpiCSeg in adults. (A) Composition in histone marks of the five chromatin states. Color code of the states is used in all other parts of this figure. (B) Proportion of the genome covered by each state in adults. Enrichment for each state around gene bodies (C), at transcription start site (TSS) (D) and transcription end site (TES) (E) of adult worms. Enrichment profiles were done on 2,428 genes longer than 5 kb.

**Supplementary Figure 4:** Principal Component Analysis (PCA) for the four studied histone marks in cercariae (A) and adults (B). Only sites with statistically significant differences (epimutations) were used. Blue circles (Bre) represent cercariae and adults from sympatric conditions (*SmBre* in *BgBre*), magenta (Gua) the allopatric conditions (*SmBre* in *BgGua*). Cercariae sample GuaFb5 for H3K4me3 was excluded from analysis (see methods).
Supplementary Figure 5: Examples of two regions (A and B) that were identified as epimutations (displayed by the red boxes) by DiffBind but did not pass the visual inspection.

Supplementary Figure 6: Example of a region that was not identified as epimutated by DiffBind but shows differences based on EpiC Seg chromatin “colors”, judged by visual inspection. The associated gene is differentially expressed and chromatin states are transmitted to adults.

References


Cosseau C, Azzi A, Smith K et al. (2009) Native chromatin immunoprecipitation (N-ChIP) and ChIP-Seq of Schistosoma mansoni: Critical experimental parameters. Molecular and Biochemical Parasitology, 166, 70–76.


Bioinformatics, 25, 1754–1760.


Wan ES, Qiu W, Baccarelli A et al. (2012) Cigarette smoking behaviors and time since quitting are associated with differential DNA methylation across the human genome. Human Molecular Genetics, 21, 3073–3082.


**Data Accessibility**

RNA-Seq and ChIP-Seq raw data are available as fastq files under study accession number SRP035609 (BioProject Accession number PRJNA236156) at the SRA of the NCBI.