



**HAL**  
open science

# Emission rythmique des cercaires de *Schistosoma mansoni*: du phénotype au(x) gène(s)

Chrystelle Lasica

► **To cite this version:**

Chrystelle Lasica. Emission rythmique des cercaires de *Schistosoma mansoni*: du phénotype au(x) gène(s). Interactions entre organismes. Université de Perpignan, 2022. Français. NNT: 2022PERP0019 . tel-03924870

**HAL Id: tel-03924870**

**<https://theses.hal.science/tel-03924870v1>**

Submitted on 5 Jan 2023

**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.



# THÈSE

Pour obtenir le grade de  
Docteur

Délivré par  
**UNIVERSITE DE PERPIGNAN VIA DOMITIA**

Préparée au sein de l'école doctorale **ED 305**  
Et de l'unité de recherche UMR5244  
Interactions Hôtes Pathogènes Environnements (IHPE)

Spécialité : Biologie

Présentée par  
Chrystelle Lasica

**Emission rythmique des cercaires de *Schistosoma mansoni* : du phénotype au(x) gène(s)**

Soutenue le 21 Septembre 2022 devant le jury composé de

Mme. Nolwenn DHEILLY, Directrice de recherche ANSES, Maisons Alfort	Rapportrice
Mme. Sylvie HURTREZ-BOUSSÈS, Professeure UM, MIVEGEC, Montpellier	Rapportrice
M. Timothy J.C ANDERSON, Directeur de recherche, Texas Biomedical Research Institute (USA)	Examinateur
M. François-Yves BOUGET, Directeur de recherche CNRS, LOMIC, Banyuls sur mer	Examinateur
M. Thierry LAGRANGE, Directeur de recherche CNRS, LGDP, UPVD, Perpignan	Examinateur
M. Christoph GRUNAU, Professeur UPVD, IHPE, Perpignan	Examinateur
Mme. Hélène MONÉ, Directrice de recherche CNRS, IHPE, UPVD, Perpignan	Directrice de thèse
M. Cristian CHAPARRO, Ingénieur d'étude CNRS, IHPE, UPVD, Perpignan	Co-directeur de thèse
M. Gabriel MOUAHID, Maître de Conférences-HDR UM, IHPE, UPVD, Perpignan	Invité



## **Remerciements**

*Je tiens tout d'abord à remercier très sincèrement Mesdames Nolwenn Dheilily et Sylvie Hurtrez-Boussès d'avoir accepté d'être les rapportrices de mes travaux de thèse. Je souhaite aussi remercier les membres du jury : Messieurs Tim Anderson, François-Yves Bouget et Thierry Lagrange (également membres de mon CSI) mais aussi Monsieur Christoph Grunau et Monsieur Gabriel Mouahid (invité). Les précieux conseils de chacun m'ont permis d'affiner ma réflexion scientifique.*

*Je remercie également mes directeurs de thèse, Madame Hélène Moné et Monsieur Cristian Chaparro. Personne ne peut rêver d'avoir un meilleur encadrement que celui que vous m'avez donné. Vous m'avez témoigné confiance, bienveillance, liberté, soutien, pédagogie et inculqué l'exigence de bien faire jusqu'au bout. Je suis très heureuse d'avoir eu des personnes aussi investies que vous comme directeurs de thèse (s'il y a une prolongation, je signe tout de suite !). Votre humour et votre bonne humeur ont été un vrai plus dans cette thèse !*

*Mes remerciements s'adressent aussi au laboratoire IHPE et aux deux directeurs qui se sont succédés Messieurs Guillaume Mitta et Christoph Grunau. Les équipes Ecoevi et Trev auxquelles j'ai appartenu m'ont apporté des échanges scientifiques de grande qualité ainsi que des moments de convivialité de grande complicité ! J'associe à ces remerciements le bureau des doctorants de l'IHPE, les anciens, maintenant docteurs, et les actuels qui, par leur sympathie et leur générosité, ont fait de ce lieu un endroit amical et divertissant.*

*Many thanks to all the members of Tim Anderson's team at the Texas Biomedical Research Institute who warmly welcomed me during three months in San Antonio and greatly contributed to this thesis.*

*Je souhaite aussi remercier Monsieur Jean-François Allienne alias Jeff pour le soutien technique au travers notamment de la plateforme bioenvironnement. Je remercie également Monsieur Bernd Schuttengruber de l'institut de génétique humaine de Montpellier pour le don d'embryons de drosophile qui ont servi pour mes travaux pour le spike-in.*

*Ces travaux de thèse ont été rendus possibles grâce aux différents financements accordés par : la région Occitanie, l'ANR ChronoGET (ANR-17-CE12-0005-01) et le NIH (N° : 1R01-AI133749-01 ; Subaward No. 53409).*

*Enfin, remerciements pleins de gratitude à ma famille et mes amis qui, de près et même de loin, ont veillé sur moi et m'ont soutenue dans cette aventure !*



# TABLE DES MATIÈRES

<b>Introduction générale.....</b>	<b>1</b>
1. Rythmes biologiques .....	1
1.1. Origine .....	1
1.2. Expériences historiques et congrès fondateur de la chronobiologie moderne .....	2
1.3. Rythmes exogènes.....	3
1.4. Rythmes endogènes .....	3
2. Bilharziose .....	5
2.1. Cycle de vie des schistosomes : l'exemple de <i>Schistosoma mansoni</i> .....	7
2.2. Rythme d'émission des cercaires .....	8
2.3. <i>Schistosoma mansoni</i> : De Porto Rico à Oman.....	10
3. Problématiques et enjeux .....	12
<b>Matériel et méthodes.....</b>	<b>13</b>
1. Matériel biologique .....	13
1.1. Les schistosomes .....	13
1.2. Les mollusques (hôtes intermédiaires) .....	13
1.3. Les rongeurs (hôtes définitifs).....	13
2. Méthodes et techniques d'entretien des cycles .....	14
2.1. Infestation des mollusques .....	14
2.2. Infestation des souris .....	14
2.3. Récupération des adultes, des œufs et des miracidia.....	15
3. Méthode de chronobiologie de l'émission cercarienne.....	16
4. Méthode du sexage des parasites <i>Schistosoma mansoni</i> et <i>Schistosoma rodhaini</i> au stade cercaire.....	17
4.1. Echantillonnage des cercaires .....	17
4.2. Extraction ADN .....	18
4.3. PCR multiplex.....	19
4.4. Electrophorèse et diagnostic.....	20
5. Méthode des croisements de schistosomes .....	20
6. Méthode histologique .....	20
7. Méthode d'analyse statistique .....	22
7.1. Statistiques descriptives .....	22
7.2. Comparaison de deux moyennes .....	22
<b>Partie I Description des profils d'émission journaliers des cercaires chez <i>Schistosoma mansoni</i> ....</b>	<b>23</b>
1.1. Description des profils d'émission journaliers des cercaires chez les deux chronotypes, diurne et nocturne, de <i>Schistosoma mansoni</i> .....	23

I.1.1. Les paramètres de description .....	24
I.1.2. Présentation des chronotypes diurne et nocturne .....	27
I.1.2.1. Protocole .....	27
I.1.2.2. Chronotype diurne .....	27
I.1.2.3. Chronotype nocturne .....	28
I.1.2.4. Comparaison des chronotypes diurne et nocturne .....	29
I.1.3. Influence du mollusque .....	29
I.1.3.1. Protocole .....	29
I.1.3.2. Comparaison des profils issus de <i>Biomphalaria pfeifferi</i> et <i>Biomphalaria glabrata</i> .....	29
I.1.4. Influence du sexe du parasite.....	31
I.1.4.1. Protocole .....	31
I.1.4.2. Chronotype diurne .....	31
I.1.4.3. Chronotype nocturne .....	33
I.1.5. Discussion .....	34
I.2. Nature des profils d'émission journaliers des cercaires chez les deux chronotypes, diurne et nocturne, de <i>Schistosoma mansoni</i> .....	35
I.2.1. Inversion de la photopériode .....	36
I.2.1.1. Protocole lumière inversée chronotype diurne .....	37
I.2.1.2. Résultats chronotype diurne .....	37
I.2.1.3. Discussion et conclusion chronotype diurne.....	41
I.2.1.4. Protocole lumière inversée chronotype nocturne .....	42
I.2.1.5. Résultats chronotype nocturne .....	42
I.2.1.7. Comparaison des chronotypes diurne et nocturne avec inversion de la photopériode .....	47
I.2.2. Zeitgeber constant.....	47
I.2.2.1. Chronotype diurne .....	48
I.2.2.1.1. Protocole lumière continue.....	48
I.2.2.1.2. Résultats. Expérience de lumière continue.....	48
I.2.2.1.4. Expérience d'histologie .....	56
I.2.2.1.5. Résultats histologie après 13 jours de photopériode équilibrée .....	57
I.2.2.1.6. Résultats histologie après 13 jours de lumière continue .....	57
I.2.2.1.7. Discussion expérience lumière continue et histologie.....	58
I.2.2.2. Chronotype nocturne (expérience d'obscurité continue).....	62
I.2.2.2.1. Protocole obscurité continue .....	62
I.2.2.2.2. Expérience d'obscurité continue.....	62
I.2.2.2.3. Discussion et conclusion.....	69

I.3. Discussion .....	70
<b>Partie II : Déterminants moléculaires du rythme d'émission des cercaires chez <i>Schistosoma mansoni</i></b> .....	<b>77</b>
II.1. : Linkage mapping.....	78
II.1.1. Méthode du linkage QTL mapping .....	80
II.1.2. Croisement intraspécifique : <i>Schistosoma mansoni</i> diurne X <i>Schistosoma mansoni</i> nocturne .....	83
II.1.2.1. Protocole .....	83
II.1.2.2. Présentation des phénotypes.....	83
II.1.2.2.1. Les phénotypes F0 .....	83
II.1.2.2.2. Le croisement A (Figure_42) .....	84
II.1.2.2.3. Le croisement B (Figure_43).....	86
II.1.2.3. QTL mapping.....	88
II.1.2.4. Gènes candidats .....	89
II.1.2.4.1. Les gènes de l'horloge et le domaine Basic Helix-Loop-Helix .....	90
II.1.2.4.2. Les rhodopsines et autres gènes impliqués dans la phototransduction .....	90
II.1.2.4.3. Les gènes impliqués dans l'acétylation des histones .....	94
II.1.2.4.4. Enrichissement GO : Processus biologiques.....	95
II.1.2.4.5. Kegg pathway .....	96
II.1.2.5. Conclusion .....	97
II.1.3. Croisement interspécifique : <i>Schistosoma mansoni</i> (diurne) X <i>Schistosoma rodhaini</i> (nocturne).....	99
II.1.3.1. Protocole .....	99
II.1.3.2. Présentation des phénotypes.....	99
II.1.3.2.1. Les phénotypes F0 .....	99
II.1.3.2.2. Le croisement A .....	100
II.1.3.2.3. Le croisement B .....	102
II.1.3.3. QTL mapping.....	104
II.1.3.3.1. Croisement A .....	104
II.1.3.3.2. Croisement B .....	106
II.1.3.4. Gènes candidats .....	108
II.1.3.4.1. Les gènes de l'horloge (Basic Helix-Loop-Helix) .....	108
II.1.3.4.2. Les rhodopsines et autres gènes impliqués dans la phototransduction .....	109
II.1.3.4.3. Les gènes impliqués dans l'acétylation et la méthylation des histones.....	112
II.1.3.4.4. Enrichissement GO : Processus biologiques.....	113
II.1.3.4.5. Kegg pathway .....	114
II.1.3.5. Conclusion .....	115



II.2. Recherche des marqueurs rythmiques .....	119
II.2.1. Echantillonnage commun pour les expériences d'épigénétique et de transcriptomique	121
II.2.1.1. Méthode d'échantillonnage pour les expériences d'épigénétique et de transcriptomique.....	121
II.2.1.2. Chronobiologie .....	124
II.2.1.2.1. Protocole .....	124
II.2.1.2.2. Chronotype diurne .....	124
II.2.1.2.3. Chronotype nocturne .....	125
II.2.1.2.4. Comparaison des chronotypes.....	125
II.2.2. CHIPmentation.....	126
II.2.2.1. Méthode et optimisations.....	127
II.2.2.1.1. Méthode .....	127
II.2.2.1.2. Optimisations .....	128
II.2.2.1.3. Article CHIPmentation .....	132
II.2.2.2. H3K4me3 .....	132
II.2.2.2.1. Rythme des pics.....	133
II.2.2.2.2. Gènes associés aux marques H3K4me3 .....	135
II.2.2.2.3. Les gènes de l'horloge (Basic Helix-Loop-Helix) .....	135
II.2.2.2.4. Les rhodopsines et autres gènes impliqués dans la phototransduction.....	135
II.2.2.2.5. Les gènes impliqués dans l'acétylation et la méthylation des histones.....	138
II.2.2.2.6. Enrichissement GO : Processus biologique .....	139
II.2.2.2.7. Kegg pathway .....	141
II.2.2.2.8. Analyses DEseq2.....	143
II.2.2.2.8.1. Comparaison des chronotypes.....	143
II.2.2.2.8.2. Gènes différentiellement marqués par H3K4me3 issus du DEseq2 entre les deux chronotypes .....	144
II.2.2.3. H3K9ac.....	148
II.2.2.3.1. Rythme des pics.....	149
II.2.2.3.2. Gènes associés aux marques H3K9ac.....	150
II.2.2.3.3. Les gènes de l'horloge (Basic Helix-Loop-Helix) .....	150
II.2.2.3.4. Les rhodopsines et autres gènes impliqués dans la phototransduction.....	150
II.2.2.3.5. Les gènes impliqués dans l'acétylation et la méthylation des histones.....	151
II.2.2.3.6. Enrichissement GO : Processus biologique .....	151
II.2.2.3.7. Kegg pathway .....	152
II.2.2.3.8. Analyses DEseq2.....	152
II.2.2.3.8.1. Comparaison des chronotypes.....	152

II.2.2.3.8.2. Gènes différentiellement marqués par H3K9ac entre les deux chronotypes .....	153
II.2.3. RNA-seq .....	158
II.2.3.1. Méthode .....	158
II.2.3.2. Gènes rythmiques .....	158
II.2.3.2.1. Les gènes communs aux deux chronotypes .....	160
II.2.3.2.2. Les gènes diurnes rythmiques .....	162
II.2.3.2.2.1. Les gènes de l'horloge (Basic Helix-Loop-Helix) .....	162
II.2.3.2.2.2. Les rhodopsines et autres gènes impliqués dans la phototransduction .....	162
II.2.3.2.2.3. Les gènes impliqués dans l'acétylation des histones .....	164
II.2.3.2.2.4. Enrichissement GO : Processus biologique .....	164
II.2.3.2.2.5. Kegg pathway .....	165
II.2.3.2.2.6. Les gènes rythmiques communs en photopériode classique et inversée....	166
II.2.3.2.3. Les gènes nocturnes .....	169
II.2.3.2.3.1. Les gènes de l'horloge (Basic Helix-Loop-Helix) .....	169
II.2.3.2.3.2. Les rhodopsines et autres gènes impliqués dans la phototransduction .....	170
II.2.3.2.3.3. Les gènes impliqués dans l'acétylation des histones .....	172
II.2.3.2.3.4. Enrichissement GO : Processus biologique .....	172
II.2.3.2.3.5. Kegg pathway .....	173
II.2.3.2.3.6. Les gènes rythmiques communs en photopériode classique et inversée....	174
II.2.3.3. Comparaison des chronotypes .....	177
II.2.3.4. Intégration des résultats de génétique et épigénétique.....	179
<b>Discussion et conclusion générales .....</b>	<b>181</b>
<b>Participation aux congrès, manifestations scientifiques et Distinctions .....</b>	<b>189</b>
<b>Références .....</b>	<b>190</b>
<b>Annexes .....</b>	<b>197</b>

# LISTE DES FIGURES

Figure_01 : Schéma des mouvements des corps célestes et leur cycle environnemental associé. ....	2
Figure_02 : Eskinogramme d'un rythme endogène synchronisé par un cycle jour/nuit. ....	3
Figure_03 : Exemples de rythmes circadiens connus dans différents règnes du vivant. ....	4
Figure_04 : Mécanisme commun des horloges internes (oscillateurs) avec des exemples de différentes espèces. ....	5
Figure_05 : Carte de la répartition de la schistosomiase dans le monde .....	6
Figure_06 : Exemple du cycle de vie de <i>Schistosoma mansoni</i> . ....	7
Figure_07 : Photo d'une coquille de mollusque <i>Biomphalaria</i> infesté par des sporocystes de schistosomes (à gauche) et photo d'une cercaire (à droite). ....	8
Figure_08 : Dessins des stades de différenciation des cercaires issus des observations de coupes histologiques de mollusque <i>Biomphalaria</i> infestés par <i>Schistosoma mansoni</i> . ....	9
Figure_09 : Schéma de la répartition de l'émission cercarienne sur 24 heures par espèce de schistosome. ....	10
Figure_10 : Carte du Sultanat d'Oman avec un zoom sur les sites de Sheer et Tibraq de la région du Dhofar d'où sont issus les deux chronotypes de <i>S. mansoni</i> . ....	11
Figure_11 : Infestation des souris (photos). ....	14
Figure_12 : Etapes de la chronobiologie (photos). ....	16
Figure_13 : Détail des filtres utilisés pour le comptage .....	17
Figure_14 : Exemple de répartition d'une plaque 96 puits pour le sexage. ....	18
Figure_15 : Détail des coupes. ....	21
Figure_16 : Exemple d'un histogramme issu d'une expérience fictive de chronobiologie. ....	24
Figure_17 : Présentation de quatre des sept paramètres de description avec un exemple fictif de chronobiologie. ....	25
Figure_18 : Illustrations du coefficient d'asymétrie (skewness). ....	26
Figure_19 : Illustrations du coefficient d'aplatissement (kurtosis). ....	26
Figure_20 : Chronobiologie des chronotypes diurne (orange) et nocturne (violet) de <i>S. mansoni</i> . .	27
Figure_21 : Chronobiologie du chronotype diurne de <i>S. mansoni</i> à partir de <i>B. pfeifferi</i> et de <i>B. glabrata</i> . ....	30
Figure_22 : Profils moyens d'émission de <i>S. mansoni</i> de chronotype diurne entre mâles (bleu) et femelles (rose). ....	32
Figure_23 : Chronobiologie de <i>S. mansoni</i> nocturne mâle (bleu) et femelle (rose). ....	33
Figure_24 : Histogrammes de l'expérience de chronobiologie en lumière inversée pour le chronotype diurne. ....	38
Figure_25 : Histogramme de l'expérience de chronobiologie en lumière inversée pour le chronotype nocturne. ....	43

Figure_26 : Histogramme de l'expérience de chronobiologie en lumière continue pour le chronotype diurne (1/2), phase lumière continue.....	49
Figure_27 : Histogramme de l'expérience de chronobiologie en lumière continue pour le chronotype diurne (2/2) : phase retour de la photopériode équilibrée.....	51
Figure_28 : Coupe histologique d'hépatopancréas de <i>B. glabrata</i> infesté avec le chronotype diurne de <i>S. mansoni</i> (Oman) après 13 jours de photopériode équilibrée (L-O) .....	57
Figure_29 : Coupe histologique d'hépatopancréas de <i>B. glabrata</i> infesté avec le chronotype diurne de <i>S. mansoni</i> (Oman) après 13 jours de lumière continue.....	58
Figure_30 : Corrélation entre la taille des mollusques et la production cercarienne aux 2ème et 3ème jours du retour de la photopériode équilibrée .....	59
Figure_31 : Schéma de l'hypothèse du devenir des cercaires en lumière continue dans les mollusques de grande taille .....	60
Figure_32 : Schéma de l'hypothèse du devenir des cercaires en lumière continue dans les mollusques de petite taille .....	61
Figure_33 : Histogramme de l'expérience de chronobiologie d'Obscurité continue pour le chronotype nocturne (1/2), phase obscurité continue. ....	63
Figure_34 : Histogramme de l'expérience de chronobiologie d'Obscurité continue pour le chronotype nocturne (2/2), phase retour à la photopériode équilibrée.....	65
Figure_35 : Schéma du modèle du sablier (« hourglass »). ....	72
Figure_36 : Schéma du modèle de la coïncidence externe à travers l'exemple simplifié de la floraison d' <i>Arabidopsis thaliana</i> . ....	73
Figure_37 : Résumé des principaux résultats de la première partie. ....	75
Figure_38 : Parallèle entre la partition musicale et l'approche moléculaire tri-disciplinaire (Génétique, Epigénétique et Transcriptomique). ....	77
Figure_39 : Schéma de la ségrégation de chromosomes autosomiques dans le cas d'un rétrocroisement et d'un croisement test.....	79
Figure_40 : Schéma de l'analyse QTL avec les fonctions utilisées dans R/qtl. ....	82
Figure_41 : Chronobiologie des chronotypes diurne (orange) et nocturne (violet) de <i>S. mansoni</i> pour l'expérience de QTL mapping.....	83
Figure_42 : Chronobiologie du croisement A intraspécifique, avec F0 diurne (orange), F0 nocturne (violet) et une sélection d'individus F2 hybrides (vert). ....	85
Figure_43 : Chronobiologie du croisement B intraspécifique, avec F0 diurne (orange), F0 nocturne (violet), F1 et une sélection d'individus F2 hybrides (vert). ....	87
Figure_44 : LOD score pour les croisements intraspécifiques de <i>Schistosoma mansoni</i> obtenu avec l'algorithme EM (Espérance-Maximisation) de la fonction scanone.....	88
Figure_45 : Transmission allélique pour les QTLs identifiés pour le croisement intraspécifique entre les deux chronotypes de <i>Schistosoma mansoni</i> . ....	89
Figure_46 : Schéma de la phototransduction du signal dans les bâtonnets rétiniens.....	92
Figure_47 : Treemap des résultats de l'analyse d'enrichissement GO des processus biologiques obtenus pour les gènes QTLs du croisement intraspécifique. ....	96

Figure_48 : Chronobiologie du chronotype diurne de <i>S. mansoni</i> (orange) et de <i>S. rodhaini</i> nocturne (violet) pour l'expérience de QTL mapping. ....	99
Figure_49 : Chronobiologie du croisement A interspécifique, avec F0 diurne (orange), F0 nocturne (violet) et F1 et une sélection d'individus F2 hybrides (vert).....	101
Figure_50 : Chronobiologie du croisement B interspécifique, avec F0 diurne (orange), F0 nocturne (violet) et F1 et une sélection d'individus F2 hybrides (vert).....	103
Figure_51 : LOD score QTL pour le croisement A interspécifique entre <i>Schistosoma mansoni</i> et <i>Schistosoma rodhaini</i> obtenu avec l'algorithme EM de la fonction scanone. ....	104
Figure_52 : Effet plot pour le QTL du chromosome 3 pour le croisement A interspécifique entre <i>Schistosoma mansoni</i> et <i>Schistosoma rodhaini</i> obtenu avec l'algorithme EM de la fonction scanone. ....	105
Figure_53 : LOD score QTL pour le croisement B interspécifique de <i>Schistosoma mansoni</i> et <i>Schistosoma rodhaini</i> obtenu avec l'algorithme EM de la fonction scanone. ....	106
Figure_54 : Effet plot pour le QTL du chromosome 3 pour le croisement B interspécifique entre <i>Schistosoma mansoni</i> et <i>Schistosoma rodhaini</i> obtenu avec l'algorithme EM de la fonction scanone. ....	107
Figure_55 : Effet plot pour le QTL du chromosome 4 pour le croisement B interspécifique entre <i>Schistosoma mansoni</i> et <i>Schistosoma rodhaini</i> obtenu avec l'algorithme EM de la fonction scanone. ....	108
Figure_56 : Treemap revigo des résultats de l'analyse d'enrichissement GO des processus biologiques obtenus pour les gènes QTL du croisement interspécifique. ....	113
Figure_57 : Bilan des QTL mapping intra et interspécifiques.....	117
Figure_58 : Schéma de l'expression d'un gène de la transcription jusqu'à la traduction. ....	119
Figure_59 : Schéma de la transcription d'un gène et des acteurs de la régulation. ....	120
Figure_60 : Les quatre principaux types de modification épigénétique. ....	120
Figure_61 : Répartition de l'échantillon entre les expériences de ChIPmentation et de RNA-seq. ....	122
Figure_62 : Description des deux séries d'échantillonnage pour les expériences d'épigénétique et transcriptomique.....	123
Figure_63 : Chronobiologie des chronotypes diurne (orange) et nocturne (violet) de <i>S. mansoni</i> pour les échantillons d'épigénétique et transcriptomique. ....	124
Figure_64 : Les quatre types de modification d'histone avec des exemples de marque.....	126
Figure_65 : Schéma récapitulatif des différentes optimisations pour l'étude des modifications d'histone chez <i>S. mansoni</i> .....	128
Figure_66 : Profils méta-gènes pour les marques H3K4me3 et H3K9ac sur le génome de <i>S. mansoni</i> , avec 0µL, 5µL, 10µL ou 15µL de spike-in drosophile. A) Marque H3K4me3, B) Marque H3K9ac. .	129
Figure_67 : Pipeline bioinformatique utilisé pour l'analyse finale des échantillons de ChIPmentation avec Snakepipe.....	131
Figure_68 : Profils des deux marques H3K4me3 rythmiques chez le chronotype nocturne .....	134
Figure_69 : Treemap revigo des résultats de l'analyse d'enrichissement GO des processus biologiques obtenus pour les gènes associés à la marque H3K4me3.....	140

Figure_70 : Analyse en composante principale (ACP) issue de l'analyse DEseq2 entre les chronotypes diurne et nocturne de <i>S. mansoni</i> (sexe confondu). .....	143
Figure_71 : MA-plots issus des analyses DEseq2 diurnes Vs nocturnes pour H3K4me3.....	144
Figure_72 : Bilan de la marque H3K4me3 pour les chronotypes diurne et nocturne de <i>S. mansoni</i> . .....	147
Figure_73 : Profil de la marque H3K9ac rythmique chez le chronotype nocturne. ....	150
Figure_74 : Treemap revigo des résultats de l'analyse d'enrichissement GO des processus biologiques obtenus pour les gènes associés à la marque H3K9ac. ....	151
Figure_75 : Analyse en composante principale (ACP) issue de l'analyse DEseq2 entre les chronotypes diurne et nocturne de <i>S. mansoni</i> (sexes confondus) pour la marque H3K9ac.....	152
Figure_76 : MA-plots issus des analyses DEseq2 diurnes Vs nocturnes pour H3K9ac. ....	153
Figure_77 : Bilan de la marque H3K9ac pour les chronotypes diurne et nocturne de <i>S. mansoni</i> . ....	157
Figure_78 : Diagramme des gènes rythmiques en fonction des logiciels et des chronotypes. ....	159
Figure_79 : Exemple de profils d'expression de gène commun aux deux chronotypes (pic d'expression à la même heure). .....	161
Figure_80 : Exemple de profils d'expression de gène commun aux deux chronotypes (pic d'expression avant le pic d'émission des cercaires).....	162
Figure_81 : Treemap des résultats de l'analyse d'enrichissement GO des processus biologiques obtenus pour les gènes diurnes rythmiques. ....	165
Figure_82 : Exemple de profils d'expression de gène commun aux chronotypes Diurnes L-O et O-L (pic d'expression inversé). .....	168
Figure_83 : Treemap des résultats de l'analyse d'enrichissement GO des processus biologiques obtenus pour les gènes nocturnes rythmiques. ....	173
Figure_84 : Exemple de profils d'expression de gène commun aux chronotypes Nocturnes L-O et O-L (pic d'expression inversé).....	175
Figure_85 : Bilan des gènes rythmiques.....	178
Figure_86 : MA-plots issus des analyses DEseq2 diurnes Vs nocturnes pour les gènes candidats et les 28 « rhodopsine-like ».....	179
Figure_87 : Proposition d'un modèle chronobiologique et moléculaire du rythme de l'émission des cercaires chez les deux chronotypes de <i>S. mansoni</i> .....	186

# LISTE DES TABLEAUX

Tableau_01 : Dénomination des rythmes exogènes en fonction de leur période.....	3
Tableau_02 : Dénomination des rythmes endogènes en fonction de leur période .....	4
Tableau_03 : Composition de la solution d’anesthésique .....	15
Tableau_04 : Composition de la solution de perfusion.....	15
Tableau_05 : Composition de la solution de Lugol .....	16
Tableau_06 : Composition de la solution de neutralisation .....	18
Tableau_07 : Séquence des amorces pour le sexage .....	19
Tableau_08 : Composition du mix PCR multiplex pour le sexage .....	19
Tableau_09 : Composition du mix des couples d’amorce pour le sexage (10X).....	19
Tableau_10 : Programme du thermocycleur pour PCR multiplex avec les couples d’amorces ScaffoLO_02739 et Rhodopsine_3 (utilisés pour le chronotype diurne de <i>S. mansoni</i> ) .....	19
Tableau_11 : Programme du thermocycleur pour PCR multiplex avec les couples d’amorces ScaffoLO_02739 et Rhodopsine_3_BIS (utilisés pour le chronotype nocturne de <i>S. mansoni</i> et <i>S. rodhaini</i> ).....	20
Tableau_12 : Composition de la solution de Halmi .....	21
Tableau_13 : Paramètres des profils moyens d’émission, accompagnés des variations individuelles, pour les parasites diurne et nocturne de <i>S. mansoni</i> .....	28
Tableau_14 : Paramètres des profils moyens d’émission, accompagnés des variations individuelles, pour les parasites diurnes de <i>S. mansoni</i> à partir de <i>B. pfeifferi</i> et de <i>B. glabrata</i> .....	30
Tableau_15 : Paramètres des profils moyens d’émission, accompagnés des variations individuelles, pour les parasites mâles et femelles du chronotype diurne de <i>S. mansoni</i> .....	32
Tableau_16 : Paramètres des profils moyens d’émission, accompagnés des variations individuelles, pour les parasites mâles et femelles du chronotype nocturne de <i>S. mansoni</i> .....	33
Tableau_17 : Paramètres des profils moyens d’émission, accompagnés des variations individuelles, pour les parasites du chronotype diurne de <i>S. mansoni</i> de l’expérience inversion de photopériode .....	39
Tableau_18 : Paramètres des profils moyens d’émission, accompagnés des variations individuelles, pour les parasites du chronotype nocturne de <i>S. mansoni</i> de l’expérience inversion de photopériode .....	44
Tableau_19 : Paramètres des profils moyens d’émission, accompagnés des variations individuelles, pour les parasites du chronotype diurne de <i>S. mansoni</i> de l’expérience de lumière continue (1/2)	50
Tableau_20 : Paramètres des profils moyens d’émission, accompagnés des variations individuelles, pour les parasites du chronotype diurne de <i>S. mansoni</i> de l’expérience de lumière continue (2/2)	52
Tableau_21 : Corrélation entre la taille des mollusques et la production cercarienne du jour 1 et 2 de la photopériode équilibrée, du 1 <sup>er</sup> jour de la lumière continue et des 2 <sup>ème</sup> jour et 3 <sup>ème</sup> jour de la photopériode équilibrée après 15 jours de lumière continue .....	59

Tableau_22 : Paramètres des profils moyens d'émission, accompagnés des variations individuelles, pour les parasites du chronotype nocturne de <i>S. mansoni</i> de l'expérience d'obscurité continue (1/2) .....	64
Tableau_23 : Paramètres des profils moyens d'émission, accompagnés des variations individuelles, pour les parasites du chronotype nocturne de <i>S. mansoni</i> de l'expérience d'obscurité continue (2/2) .....	66
Tableau_24 : Paramètres des profils moyens d'émission, accompagnés des variations individuelles, pour les parasites diurne et nocturne de <i>S. mansoni</i> pour l'expérience de QTL mapping.....	84
Tableau_25 : Liste des gènes QTLs avec un motif BHLH (Basic Helix Loop Helix).....	90
Tableau_26 : Liste des gènes QTLs rhodospine et rhodospine_like .....	91
Tableau_27 : Liste des gènes QTLs en lien avec la phototransduction .....	93
Tableau_28 : Liste des gènes QTLs en lien avec les histones et l'acétylation/désacétylation des histones .....	95
Tableau_29 : Liste des gènes QTLs correspondant aux Kegg pathways sélectionnés.....	96
Tableau_30 : Paramètres des profils moyens d'émission, accompagnés des variations individuelles, pour les parasites diurne <i>S. mansoni</i> et nocturne <i>S. rodhaini</i> pour l'expérience de QTL mapping	100
Tableau_31 : Tableau récapitulatif du scantwo pour déterminer l'interaction (épistasie et l'effet additif) des QTLs pour le croisement B interspécifique .....	106
Tableau_32 : Explication de la variance du phénotype (moyenne du pourcentage de cercaires nocturne) par le modèle additif des QTLs et les QTLs séparés pour le croisement B interspécifique .....	107
Tableau_33 : Liste des gènes QTLs avec un motif BHLH (Basic Helix Loop Helix) pour les QTLs des croisements interspécifiques .....	109
Tableau_34 : Liste des gènes QTLs rhodopsine et rhodopsine-like pour le QTL interspécifique....	109
Tableau_35 : Liste des gènes QTLs potentiellement liés à la phototransduction.....	110
Tableau_36 : Liste des gènes QTLs en lien avec les histones, l'acétylation des histones et la méthylation/déméthylation des histones .....	112
Tableau_37: Liste des gènes QTLs correspondant aux Kegg pathways sélectionnés. ....	114
Tableau_38 : Description des deux séries d'échantillonnage pour les expériences d'épigénétique et de transcriptomique.....	123
Tableau_39 : Paramètres des profils moyens d'émission, accompagnés des variations individuelles, pour les parasites diurne et nocturne de <i>S. mansoni</i> pour les échantillons de l'expérience épigénétique et transcriptomique.....	125
Tableau_40 : Profondeur moyenne des pics uniques à <i>Drosophila melanogaster</i> en fonction du volume de spike-in pour la marque H3K4me3 .....	130
Tableau_41 : Profondeur moyenne des pics uniques à <i>Drosophila melanogaster</i> en fonction du volume de spike-in pour la marque H3K9ac.....	130
Tableau_42 : Nombre et pourcentage d'alignement des reads moyens, accompagnés des variations individuelles, pour les parasites diurne et nocturne de <i>S. mansoni</i> pour les échantillons de l'expérience épigénétique H3K4me3 .....	132



Tableau_43 : Nombre de pics H3K4me3 détectés comme rythmiques et nombre de gènes associé en fonction du logiciel et du chronotype du parasite ( $p < 0,05$ ). .....	134
Tableau_44 : Liste des gènes associés aux pics H3K4me3, rhodopsine et rhodopsine-like .....	135
Tableau_45 : Liste des gènes associés aux pics H3K4me3 potentiellement liés à la phototransduction .....	135
Tableau_46 : Liste des gènes associés avec la marque H3K4me3 en lien avec les histones, l'acétylation/désacétylation des histones et la méthylation/déméthylation des histones.....	138
Tableau_47 : Liste des gènes associés avec la marque H3K4me3 en lien avec l'annotation GO processus métabolique des pigments et processus de synthèse des pigments.....	140
Tableau_48 : Liste des gènes associés avec la marque H3K4me3 en lien avec les Kegg pathway sélectionnés .....	141
Tableau_49 : Tableau récapitulatif des analyses DEseq2 entre les chronotypes diurne et nocturne de <i>S. mansoni</i> pour la marque H3K4me3 .....	144
Tableau_50 : Liste des gènes associés aux marques H3K4me3 différenciellement présentes entre les deux chronotypes et communs aux gènes QTL .....	145
Tableau_51 : Nombres et pourcentages d'alignement des reads moyens, accompagnés des variations individuelles, pour les parasites diurne et nocturne de <i>S. mansoni</i> pour les échantillons de l'expérience épigénétique H3K9ac.....	148
Tableau_52 : Nombre de pics H3K9ac détectés comme rythmiques et nombre de gènes associés en fonction du logiciel et du chronotype du parasite ( $p < 0,05$ ).....	149
Tableau_53 : Tableau récapitulatif des analyses DEseq2 entre les chronotypes diurne et nocturne de <i>S. mansoni</i> pour la marque H3K9ac .....	153
Tableau_54 : Liste des gènes associés aux marques H3K9ac différenciellement présentes entre les deux chronotypes et communs aux gènes QTL et/ou sélectionnés dans les points précédents....	154
Tableau_55 : Liste des gènes associés aux marques H3K9ac différenciellement présentes entre les deux chronotypes qui sont co-marqués avec H3K4me3 .....	154
Tableau_56 : Nombre de gènes détectés comme rythmiques en fonction du logiciel et du chronotype du parasite ( $p < 0,05$ ) pour les échantillons en photopériode équilibrée .....	159
Tableau_57 : Liste des gènes communs aux chronotypes diurne et nocturne avec le pic d'expression des gènes .....	160
Tableau_58 : Répartition des pics d'expression des gènes rythmiques dans les deux chronotypes .....	161
Tableau_59 : Liste des gènes diurnes rythmiques rhodopsine-like.....	163
Tableau_60 : Liste des gènes diurnes rythmiques liés à la phototransduction.....	163
Tableau_61 : Liste des gènes diurnes rythmiques issus des annotations GO sélectionnées.....	164
Tableau_62 : Liste des gènes diurnes rythmiques issus des Kegg pathways sélectionnées.....	166
Tableau_63 : Liste des gènes diurnes rythmiques communs aux échantillons photopériode classique (L-O) et photopériode inversée (O-L). Les gènes aux pics d'expression diamétralement opposés sont marqués en rouge. ....	167
Tableau_64 : Répartition des pics d'expression des gènes rythmiques aux deux photopériodes du chronotype diurne.....	169

Tableau_65 : Liste des gènes nocturnes rythmiques BHLH .....	170
Tableau_66 : Liste des gènes nocturnes rythmiques rhodopsine-like.....	170
Tableau_67 : Liste des gènes nocturnes rythmiques liés à la phototransduction .....	171
Tableau_68 : Liste des gènes nocturnes rythmiques issus des annotations GO sélectionnées.....	172
Tableau_69 : Liste des gènes diurnes rythmiques issus des Kegg pathways sélectionnés.....	173
Tableau_70 : Liste des gènes nocturnes rythmiques communs aux échantillons photopériode classique (L-O) et photopériode inversée (O-L). .....	175
Tableau_71 : Répartition des pics d'expression des gènes rythmiques aux deux photopériodes du chronotype nocturne.....	176
Tableau_72 : Répartition des pics d'expression des gènes rythmiques des deux chronotypes en photopériode classique (L-O) .....	177
Tableau_73 : Tableau récapitulatif des analyses DEseq2 entre les chronotypes diurne et nocturne de <i>S. mansoni</i> pour les gènes candidats sélectionnés.....	180

# LISTE DES ANNEXES

**Annexe 1 : Article sur la chronobiologie comme outil de mise en évidence d'un événement d'hybridation**

**Annexe 2 : Test de la PCR multiplex pour le sexage avec les couples d'amorces Scaff\_02739 et Rhodospine 3\_BIS**

**Annexe 3 : Répartition des profils chronobiologiques des générations F0, F1 et F2 des croisements intraspécifiques**

**Annexe 4 : Liste des gènes candidats issus du QTL mapping des croisements intraspécifiques**

**Annexe 5 : Répartition des profils chronobiologiques des générations F0, F1 et F2 des croisements interspécifiques**

**Annexe 6 : Liste des gènes candidats issus du QTL mapping des croisements interspécifiques**

**Annexe 7 : Article méthodologique sur la technique de ChIPmentation sur échantillons *S. mansoni***

**Annexe 8 : Liste des gènes différenciellement marqués par H3K4me3 issus du DEseq2 entre les chronotypes diurne et nocturne de *S. mansoni***

**Annexe 9 : Liste des gènes différenciellement marqués par H3K9ac issus du DEseq2 entre les chronotypes diurne et nocturne de *S. mansoni***

**Annexe 10 : Liste des gènes diurnes rythmiques**

**Annexe 11 : Liste des gènes nocturnes rythmiques**

**Annexe 12 : Liste des gènes candidats sélectionnés pour être potentiellement impliqués dans le rythme d'émission des cercaires**

**Annexe 13 : Liste des gènes candidats différenciellement exprimés**

## Introduction générale

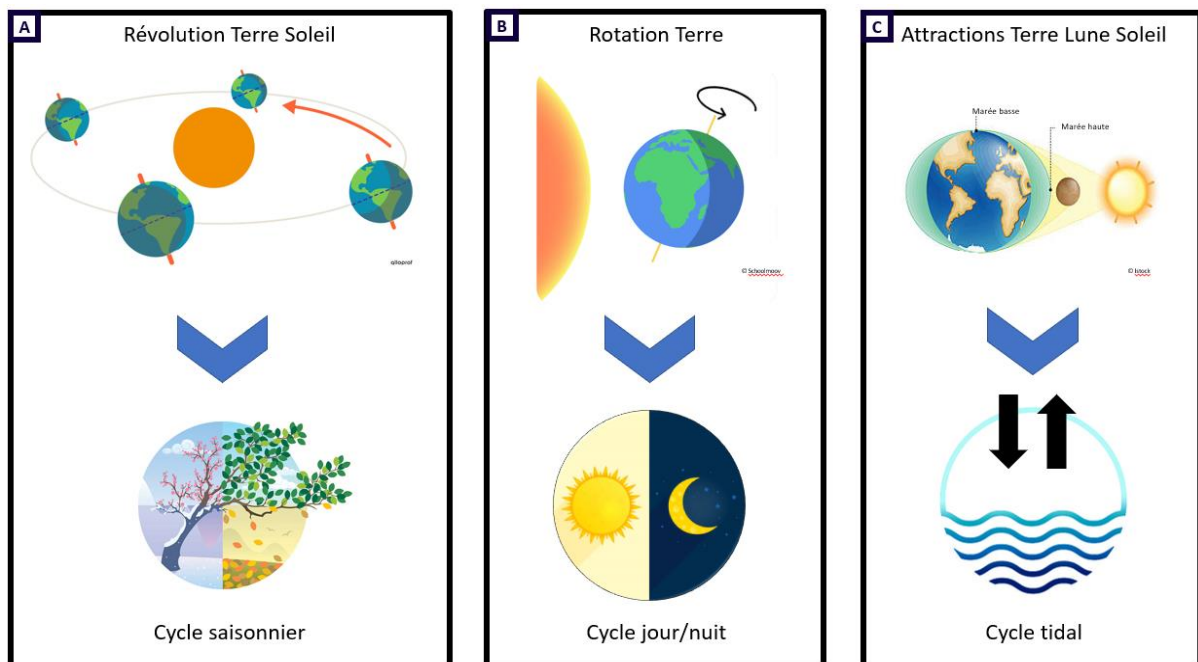
### 1. Rythmes biologiques

#### 1.1. Origine

L'environnement est soumis à des phénomènes physiques cycliques causés par les mouvements et attractions des corps célestes. La rotation de la terre autour du soleil entraîne un cycle annuel de 365 jours, c'est-à-dire le temps pour effectuer une révolution complète. La combinaison de cette révolution autour du soleil ainsi que de l'angle de rotation de la terre sur elle-même ( $23,5^\circ$ ) entraîne les cycles saisonniers (Figure\_01\_A). En fonction de la position de la terre par rapport au soleil et de l'angle d'arrivée des rayons du soleil, leur diffusion sur la surface de la terre diffère. L'angle d'arrivée des rayons du soleil détermine la chaleur qui sera diffusée à la surface de la terre. Plus l'angle est droit, plus les rayons sont concentrés et plus la chaleur est importante. La saison la plus chaude correspond donc au moment où les rayons du soleil sont le plus concentrés sur la surface donnée et inversement pour la période la plus froide.

La rotation de la terre sur elle-même entraîne un cycle journalier de 24h correspondant au temps pour effectuer cette révolution. Ce cycle voit succéder une alternance entre la période d'éclaircissement (jour) et de non-éclairage (nuit) de la surface terrestre (Figure\_01\_B). Cette alternance s'accompagne par des variations de la température causées par la présence ou l'absence des rayons du soleil.

La rotation de la lune autour de la terre ainsi que les attractions mutuelles entre la lune, le soleil et la terre, entraînent le cycle tidal (Figure\_01\_C). Les forces d'attractions exercées entraînent une légère déformation de la terre avec un flux et reflux des marées. Un cycle tidal dure environ 12h, avec 6h de montée des eaux et 6h de descente, entraînant ainsi des modifications importantes des rivages.



## Introduction générale

**Figure 01 : Schéma des mouvements des corps célestes et leur cycle environnemental associé.** Les distances entre les différents corps célestes ne sont pas à l'échelle. A) La révolution de la terre autour du soleil entraîne le cycle saisonnier. B) La rotation de la terre sur elle-même entraîne le cycle jour/nuit. C) Les forces d'attraction de la terre, de la lune et du soleil entraînent le cycle tidal.

L'ensemble de ces continues variations cycliques de l'environnement ont imposé aux êtres vivants de toute forme de s'y adapter pour optimiser leur survie. Certains se sont émancipés de cette contrainte en vivant dans des environnements stables comme les profondeurs océaniques ou les caves naturelles. Les autres se sont adaptés à ces contraintes en adoptant des rythmes biologiques synchronisés avec leur environnement leur permettant ainsi d'anticiper les changements (Sharma, 2003). On définit alors comme rythme biologique tous les phénomènes biologiques qui se répètent à intervalles réguliers. Il se caractérise par une période, une phase et une amplitude. Ces phénomènes peuvent être présents à toutes les échelles, allant d'un trait d'histoire de vie à une fonction physiologique jusqu'au niveau cellulaire avec des interactions biochimiques (Marsaudon, 2006).

### 1.2. Expériences historiques et congrès fondateur de la chronobiologie moderne

La discipline scientifique qui étudie les phénomènes de rythme chez les êtres-vivants se nomme chronobiologie. Ce mot qui tire son étymologie du grec, se compose des racines χρόνος (*khrónos*) le temps, de βίος (*bíos*) la vie et de λόγος (*logos*) étude. Par définition, la chronobiologie va rassembler l'étude de l'organisation temporelle des êtres-vivants ainsi que ses altérations et l'étude des mécanismes impliqués (Reinberg, 2003).

Si la discipline à proprement parler est récente puisque le premier emploi du terme « chronobiologie » date de 1967 (Klarsfeld, 2013), les premières observations rapportées de rythme chez des êtres vivants remontent au IV<sup>e</sup> siècle avant J-C. Ainsi, Androsthène qui servit sous le règne d'Alexandre le grand, rapporte les premières observations d'un rythme journalier dans l'ouverture et la fermeture des feuilles d'un arbre (Dunlap, Loros & DeCoursey, 2004). Il faudra attendre 1729 et l'astronome français Jean-Jacques Dortous de Mairan pour qu'une réelle expérience soit consacrée aux rythmes des êtres-vivants (Dortous de Mairan, 1729). C'est avec l'espèce *Mimosa pudica* que l'astronome observe que les mouvements journaliers des feuilles persistent en obscurité prolongée. Il pose ainsi les premières bases scientifiques qui permettront d'aboutir à l'existence d'un rythme biologique ne dépendant pas uniquement de l'alternance jour/nuit. En 1832, les travaux du botaniste suisse Augustin-Pyramus de Candolle complètent les observations de Dortous de Mairan en montrant que non seulement ces mouvements journaliers persistent aussi en condition d'éclairage continu mais que le rythme de ces mouvements foliaires ne fait pas exactement 24h comme la rotation terrestre mais plutôt 22-23h (McClung, 2006 ; Klarsfeld, 2013). C'est en 1932 qu'Erwin Bünning, biologiste allemand, va mettre en évidence chez des feuilles de haricots que les mouvements foliaires persistants en l'absence de cycle jour/nuit sont héréditaires et sont donc déterminés génétiquement (Queiroz, 1979). Il faudra attendre 1959 pour que Franz Halberg, biologiste roumain travaillant à l'université du Minnesota, propose le terme de *circadien* pour qualifier ce genre de rythme (*in* Erren, 2018). Issu du latin, le mot circadien, signifie presque (*circa*) une journée (*diem*).

Le milieu du XX<sup>ème</sup> siècle marque l'explosion des expériences sur les rythmes biologiques et met en lumière le manque de terminologie commune. Le 25<sup>ème</sup> "Cold Spring Harbor Symposium of Quantitative Biology" consacré aux horloges biologiques («biological clocks») va y remédier. Ce

## Introduction générale

congrès est considéré comme fondateur de la chronobiologie moderne puisqu'il a réuni au même endroit tous les futurs grands noms du domaine comme Jürgen Aschoff, Erwin Bünning, Franz Halberg et Colin Pittendrigh. C'est aussi et surtout lors de ce congrès qu'ont été conceptualisées et définies la notion d'horloge biologique mais aussi la distinction entre les rythmes endogènes et exogènes (Dunlap, Loros & DeCoursey, 2004; Erren, 2018; Kuhlman, Craig & Duffy, 2018). Ces concepts, notions et définitions, sont toujours valables aujourd'hui.

### 1.3. Rythmes exogènes

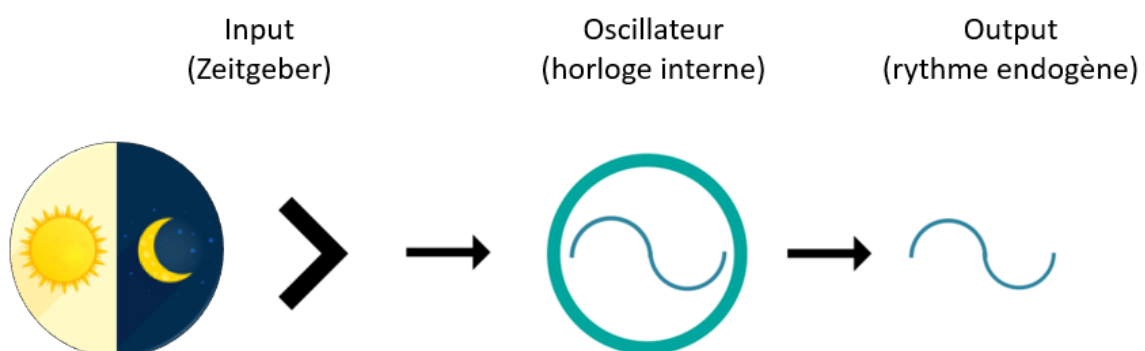
Les rythmes biologiques exogènes sont l'ensemble des rythmes causés exclusivement par un *zeitgeber* ou synchroniseur environnemental. Ce terme allemand démocratisé par Aschoff peut être traduit par « donneur de temps ». Le principal *zeitgeber* est le cycle jour/nuit mais il existe des *zeitgebers* non photiques. C'est le cas de la température, de l'humidité ou encore des interactions sociales (Dunlap, Loros & DeCoursey, 2004). La période du rythme exogène correspond à celle du *zeitgeber* (Tableau\_01) (Aschoff, 1960; Queiroz, 1979). Dès lors qu'il est privé de son *zeitgeber*, le rythme biologique exogène cesse.

**Tableau\_01 : Dénomination des rythmes exogènes en fonction de leur période**

Période	1 an	24h	1 mois	12h
Rythme exogène	Annuel	Nyctéméral	Mensuel	Tidal

### 1.4. Rythmes endogènes

Les rythmes endogènes sont l'ensemble des rythmes sous le contrôle d'une horloge interne (Figure\_02) ; c'est-à-dire des rythmes qui sont dus à l'organisme et pas seulement à l'environnement (Aschoff, 1960).



**Figure\_02 : Eskinogramme d'un rythme endogène synchronisé par un cycle jour/nuit.** Dans sa représentation la plus simplifiée, un rythme endogène est déclenché par l'action conjointe du *zeitgeber* et de l'horloge interne. L'horloge interne ou oscillateur va se synchroniser avec le *zeitgeber* (ici le cycle jour/nuit) et commander le rythme endogène (output, exemple mouvement des feuilles).

## Introduction générale

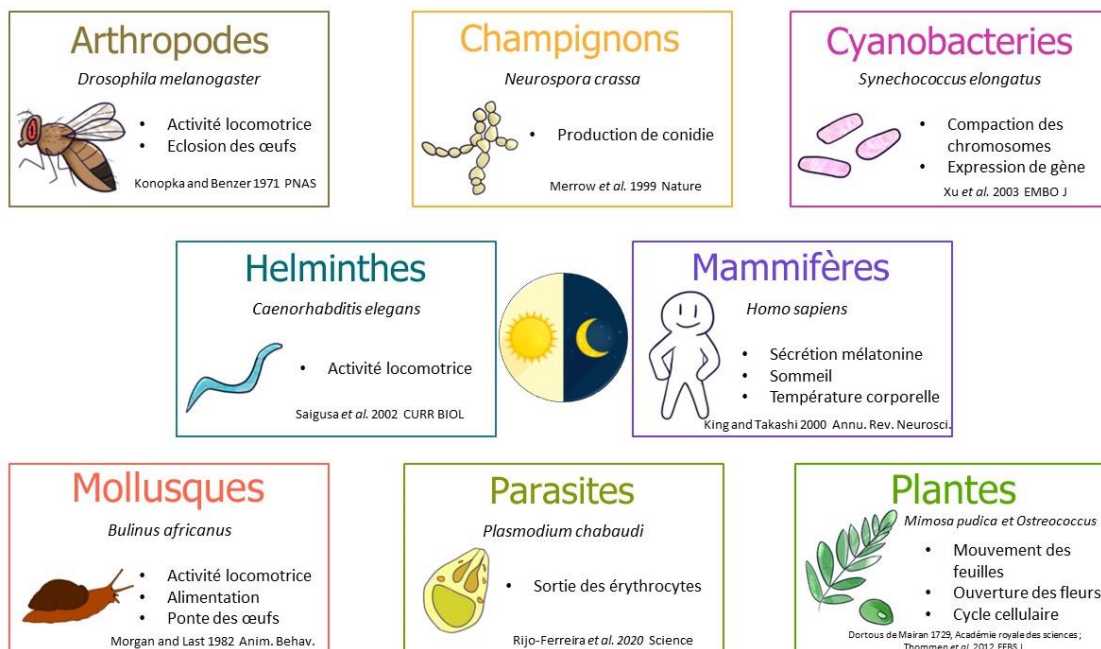
Ces rythmes reposent sur trois caractéristiques fondamentales (Queiroz, 1979; Rijo-Ferreira, Takahashi & Figueiredo, 2017).

- 1) Les rythmes endogènes persistent en l'absence de synchroniseurs environnementaux (zeitgebers). On parle alors de « free running period » (FRP) ou période de libre cours. C'est-à-dire la période que va prendre le rythme endogène lorsque l'organisme est dans un environnement constant.
- 2) Les rythmes endogènes sont entraînaibles. C'est-à-dire que l'horloge interne se synchronise avec les zeitgebers. C'est par exemple l'adaptation à un décalage horaire.
- 3) Les rythmes endogènes font une compensation de la température. C'est-à-dire que la période de rythme endogène restera la même à différentes températures.

Les rythmes endogènes sont divisés en plusieurs types en fonction de leur période (Tableau\_02). Pour cette thèse nous allons uniquement nous intéresser au rythme de type circadien. Le rythme circadien est un rythme endogène avec une FRP d'environ 24 heures et existe dans tous les règnes du vivant (Figure\_03).

**Tableau\_02 : Dénomination des rythmes endogènes en fonction de leur période**

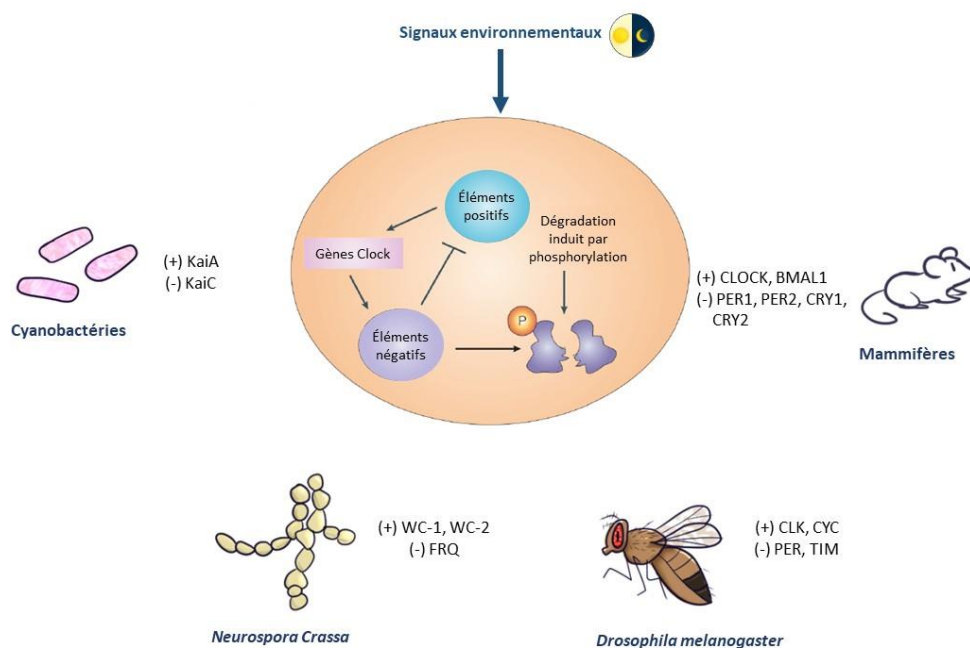
Période	Environ 1an	De 28h à 1 an	Environ 24h	Inférieur à 20h	Environ 12h
Rythme endogène	Circannuel	Infradien	Circadien	Ultradien	Circatidal



**Figure\_03 : Exemples de rythmes circadiens connus dans différents règnes du vivant.** (Dortous de Mairan, 1729; Konopka & Benzer, 1971 ; Morgan & Last, 1982 ; Mellow, Brunner & Roenneberg, 1999 ; King & Takahashi, 2000 ; Saigusa, Ishizaki et al., 2002; Xu, Mori & Johnson, 2003 ; Thommen, Pfeuty et al., 2012 ; Rijo-Ferreira, Acosta-Rodriguez et al., 2020).

## Introduction générale

Si c'est vers le second quart du XX<sup>ème</sup> siècle que la base génétique du rythme circadien a été mise en évidence, il faut attendre la fin du XX<sup>ème</sup> avec les progrès de la biologie moléculaire pour que les gènes responsables commencent à être identifiés. La toute première découverte revient à Ronald Konopa et Seymour Benzer en 1971 avec le gène *PERIOD* chez *Drosophila melanogaster* (Konopka & Benzer, 1971). D'autres gènes sont découverts à travers l'arbre du vivant, comme le gène *FREQUENCY* chez *Neurospora crassa* (Feldman & Hoyle, 1973), le gène *CLOCK* chez la souris (King, Zhao et al., 1997) ou encore les gènes *KAI* chez les cyanobactéries (Nishimura, Nakahira et al., 2002). Tous ces gènes, dits de l'horloge ou gènes « clock », ont permis de faire ressortir un schéma commun de boucle de rétrocontrôle durant approximativement 24h (Figure\_04). De façon générale, ces boucles se composent d'éléments positifs qui activent les gènes de l'horloge qui activent eux-mêmes les éléments négatifs qui inhibent l'activité des éléments positifs. La phosphorylation des éléments négatifs entraîne leur dégradation et ainsi le retour des éléments positifs pour démarrer un nouveau cycle (Bell-Pedersen, Cassone et al., 2005).



**Figure 04 : Mécanisme commun des horloges internes (oscillateurs) avec des exemples de différentes espèces. (Figure modifiée à partir de Bell-Pedersen, Cassone et al., (2005)).** Les principaux éléments des horloges sont indiqués pour chaque espèce. Les éléments positifs, indiqués par des (+), sont KaiA, WHITE COLLAR-1 (WC-1), WHITE COLLAR-2 (WC-2), CLOCK (CLK pour *D. melanogaster*), CYCLE (CYC) et brain and muscle Anrt-like protein 1 (BMAL1). Les éléments négatifs, indiqués par des (-), sont KaiC, FREQUENCY (FRQ), PERIOD (PER), TIMELESS (TIM) et CRYPTOCHROME (CRY).

## 2. Bilharziose

L'hématurie d'Égypte représentée sur les bas-reliefs de l'époque par le dieu égyptien Seth (Ziskind, 2009) est aujourd'hui connue sous le nom de bilharziose ou schistosomiase. Cette parasitose toujours d'actualité est la plus importante maladie causée par un helminthe avec plus de 230 millions de personnes infestées dans le monde (WHO, 2022). Le schistosome, parasite sanguin responsable, a été découvert en 1851 par Théodore Bilharz qui a ainsi donné son nom à la pathologie (Tan & Ahana, 2007). La plus ancienne trace de cette infestation parasitaire a été découverte en 1910 sur une momie

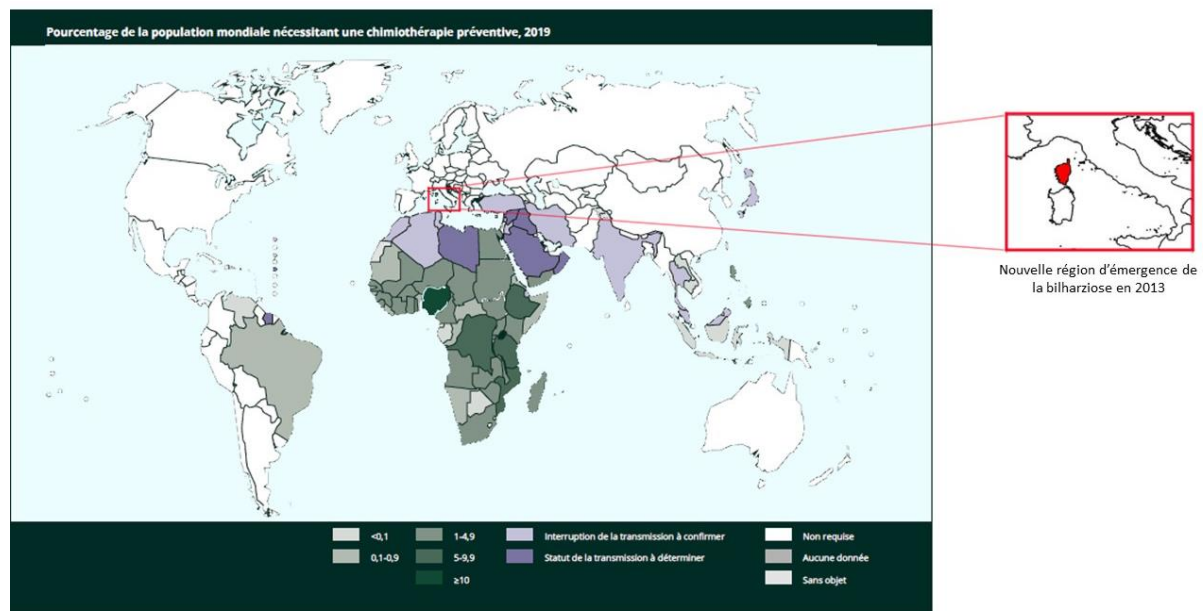


## Introduction générale

égyptienne datant de plus de 10 siècles avant notre ère (XX<sup>ème</sup> dynastie). L'auteur de cette découverte, Sir Marc Armand Ruffer, posera ainsi les bases d'une nouvelle discipline scientifique, la paléoparasitologie (Ziskind, 2009).

Il existe deux formes de bilharziose chez l'Homme causées par six espèces de schistosomes. L'espèce *Schistosoma haematobium* cause la forme urinaire. Cette forme se caractérise par la présence de sang dans les urines (hématurie) ainsi qu'une fibrose des voies urinaires. Les cinq autres espèces, *S. guineensis*, *S. intercalatum*, *S. japonicum*, *S. mansoni* et *S. mekongi* causent la forme intestinale. Cette forme se caractérise par du sang dans les fèces (rectorragie), des douleurs abdominales, de la fièvre et une hépatomégalie. Dans les deux formes, les symptômes de la maladie sont causés par l'accumulation des œufs du parasite à l'intérieur du système urogénital (forme urinaire) ou du foie et des intestins (forme intestinale). Les pathologies les plus avancées peuvent évoluer en cancer. Côté traitement, si à l'époque des pharaons le remède préconisé était une incantation rituelle et une perle de coraline en guise de suppositoire, aujourd'hui nous pouvons compter sur l'efficacité du Praziquantel. Cette molécule, dont le mécanisme d'action est inconnu, agit sur la perméabilité des membranes des schistosomes en provoquant une entrée de Ca<sup>2+</sup> et une paralysie du parasite (Cupit & Cunningham, 2015). Elle reste à l'heure actuelle le traitement curatif le plus efficace pour se débarrasser des schistosomes.

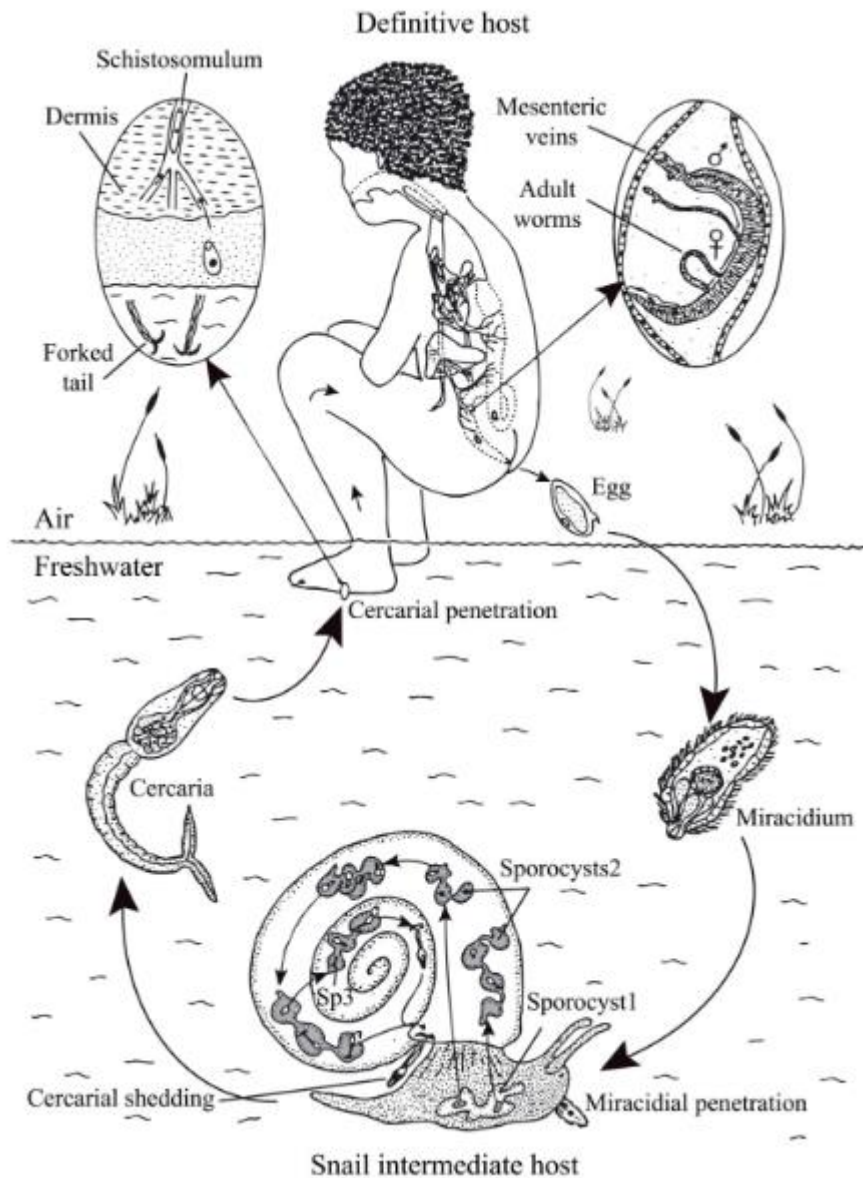
La bilharziose étant une parasitose plus handicapante que mortelle, elle entraîne ainsi des conséquences négatives sanitaires mais aussi économiques et sociales. Ces conséquences sont d'autant plus importantes que la bilharziose sévit principalement dans les zones tropicales et subtropicales incluant des pays émergents et des pays en développement (Figure\_05). Considérée comme une maladie tropicale négligée, la bilharziose touche pourtant aussi l'Europe depuis 2013 avec l'émergence de cas en Corse dans la rivière du Cavu (Holtfreter, Moné et al., 2014 ; Berry, Moné et al., 2014).



**Figure 05 : Carte de la répartition de la schistosomiase dans le monde (Figure modifiée à partir de Boissier, Mouahid & Moné (2019) et de WHO (2021)).** Indiqués par dégradé de vert, les pays dont la population nécessite une prise en charge thérapeutique contre la bilharziose. Les pays en violet foncé indiquent les pays où le statut de la transmission est à déterminer. Les pays en violet pâle sont les pays où l'interruption de la transmission est à confirmer. Le rouge indique la Corse où a eu lieu une émergence de la bilharziose en 2013.

## 2.1. Cycle de vie des schistosomes : l'exemple de *Schistosoma mansoni*

Les parasites du genre *Schistosoma* sont des vers trématodes gonochoriques. Leur cycle de vie nécessite deux hôtes : un hôte mammifère dans lequel le parasite fait sa reproduction sexuée et un hôte mollusque d'eau douce dans lequel le parasite se multiplie de façon clonale (multiplication asexuée) (Figure\_06).



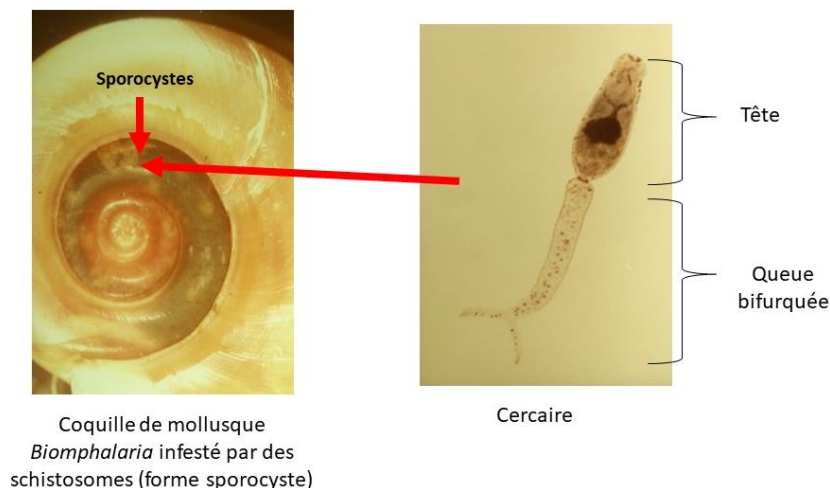
Figure\_06 : Exemple du cycle de vie de *Schistosoma mansoni*. Figure issue de Boissier, Mouahid & Moné (2019) (cf description ci-après).

## Introduction générale

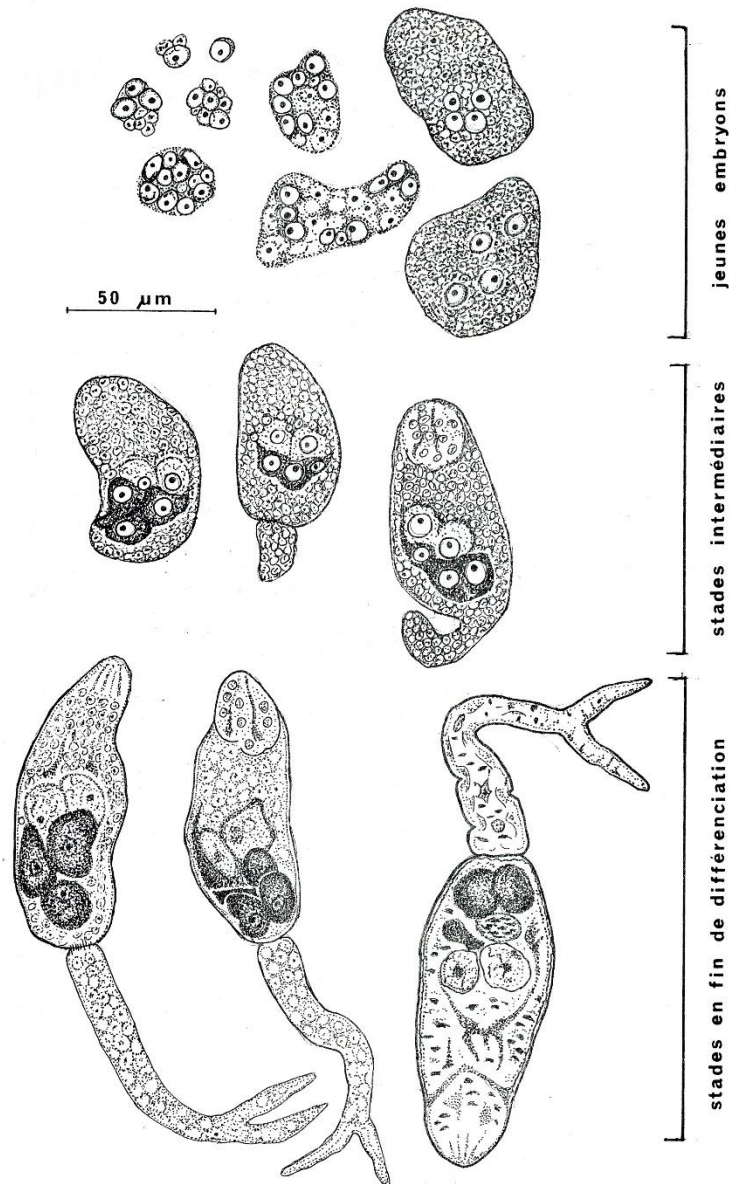
Au stade vers adulte, les parasites sont dans l'hôte mammifère. Le mâle, accroché à la paroi de la veine porte de l'hôte, porte dans son canal gynécophore la femelle avec laquelle il se reproduit. Le couple va migrer jusqu'aux veinules mésentériques pour que la femelle puisse pondre. Vingt à trente pourcents des œufs vont passer la barrière intestinale et se retrouver dans la lumière où ils seront excrétés avec les fèces de l'hôte. Une fois arrivé dans un milieu aquatique favorable, un œuf va éclore en un miracidium. Cette larve nageante va partir à la recherche de l'hôte mollusque. Une fois le mollusque trouvé, le miracidium va pénétrer activement à l'intérieur et se transformer en sporocyste primaire. Le sporocyste va effectuer une reproduction clonale pour donner des sporocystes secondaires qui vont migrer vers le centre du mollusque au niveau de l'hépatopancréas. Les sporocystes secondaires génèrent des cercaires qui sortent du mollusque pour se retrouver dans le milieu aquatique. La cercaire est le stade larvaire nageant qui infeste l'hôte mammifère. La cercaire pénètre activement la peau de l'hôte en contact direct avec l'eau infestée. Lors de ce processus la cercaire laisse sa queue à l'extérieur de l'hôte et se transforme en schistosomule. Le schistosomule va passer dans le système circulatoire sanguin où il va évoluer morphologiquement pour se transformer en ver adulte mâle ou femelle et ainsi recommencer le cycle.

### 2.2. Rythme d'émission des cercaires

Cette thèse s'intéresse à l'une des étapes critiques du cycle de vie des schistosomes, la transmission à l'hôte mammifère, au travers du rythme d'émission des cercaires. La cercaire est une larve nageante composée d'une tête et d'une queue bifurquée (furcocercaire) (Figure\_07). Elle est issue d'un sporocyste secondaire qui est composé d'une suite de chambres blanchâtres contenant des cercaires à différents stades de différenciation : cellules germinales, cellules embryonnaires, cercaires matures (Figures\_07 et \_08).



**Figure 07 : Photo d'une coquille de mollusque *Biomphalaria* infesté par des sporocystes de schistosomes (à gauche) et photo d'une cercaire (à droite).** Les sporocystes sont les agglomérats blancs au centre de la coquille ; ils contiennent les cercaires.



**Figure 08 : Dessins des stades de différenciation des cercaires issus des observations de coupes histologiques de mollusque *Biomphalaria* infestés par *Schistosoma mansoni*. (Figure issue de Théron (1982)).** L'évolution de la cercaire va du stade embryonnaire (en haut à gauche) jusqu'au stade de la cercaire mature (en bas à droite). Pour les trois types de stades, la forme la moins différenciée est à gauche et la plus différenciée à droite.

L'émission des cercaires est donc la sortie des cercaires matures des sporocystes et de l'hôte mollusque pour se retrouver dans l'environnement aquatique à la recherche de l'hôte mammifère. Afin de maximiser la rencontre parasite-hôte, les schistosomes, et plus largement les trématodes, ont adopté un rythme d'émission des cercaires journalier, calqué sur la présence de l'hôte dans le milieu. En d'autres termes, le pic d'émission des cercaires est synchronisé avec le pic d'activité aquatique de l'hôte mammifère correspondant. Chaque espèce de schistosome possède son propre rythme d'émission des cercaires (Figure\_09).

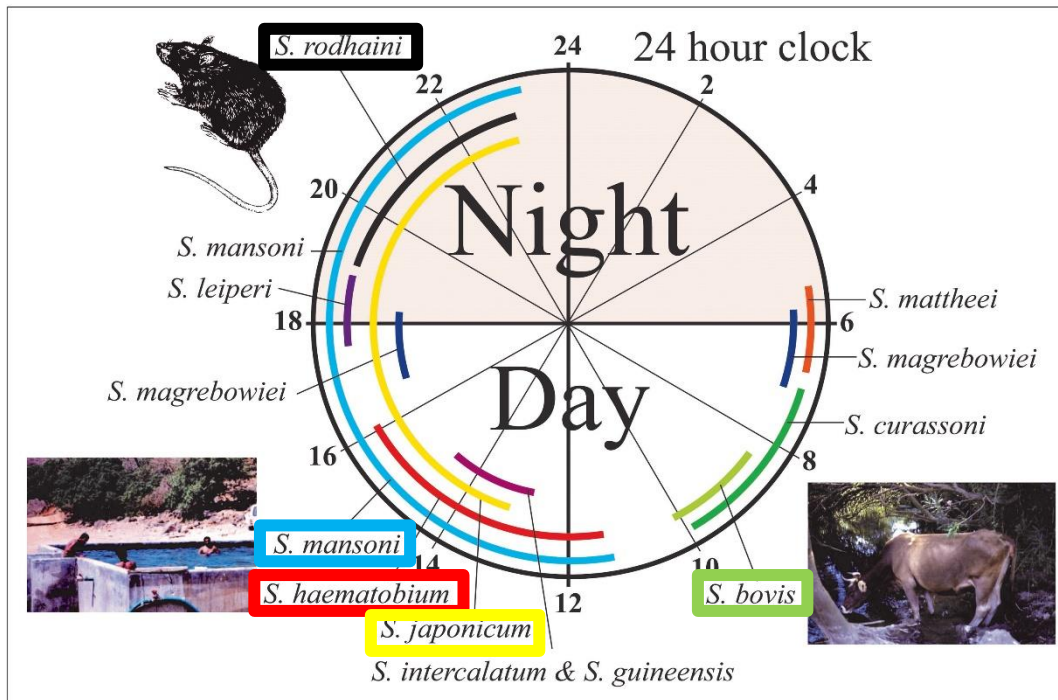


Figure 09 : Schéma de la répartition de l'émission cercarienne sur 24 heures par espèce de schistosome. Figure issue de Boissier, Mouahid, & Moné (2019). Chaque couleur correspond à une espèce de schistosomes. (Traduction de l'anglais : 24 hour clock = les 24 heures de la journée, Night = nuit et Day = jour).

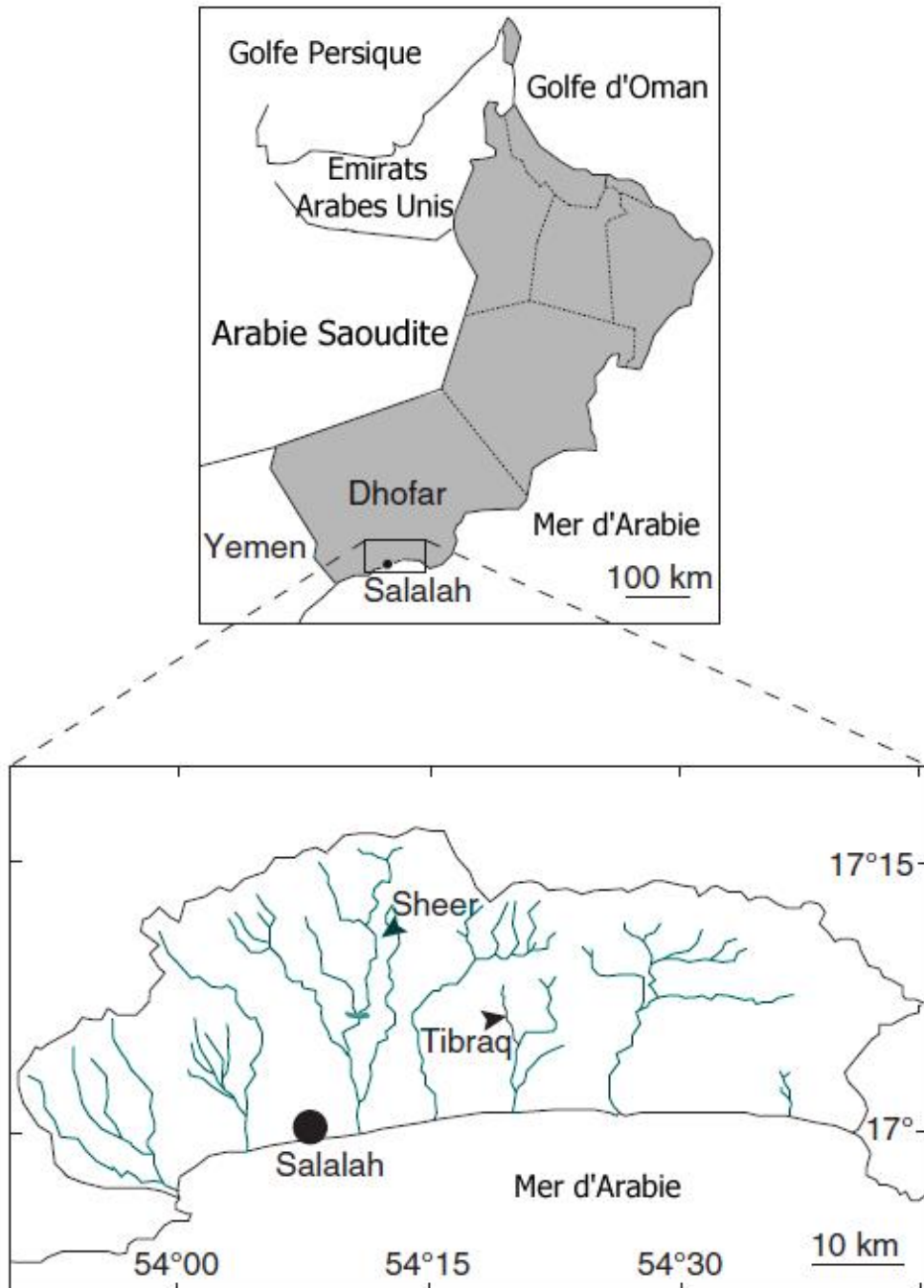
Un schistosome qui infeste les bovins aura son pic d'émission tôt le matin, période à laquelle les bovins vont s'abreuver (cas de *S. bovis*). Dans le cas d'un schistosome qui infeste les humains, son pic aura lieu en fin de matinée début après-midi, période à laquelle les humains vont se baigner (cas de *S. haematobium*). Enfin un schistosome qui infeste les rongeurs aura son pic la nuit (cas de *S. rodhaini*), période à laquelle les rongeurs sont actifs. Une exception est *Schistosoma japonicum* qui a un spectre d'hôtes très large, et pourra avoir une émission cercarienne diurne ou nocturne en fonction de l'hôte visé.

### 2.3. *Schistosoma mansoni* : De Porto Rico à Oman

Sur la figure\_09, le schéma d'émission cercarienne journalier de *S. mansoni* s'étend de la fin de matinée au milieu de la nuit. Le premier *Schistosoma mansoni* a été découvert chez l'Homme en 1904 par Isaac Gonzalez-Martinez à Maraguez (Puerto Rico) (Ritchie, Jachowski & Ferguson, 1960). L'émission cercarienne de cette souche était strictement diurne et le pic avait lieu entre 10h et 13h (Asch, 1972; Nojima & Sato, 1982). Cette émission diurne a été considérée comme le chronotype classique de l'espèce *S. mansoni*. De plus, c'est la souche NMRI Porto Rico qui sert de génome de référence pour *S. mansoni* (Protasio, Tsai et al., 2012). En 1984, André Théron découvre un nouveau chronotype dit « tardif » (Théron, 1984) : isolé en Guadeloupe à partir de *Rattus rattus* dans la région des Trois-Rivières, ce chronotype se différencie par un pic d'émission à 16h et un nouvel hôte mammifère. Cette découverte illustre la capacité d'adaptation de *S. mansoni* au comportement

## Introduction générale

aquatique de son hôte. En 2012, Gabriel Mouahid découvre à Oman, dans la région du Dhofar, deux chronotypes diamétralement opposés de *S. mansoni* (Mouahid, Idris et al., 2012): le premier chronotype, isolé chez l'Homme à Sheer (Figure\_10), correspond au chronotype classique diurne de *S. mansoni*, avec un pic d'émission à 11h ; le second chronotype, isolé chez le rat à Tibraq (Figure\_10), a une émission cercarienne nocturne avec un pic d'émission à 19h.



**Figure 10 :** Carte du Sultanat d'Oman avec un zoom sur les sites de Sheer et Tibraq de la région du Dhofar d'où sont issus les deux chronotypes de *S. mansoni*. (Figure modifiée à partir de Mouahid, Idris et al., (2012)). Le sultanat d'Oman correspond à la partie grisée (carte du haut). Les rivières sont symbolisées par des traits bleus (carte du bas).

La découverte ces deux chronotypes opposés appartenant à la même espèce (*S. mansoni*) dans une zone géographique proche (Dhofar, Oman) est un modèle biologique idéal pour comprendre l'écologie et les mécanismes impliqués dans le rythme d'émission des cercaires. En effet, le rythme d'émission des cercaires de *S. mansoni* est déterminé génétiquement (Théron & Combes, 1988) et leur gonochorisme rend possible les croisements intra- et inter-espèces permettant des analyses moléculaires.

### 3. Problématiques et enjeux

Le traitement curatif de la bilharziose est un traitement efficace pour tuer les parasites. Toutefois il s'avère inutile sur le long-terme si les populations humaines continuent d'être perpétuellement en contact avec les eaux contaminées. A ceci s'ajoute la capacité de certaines espèces de schistosomes à pouvoir s'hybrider entre elles et ainsi à faire tomber la barrière de l'espèce hôte. C'est l'exemple du parasite responsable de l'émergence des cas en Corse qui est un hybride entre *S. haematobium* (parasite de l'Homme) et *S. bovis* (parasite des bovins) (Moné, Holtfreter et al., 2015). Les autres mammifères sauvages (rongeurs) ou domestiqués (bétails) présentent ainsi le risque d'être des réservoirs de la maladie et des lieux d'hybridation.

Pour traiter efficacement cette maladie tropicale négligée, la stratégie One Health est préconisée par l'OMS. Elle vise à mettre un terme à la vision anthropocentrée et à appréhender la maladie comme l'interaction d'un environnement avec l'Homme et les animaux (sauvages et domestiques). La meilleure stratégie de lutte contre la bilharziose est donc l'association du traitement curatif avec une stratégie de prévention pour rompre le cycle de vie du parasite.

La compréhension de la transmission du schistosome à l'hôte mammifère au travers du rythme d'émission des cercaires sera l'objet de cette thèse. Elle vise entre autres à mettre en lumière la chronobiologie comme paramètre écologique à prendre en compte dans les stratégies de prévention notamment comme outil pour identifier les hybrides (Savassi, Mouahid, Lasica et al., 2020 (Annexe 1) ; Savassi, Dobigny et al., 2021). A l'image de la stratégie One Health et de la chronobiologie, la transversalité des approches sera le fil conducteur de ce manuscrit. Comme le titre l'indique, l'émission rythmique des cercaires y sera étudiée de son phénotype jusqu'aux gènes.

La première partie de cette thèse aura pour but de décrire le rythme des profils d'émission journaliers des cercaires chez les deux chronotypes de *Schistosoma mansoni* (Oman) et de déterminer quelle est la nature de ces rythmes. La deuxième partie aura pour but d'identifier les déterminants moléculaires impliqués, au travers d'approches génétique, épigénétique et transcriptomique.

---

## Matériel et méthodes

Le Matériel et méthodes, présenté ci-dessous, regroupe les techniques communes aux divers points abordés dans ce manuscrit. Pour les besoins de cette thèse, certaines techniques ont dû être optimisées et/ou améliorées. Le détail de ces techniques et améliorations fera ainsi l'objet d'un matériel et méthodes spécifique dans les parties dédiées plus tard dans ce manuscrit.

### 1. Matériel biologique

#### 1.1. Les schistosomes

Deux espèces de schistosomes ont été utilisées pour les travaux de cette thèse : l'espèce *Schistosoma mansoni*, avec les deux chronotypes, diurne et nocturne, originaires de la région du Dhofar à Oman (Mouahid, Idris et al., 2012) et l'espèce *Schistosoma rodhaini* originaire du Burundi. Les cycles de vie de ces parasites sont maintenus en animalerie au laboratoire, en utilisant des mollusques du genre *Biomphalaria* (*B. pfeifferi* et *B. glabrata*) comme hôtes intermédiaires et des souris Swiss OF1 comme hôtes définitifs.

#### 1.2. Les mollusques (hôtes intermédiaires)

Deux espèces de mollusques ont été utilisées pour les expérimentations de cette thèse, *Biomphalaria pfeifferi* et *Biomphalaria glabrata*. L'espèce *B. pfeifferi* est originaire de la région du Dhofar à Oman et l'espèce *B. glabrata* est originaire de Guadeloupe. Les mollusques sont élevés dans des pièces climatisées sous une température maintenue constante à 26°C afin de se rapprocher au maximum des conditions du terrain (climat tropical) et sous photopériode équilibrée (Lumière Obscurité (LO) 12-12). Ils sont nourris avec de la laitue fraîche pour les mollusques adultes, et de la laitue déshydratée pour les jeunes mollusques ; quelques copeaux de spiruline sont rajoutés une fois par semaine. L'eau utilisée provient d'un forage situé sur le site de l'UPVD. Toutes les expériences ont été réalisées avec des mollusques *B. pfeifferi*. Seule l'expérience de chronobiologie pour tester les conditions lumière continue et lumière inversée pour le chronotype diurne ont été réalisées sur des *B. pfeifferi* ET des *B. glabrata*.

#### 1.3. Les rongeurs (hôtes définitifs)

Pour maintenir le cycle de vie des parasites, des souris femelles Swiss OF1 ont été utilisées. Les souris sont dans des cages de maximum 10 individus et sont nourries avec des granulés. L'élevage, le soin et ainsi que les protocoles sur les animaux ont été approuvés par l'agence vétérinaire française de la DRAAF (Direction Régionale de l'Alimentation, de l'Agriculture et de la Forêt) Languedoc-Roussillon (Numéros d'agrément 007083 puis E6613601). Les autorisations éthiques sur l'expérimentation animale ont été délivrées par le Ministère de l'Éducation Nationale, de la Recherche et de la Technologie, du Ministère de l'Agriculture et de la Pêche (convention n° A 66040), et de la Direction Départementale de la Protection des Populations (n° C 66-136-01 avec arrêté préfectoral n° 2012- 201-0008). Ma co-directrice de thèse, Hélène Moné, possède le Certificat d'autorisation d'expérimenter sur animaux vertébrés vivants numéro C 66.11.01.



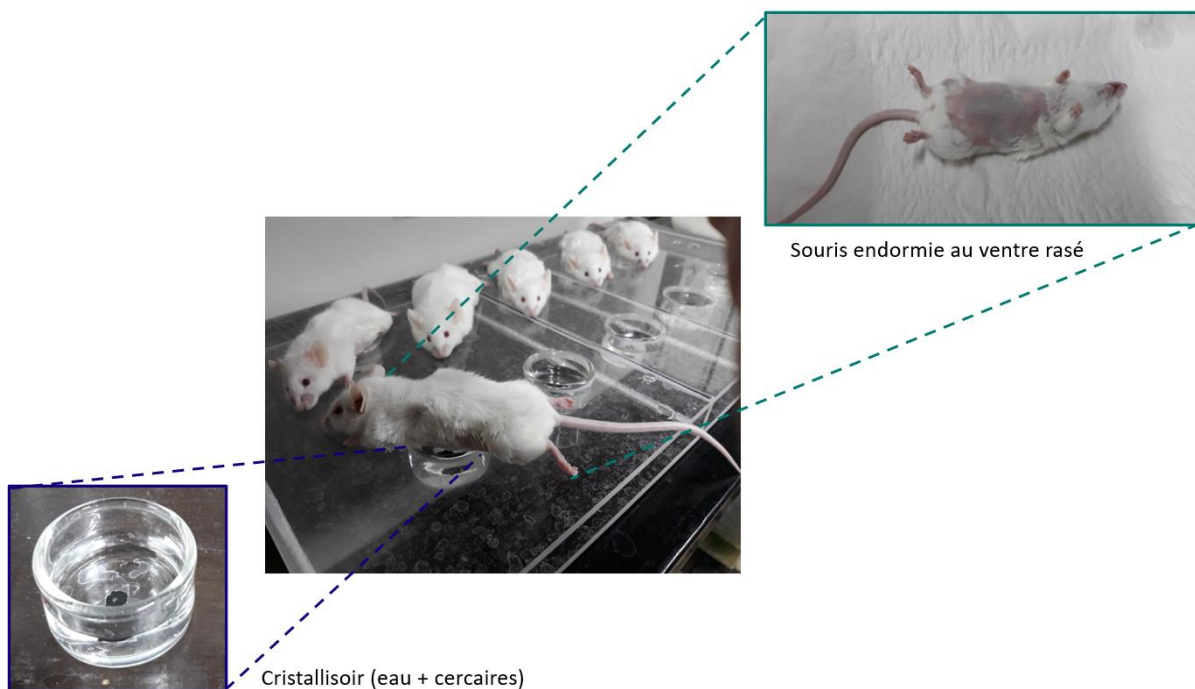
## 2. Méthodes et techniques d'entretien des cycles

### 2.1. Infestation des mollusques

Les *B. pfeifferi* sont répartis individuellement dans des plaques de 24 puits contenant 1 miracidium et de l'eau de forage (la présence du, et d'un seul, miracidium est vérifiée). Les *B. glabrata* sont répartis individuellement dans de petits piluliers contenant 10 miracidia et de l'eau de forage ; les mollusques sont mis en contact avec le ou les miracidia durant 24 heures. Les mollusques sont ensuite placés dans des bacs et nourris *ad libitum*. Trente jours post-infestation, les mollusques sont transférés dans de petits piluliers individuels remplis d'eau de forage. Les mollusques positifs à l'infestation, c'est-à-dire les mollusques à partir desquels des cercaires ont pu être observées à la loupe binoculaire, sont isolés.

### 2.2. Infestation des souris

L'infestation consiste à mettre des souris en contact avec des cercaires pour obtenir une nouvelle génération de parasite. Les cercaires sont comptées (80 cercaires par souris pour le cycle) et déposées dans un cristallisoir contenant de l'eau de forage à 26°C. Les cristallisoirs sont ensuite insérés dans un grand portoir en plastique (Figure\_11). Chaque souris est pesée pour déterminer le volume d'anesthésique (Tableau\_03) à lui injecter (0,11mL pour 10 g). Une fois les souris anesthésiées, leur ventre est rasé et humidifié ; les souris sont placées, ventre en bas, au-dessus d'un cristallisoir pour que leur peau nue soit directement en contact avec l'eau et ainsi faciliter la pénétration des cercaires. Les souris sont ainsi laissées 1h en contact avec l'eau infestée avant d'être réintégrées à leur cage.



**Figure 11 : Infestation des souris (photos).** Les souris anesthésiées dont le ventre est rasé sont déposées sur un portoir en plastique percé contenant les cristallisoirs d'infestation. Les cristallisoirs contiennent de l'eau de forage et des cercaires.

**Tableau 03 : Composition de la solution d'anesthésique**

Solution anesthésiante	Pour 50mL
NaCl physiologique	42,5mL
Rompun®	2,5mL
Imalgène®	5mL

### 2.3. Récupération des adultes, des œufs et des miracidia

A partir des semaines 7 et 8 post-infestation, on peut procéder à la récupération des adultes et des œufs. Pour cela, la souris subit une injection de 0,05mL d'héparine (anticoagulant) puis est euthanasiée avec une injection intrapéritonéale d'une dose létale de 1ml/kg de doléthal® (pentobarbital sodique dilué). Lorsque la souris arrête de respirer (poumons totalement immobiles), son ventre est rasé. Elle est alors suspendue dos à une potence par les pattes avant et arrière gauches. Les organes digestifs sont mis à nu et la cage thoracique est ouverte afin de laisser l'accès au cœur et aux poumons. La veine cave est coupée et la solution de perfusion (Tableau\_04) est injectée aussitôt dans le ventricule gauche grâce à une pompe péristaltique.

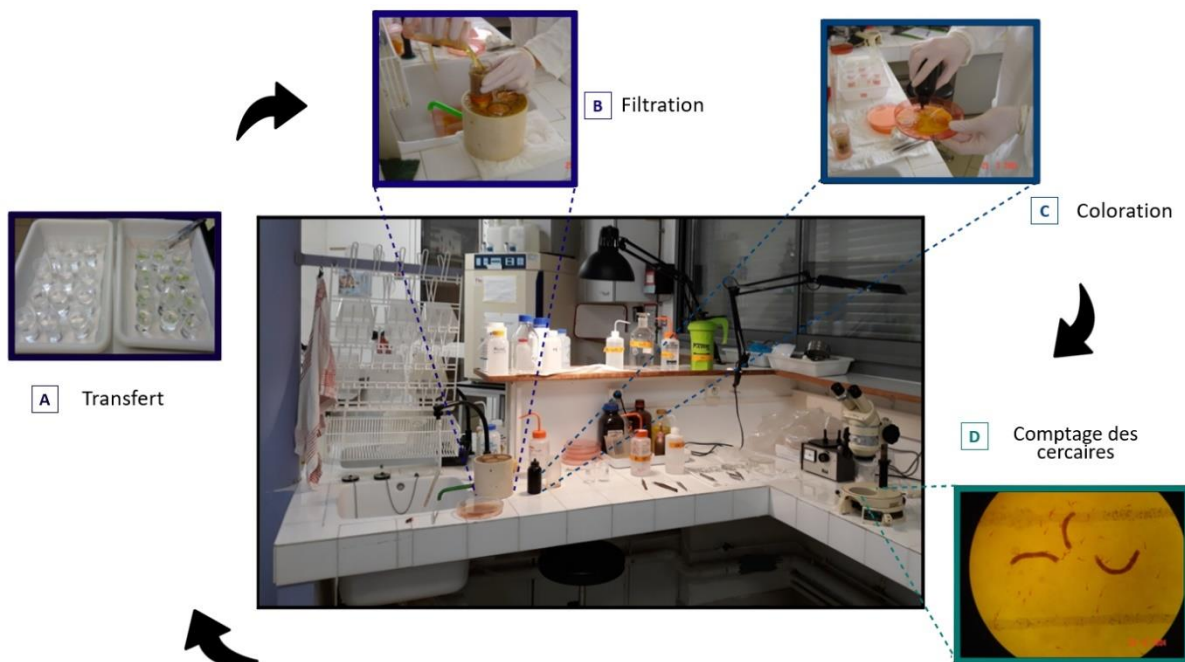
**Tableau 04 : Composition de la solution de perfusion**

Solution perfusion	Pour 1L d'eau milliQ
Chlorure de sodium	8,5g
Citrate de sodium	7,5g

Le sang qui contient les parasites est recueilli à l'aide d'un verre à pied placé sous la souris. La perfusion est arrêtée dès que les poumons sont totalement décolorés, signe qu'il n'y a plus de sang dans l'animal. Les organes sont rincés puis les intestins et le foie qui contiennent les œufs sont alors réservés dans un cristalliseur avec la solution de perfusion. Le sang recueilli est filtré à travers un filtre de polyamide nitrile de 25 µm qui retient les parasites. Les vers sont alors transférés dans un verre de montre contenant une solution de NaCl physiologique (0,9%). Pour récupérer d'éventuels autres adultes pris dans les organes, le foie est décortiqué à l'aide de pinces à dissection et les vers sont transférés dans le verre de montre. Les adultes sont amalgamés en boule puis introduits dans des cryotubes pour ensuite être conservés à -80°C. Pour la récupération des œufs, le broyat de foie est passé à travers une série de tamis (425, 180, 106 et 45µm) à l'aide d'un pulvérisateur à pompe contenant une solution physiologique de NaCl à 0,9%. Les œufs sont récupérés sur le tamis 45µm et réservés dans la solution saline pour empêcher l'éclosion. Il est aussi possible de récupérer des œufs dans les intestins. Pour cela ces derniers sont vidés, découpés et transférés dans un mortier, puis broyés. Comme pour le foie, le broyat d'intestin est passé à travers la série de filtres et les œufs sont récupérés sur le tamis 45µm. Pour déclencher l'éclosion des œufs, les œufs sont transférés dans de l'eau de forage à 26°C sous une source de lumière (lampe). Les œufs sont laissés entre 30 à 60 minutes afin de laisser éclore les miracidia qui pourront être utilisés pour l'infestation des mollusques.

### 3. Méthode de chronobiologie de l'émission cercarienne

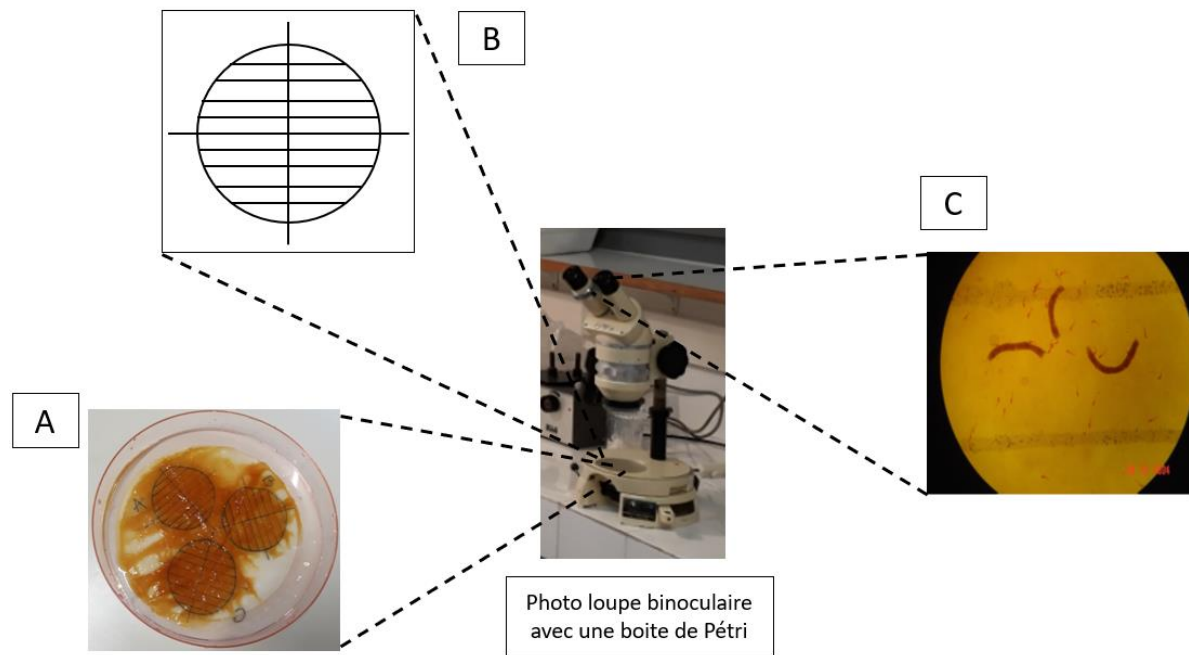
La chronobiologie se divise en 4 étapes : Le transfert, la filtration, la coloration et le comptage des cercaires (Figure\_12). Les mollusques positifs aux cercaires sont maintenus individuellement dans des piluliers remplis de 40mL d'eau de forage. Chaque heure, chaque mollusque et sa salade sont transférés dans un nouveau pilulier à l'aide d'une pince souple, c'est l'étape de transfert. Les piluliers de l'heure précédente sont alors récupérés. L'étape de filtration consiste à passer au travers d'un filtre de polyamide de nitrile de 25µm le contenu des piluliers. Les filtres sont ensuite déposés dans des boîtes de pétri (150mm de diamètre) et colorés avec quelques gouttes de Lugol, c'est l'étape de coloration (Tableau\_05). Chaque filtre contient trois cercles quadrillés (Figure\_13) et chaque cercle correspond au contenu d'un pilulier. Le comptage des cercaires s'effectue pour chaque mollusque et à chaque heure en comptant toutes les cercaires qui sont contenues dans un cercle.



**Figure 12 : Etapes de la chronobiologie (photos).** A) Transfert : Les mollusques ont été transférés du bac de gauche à celui de droite. Le bac de gauche est récupéré pour passer à l'étape suivante. B) Filtration : L'eau contenue dans chaque pilulier est passée au travers d'un cercle du filtre. C) Coloration : Chaque cercle du filtre est coloré avec quelques gouttes de Lugol. D) Comptage des cercaires : Chaque filtre est observé sous la loupe binoculaire. Ces quatre étapes sont répétées toutes les heures.

**Tableau 05 : Composition de la solution de Lugol**

Solution de Lugol	Pour 1 L d'eau milliQ
Iodure de potassium	20g
Iode	10g



**Figure 13 : Détail des filtres utilisés pour le comptage.** A) Filtre entier. B) Cercle de comptage pour un profil. C) Détail de ce qui est observé à la loupe binoculaire : cercaires en rose et 3 fèces de mollusque en marron.

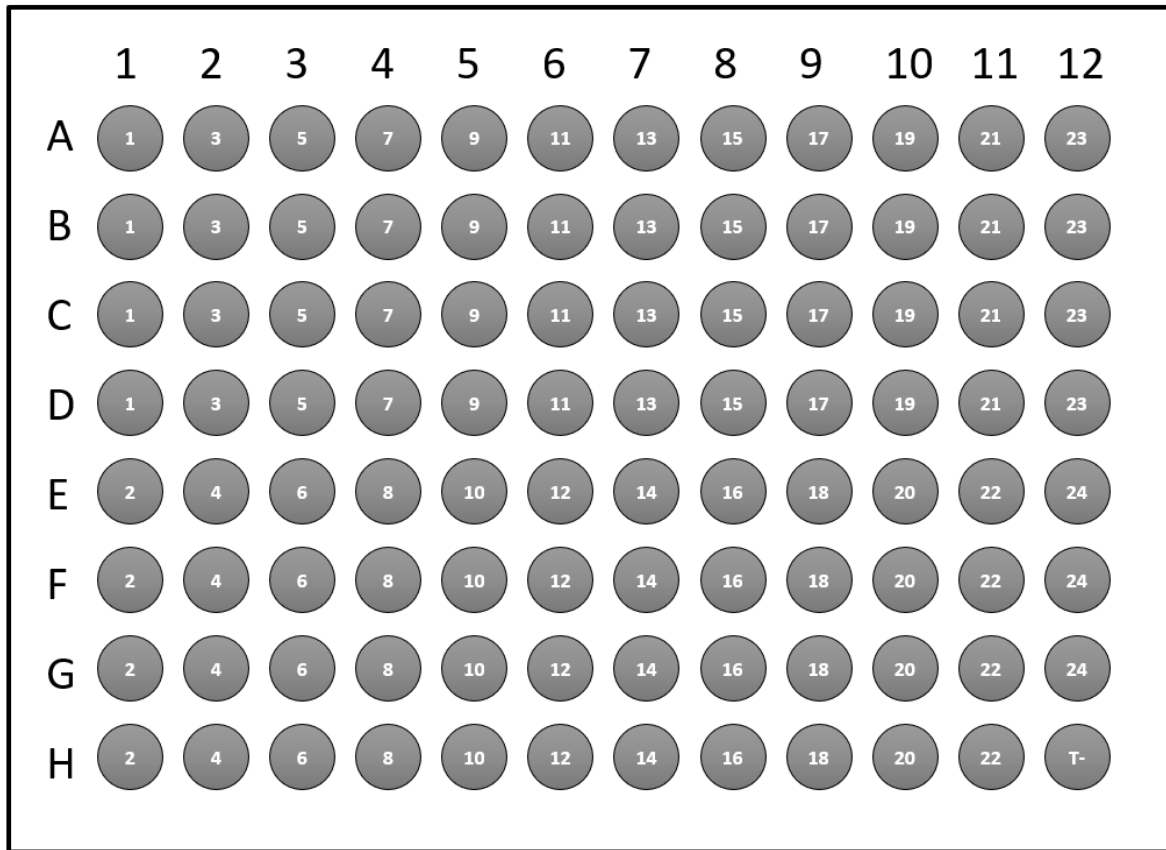
Pour chaque expérience de chronobiologie, il y a une réserve de mollusques remplaçants pour éventuellement changer un mollusque qui décèderait au cours de l'expérience. Les remplaçants subissent l'expérience de chronobiologie en même temps que les autres individus, sans que leurs cercaires ne soient comptées, jusqu'à ce qu'ils intègrent l'expérience.

#### 4. Méthode du sexage des parasites *Schistosoma mansoni* et *Schistosoma rodhaini* au stade cercaire

La détection du sexe des cercaires nécessite plusieurs étapes qui sont : l'échantillonnage des cercaires, l'extraction de l'ADN, la préparation et réalisation de la PCR multiplex puis le diagnostic par gel d'électrophorèse.

##### 4.1. Echantillonnage des cercaires

Les mollusques utilisés sont infestés avec un seul miracidium. En d'autres termes, un mollusque correspond à un seul génome de parasite. Les cercaires sont récupérées directement dans l'eau des mollusques à l'aide d'une pipette pasteur munie d'une poire. Une à 3 cercaires sont prélevées pour chaque test et déposée(s) dans le puit d'une plaque 96 puits. Pour chaque mollusque on réalise dans la mesure du possible quatre prélèvements (quatre tests) afin de n'avoir aucun doute sur le résultat de la réaction PCR (Figure\_14).



**Figure 14 : Exemple de répartition d’une plaque 96 puits pour le sexage.** Un puit, symbolisé par un cercle gris, contient de 1 à 3 cercaires. Un parasite correspond à 4 puits, à l’exception du 24<sup>ème</sup> qui en a seulement trois car le dernier puit correspond au contrôle négatif de la plaque (puits sans cercaire). Avec une plaque 96 puits, on réalise les tests pour 23 parasites avec 4 réactions PCR, 1 parasite avec 3 réactions PCR et un contrôle négatif (puits sans cercaire).

## 4.2. Extraction ADN

Une fois les cercaires déposées dans la plaque 96 puits, 20 µL de NaOH 250mM stérile sont ajoutés par puits pour perméabiliser les membranes. Après 20 minutes à température ambiante, la plaque est mise 3 minutes à 99°C dans un thermocycleur. 20µL de la solution de neutralisation (Tableau\_06) sont ensuite ajoutés et la plaque est une nouvelle fois incubée dans un thermocycleur durant 3 minutes à 99°C. Dans une nouvelle plaque, on dépose 54 µL d’eau milliQ par puit et on ajoute 6 µL de l’ADN extrait (dilution au 10<sup>ème</sup>). La plaque d’ADN « mère » est conservée à -20°C.

**Tableau\_06 : Composition de la solution de neutralisation**

Solution de neutralisation	Pour 1 puit d’une plaque 96 puits
HCl (250 mM)	10µL
Tris/HCl (500 mM)	5µL
Triton 2%	5µL

### 4.3. PCR multiplex

Chaque test de sexage repose sur le principe de la PCR multiplex, c'est-à-dire une PCR qui implique plusieurs couples d'amorces avec différentes cibles en une seule réaction. Pour le sexage, deux couples d'amorces sont utilisés. Un premier couple d'amorces cible la rhodopsine\_3, gène autosomal (témoin positif de la réaction PCR) dont le produit mesure 200 pb. Le second couple d'amorces cible le scaffoLO\_02739 qui est spécifique au chromosome W (femelle) dont le produit mesure 400pb (Tableau\_07). L'ADN utilisé pour les réactions PCR est dilué au 10<sup>ème</sup>. Pour préparer la PCR multiplex, 6 µL de mix PCR (Tableau\_08 et Tableau\_09) sont déposés dans une troisième plaque 96 puits avec 4 µL de l'ADN dilué au 10<sup>ème</sup>. La plaque est ensuite mise dans le thermocycleur (Tableau\_10 et Tableau\_11). Suite à des problèmes de puits vides positifs à la Rhodopsine\_3, le couple d'amorces Rhodopsine\_3 a été remplacé par un nouveau couple d'amorces appelé Rhodopsine\_3\_BIS (cf Annexe 2 : test de la PCR multiplex pour le sexage avec les couples d'amorces Scaff\_02739 et Rhodopsine\_3\_BIS).

**Tableau\_07 : Séquence des amorces pour le sexage**

Amorce	Séquence
Scaff_02739_F	TGTTTCGAATTTCACTTCA
Scaff_02739_R	CATTCACAGTTTGGCGAACA
Rhodopsin_3_F	GACGGCCCACTAAAG
Rhodopsin_3_R	AGTAAAATGGTCACTGCTAT
Rhodopsin_3_F BIS	GGAATTGGTACCGTCAGC
Rhodopsin_3_R BIS	AGGTTCACTTTAGTGTGGC

**Tableau\_08 : Composition du mix PCR multiplex pour le sexage**

Mix PCR multiplex pour le sexage	Pour 1 puit d'une plaque 96 puits
Qiagen multiplex PCR master mix (2X)	5µL
Mix couples d'amorces (cf tableau m&m_07)	1µL
ADN dilué au 10ème	4µL

**Tableau\_09 : Composition du mix des couples d'amorce pour le sexage (10X)**

Mix couples amorces pour le sexage (10X)	Pour 500µL
Scaff_02739_F	10µL
Scaff_02739_R	10µL
Rhodopsin_3_F ou Rhodopsin_3_F BIS	10µL
Rhodopsin_3_R ou Rhodopsin_3_R BIS	10µL
H <sub>2</sub> O milliQ	460µL

**Tableau\_10 : Programme du thermocycleur pour PCR multiplex avec les couples d'amorces ScaffoLO\_02739 et Rhodopsine\_3 (utilisés pour le chronotype diurne de *S. mansoni*)**

Température	Temps	Nombre de cycles
95°C	15min	
94°C	30s	45X
56°C	1min	
72°C	40s	
72°C	2min	
10°C	∞	

**Tableau 11 : Programme du thermocycleur pour PCR multiplex avec les couples d'amorces ScaffoLO\_02739 et Rhodopsine\_3\_BIS (utilisés pour le chronotype nocturne de *S. mansoni* et *S. rodhaini*)**

Température	Temps	Nombre de cycles
95°C	15min	
94°C	30s	40X
54°C	40s	
71°C	40s	
71°C	2min	
10°C	∞	

#### 4.4. Electrophorèse et diagnostic

Pour faire le gel, 2g d'agarose sont ajoutés à 100mL de solution de TBE (Tris/Borate/EDTA) et portés à ébullition jusqu'à totale dissolution de la poudre. Tandis que le mélange tiédit, 10µL de Midori green<sup>®</sup> (coloration d'acide nucléique) sont ajoutés. Le gel coloré est ensuite coulé dans un portoir puis des peignes sont ajoutés pour creuser les puits. Lorsque le gel est totalement solidifié, les peignes sont retirés puis le gel est déposé dans une cuve d'électrophorèse remplie de TBE. Une fois la PCR terminée, 2µL de tampon de charge sont ajoutés à chaque puit de la plaque 96 puits. Les puits du gel sont chargés avec 8µL du mélange échantillon-tampon de charge. La migration du gel se fait à 135V pendant 40 minutes. Enfin, la lecture du gel se fait grâce à une table à UV. Pour le diagnostic, les parasites mâles présentent une seule bande à 200 pb (Rhodopsine\_3/\_BIS) dans le gel d'électrophorèse. Les femelles présentent 2 bandes, une à 200 pb (Rhodopsine\_3/\_BIS) et une à 400 pb (scaff\_02739).

### 5. Méthode des croisements de schistosomes

Le croisement s'opère lors de la reproduction sexuelle du parasite qui a lieu dans l'hôte mammifère. Pour éviter un biais sexuel, les croisements et leurs réciproques ont été faits. La méthode expérimentale reste la même entre le croisement et son réciproque mais aussi entre les croisements intra et interspécifique. Un croisement se divise en plusieurs étapes : l'infestation des mollusques, le sexage des cercaires, la chronobiologie des cercaires, l'infestation des souris puis la récupération des œufs de la nouvelle génération. Toutes les méthodes employées ont été détaillées précédemment dans le « Méthodes et techniques d'entretien des cycles » et le « Méthode de chronobiologie de l'émission cercarienne ». Pour les besoins de ces travaux, il y a trois générations de parasites par croisement. La génération F0, la génération F1 (descendance hybride de F0) et la génération F2 (descendance issue du croisement entre deux individus F1). Seule la génération F2 du croisement (*S. mansoni* X *S. rodhaini*) a été générée pendant cette thèse.

### 6. Méthode histologique

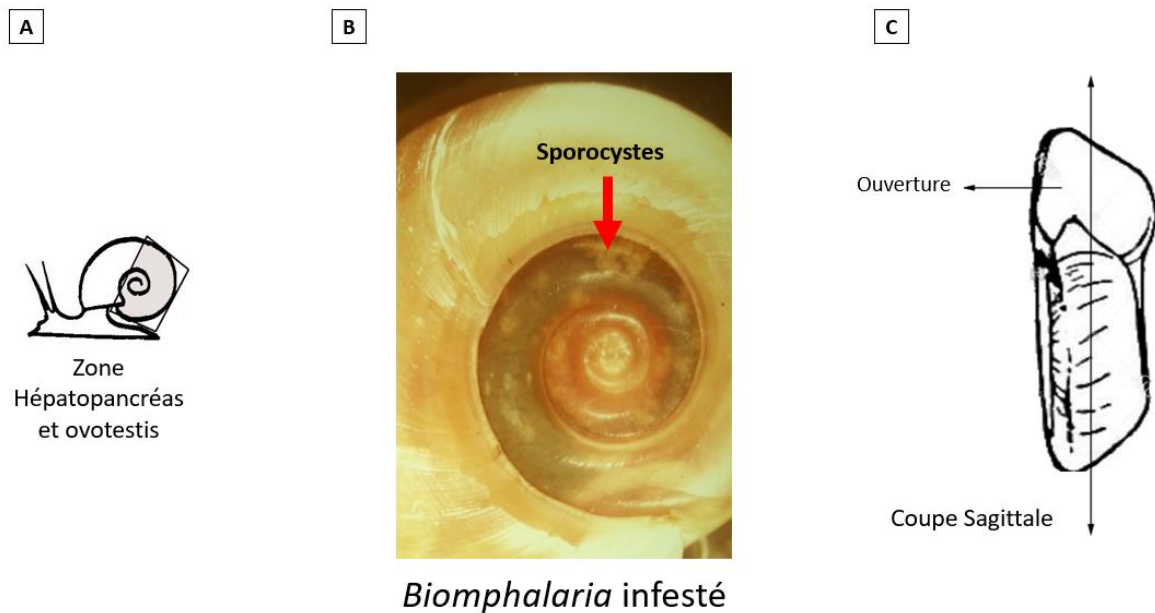
La technique se divise en plusieurs étapes qui sont la fixation, la déshydratation, l'inclusion dans la paraffine, la coupe, la confection des lames, la coloration puis le montage des lames. Dans un premier temps les mollusques entiers sont transférés dans un pilulier rempli de 10mL de solution de Halmi (Tableau\_12) pendant 10 jours (fixation). Ils subissent ensuite une série de bains d'éthanol à 70% jusqu'à la disparition de toute trace de Halmi (déshydratation). La coquille des mollusques est ensuite

## Matériel et méthodes

retirée ainsi que la tête, le pied et l'estomac pour garder uniquement l'hépatopancréas et l'ovotestis pour les coupes histologiques. Ces tissus sont inclus dans la paraffine, mis en bloc et coupés (coupes sagittales, Figure\_15).

**Tableau 12 : Composition de la solution de Halmi**

Solution de Halmi	Pour 100mL
Solution aqueuse saturée d'acide picrique	10ml
Susa de Heidenhain	90ml



**Figure 15 : Détail des coupes.** A) La zone grisée représente où se situent l'hépatopancréas et l'ovotestis dans le mollusque. B) Photographie de la coquille d'un *Biomphalaria* infesté au niveau de l'hépatopancréas et de l'ovotestis. Les granulomes blancs visibles au travers de la coquille sont les sporocystes. C) L'axe vertical représente la coupe sagittale effectuée pour les échantillons.

Les lames sont ensuite confectionnées, déparaffinées et subissent une coloration HES (Hématoxyline Eosine Safran). L'ultime étape consiste à faire le montage lame/lamelle puis de procéder à l'observation au microscope. L'inclusion paraffine jusqu'au montage des lames a été faite par la plateforme RHEM (Réseau d'histologie expérimentale de Montpellier).



## 7. Méthode d'analyse statistique

### 7.1. Statistiques descriptives

Les statistiques descriptives telles que la moyenne, l'écart-type, l'écart-standard, le coefficient d'asymétrie et le coefficient d'aplatissement sont calculées à partir d'Excel version 2201.

### 7.2. Comparaison de deux moyennes

Les moyennes sont données avec l'écart standard. Elles ont été comparées avec la fonction `wilcox.test` et la fonction `t.test` de R (version 4.04). Préalablement à cela, la normalité des données a été testée avec la fonction `shapiro.test`. Si la  $p$ -value est significative ( $p$ -value < 0.01), les données ne sont pas normales et le test non-paramétrique de Wilcoxon-Mann-Whitney (`wilcox.test`) est réalisé. Dans le cas contraire, le test paramétrique de Student ou de Welch (`t.test`) est effectué. Si l'effectif est inférieur à 30, le test non-paramétrique de Wilcoxon-Mann-Whitney (`wilcox.test`) est directement réalisé.

---

## Partie I Description des profils d'émission journaliers des cercaires chez

### *Schistosoma mansoni*

Les travaux de cette thèse reposent principalement sur l'étude des chronotypes diurne et nocturne de *Schistosoma mansoni* découverts dans la région du Dhofar à Oman (Mouahid, Idris et al., 2012). Cette proximité géographique, leurs phénotypes bien distincts ainsi que leur gonochorisme sont des avantages pour comprendre et décortiquer les déterminants moléculaires impliqués dans les profils d'émission journaliers des cercaires. L'étude de cette émission est critique puisque c'est l'étape de transmission du parasite aux mammifères. Il est donc essentiel de caractériser dans un premier temps le profil d'émission avant de vouloir chercher à comprendre les mécanismes moléculaires sous-jacents.

Cette première partie de thèse vise à établir la base des phénotypes des émissions journalières qui serviront de point de départ pour les études moléculaires. Cela se fera au travers de deux chapitres avec des objectifs bien distincts.

Le premier chapitre sera consacré à la présentation des caractéristiques des profils d'émission journaliers de chaque chronotype. Il permettra de mettre en lumière ce qui distingue et rassemble ces deux comportements d'émission en condition classique.

Le deuxième chapitre de cette partie servira à déterminer la nature endogène ou exogène du rythme d'émission et ainsi orienter les recherches des mécanismes moléculaires impliqués.

### I.1. Description des profils d'émission journaliers des cercaires chez les deux chronotypes, diurne et nocturne, de *Schistosoma mansoni*

L'étude des profils d'émission se fait au travers des expériences de chronobiologie. Elles consistent à observer le comportement du rythme d'émission en fonction de différentes conditions expérimentales.

Un grand nombre d'expériences de chronobiologie ont été menées dans la littérature chez *S. mansoni* diurne. Toutefois, la description des rythmes d'émission des cercaires et la méthodologie répondent à la discrétion de chaque auteur. Cette diversité se retrouve dès l'étape de comptage des cercaires et peut ainsi influencer sur la finesse des analyses. Par exemple, le comptage peut être effectué toutes les 12 (Valle, Pellegrino et al., 1973), 6 (Glaudel & Etges, 1973), 2 (Williams, Wessels & Gilbertson, 1984 ; Wang, Zhu et al., 2015) ou 1 heure(s) (Nojima & Sato 1982 ; Théron, Mouahid & Moné, 1997 ; Mouahid, Idris et al., 2012) avec des méthodes différentes, à partir d'aliquots (Luttermoser, 1955 ; Nojima & Sato 1982), à partir de récipient contenant plusieurs mollusques (Asch, 1972), à partir de mollusques individuels (Mouahid, Idris et al., 2012) ou encore avec une machine (Williams, Wessels & Gilbertson, 1984 ; Théron, 1989 ; Théron, Mouahid & Moné, 1997).

En plus de cette disparité dans les méthodes, il est naturel de s'interroger sur les composantes des variabilités biologiques qui pourraient entrer en compte dans l'étude des profils d'émission des

## Partie I

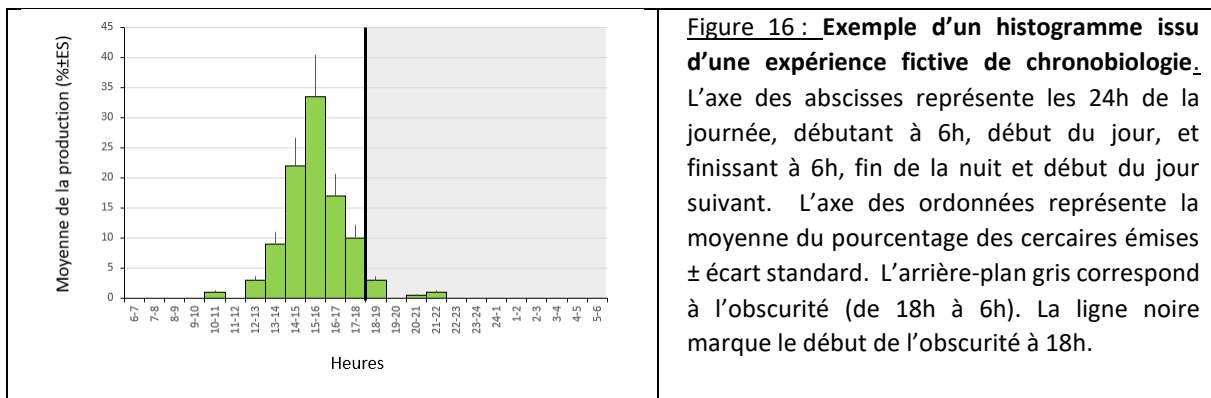
cercaires. La première inhérente à l'étude d'un parasite est de s'interroger sur l'influence de l'hôte. Un certain nombre d'études ont été menées à ce sujet et tendent à montrer que l'influence de l'hôte est négligeable. C'est le cas d'études menées sur le couple *S. mansoni*/*Biomphalaria* qui montrent que ni la perturbation du rythme de l'hôte (activité locomotrice et battements de cœur de *B. glabrata*) (Williams & Gilbertson, 1983), ni la pigmentation de l'hôte (*B. glabrata* pigmenté Vs albinos) (Théron, 1980), ni l'espèce d'hôte (*Biomphalaria sudanica* et *Biomphalaria stanleyi*) (Kazibwe, Makanga et al., 2010), n'ont d'effet sur le rythme d'émission des cercaires. L'autre composante biologique, cette fois-ci propre au parasite, est la potentielle influence du sexe sur le rythme d'émission des cercaires. Toutefois aucune étude ne s'est penchée spécifiquement sur cette problématique.

L'objectif de ce premier chapitre va être d'établir un profil clair d'émission journalier des cercaires pour chaque chronotype de *S. mansoni* et de définir les variations biologiques à prendre en compte ou non pour les expériences suivantes.

Pour pallier à la variation dans les études, le premier point de ce chapitre sera consacré à la définition des paramètres de description qui servira tout au long de la thèse. Le second point sera une présentation des chronotypes diurne et nocturne de *Schistosoma mansoni* Oman avec ces nouveaux paramètres. Le troisième point servira à déterminer l'influence de l'hôte mollusque sur le rythme d'émission des cercaires. Enfin, le quatrième et dernier point s'intéressera à l'influence du sexe du parasite sur le rythme d'émission des cercaires.

### I.1.1. Les paramètres de description

Les résultats des expériences de chronobiologie sont représentés sous la forme d'histogrammes (Figure\_16). Les abscisses indiquent les 24 heures de la journée avec une alternance d'un jour et d'une nuit correspondant à un cycle biologique de 24 heures (nyctémère). Le jour est symbolisé par un fond blanc commençant à 6h et se terminant à 18h. La nuit est symbolisée par un fond gris qui commence à 18h et finit à 6h. Ces horaires d'alternance 12-12 correspondent aux horaires retrouvés dans les pays tropicaux. Les ordonnées de l'histogramme représentent les pourcentages moyens d'émission des cercaires.

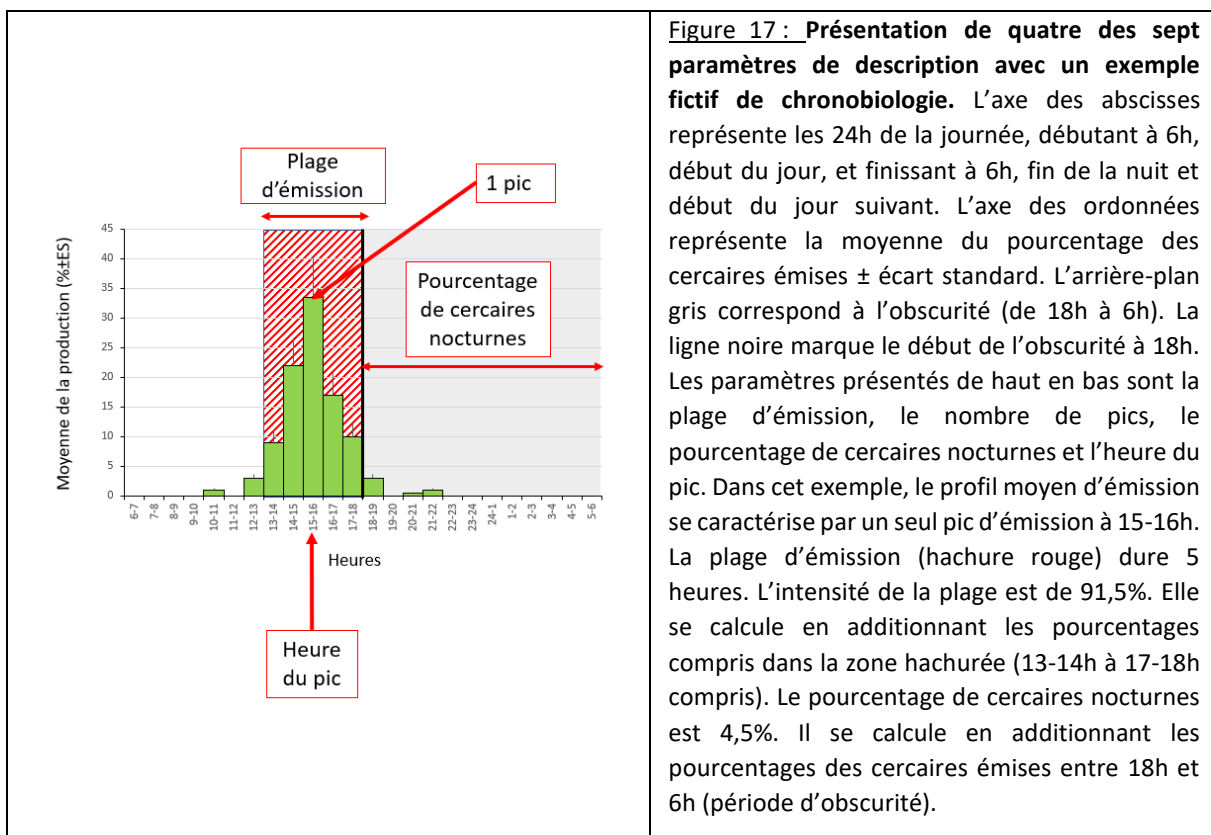


## Partie I

Toutes les expériences de chronobiologie de cette thèse reposent sur un comptage des cercaires toutes les heures pour chaque individu. Les résultats présentés sont la moyenne de l'émission de tous les profils  $\pm$  l'écart-standard (ES). Pour la significativité des analyses, seuls les profils d'émission basés sur un minimum de 20 cercaires émises par 24h sont utilisés. Le pallier des 5%, abordé ci-après, correspond au minimum à l'émission d'une cercaire.

La description des profils d'émission se fera en utilisant sept paramètres : deux liés directement au(x) pic(s) : le nombre de pics et l'heure du pic (Figure\_17) ; quatre autres paramètres liés à la forme du profil : la plage horaire d'émission (Figure\_17), l'intensité de la plage horaire d'émission, le coefficient d'asymétrie (Figure\_18) et le coefficient d'aplatissement (Figure\_19) ; et enfin un dernier paramètre lié au positionnement dans le nyctémère : le pourcentage de cercaires émises la nuit (Figure\_17).

Pour le besoin de certaines expériences, le nombre total de cercaires émises sur 24h pourra être pris compte comme un huitième paramètre de description.

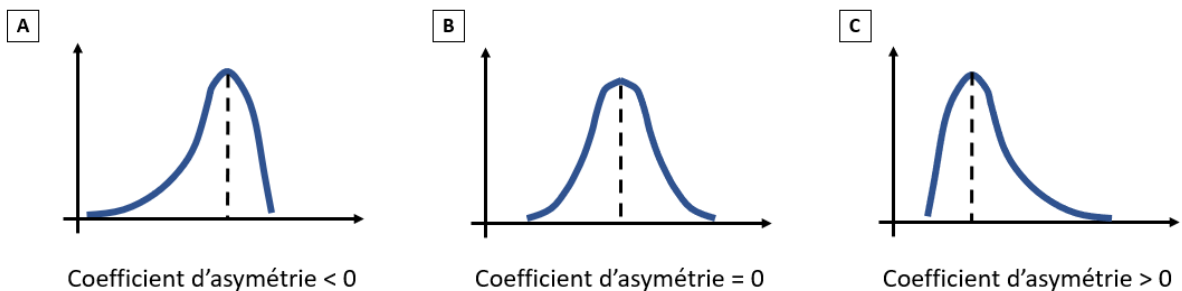


Une explication plus approfondie de ces paramètres suit :

- Nombre de pics : le chronotype pourra être soit unimodal soit multimodal.
- Heure du pic : heure d'émission durant laquelle le pourcentage de cercaires émises est le plus élevé ; en d'autres termes, l'heure à laquelle les risques de transmission du parasite sont les plus fortes. L'heure moyenne du pic se détermine en sélectionnant l'heure où le plus de profils ont eu leur pic, et non pas en sélectionnant l'heure où la moyenne des pourcentages d'émission est le plus élevé.
- La plage horaire d'émission : correspond au nombre d'heures entourant le pic pendant lesquelles le pourcentage des cercaires émises est supérieur à 5%.

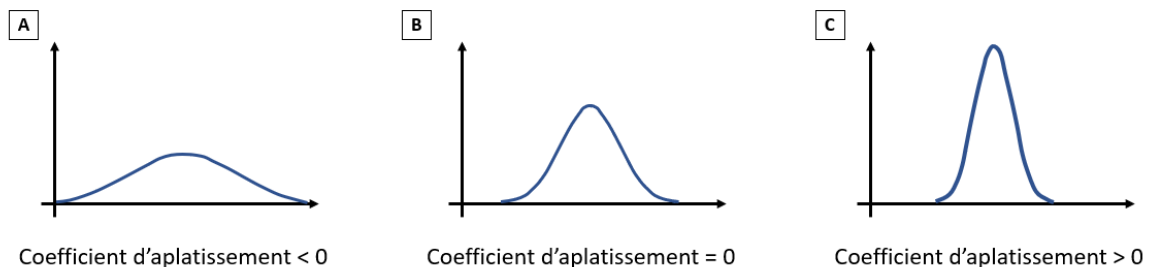
## Partie I

- Intensité de la plage horaire d'émission : elle se calcule en additionnant les pourcentages d'émission compris dans la plage d'émission. Cela comprend le pic, l'heure avant le pic, l'heure après le pic et toutes les heures entourant le pic où la production cercarienne dépasse 5%.
- Coefficient d'asymétrie ou skewness (Figure\_18) : mesure le degré d'asymétrie d'un profil. Cela permet de déterminer si plus de cercaires sont émises avant ou après le pic d'émission. Si le coefficient est nul, il y a autant de cercaires qui sont émises avant et après le pic (distribution symétrique). Un coefficient positif indique que plus de cercaires sont émises après le pic (distribution étalée sur la droite). A l'inverse, un coefficient négatif indique que plus de cercaires sont émises avant le pic (distribution étalée sur la gauche). Il est convenu que les coefficients d'asymétrie dont les valeurs se situent entre  $-0,5$  et  $+0,5$ ,  $-1$  et  $+1$  et  $<-1$  et  $>+1$  rendent compte respectivement d'une symétrie d'émission, d'une asymétrie modérée et d'une asymétrie élevée. Ce coefficient se calcule uniquement sur la plage d'émission.



**Figure 18 : Illustrations du coefficient d'asymétrie (skewness).** Chaque schéma représente l'allure de la distribution en fonction de la valeur du coefficient d'asymétrie. La ligne pointillée représente le mode de la distribution. A) La distribution s'étale sur la gauche (coefficient d'asymétrie  $< 0$ ). B) La distribution est symétrique (coefficient d'asymétrie = 0). C) La distribution s'étale sur la droite (coefficient d'asymétrie  $> 0$ ).

- Coefficient d'aplatissement ou kurtosis (Figure\_19) : détermine le centrage des données autour du pic. Dans le cas d'une distribution normale, la valeur de la statistique est égale à zéro. Un coefficient positif indique ainsi que l'émission des cercaires est concentrée autour du pic. A l'inverse, un coefficient négatif indique que l'émission des cercaires se fait de façon plus étalée autour du pic. Ce coefficient se calcule uniquement sur la plage d'émission.



**Figure 19 : Illustrations du coefficient d'aplatissement (kurtosis).** Chaque schéma représente l'allure de la distribution en fonction de la valeur du coefficient d'asymétrie. A) Distribution aplatie (coefficient d'aplatissement  $< 0$ ). B) Distribution normale (coefficient d'aplatissement = 0). C) Distribution centrée (coefficient d'aplatissement  $> 0$ ).

## Partie I

- Le pourcentage de cercaires émises la nuit : ce sont les cercaires émises entre 18h et 6h. Pour les expériences de photopériode continue, nous parlerons de « nuit subjective », c'est-à-dire les douze dernières heures d'une journée qui correspondraient à la nuit en condition de photopériode équilibrée L-O (12-12).

### I.1.2. Présentation des chronotypes diurne et nocturne

Les deux chronotypes de *S. mansoni* utilisés dans cette thèse ont été décrits pour la première fois dans l'article de Mouahid, Idris et al. (2012). L'objectif de ce point est de décrire ces deux chronotypes avec nos paramètres de description des profils et de poser les bases pour la suite de la thèse.

#### I.1.2.1. Protocole

La chronobiologie a été effectuée à partir de *B. pfeifferi* infestés avec 1 miracidium de *S. mansoni* diurne ou nocturne et de *B. glabrata* infestés avec 10 miracidia de *S. mansoni* diurne. Nous avons analysé, pendant deux jours consécutifs, les profils de 42 infestations pour chaque chronotype ; soit au total 84 profils diurnes (60 à partir de *B. pfeifferi* et 24 à partir de *B. glabrata*) et 84 profils nocturnes à partir de *B. pfeifferi*.

#### I.1.2.2. Chronotype diurne

Les résultats moyens, accompagnés des variations individuelles, sont présentés dans la Figure\_20 et le Tableau\_13. Sur les 84 profils diurnes réalisés, 74 ont pu être utilisés car leurs émissions totales sur les 24 heures étaient égales ou supérieures à 20 cercaires (50 profils à partir de *B. pfeifferi* et 24 à partir de *B. glabrata*). Chez le chronotype nocturne, 75 profils sur 84 ont rempli cette condition des 20 cercaires.

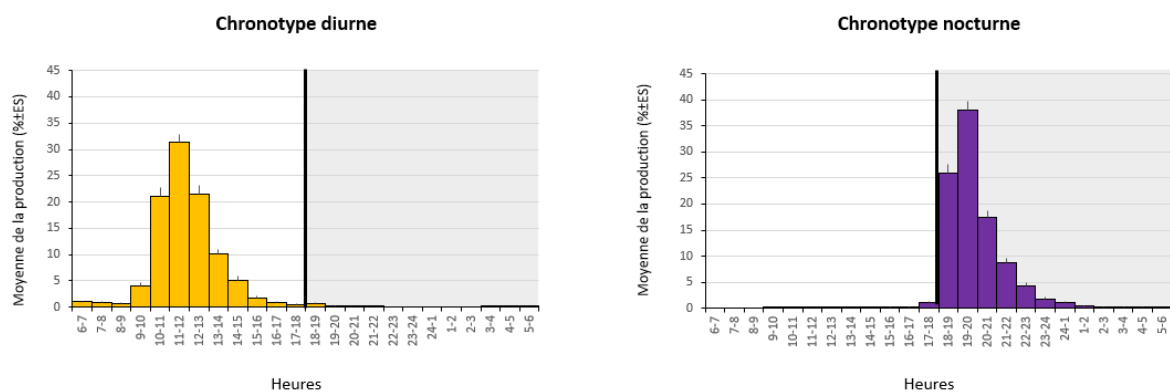


Figure 20 : Chronobiologie des chronotypes diurne (orange) et nocturne (violet) de *S. mansoni*. Chaque histogramme a, en ordonnées, la moyenne du pourcentage des cercaires émises ± écart standard. L'axe des abscisses représente les 24h de la journée, débutant à 6h, début du jour, et finissant à 6h, fin de la nuit. L'arrière-plan gris correspond à l'obscurité. La ligne noire marque le début de l'obscurité à 18h.

**Tableau 13 : Paramètres des profils moyens d'émission, accompagnés des variations individuelles, pour les parasites diurne et nocturne de *S. mansoni***

Paramètres	Moyenne $\pm$ ES <i>Variations individuelles</i>		Test de Wilcoxon- Mann-Whitney
	Diurne	Nocturne	
Effectif	N=74	N=75	
Photopériode	L-O	L-O	
Nombre de pics	1 1	1 1	
Heure du pic	11-12h 10-11h à 14-15h	19-20h 18-19h à 22-23h	$\neq$
Plage d'émission (h)	4,1 $\pm$ 0,1 3 à 7	4,1 $\pm$ 0,1 3 à 6	W = 2686, p = 0,72
Intensité de la plage (%)	90,1 $\pm$ 0,9 57,1 à 100	93,2 $\pm$ 0,6 71,3 à 100	W = 2003, p = 0,003
Asymétrie	+0,7 $\pm$ 0,1 -1,4 à +2,1	+0,5 $\pm$ 0,1 -1,8 à +1,9	W = 3069, p = 0,27
Aplatissement	+0,3 $\pm$ 0,3 (N=57) -5,9 à +4,6	-0,3 $\pm$ 0,4 (N=55) -5,7 à +4	W = 1777, p = 0,23
Cercaires la nuit (%)	1,06 $\pm$ 0,2 0,0 à 10,1	98,3 $\pm$ 0,3 89,3 à 100	W = 0, p < 2,2e-16

Le chronotype diurne a une émission unimodale dont l'heure du pic moyen est à 11-12h. La plage horaire moyenne d'émission est de 4h avec une intensité de 90%. La plage d'émission a une asymétrie modérée sur la droite et une distribution normale. L'émission nocturne est quasi inexistante, avec en moyenne 1% des cercaires émises la nuit. Du point de vue de la transmission, cela signifie que le risque le plus fort se concentre sur 4 heures de la journée, avec le pic à 11-12h, puisque 90% des cercaires sont émises sur cette période. La nuit est une période de risque faible puisque uniquement 1% de cercaires sont émises.

### I.1.2.3. Chronotype nocturne

Les résultats moyens, accompagnés des variations individuelles, sont présentés dans la Figure\_20 et le Tableau\_13. Sur les 84 profils nocturnes réalisés, 75 profils sur 84 ont rempli cette condition des 20 cercaires.

Le chronotype nocturne a une émission unimodale dont l'heure du pic moyen est à 19-20h. La plage horaire moyenne d'émission est de 4h avec une intensité de 93%. La plage d'émission moyenne est symétrique avec une distribution normale. L'émission nocturne est majoritaire avec en moyenne 98% des cercaires émises la nuit. Du point de vue de la transmission cela signifie que le risque le plus fort se concentre sur 4 heures de la journée avec le pic à 19-20h, puisque 93% des cercaires sont émises sur cette période. La nuit est la période de risque la plus élevée puisque 98% de cercaires sont émises. Le jour est donc une période de risque faible.

## Partie I

### I.1.2.4. Comparaison des chronotypes diurne et nocturne

Les comparaisons sont présentées dans le Tableau\_13. Les profils moyens des deux chronotypes se différencient par l'heure du pic (11-12h pour les diurnes et 19-20h pour les nocturnes), et le pourcentage de cercaires émises la nuit (autour de 1% pour les diurnes et autour de 98% pour les nocturnes ;  $p < 0,01$ ). Il est à noter que l'intensité de la plage horaire d'émission est significativement plus forte chez le chronotype nocturne ( $p < 0,01$ ) ; ce qui signifie que l'émission cercarienne du chronotype nocturne est plus concentrée vers l'heure du pic que chez le diurne. Les deux chronotypes ont tout de même en commun la présence d'un seul pic. De plus, les plages d'émission et leurs coefficients d'asymétrie et d'aplatissement ne sont pas significativement différents entre les diurnes et les nocturnes ( $p > 0,01$ ).

En conclusion, les résultats pour les deux chronotypes sont en accord avec Mouahid, Idris et al. (2012). Malgré quatre paramètres non discriminants, les périodes de haut et faible risques de transmission sont inversées (jour et nuit), ce qui illustre la singularité des chronotypes étudiés.

### I.1.3. Influence du mollusque

L'hôte mollusque naturel des deux chronotypes de *S. mansoni* Oman est *Biomphalaria pfeifferi*. *In natura*, certaines espèces de schistosomes peuvent utiliser comme hôtes intermédiaires des mollusques d'espèces (*S. mansoni* en Ouganda avec *B. stanleyi* et *B. sudanica*) (Kazibwe, Makanga et al., 2010) et même de genres différents (*S. bovis* avec *Bulinus truncatus* et *Planorbarius metidjensis*) (Mouahid & Théron, 1986). En laboratoire, il est possible de maintenir les parasites dans des mollusques non naturellement présents pour peu qu'ils soient compatibles. L'enjeu de ce point est de déterminer si l'utilisation d'un hôte non naturellement présent mais compatible (ici *B. glabrata*) a une influence sur le profil d'émission des cercaires, en comparaison avec un hôte naturel (*B. pfeifferi*).

#### I.1.3.1. Protocole

La chronobiologie a été effectuée à partir de *B. pfeifferi* infestés avec 1 miracidium de *S. mansoni* et de *B. glabrata* infestés avec 10 miracidia de *S. mansoni*. Nous avons analysé, pendant deux jours consécutifs, les profils de 42 infestations pour le chronotype diurne ; soit au total 84 profils diurnes (60 à partir de *B. pfeifferi* et 24 à partir de *B. glabrata*).

#### I.1.3.2. Comparaison des profils issus de *Biomphalaria pfeifferi* et *Biomphalaria glabrata*

Les résultats moyens, accompagnés des variations individuelles, sont présentés dans la Figure\_21 et le Tableau\_14. Sur les 84 profils diurnes réalisés, 74 ont pu être utilisés car leurs émissions totales sur les 24 heures étaient égales ou supérieures à 20 cercaires (50 à partir de *B. pfeifferi* et 24 à partir de *B. glabrata*). La même expérience n'a pas pu être effectuée avec le chronotype nocturne car les parasites nocturnes ont été incapables de se développer dans les mollusques *B. glabrata*.



Partie I

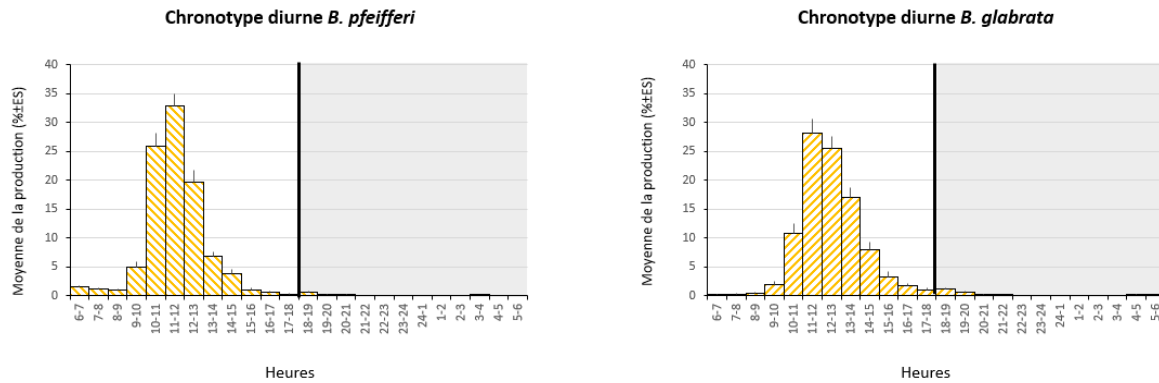


Figure 21 : Chronobiologie du chronotype diurne de *S. mansoni* à partir de *B. pfeifferi* et de *B. glabrata*. Chaque histogramme a, en ordonnées, la moyenne du pourcentage des cercaires émises  $\pm$  écart standard. L’axe des abscisses représente les 24h de la journée, débutant à 6h, début du jour, et finissant à 6h, fin de la nuit. L’arrière-plan gris correspond à l’obscurité. La ligne noire marque le début de l’obscurité à 18h.

Tableau 14 : Paramètres des profils moyens d’émission, accompagnés des variations individuelles, pour les parasites diurnes de *S. mansoni* à partir de *B. pfeifferi* et de *B. glabrata*

Paramètres	Moyenne $\pm$ ES Variations individuelles		Test de Wilcoxon-Mann-Whitney
	Diurne <i>B. pfeifferi</i>	Diurne <i>B. glabrata</i>	
Effectif	N=50	N=24	
Photopériode	L-O	L-O	
Nombre de pics	1 1	1 1	
Heure du pic	11-12h 10-11h à 14-15h	12-13h 10-11h à 13-14h	
Plage d’émission (h)	3,9 $\pm$ 0,1 3 à 6	4,6 $\pm$ 0,2 3 à 7	W = 329, p = 0,0008
Intensité de la plage (%)	89,9 $\pm$ 1,2 57,1 à 100	90,3 $\pm$ 1,1 72,5 à 97,6	W = 660, p = 0,50
Asymétrie	+0,8 $\pm$ 0,1 -1,4 à +2,1	+0,5 $\pm$ 0,2 -1,3 à +2	W = 758, p = 0,07
Aplatissement	+0,8 $\pm$ 0,3 (N=34) -3,1 à +4,6	-0,4 $\pm$ 0,5 (N=23) -5,9 à +3,9	W = 484, p = 0,13
Cercaires la nuit (%)	0,6 $\pm$ 0,2 0 à 10,1	1,9 $\pm$ 0,4 0 à 7,4	W = 223, p = 4,61e-06
Nombre total de cercaires	194,8 $\pm$ 37,8 21 à 1371	672,8 $\pm$ 72,6 126 à 1589	W = 119, p = 2,79e-08

Les profils moyens issus de *B. pfeifferi* et issus de *B. glabrata* ont en commun la présence d’un seul pic dont l’heure moyenne varie d’une heure (respectivement 11-12h et 12-13h). Ils partagent aussi un coefficient d’asymétrie modéré positif et une émission nocturne quasi inexistante (moins de 2%). L’intensité de la plage et les coefficients d’asymétrie et d’aplatissement ne sont pas significativement différents entre les profils issus de *B. pfeifferi* et ceux issus de *B. glabrata* ( $p > 0,01$ ).

## Partie I

A l'inverse, la plage horaire d'émission des cercaires et le nombre total de cercaires émises sur 24h sont significativement plus grands dans les profils issus de *B. glabrata* ( $p < 0,01$ ). A noter que, même si le nombre de cercaires nocturnes est significativement plus grand chez les profils issus de *B. glabrata*, cette émission moyenne reste anecdotique puisqu'elle représente moins de 2% d'émission.

La différence majeure entre les deux types de profils s'observe au travers d'un paramètre qui n'était pas pris en compte jusque-là, le nombre total de cercaires émises sur 24h. Cette différence peut s'expliquer par le fait que les *B. glabrata* sont naturellement plus grands que les *B. pfeifferi*. Les parasites ayant plus de place pour se développer sont plus nombreux à sortir. Cette différence quantitative de production cercarienne ne modifie en rien l'heure de sortie des cercaires.

D'un point de vue de la transmission, le risque est plus grand avec *B. glabrata* car plus de cercaires sont émises. Cependant, la période de risque ne change pas entre les deux espèces de mollusques. En effet, la différence d'une heure dans l'heure du pic moyen n'est pas importante, car ce genre d'écart est observé entre les profils appartenant à une même espèce.

En conclusion, le changement d'hôte ne change pas le rythme d'émission du chronotype diurne. Seule la production cercarienne (paramètre quantitatif) varie significativement mais n'a pas d'influence sur le rythme en lui-même. De ce fait, l'espèce de l'hôte mollusque n'est pas un paramètre qui influence le rythme des profils d'émission des cercaires.

### I.1.4. Influence du sexe du parasite

Les parasites du genre *Schistosoma* sont gonochoriques. Contrairement au stade ver adulte, il n'est pas possible de distinguer morphologiquement le sexe du parasite pour les autres stades. Même s'il n'y a pas de différence sur la morphologie des cercaires et sporocystes, il est nécessaire d'explorer si des différences comportementales existent entre les deux formes. L'objectif de ce point va être de déterminer si le sexe du parasite a une influence sur le profil d'émission des cercaires des deux chronotypes de *S. mansoni*.

#### I.1.4.1. Protocole

La chronobiologie a été effectuée à partir de *B. pfeifferi* infestés individuellement avec 1 miracidium de *S. mansoni*. Nous avons analysé les profils d'émission de 15 infestations mâles et de 15 infestations femelles pour chaque chronotype diurne et nocturne de *S. mansoni*, pendant deux jours consécutifs ; soit au total 30 profils mâles et 30 profils femelles.

#### I.1.4.2. Chronotype diurne

Les résultats moyens, accompagnés des variations individuelles, sont présentés dans la Figure\_22 et le Tableau\_15. Sur les 30 profils mâles réalisés, 25 ont pu être utilisés car leurs émissions totales sur les 24 heures étaient égales ou supérieures à 20 cercaires. Chez les femelles, 25 profils sur 30 remplissaient cette condition des 20 cercaires.

## Partie I

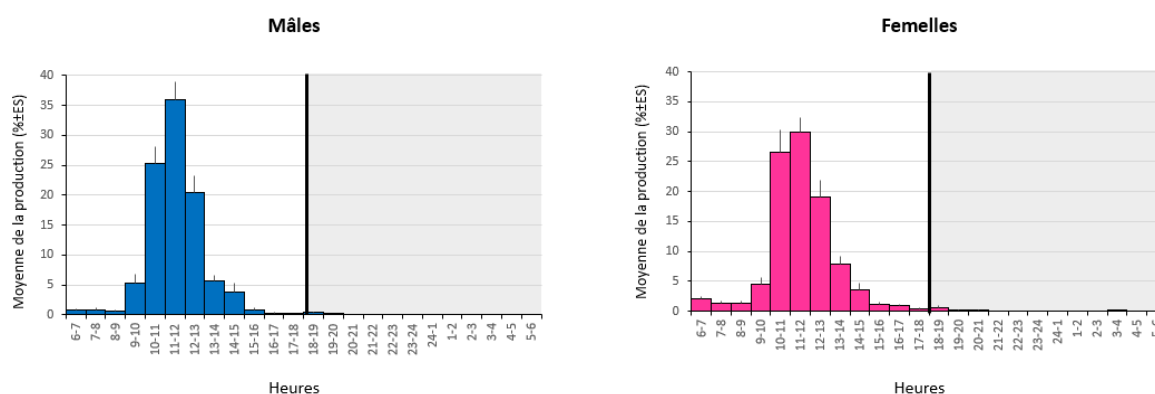


Figure 22 : Profils moyens d'émission de *S. mansoni* de chronotype diurne entre mâles (bleu) et femelles (rose). Chaque histogramme a, en ordonnées, la moyenne du pourcentage des cercaires émises  $\pm$  écart standard. L'axe des abscisses représente les 24h de la journée, débutant à 6h, début du jour, et finissant à 6h, fin de la nuit. L'arrière-plan gris correspond à l'obscurité. La ligne noire marque le début de l'obscurité à 18h.

Tableau 15 : Paramètres des profils moyens d'émission, accompagnés des variations individuelles, pour les parasites mâles et femelles du chronotype diurne de *S. mansoni*

Paramètres	Moyenne $\pm$ ES Variations individuelles		Test de Wilcoxon- Mann-Whitney
	Mâle diurne	Femelle diurne	
Effectif	N=25	N=25	
Photopériode	L-O	L-O	
Nombre de pics	1 1	1 1	
Heure du pic	11-12h 10-11h à 14-15h	11-12h 10-11h à 12-35h	
Plage d'émission (h)	3,8 $\pm$ 0,1 3 à 5	4,0 $\pm$ 0,2 3 à 6	W = 190, $p = 0,64$
Intensité de la plage (%)	92,5 $\pm$ 1,1 76,3 à 100	87,3 $\pm$ 2,1 57,1 à 100	W = 420, $p = 0,039$
Asymétrie	+0,9 $\pm$ 0,2 -1,4 à +2,1	+0,8 $\pm$ 0,2 -1,4 à +2	W = 312, $p = 1$
Aplatissement	+0,5 $\pm$ 0,6 (N=17) -3,1 à +4,6	+1,2 $\pm$ 0,6 (N=17) -3,1 à +4,1	W = 121, $p = 0,43$
Cercaires la nuit (%)	0,4 $\pm$ 0,1 0 à 2,7	0,9 $\pm$ 0,5 0 à 10,1	W = 307, $p = 0,90$

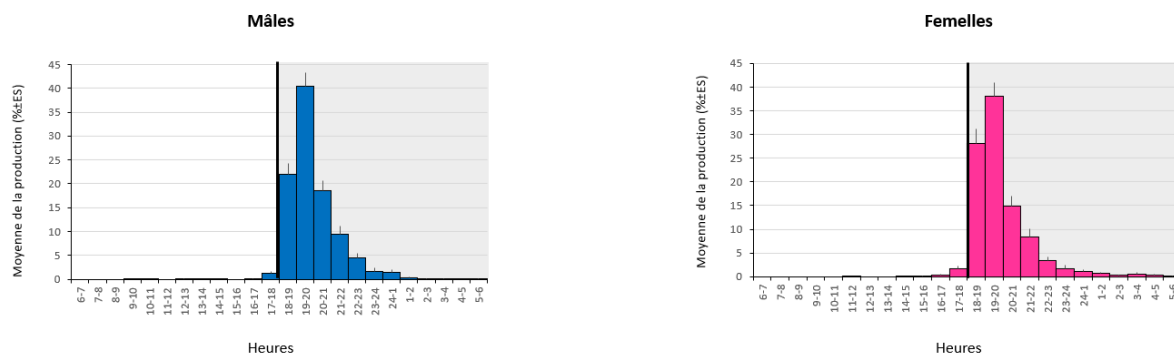
Les profils moyens, mâle et femelle, ont en commun la présence d'un seul pic situé à 11-12h et une plage d'émission d'environ 4h avec une intensité moyenne supérieure à 87%. Ils partagent aussi un coefficient d'asymétrie modéré positif et un coefficient d'aplatissement positif, ainsi qu'une émission nocturne quasi inexistante. La plage d'émission, son intensité, les coefficients d'asymétrie et d'aplatissement et le pourcentage de cercaires nocturnes ne sont pas significativement différents entre les mâles et les femelles ( $p > 0,01$ ).

En conclusion, les résultats montrent qu'il n'y a pas d'influence du sexe sur le profil d'émission des cercaires chez le chronotype diurne de *S. mansoni*.

## Partie I

### I.1.4.3. Chronotype nocturne

Les résultats moyens, accompagnés des variations individuelles, sont présentés dans la Figure\_23 et le Tableau\_16. Sur les 30 profils mâles réalisés, 29 ont pu être utilisés car leurs émissions totales sur les 24 heures étaient égales ou supérieures à 20 cercaires. Chez les femelles, 26 profils sur 30 remplissent cette condition des 20 cercaires.



**Figure 23 : Chronobiologie de *S. mansoni* nocturne mâle (bleu) et femelle (rose).** Chaque histogramme a, en ordonnées, la moyenne du pourcentage des cercaires émises  $\pm$  écart standard. L'axe des abscisses représente les 24h de la journée, débutant à 6h, début du jour, et finissant à 6h, fin de la nuit. L'arrière-plan gris correspond à l'obscurité. La ligne noire marque le début de l'obscurité à 18h.

**Tableau 16 : Paramètres des profils moyens d'émission, accompagnés des variations individuelles, pour les parasites mâles et femelles du chronotype nocturne de *S. mansoni***

Paramètres	Moyenne $\pm$ ES <i>Variations individuelles</i>		Test de Wilcoxon-Mann-Whitney
	Mâle nocturne	Femelle nocturne	
Effectif	N=29	N=26	
Photopériode	L-O	L-O	
Nombre de pics	1 1	1 1	
Heure du pic	19-20h 18-19h à 21-22h	19-20h 18-19h à 21-22h	
Plage d'émission (h)	4,2 $\pm$ 0,2 3 à 6	4,1 $\pm$ 0,1 3 à 4	W = 396, $p = 0,74$
Intensité de la plage (%)	94,4 $\pm$ 0,7 83,4 à 100	92,3 $\pm$ 1,3 71,3 à 100	W = 462, $p = 0,15$
Asymétrie	+0,5 $\pm$ 0,2 -1,7 à +1,7	+0,7 $\pm$ 0,2 -1,2 à +1,9	W = 326, $p = 0,39$
Aplatissement	-0,6 $\pm$ 0,6 (N=19) -5,7 à +2,7	-0,1 $\pm$ 0,6 (N=20) -5,7 à +4	W = 166, $p = 0,51$
Cercaires la nuit (%)	98,6 $\pm$ 0,3 92,7 à 100	97,8 $\pm$ 0,6 89,3 à 100	W = 380, $p = 0,97$

Les profils moyens mâles et femelles ont en commun la présence d'un seul pic situé à 19-20h et une plage d'émission d'environ 4h avec une intensité moyenne au-dessus de 92%. Ils partagent aussi un coefficient d'asymétrie positif, un coefficient d'aplatissement négatif, ainsi qu'une forte émission nocturne. La plage d'émission et son intensité, les coefficients d'asymétrie et d'aplatissement et le pourcentage de cercaires nocturnes ne sont pas significativement différents entre les mâles et les femelles ( $p > 0,01$ ).

## Partie I

En conclusion, les résultats montrent qu'il n'y a pas d'influence du sexe sur le profil d'émission des cercaires chez le chronotype nocturne de *S. mansoni* (Oman).

### I.1.5. Discussion

Afin d'éliminer toute potentielle variation expérimentale, une méthodologie et des paramètres d'analyses communs ont été adoptés pour toutes les expériences de chronobiologie de cette thèse. Afin de refléter au mieux les caractéristiques des profils journaliers d'émission de cercaires, sept paramètres ont été choisis. Ces paramètres visent, en plus de la description du pic de l'émission (heure et nombre), à caractériser la forme du profil (plage horaire, intensité, coefficients d'asymétrie et d'aplatissement) et la position dans le nyctémère (pourcentage de cercaires nocturnes). En effet, en fonction des espèces et chronotypes de schistosomes l'émission des cercaires peut être plus ou moins concentrée autour du pic d'émission. Par exemple un schistosome qui infeste l'Homme a un intérêt stratégique à émettre des cercaires sur une plage horaire plus longue qu'un schistosome qui infeste les rongeurs ; en effet, l'Homme, de par la multitude de ses activités aquatiques (baignade, pêche, élevage, lavage, etc.) est en contact avec l'eau sur des périodes plus longues que les rongeurs. De plus, les rongeurs sauvages ont un contact avec l'eau plus court pour éviter les prédateurs. Voilà pourquoi il est important d'un point de vue écologique, sanitaire et aussi expérimental de ne pas uniquement s'intéresser au pic mais d'aussi prendre en compte la variabilité de la forme et la position de l'émission.

Le nombre de cercaires est un huitième paramètre qui n'est pas directement lié à la notion de rythme d'émission. Bien qu'il faille tout de même un nombre conséquent de cercaires pour caractériser correctement un profil d'émission (fixé à un minimum de 20 cercaires), ce paramètre n'est utilisé que lorsque son analyse est pertinente dans l'expérimentation.

Les résultats obtenus avec l'utilisation des sept paramètres pour décrire les profils d'émission journalière des cercaires des chronotypes diurne et nocturne de *S. mansoni* (Oman) sont en accord avec ce qui avait été décrit par Mouahid, Idris et al. (2012). Les différences observées entre les deux chronotypes sont l'heure du pic (11-12h pour les diurnes et 19-20h pour les nocturnes), la place de l'émission dans le nyctémère avec le pourcentage de cercaires nocturnes (autour de 1% pour les diurnes et autour de 98% pour les nocturnes ;  $p < 0,01$ ). Une autre différence observée est l'intensité de plage d'émission qui est significativement plus élevée chez le chronotype nocturne. Les quatre autres paramètres n'ont pas montré de différence significative. D'autres similitudes et différences ont été rapportées entre les deux chronotypes (Mouahid, Mintsu Nguema et al., 2019) ; par exemple, le taux d'infestation chez les mollusques et le taux de survie chez les souris sont similaires entre les deux chronotypes. Ces deux chronotypes diffèrent aussi par la morphologie de leurs œufs (légère courbure à l'opposé de l'éperon latéral chez le chronotype nocturne). De plus, le chronotype diurne présente une période pré-patente chez le mollusque plus courte et une production cercarienne significativement plus élevée. Ces observations supplémentaires confirment la singularité de ces deux chronotypes de *S. mansoni* (Oman).

Avec les profils d'émission établis, les potentiels paramètres biologiques (hôte et parasite) qui pourraient influencer le rythme en dehors des conditions expérimentales ont été testés. Cette expérience sur l'influence de l'hôte a uniquement pu être menée sur le chronotype diurne. Les différentes tentatives d'infestation sur *B. glabrata* par le chronotype nocturne se sont révélées infructueuses, illustrant ainsi l'incompatibilité du couple *S. mansoni* nocturne/ *B. glabrata* et la singularité du chronotype diurne. La comparaison entre les profils d'émission issus de *B. glabrata* et *B. pfeifferi* chez le chronotype diurne a montré qu'il n'y avait pas de différence significative dans le rythme d'émission des cercaires. Ces résultats sont en accord avec ceux qui avaient été montrés chez d'autres souches de *S. mansoni* (Théron, 1980 ; Williams & Gilbertson, 1983 ; Kazibwe, Makanga et al.,

## Partie I

2010). Si l'hôte mammifère associé au parasite joue un rôle crucial dans le rythme d'émission des cercaires des schistosomes (adaptation à l'hôte définitif), l'hôte mollusque n'exerce pas une influence notable. Cette observation est appuyée par une étude menée sur *Schistosoma bovis* qui infeste naturellement des mollusques de genres différents (*Planorbarius* et *Bulinus*) et montre que le rythme d'émission des cercaires n'est pas impacté (Mouahid & Théron, 1986). Nous pouvons donc conclure que ni le genre, ni l'espèce, ni la pigmentation, ni la perturbation du rythme cardiaque et de l'activité locomotrice de l'hôte intermédiaire n'ont d'influence sur le rythme d'émission des cercaires chez les schistosomes.

S'il n'y a pas de différence pour le rythme, nos résultats ont tout de même montré une différence significative pour le nombre de cercaires émises et la plage horaire d'émission qui sont plus élevées chez *B. glabrata*. Cette différence peut s'expliquer par la taille des *B. glabrata* qui sont plus grands. Les parasites ont plus d'espace pour se développer et la plage d'émission s'agrandit. Cette différence de plage horaire n'est pas assez importante (moins d'une heure) pour être considérée comme un paramètre biologique de variation dans l'étude des profils d'émission des cercaires.

Si une potentielle influence de l'hôte mollusque est écartée, des conclusions similaires ont été trouvées pour le sexe du parasite. En effet, que ce soit pour le chronotype diurne ou nocturne, aucun des sept paramètres de description ne s'est montré significativement différent entre les sexes. Cette non-différence entre les sexes vient s'ajouter à d'autres traits de vie déjà répertoriés comme l'éclosion des œufs, la durée de vie et l'infectivité des miracidia ou encore la période pré-patente dans le mollusque (Boissier, Morand & Moné, 1999). Cette similitude biologique s'explique naturellement d'un point de vue de la transmission puisqu'il n'y a pas d'intérêt écologique à ce que les partenaires sexuels sortent à des heures différentes. Au contraire, l'intérêt du parasite est de faire en sorte que les formes mâles et femelles infestent le même hôte mammifère et donc sortent du mollusque à la même heure. A noter qu'il existe tout de même un biais sexuel dans l'infectivité des cercaires (en infestation unisexe) et leur durée de vie (Boissier & Moné, 2000).

Ce chapitre illustre bien la singularité des deux chronotypes de *S. mansoni* Oman et écarte l'influence de l'hôte mollusque et du sexe du parasite dans le rythme des profils journaliers d'émission des cercaires.

### 1.2. Nature des profils d'émission journaliers des cercaires chez les deux chronotypes, diurne et nocturne, de *Schistosoma mansoni*

Les schistosomes et plus largement les trématodes ne sont pas les seuls parasites à avoir développé des comportements basés sur un rythme journalier. Par exemple, chez les nématodes filaires agents de filarioses lymphatiques (*Brugia malayi* et *Wuchereria bancrofti*), le nombre de microfilaries des parasites est plus élevé dans la circulation périphérique sanguine de l'hôte humain la nuit (Hawking, 1967). Ce comportement rythmique augmente les chances de rencontre avec le vecteur anophèle qui pique les humains la nuit. Chez les parasites sanguins du genre *Plasmodium* (agents du paludisme), le développement intra-érythrocytaire suit un rythme de 24h (Hawking, Worms & Gammage, 1968) créant ainsi des fièvres rythmiques chez l'hôte. Ces rythmes journaliers ne se limitent pas aux parasites de mammifères. En effet, le parasite des plantes, *Botrytis cinerea*, augmente sa virulence au crépuscule pour esquiver le système immunitaire de la plante (Hevia, Canessa et al., 2015).

## Partie I

De façon générale ces divers comportements rythmiques chez les parasites leur permettent d'optimiser les conditions de rencontre avec l'hôte, d'optimiser l'infestation ou encore d'échapper au système immunitaire de l'hôte (Reece, Prior & Mideo, 2017). Le questionnement naturel est donc de savoir si ces avantages évolutifs rythmiques sont dus à la présence d'une horloge interne chez le parasite (rythme endogène) ou sont dépendants du rythme de l'hôte en tant que zeitgeber et/ou directement de l'environnement extérieur (rythme exogène).

Cette question a été résolue pour quelques espèces puisqu'une horloge interne a été découverte, entre autres, pour les parasites *Trypanosoma brucei* (Rijo-Ferreira, Pinto-Neves et al., 2017), *Plasmodium chabaudi* (Rijo-Ferreira, Acosta-Rodriguez et al., 2020) et *Botrytis cinerea* (Hevia, Canessa et al., 2015). Chez les helminthes, et plus particulièrement les nématodes, des rythmes circadiens dans la locomotion et la résistance au stress hyperosmotique ont été identifiés chez *Caenorhabditis elegans* (organisme libre), sans pour autant clairement identifier les acteurs (Hasegawa, Saigusa & Tamai, 2005).

L'objectif de ce deuxième chapitre va être de déterminer quelle est la nature du rythme journalier d'émission des cercaires des deux chronotypes du plathelminthe *Schistosoma mansoni* ?

Le rythme d'émission des cercaires est génétiquement déterminé (Théron & Combes, 1988). Pour savoir si le rythme est endogène, il doit remplir trois caractéristiques ; être entraînable par un zeitgeber, persister en conditions constantes et avoir une compensation à la température. Dans la littérature, seule l'entraînabilité du rythme et la persistance du rythme en condition constante ont été testées pour la souche Porto Rico de *S. mansoni*. Pour répondre à la question sur la nature du rythme, comparer nos résultats avec la souche Porto Rico et agrandir la connaissance sur le chronotype nocturne de *S. mansoni*, nous avons testé les deux mêmes caractéristiques du rythme circadien (entraînabilité et persistance). Le premier point sera consacré à déterminer si le rythme d'émission suit les variations du synchroniseur (zeitgeber). Le deuxième point étudiera le comportement du rythme d'émission des cercaires en conditions constantes.

### I.2.1. Inversion de la photopériode

Une des caractéristiques du rythme circadien est qu'il est entraînable par les synchroniseurs environnementaux. Pour le rythme d'émission des cercaires, le zeitgeber principal est la photopériode (Asch, 1972 ; Valle, Pellegrino & Alvarenga, 1973). La variation de température peut maintenir le rythme en obscurité continue, en entraînant toutefois une augmentation des cercaires émises en période de nuit subjective par rapport au rythme en condition de photopériode équilibrée (Valle, Pellegrino, & Alvarenga, 1973). L'influence de la température reste moins importante que la photopériode équilibrée comme zeitgeber car la température interne du mollusque varie d'à peine 0,25°C entre obscurité et pleine lumière (Asch, 1972).

Les expériences d'inversion de la photopériode qui ont été menées uniquement sur la souche *S. mansoni* de Porto Rico (chronotype diurne classique) montrent que le rythme d'émission des cercaires s'adapte très rapidement au changement de photopériode (Luttermoser, 1955 ; Asch, 1972 ; Glaudel & Etges, 1973).

L'objectif de ce point, au travers des expériences de chronobiologie, va être de déterminer en combien de temps s'adapte le rythme d'émission des cercaires à l'inversion de photopériode pour les deux chronotypes de *S. mansoni* (Oman).

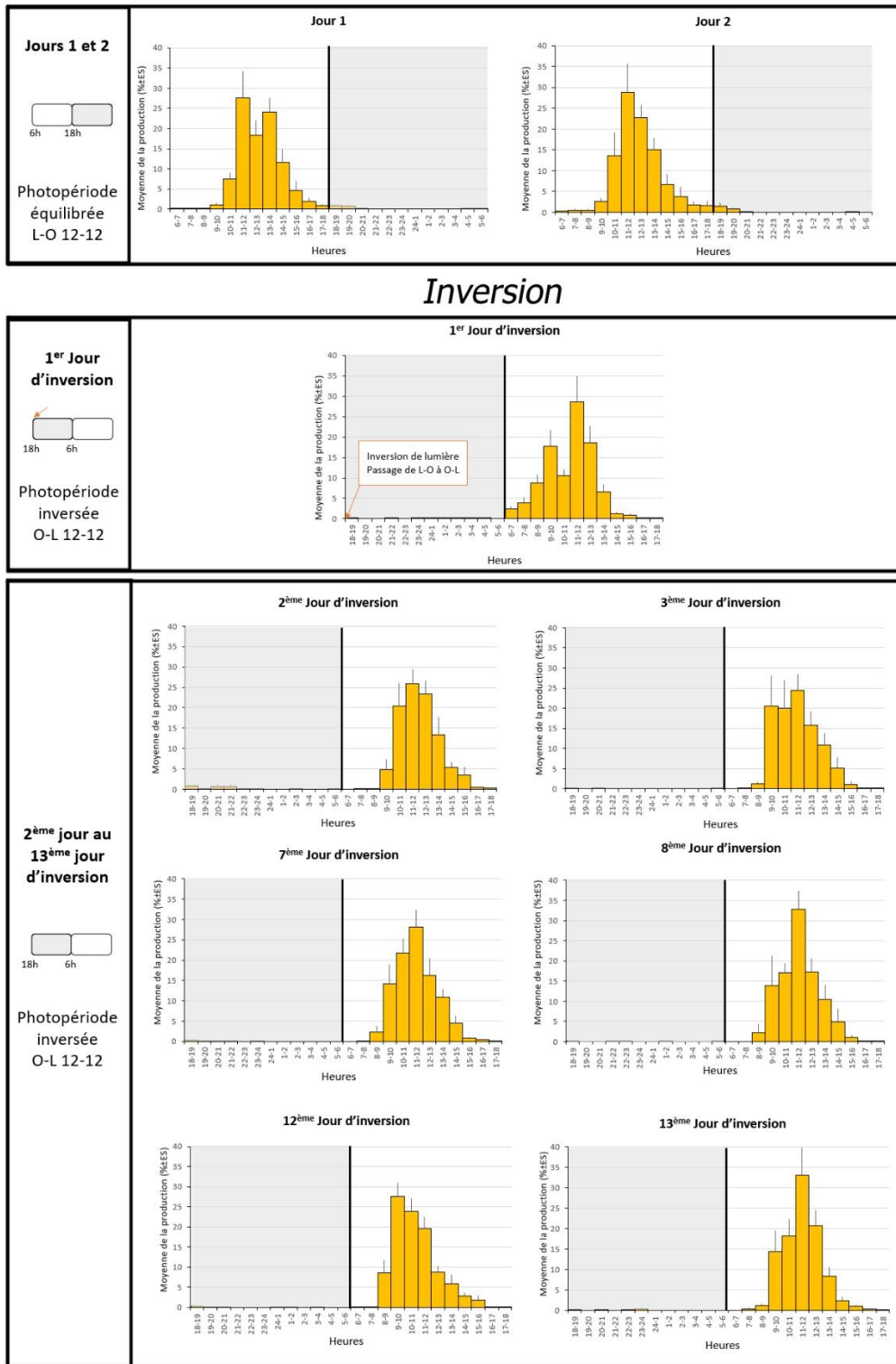
### I.2.1.1. Protocole lumière inversée chronotype diurne

L'expérience de chronobiologie en lumière inversée a été effectuée sur une période de 15 jours et réalisée sur 6 *Biomphalaria glabrata* infestés par le chronotype diurne de *S. mansoni*. Ces jours se divisent en deux phases d'expérimentation. Les jours 1 et 2 représentent les conditions classiques avec une photopériode équilibrée L-O 12-12 ; la lumière s'allumant à 6h du matin et s'éteignant à 18h le soir. A partir du jour 3 et jusqu'à la fin de l'expérimentation la photopériode est inversée. Le jour 3 correspond au 1<sup>er</sup> jour d'inversion de la photopériode.

### I.2.1.2. Résultats chronotype diurne

Les résultats moyens, accompagnés des variations individuelles, sont présentés dans la Figure\_24 et le Tableau\_17. Tous les profils réalisés ont été utilisés car leurs émissions totales sur les 24 heures étaient égales ou supérieures à 20 cercaires. Pour cette expérience, le paramètre du nombre total de cercaires émises est pris en compte.





**Figure 24 : Histogrammes de l'expérience de chronobiologie en lumière inversée pour le chronotype diurne.** L'expérimentation voit se succéder deux conditions : Photopériode équilibrée L-O 12-12 (jours 1 et 2) et photopériode inversée (1<sup>er</sup> au 13<sup>ème</sup> jour d'inversion). L'inversion (flèche rouge) se fait après 24h d'obscurité et l'allumage de la lumière à 18h. En blanc, période de lumière (indiquée de 6h à 18h), en gris, période d'obscurité (indiquée de 18h à 6h).

## Partie I

**Tableau 17 : Paramètres des profils moyens d'émission, accompagnés des variations individuelles, pour les parasites du chronotype diurne de *S. mansoni* de l'expérience inversion de photopériode**

Paramètres	Moyenne ± ES Variations individuelles								
	Jour 1	Jour 2	1 <sup>er</sup> Jour d'inversion	2 <sup>ème</sup> Jour d'inversion	3 <sup>ème</sup> Jour d'inversion	7 <sup>ème</sup> Jour d'inversion	8 <sup>ème</sup> Jour d'inversion	12 <sup>ème</sup> Jour d'inversion	13 <sup>ème</sup> Jour d'inversion
Effectif	6/6	6/6	6/6	6/6	6/6	6/6	6/6	6/6	6/6
Photopériode	L-O	L-O	O-L	O-L	O-L	O-L	O-L	O-L	O-L
Nombre de pics	1 1	1 1	1 1	1 1	1 1	1 1	1 1	1 1	1 1
Heure du pic	11-12h 10-11h à 13-14h	11-12h 10-11h à 13-14h	11-12h 9-10h à 12-13h	11-12h 10-11h à 13-14h	11-12h 10-11h à 12-13h	11-12h 9-10h à 11-12h	11-12h 9-10h à 13-14h	9-10h 9-10h à 11-12h	11-12h 9-10h à 12-13h
Plage d'émission (h)	4,7 ± 0,4 4 à 6	4,5 ± 0,2 4 à 5	5,7 ± 0,2 5 à 6	4,3 ± 0,5 3 à 6	4,8 ± 0,3 4 à 6	5,5 ± 0,2 5 à 6	4,7 ± 0,2 4 à 5	5,3 ± 0,5 4 à 7	4,8 ± 0,2 4 à 5
Intensité de la plage (%)	90,4 ± 1,7 84,9 à 94,6	87,3 ± 3,6 72,5 à 95,7	90,5 ± 2,5 81,5 à 96,6	87,9 ± 4,3 70,0 à 95,8	93,6 ± 0,9 91,3 à 95,7	95,6 ± 1 91,6 à 98	94,6 ± 1 91,0 à 97,1	92,7 ± 1,5 88,0 à 97,6	93,9 ± 1,2 88,6 à 95,9
Asymétrie	+0,8 ± 0,5 -0,6 à +2	+0,8 ± 0,3 0 à +1,4	+1,3 ± 0,3 +0,8 à +2,4	+0,6 ± 0,5 -1,1 à +1,7	+0,8 ± 0,4 -0,1 à +2	+0,3 ± 0,4 -0,8 à +1,3	+0,7 ± 0,3 -0,3 à +1,8	+0,5 ± 0,3 -0,2 à +1,6	+0,1 ± 0,5 -0,9 à +1,9
Aplatissement	-0,2 ± 1,6 -4,5 à +3,9	-0,6 ± 1,4 -5,9 à +2	+1,1 ± 1,2 -1,4 à +5,7	-0,7 ± 1,5 (N=4) -2,9 à +3,5	+0,7 ± 1,5 -2,8 à +4,8	-0,3 ± 0,9 -2,9 à +1,8	-0,4 ± 0,9 -2,6 à +3,2	-0,9 ± 1,2 -2,8 à +3,4	-0,3 ± 1,1 -3,1 à +3,7
Cercaires la nuit (%)	1,9 ± 0,5 0,2 à 3,6	2,5 ± 1,3 0,1 à 7,4	0,2 <sup>a</sup> ± 0,2 0 à 1	2,4 ± 1,1 0,5 à 5,7	0,3 ± 0,2 0 à 1,1	0,7 ± 0,1 0,3 à 1,2	0,3 ± 0,2 0 à 1	0,6 ± 0,3 0 à 2	0,6 ± 0,4 0 à 1,8
Nombre total de cercaires	711,8 ± 156,1 167 à 1194	550,3 ± 155,6 126 à 1009	3361,7 <sup>b</sup> ± 649,6 1785 à 5026	394,3 <sup>c</sup> ± 80,9 194 à 663	302,3 ± 51,4 156 à 488	939,5 <sup>d</sup> ± 169,9 510 à 1478	684,5 ± 176 250 à 1284	1091 ± 360,2 83 à 2191	303,7 ± 83,5 57 à 589

ES : Ecart-standard

L-O : Lumière-Obscurité (12-12) ; O-L : Obscurité-Lumière (12-12)

Tests Wilcoxon Mann Whitney significatifs ( $p$ -value < 0,01) pour les effectifs supérieurs à 4 individus :

Comparaisons avec le jour précédent : <sup>a</sup>  $W = 34, p = 0,009$  ; <sup>b</sup>  $w=0, p = 0,0022$  ; <sup>c</sup>  $W = 36, p = 0,0022$  ; <sup>d</sup>  $W = 0, p = 0,0022$

## Partie I

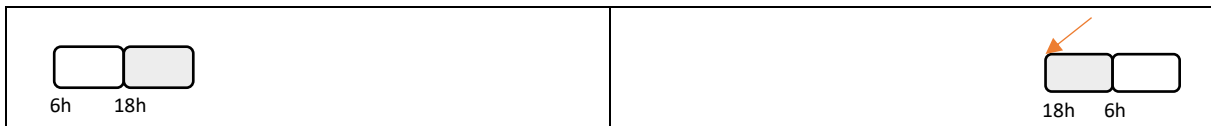
Chacun des paramètres d'un jour a été comparé avec le même paramètre du jour précédent et les résultats sont présentés ci-dessous.

### Jour 1 versus Jour 2



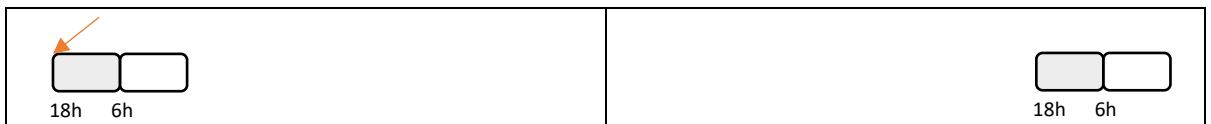
Aucun des paramètres de description n'est significativement différent entre le profil du jour 1 et le profil du jour 2 ( $p > 0,01$ ). Ces profils ont les caractéristiques du chronotype diurne avec la présence d'un seul pic à 11-12h et une plage d'environ 4h avec une intensité moyenne supérieure à 87%. Ils partagent aussi un coefficient d'asymétrie modéré positif et un coefficient d'aplatissement négatif, ainsi qu'une émission nocturne quasi inexistante. Le nombre total de cercaires émises est d'environ 600.

### Jour 2 versus 1<sup>er</sup> jour d'inversion



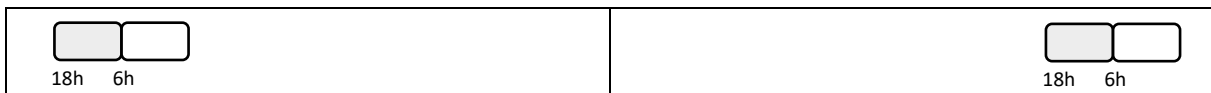
L'inversion de photopériode (flèche rouge) entraîne une différence significative sur deux des neuf paramètres de description qui sont le nombre total de cercaires émises et le pourcentage de cercaires nocturnes ( $p < 0,01$ ). Le profil du 1<sup>er</sup> jour d'inversion a émis en moyenne 6 fois plus de cercaires que celui du jour 2 (respectivement 3362 et 550) et presque aucune cercaire sur la période nocturne. Même si le pourcentage de cercaires nocturnes est significativement plus grand dans le profil du jour 2, il reste faible (moins de 2,5% d'émission). Tout comme les profils des jours 1 et 2, l'heure du pic moyen a lieu 6h après le début de la lumière, soit à 11-12h.

### 1<sup>er</sup> jour d'inversion versus 2<sup>ème</sup> jour d'inversion



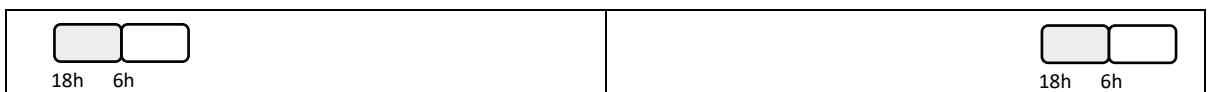
Le nombre de cercaires est le seul paramètre qui est significativement différent avec le 1<sup>er</sup> jour d'inversion ( $p < 0,01$ ). Le profil du 2<sup>ème</sup> jour d'inversion a émis en moyenne 8,5 fois moins de cercaires que celui du 1<sup>er</sup> jour d'inversion (respectivement 394 et 3361). L'heure du pic moyen a lieu 6h après le début de la lumière comme pour les jours précédents, soit à 11-12h.

### 2<sup>ème</sup> jour d'inversion versus 3<sup>ème</sup> jour d'inversion



Aucun des paramètres de description n'est significativement différent entre le 2<sup>ème</sup> jour d'inversion et le 3<sup>ème</sup> jour d'inversion ( $p > 0,01$ ). L'heure du pic moyen se situe 6h après le début de la lumière, comme pour les jours précédents, soit à 11-12h.

### 3<sup>ème</sup> jour d'inversion versus 7<sup>ème</sup> jour d'inversion



Le nombre total de cercaires est le seul paramètre qui est significativement différent avec le 3<sup>ème</sup> jour d'inversion ( $p < 0,01$ ). Le profil du 7<sup>ème</sup> jour d'inversion a émis en moyenne 3 fois plus de cercaires

## Partie I

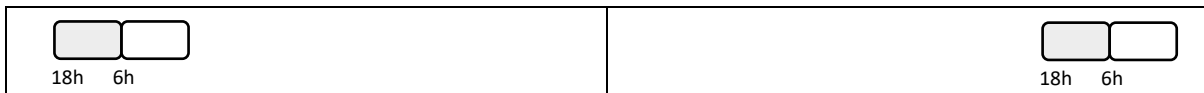
que celui du 3<sup>ème</sup> jour d'inversion (respectivement 939 et 302). L'heure du pic moyen se situe 6h après le début de la lumière comme pour les jours précédents, soit à 11-12h.

### 7<sup>ème</sup> jour d'inversion versus 8<sup>ème</sup> jour d'inversion



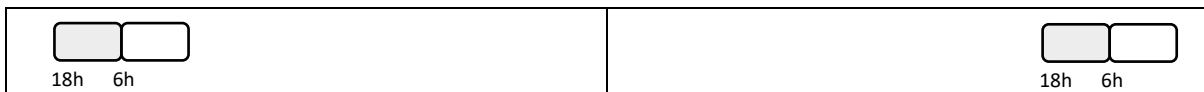
Aucun des paramètres de description n'est significativement différent entre le 7<sup>ème</sup> jour d'inversion et le 8<sup>ème</sup> jour d'inversion ( $p > 0,01$ ). L'heure du pic moyen se situe 6h après le début de la lumière comme pour les jours précédents, soit à 11-12h.

### 8<sup>ème</sup> jour d'inversion versus 12<sup>ème</sup> jour d'inversion



L'heure du pic est le seul paramètre qui est différent avec les profils des autres jours. Le profil du 12<sup>ème</sup> jour d'inversion a eu son pic moyen 2 heures avant celui des jours précédents (respectivement à 9-10h et 11-12h).

### 12<sup>ème</sup> jour d'inversion versus 13<sup>ème</sup> jour d'inversion



L'heure du pic est le seul paramètre qui est différent avec le 12<sup>ème</sup> jour d'inversion. Le profil du 13<sup>ème</sup> jour d'inversion a eu son pic moyen 6h après le début de la lumière comme pour les jours précédents à l'exception du 12<sup>ème</sup> jour d'inversion, soit à 11-12h.

#### I.2.1.3. Discussion et conclusion chronotype diurne

Les jours 1 et 2 représentent le chronotype classique de *S. mansoni*. Le profil du 1<sup>er</sup> jour d'inversion a eu 24h d'obscurité (12h du jour 2 suivies des 12 premières heures du 1<sup>er</sup> jour d'inversion). Durant cette période, significativement moins de cercaires nocturnes ont été émises par rapport aux conditions normales. L'inversion de la photopériode a entraîné une adaptation directe au changement du rythme car le pic moyen d'émission des cercaires a eu lieu 6h après le début de la lumière comme pour les profils en condition classique (Jours 1 et 2).

Cette inversion a aussi eu des effets sur le nombre total de cercaires puisque 6 fois plus de cercaires ont été émises le 1<sup>er</sup> jour d'inversion par rapport au jour 2 (L-O). Le 2<sup>ème</sup> jour d'inversion diffère du 1<sup>er</sup> jour d'inversion par le nombre total de cercaires qui est cette fois-ci 8,5 fois inférieur à celui de la veille. Ces différentes variations en termes de nombre de cercaires peuvent s'expliquer par le fait que les cercaires prêtes pour le 1<sup>er</sup> jour d'inversion ont dû attendre 12h supplémentaires le retour de la lumière. Ce temps d'attente supplémentaire aurait induit un stress qui aurait poussé les parasites à émettre un maximum de cercaires (celles prêtes pour l'émission du jour et celles en préparation pour les jours suivants). Ceci est appuyé par le fait que le profil du 2<sup>ème</sup> jour d'inversion (O-L) voit une chute drastique du nombre total de cercaires émises. Toutefois cette production cercarienne du 2<sup>ème</sup> jour d'inversion n'est pas significativement différente des jours 1 et 2 de photopériode classique (Jour 1 :  $W = 28$ ,  $p = 0,13$  et Jour 2  $W = 22$ ,  $p = 0,59$ ). Le profil du 3<sup>ème</sup> jour d'inversion (O-L) montre une stabilisation du rythme et de la production cercarienne puisqu'aucun paramètre ne diffère avec le 2<sup>ème</sup> jour d'inversion (O-L).

## Partie I

A partir du 7<sup>ème</sup> jour d'inversion, le nombre total de cercaires connaît une augmentation significative ce qui illustre que les parasites se sont adaptés et remis du stress d'inversion de la photopériode. Ce nombre total de cercaires n'est pas différent significativement des jours 1, 2 (L-O) et 2<sup>ème</sup> jour d'inversion (Jour 1 :  $W = 13, p = 0,49$  ; Jour 2  $W = 8, p = 0,13$  ; 2<sup>ème</sup> jour d'inversion  $W = 3, p = 0,02$ ). Le nombre total de cercaires reste par la suite stable jusqu'à la fin de l'expérience ( $p > 0.01$ ). L'heure du pic moyen se situe 6h après le début de la lumière pour tous les jours de l'expérience à l'exception du 12<sup>ème</sup> jour d'inversion (9-10h). Cet écart de deux heures avec le 12<sup>ème</sup> jour d'inversion n'est pas important car des écarts similaires sont vus au sein des profils de même jour.

En conclusion, le rythme d'émission des cercaires du chronotype diurne s'adapte immédiatement à l'inversion de photopériode lorsque celle-ci est déclenchée après 24h d'obscurité.

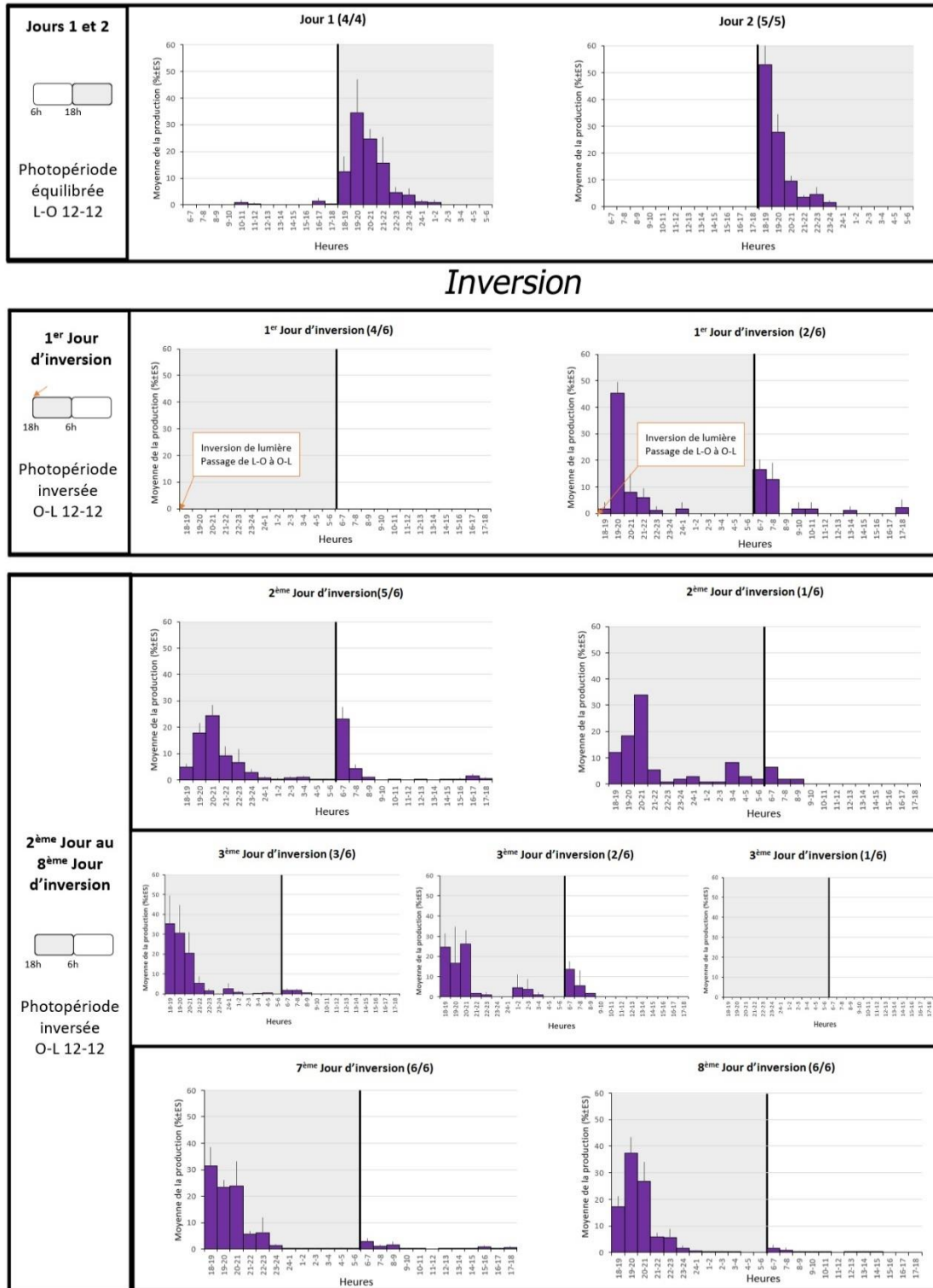
### I.2.1.4. Protocole lumière inversée chronotype nocturne

L'expérience de chronobiologie en lumière inversée a été effectuée sur une période de 10 jours et réalisée sur 6 *Biomphalaria pfeifferi* infestés par le chronotype nocturne de *S. mansoni*. Ces jours se divisent en deux phases d'expérimentation. Les jours 1 et 2 représentent les conditions classiques avec une photopériode équilibrée LO 12-12 ; la lumière s'allumant à 6h du matin et s'éteignant à 18h le soir. A partir du jour 3 et jusqu'à la fin de l'expérimentation la photopériode est inversée. Le jour 3 correspond au 1<sup>er</sup> jour d'inversion de la photopériode.

### I.2.1.5. Résultats chronotype nocturne

Les résultats moyens, accompagnés des variations individuelles, sont présentés dans la Figure\_25 et le Tableau\_18. Pour les jours 1 et 2, seuls les profils avec une émission supérieure à 20 cercaires sont pris en compte. Pour les autres jours, les profils avec une émission inférieure à 20 cercaires sont pris en compte pour la caractérisation des comportements mais sont exclus des analyses statistiques.

Partie I



*Inversion*

Figure 25 : Histogramme de l'expérience de chronobiologie en lumière inversée pour le chronotype nocturne. L'expérimentation voit se succéder deux conditions : Photopériode équilibrée L-O 12-12 (jours 1 et 2) et photopériode inversée (1<sup>er</sup> au 8<sup>ème</sup> jour d'inversion). L'inversion (flèche rouge) se fait après 24h d'obscurité et l'allumage de la lumière à 18h. En blanc, période de lumière (indiquée de 6h à 18h), en gris, période d'obscurité (indiquée de 18h à 6h).

Partie I

**Tableau 18 : Paramètres des profils moyens d'émission, accompagnés des variations individuelles, pour les parasites du chronotype nocturne de *S. mansoni* de l'expérience inversion de photopériode**

Paramètres	Moyenne ± ES Variations individuelles													
	Jour 1	Jour 2	1 <sup>er</sup> Jour d'inversion		2 <sup>ème</sup> Jour d'inversion			3 <sup>ème</sup> Jour d'inversion			7 <sup>ème</sup> Jour d'inversion	8 <sup>ème</sup> Jour d'inversion		
Effectif	4/4	5/5	4/6	2/6		5/6		1/6	3/6	2/6		1/6	6/6	6/6
Photopériode	L-O	L-O	O-L		O-L			O-L			O-L	O-L		
Nombre de pics	1 1	1 1		2 2		2 2		1 1	1 1	2 2			1 1	1 1
Heure du pic	19-20h 19-20h à 20-21h	18-19h 18-19h à 19-20h		19-20h 19-20h	6-7h et 7-8h 6-7h à 7-8h	20-21h 20-21h à 6-7h	6-7h 19-20h à 6-7h	20-21h 20-21h	18-19h 18-19h à 22-23h	18-19h et 20-21h 18-19h à 20-21h	6-7h 6-7h		18-19h 18-19h à 22-23h	19-20h 19-20h à 20-21h
Plage d'émission (h)	4,5 ± 0,6 4 à 6	3,6 ± 0,2 3 à 4		3,5 ± 0,7 3 à 4	3 3	4 ± 1,2 3 à 6	3 3	4 4	4,0 ± 0,6 2 à 6	3,5 ± 0,7 3 à 4	3 ± 0 3		4,0 ± 0,6 2 à 6	4,0 ± 0,9 3 à 5
Intensité de la plage (%)	91,7 ± 2,3 86,2 à 95,2	89,8 ± 3,4 78,7 à 97,8		59,5 ± 6,1 55,2 à 63,8	29,3 ± 2,4 27,7 à 31,0	61,4 ± 5,0 46,6 à 72,6	26,6 ± 5 15,1 à 41,7	69,7 69,7	88,5 ± 4 71,8 à 96,3	68,3 ± 16,9 56,4 à 80,3	19,1 ± 3,8 16,4 à 21,8		88,5 ± 4 71,8 à 96,3	91,3 ± 3,4 77,6 à 98,9
Asymétrie	+0,6 ± 0,6 -0,8 à +1,8	+0,8 ± 0,4 -0,1 à +1,6		+1,6 ± 0,2 +1,5 à +1,7	-0,6 ± 1,3 -1,5 à +0,3	+0,6 ± 0,4 -0,3 à +1,4	+0,9 ± 0,7 -1,4 à +1,7	+1 +1	+0,7 ± 0,3 -0,3 à +1,5	-0,6 ± 1,6 -1,7 à +0,5	0 ± 2,4 -1,7 à +1,7		+0,7 ± 0,3 -0,3 à +1,5	+0,9 ± 0,3 0 à +1,6
Aplatissement	+0,5 ± 1,5 -2,7 à +3,5	0 ± 3,3 (N=3) -5,4 à +2,9		2,6 (N=1) 2,6	(N=0)	-2,1 ± 2,8 (N=3) -5,5 à +2,3	(N=0)	+0,9 +0,9	0 ± +1,1 (N=5) -2,9 à +2,2	-3,1 (N=1) -3,1	(N=0)		0 ± +1,1 (N=5) -2,9 à +2,2	+0,5 ± 0,9 (N=4) -1,1 à +2,4
Cercaires la nuit (%)	97,2 ± 0,8 95,5 à 98,5	100 ± 0 100		64 ± 2,7 62,1 à 66		68,6 ± 7,1 47,5 à 83,6		89,9 89,9	96,3 ± 2,1 92,9 à 98,8	79,2 ± 4 76,4 à 82			92,6 ± 3,1 81,3 à 99,4	96,3 ± 2 87,9 à 100

ES : Ecart-standard

L-O : Lumière-Obscurité (12-12) ; O-L : Obscurité-Lumière (12-12)

Tests Wilcoxon-Mann-Whitney significatifs ( $p$ -value < 0,01) pour les effectifs supérieurs à 4 individus : comparaison avec le jour précédent.

## Partie I

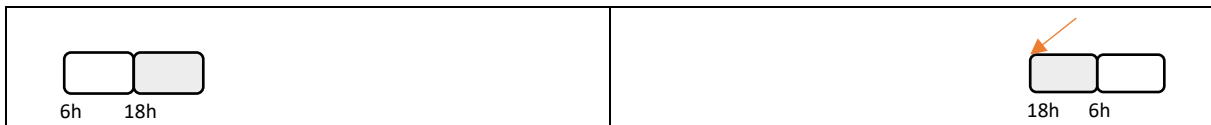
Chacun des paramètres d'un jour a été comparé avec le même paramètre du jour précédent et les résultats sont présentés ci-dessous.

### Jour 1 versus Jour 2



Aucun des paramètres de description n'est significativement différent entre le profil du jour 1 et le profil du jour 2 ( $p > 0,01$ ). Ces profils ont les caractéristiques du chronotype nocturne avec la présence d'un seul pic et une plage d'environ 4h avec une intensité moyenne aux alentours de 90%. Ils partagent aussi un coefficient d'asymétrie modéré positif ainsi qu'une émission nocturne dépassant les 97%.

### Jour 2 versus 1<sup>er</sup> jour d'inversion



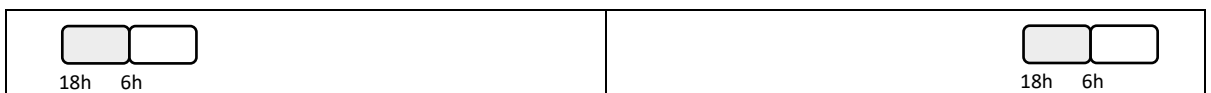
L'inversion de photopériode entraîne une différence de comportement au sein des profils du 1<sup>er</sup> jour d'inversion. Sur les 6 profils, 4 ont émis moins de 20 cercaires. Les deux profils restants sont multimodaux. Le pic avec la plus grande intensité est à 19-20h (obscurité) et le second pic a lieu à 6-7h ou 7-8h (lumière) en fonction du profil. Aucun paramètre des deux profils multimodaux du 1<sup>er</sup> jour d'inversion n'est significativement différent avec le profil du jour 2 ( $p > 0,01$ ).

### 1<sup>er</sup> jour d'inversion versus 2<sup>ème</sup> jour d'inversion



Sur les 6 profils du 2<sup>ème</sup> jour d'inversion, tous ont émis plus de 20 cercaires avec deux comportements qui se distinguent. La majorité (5 sur 6) a une émission multimodale avec le pic principal moyen à 20-21h et le second pic à 6-7h en période de lumière. Le profil restant du 2<sup>ème</sup> jour d'inversion est unimodal avec son pic à 20-21h. Cette heure de pic principal est décalée d'une à deux heures par rapport aux jours 1 et 2 mais est comprise dans leur variabilité individuelle. Aucun paramètre des profils unimodal et multimodal du 2<sup>ème</sup> jour d'inversion n'est significativement différent du profil multimodal du 1<sup>er</sup> jour d'inversion ( $p > 0,01$ ).

### 2<sup>ème</sup> jour d'inversion versus 3<sup>ème</sup> jour d'inversion



Sur les six profils du 3<sup>ème</sup> jour d'inversion, trois comportements différents sont observés. Trois profils sur six sont unimodaux avec le pic à 18-19h. Deux profils sur six sont encore multimodaux avec le pic principal à 18-19h ou 20-21h en fonction du profil et le pic secondaire à 6-7h. Le dernier profil a émis moins de 20 cercaires. Aucun des paramètres de description n'est significativement différent entre le 2<sup>ème</sup> jour d'inversion et le 3<sup>ème</sup> jour d'inversion ( $p > 0,01$ ).

### 3<sup>ème</sup> jour d'inversion versus 7<sup>ème</sup> jour d'inversion

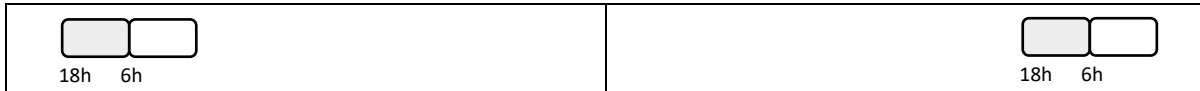




## Partie I

Les six profils du 7<sup>ème</sup> jour d'inversion sont unimodaux avec le pic moyen à 18-19h. Aucun des paramètres de description n'est significativement différent entre le 7<sup>ème</sup> jour d'inversion et le 3<sup>ème</sup> jour d'inversion ( $p > 0,01$ ) ni entre le jour 2 de photopériode classique et le 7<sup>ème</sup> jour d'inversion ( $p > 0,01$ ). Tout comme le jour 2, l'heure du pic moyen du 7<sup>ème</sup> jour d'inversion se fait dès le début de l'obscurité (18-19h).

### 7<sup>ème</sup> jour d'inversion versus 8<sup>ème</sup> jour d'inversion



Tout comme le 7<sup>ème</sup> jour d'inversion, les six profils du 8<sup>ème</sup> jour sont unimodaux avec le pic moyen à 19-20h. Aucun des paramètres de description n'est significativement différent entre le 7<sup>ème</sup> jour d'inversion et le 8<sup>ème</sup> jour d'inversion ( $p > 0,01$ ) ni entre le jour 2 de photopériode classique et le 8<sup>ème</sup> jour d'inversion ( $p > 0,01$ ). Tout comme le jour 2, l'heure du pic moyen du 8<sup>ème</sup> jour d'inversion se fait une heure après le début de l'obscurité (19h-20h).

#### I.2.1.6. Discussion et conclusion chronotype nocturne

Les jours 1 et 2 représentent le chronotype classique de *S. mansoni* nocturne. Les profils du 1<sup>er</sup> jour d'inversion ont eu 24 heures d'obscurité (12h du jour 2 suivies des 12 premières heures du 1<sup>er</sup> jour d'inversion).

L'inversion de la photopériode a mis en lumière des différences de comportement dans l'adaptation au changement de photopériode. Le premier comportement suivi par la majorité des profils du 1<sup>er</sup> jour d'inversion est une réduction de l'émission en dessous de la limite des 20 cercaires. Tandis que le second comportement est l'apparition d'un profil d'émission multimodal. Le pic principal (intensité aux alentours de 60%) s'adapte immédiatement au changement puisqu'il apparaît dès la deuxième heure d'obscurité du 1<sup>er</sup> jour d'inversion (19-20h). Le second pic a lieu dans les deux premières heures de lumière qui correspond à l'heure du pic de la veille (Jour 2, L-O). La présence de ces deux pics illustre que la majorité de l'émission fait la bascule mais que l'adaptation n'est pas complète car des cercaires ont été émises durant la période lumineuse à l'heure du pic en photopériode classique. A partir du 2<sup>ème</sup> jour d'inversion tous les profils émettent de nouveau plus de 20 cercaires, mais la majorité reste multimodale. Il faut attendre le 7<sup>ème</sup> jour d'inversion pour que tous les profils soient de nouveau unimodaux comme les jours 1 et 2 de photopériode classique. A l'exception de l'heure du pic pour les profils multimodaux, aucun des paramètres de description des profils n'est significativement différent lorsque l'on effectue une comparaison avec le jour précédent. Les tests impliquant les 1<sup>er</sup> et 3<sup>ème</sup> jours d'inversion peuvent être discutés car la taille des échantillons est très petite ( $\leq 3$ ). Il peut toutefois être dit que l'inversion de photopériode induit dans la majorité des cas une baisse de l'émission cercarienne. Cela peut être apparenté à un statut d'attente pour le retour d'une alternance jour/nuit puisque dès le 2<sup>ème</sup> jour d'inversion tous les profils sont de nouveau au-dessus de la limite. L'adaptation pour les profils dont l'émission reste au-dessus de la limite passe par un stade intermédiaire de deux pics qui ont lieu à la nouvelle et à l'ancienne heure du pic. Le fait qu'en photopériode classique le pic se passe à la transition jour/nuit rend plus difficilement interprétable la nature du pic diurne à l'ancienne heure d'émission. Ce pic diurne peut être soit assimilé à un pic résiduel qui n'a pas encore fait la bascule face à l'inversion, soit à une réponse mécanique à la lumière due à la perturbation.

En conclusion, le rythme d'émission des cercaires du chronotype nocturne commence à s'adapter immédiatement à l'inversion de photopériode lorsque celle-ci est déclenchée après 24h d'obscurité. Il faut toutefois attendre le 7<sup>ème</sup> jour d'inversion pour que l'inversion concerne l'ensemble des individus.

### I.2.1.7. Comparaison des chronotypes diurne et nocturne avec inversion de la photopériode

Les chronotypes diurne et nocturne de *S. mansoni* ont des façons différentes de réagir à l'inversion de photopériode induite par 24h d'obscurité. Le chronotype diurne s'adapte immédiatement et totalement à l'inversion de photopériode tandis que le chronotype nocturne doit soit passer par 24h de pause puis une émission bimodale (aux heures de l'ancien et du nouveau pic), soit directement être bimodal pour finalement s'adapter complètement à la nouvelle photopériode. Cette différence entre chronotypes peut s'expliquer par la place du pic nocturne dans le nyctémère et par l'inversion de 12 heures qui peut représenter une grande perturbation pour les organismes.

Une autre expérience d'inversion de 3 jours (résultats non montrés) a été menée avec cette fois-ci l'inversion induite après 24h de lumière. L'adaptation du chronotype nocturne avec une étape intermédiaire bimodale (ancienne et nouvelle heures du pic) reste inchangée. Pour le chronotype diurne, l'adaptation est totale dès le 1<sup>er</sup> jour d'inversion. Il est intéressant de noter qu'aux 2<sup>ème</sup> et 3<sup>ème</sup> jour d'inversion, il y a une émission cercarienne à 6h du matin qui correspond au début de l'obscurité (O-L). Cette émission à la transition jour/nuit observée chez le diurne pourrait signifier que l'émission observée chez le nocturne est plus une réponse à la transition jour/nuit qu'un pic qui n'est pas encore adapté à la nouvelle photopériode.

### I.2.2. Zeitgeber constant

Une autre des caractéristiques d'un rythme circadien est qu'il persiste en l'absence de zeitgeber. Privé du zeitgeber, le rythme adopte une période dite de libre cours (« free running period ») d'approximativement 24 heures (entre 20 et 28 heures) dictée par l'horloge interne. Si en conditions continues, la période de libre cours est supérieure à 20 heures et inférieure à 24h, l'horloge interne avance par rapport à un rythme de 24h (avancement du rythme). A l'inverse, si la période de libre cours est supérieure à 24 heures et inférieure à 28h, l'horloge interne retarde par rapport à un rythme de 24h (retard du rythme). Dans le cas de *S. mansoni*, il s'agit de voir si le rythme d'émission des cercaires persiste dans un environnement à photopériode continue (lumière ou obscurité) et température constante. Sa période de libre cours correspond à la période qu'il y a entre deux pics d'émission. Selon la littérature, des expériences ont été menées uniquement sur le chronotype diurne classique de *S. mansoni*. Que ce soit sous luminosité continue ou sous obscurité continue, les schistosomes ont perdu leur rythme d'émission. Cette perte de rythme intervient soit immédiatement après (Williams, Wessels & Gilbertson, 1984), soit deux jours après le changement de photopériode (Valle, Pellegrino & Alvarenga, 1973). Dans les deux cas, le rythme d'émission des cercaires revient à la normale dès le retour d'un cycle jour/nuit.

L'objectif de ce point, au travers des expériences de chronobiologie, va être de déterminer si le rythme d'émission des cercaires persiste en lumière continue pour le chronotype diurne et en obscurité continue pour le chronotype nocturne de *S. mansoni* (Oman). S'il y a persistance du rythme, le second objectif sera de déterminer la période de libre cours du rythme d'émission des cercaires, et ainsi si l'horloge interne avance (20h > période de libre cours > 24h) ou retarde (24h > période de libre cours > 28h) par rapport aux rythmes de 24h imposés par le cycle jour/nuit.

## Partie I

### I.2.2.1. Chronotype diurne

#### I.2.2.1.1. Protocole lumière continue

L'expérience de chronobiologie en lumière continue a été effectuée sur une période de 26 jours et réalisée sur 6 *Biomphalaria glabrata* infestés par le chronotype diurne de *S. mansoni*. Ces jours se divisent en trois phases d'expérimentation. Les jours 1 et 2 représentent les conditions classiques, avec une photopériode équilibrée L-O 12-12 ; la lumière s'allumant à 6h du matin et s'éteignant à 18h le soir. A partir du jour 3, la lumière est laissée allumée à 18h. Le jour 3 correspond au 1<sup>er</sup> jour de lumière continue et le jour 18 au 16<sup>ème</sup> et dernier jour de lumière continue L-L 12-12. A partir du jour 19 et jusqu'à la fin de l'expérimentation, la lumière est arrêtée à 18h pour un retour de la photopériode équilibrée L-O 12-12 identique aux conditions classiques.

#### I.2.2.1.2. Résultats. Expérience de lumière continue

Les résultats moyens, accompagnés des variations individuelles, sont présentés dans la Figure\_26 et le Tableau\_19 pour la partie de lumière continue et la Figure\_27 et le Tableau\_20 pour le retour de la photopériode équilibrée. Pour les jours 1 et 2, seuls les profils avec une émission supérieure à 20 cercaires sont pris en compte. Pour les autres jours, les profils avec une émission inférieure à 20 cercaires sont pris en compte pour la caractérisation des comportements mais sont exclus des analyses statistiques. Pour cette expérience, le paramètre du nombre total de cercaires émises est pris en compte. Pour les jours de lumière continue (L-L), le pourcentage de cercaires nocturnes correspondra au pourcentage de cercaires émises durant les douze dernières heures de la journée, c'est-à-dire pendant la nuit subjective.

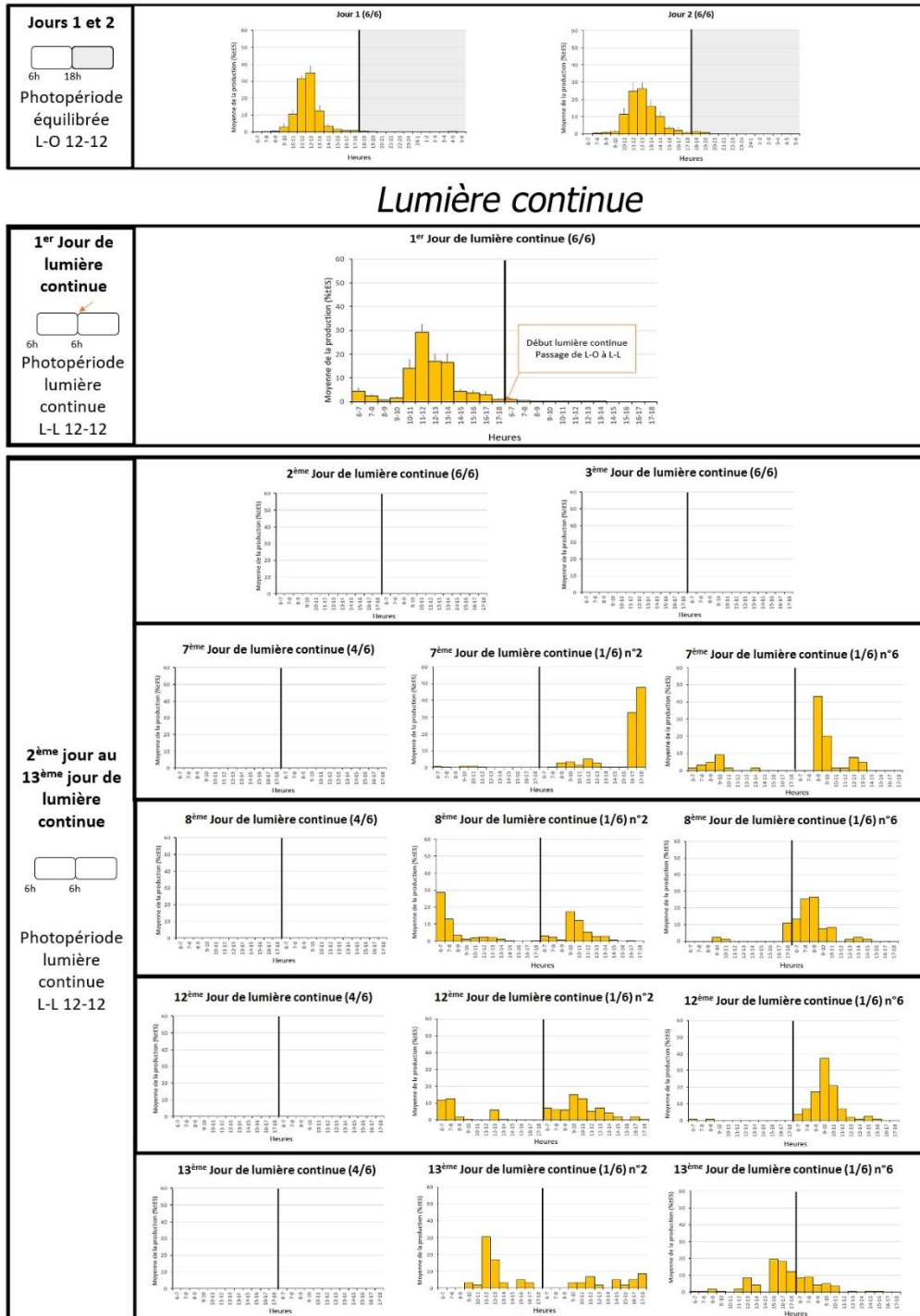


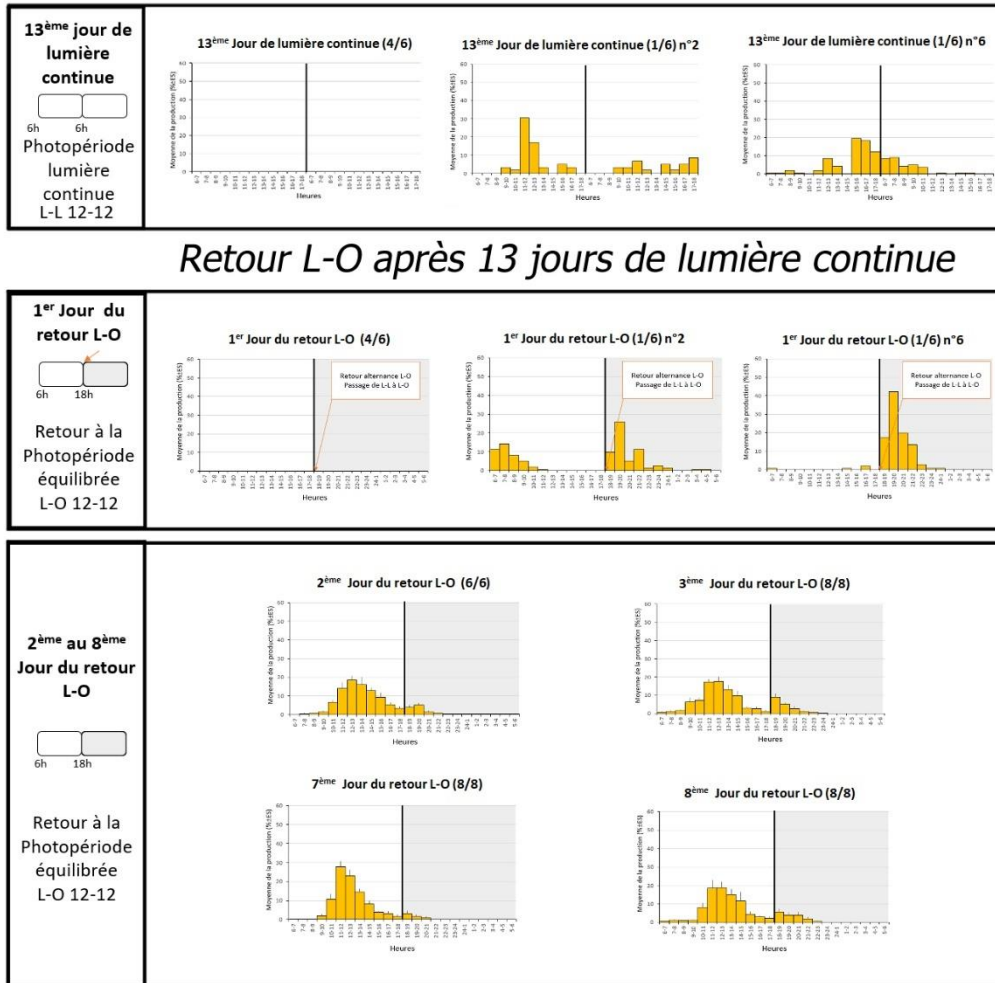
Figure 26 : Histogramme de l'expérience de chronobiologie en lumière continue pour le chronotype diurne (1/2), phase lumière continue. L'expérimentation voit se succéder trois conditions : Photopériode équilibrée L-O 12-12 (jours 1 et 2), lumière continue (1<sup>er</sup> au 13<sup>ème</sup> jour) et le retour à la photopériode équilibrée en figure 27 (1<sup>er</sup> au 8<sup>ème</sup> jour). Le début de la lumière continue se fait à partir de 18h. En blanc, période de lumière (indiquée de 6h à 18h) ; en gris, période d'obscurité (indiquée de 18h à 6h).

Partie I

**Tableau 19 : Paramètres des profils moyens d'émission, accompagnés des variations individuelles, pour les parasites du chronotype diurne de *S. mansoni* de l'expérience de lumière continue (1/2)**

Paramètres	Moyenne ± ES Variations individuelles																
	Jour 1	Jour 2	1 <sup>er</sup> jour de L-L	2 <sup>ème</sup> Jour de L-L	3 <sup>ème</sup> Jour de L-L	7 <sup>ème</sup> Jour de L-L			8 <sup>ème</sup> Jour de L-L			12 <sup>ème</sup> Jour de L-L			13 <sup>ème</sup> Jour de L-L		
Effectif	6/6	6/6	6/6	6/6	6/6	4/6	1/6	1/6	4/6	1/6	1/6	4/6	1/6	1/6	4/6	1/6	1/6
Photopériode	L-O	L-O	L-L	L-L	L-L	L-L			L-L			L-L			L-L		
Nombre de pics	1 1	1 1	1 1				1	1		1	1						
Heure du pic	12-13h 11-12h à 12-13h	12-13h 11-12h à 13-14h	11-12h 11-12h à 13-14h				17-18 (pm)	8-9 (pm)		6-7 (am)	8-9 (pm)		9-10 (pm)	9-10 (pm)		11-12(a m)	15-16 (am)
Plage d'émission (h)	4,0 ± 0,3 3 à 5	5,2 ± 0,4 4 à 7	4,3 ± 0,2 4 à 5				2	2		2	6		7	5		2	5
Intensité de la plage (%)	91,4 ± 1,0 88,7 à 93,9	92,1 ± 2,2 83,6 à 97,6	79,6 ± 3,5 70 à 92,8				80,6	63,1		41,6	91,6		58,8	88,8		47,5	67,3
Asymétrie	-0,3 ± 0,4 -1,3 à +0,6	+0,4 ± 0,3 -0,6 à +1	+0,3 ± 0,4 -1,2 à +1,7				N=0	N=0		N=0	0,7		+1,2	+1		N=0	+0,3
Aplatissement	-0,8 ± 0,8 (N=5) -2,4 à +1,3	0 ± 0,7 -2,3 à +1,7	+0,3 ± 0,9 -2,7 à +2,8				N=0	N=0		N=0	-1,9		-0,1	+0,8		N=0	-2,9
Cercaires la nuit subjective (%)	0,7 ± 0,3 0 à 1,3	2,6 ± 1,0 0,4 à 6,9	2,1 ± 0,8 0,7 à 4,6				97,2	78,5		46,5	85,5		67	98,3		35,6	32,1
Nombre total de cercaires	749,3 ± 215,4 274 à 1589	679,5 ± 106,2 464 à 986	830,2 ± 110,4 671 à 1313	3 <sup>a</sup> ± 1,02 0 à 6	1,17 ± 0,87 0 à 5	2,25 ± 0,87 0 à 3	607	65	7,5 ± 2,56 4 à 14	649	83	8,5 ± 2,0 5 à 12	294	116	3,8 ± 1,6 1 à 7	59	165
N° mollusque							N°2	N°6		N°2	N°6		N°2	N°6		N°2	N°6

ES : Ecart-standard ; L-O : Lumière-Obscurité (12-12) ; O-L : Obscurité-Lumière (12-12) ; Tests Wilcoxon Mann Whitney significatifs ( $p$ -value <0,01) pour les effectifs supérieurs à 4 individus : Comparaison avec le jour précédent ( $p$ <0,01) : a W = 36,  $p$  = 0,005



**Figure 27 : Histogramme de l'expérience de chronobiologie en lumière continue pour le chronotype diurne (2/2) : phase retour de la photopériode équilibrée.** L'expérimentation voit se succéder trois conditions : Photopériode équilibrée L-O 12-12 (jours 1 et 2) en figure 26, lumière continue en figure 26 (1<sup>er</sup> au 13<sup>ème</sup> jour) et le retour à la photopériode équilibrée (1<sup>er</sup> au 8<sup>ème</sup> jour). Le retour de la photopériode équilibrée s'opère avec le retour de l'obscurité à 18h. En blanc, période de lumière (indiquée de 6h à 18h) ; en gris, période d'obscurité (indiquée de 18h à 6h).

## Partie I

**Tableau 20 : Paramètres des profils moyens d'émission, accompagnés des variations individuelles, pour les parasites du chronotype diurne de *S. mansoni* de l'expérience de lumière continue (2/2)**

Paramètres	Moyenne ± ES Variations individuelles						
	1 <sup>er</sup> jour du retour L-O			2 <sup>ème</sup> jour du retour L-O	3 <sup>ème</sup> jour du retour L-O	7 <sup>ème</sup> jour du retour L-O	8 <sup>ème</sup> jour du retour L-O
Effectif	4/6	1/6	1/6	6/6	8/8	8/8	8/8
Photopériode	L-O			L-O	L-O	L-O	L-O
Nombre de pics				1 1	1 1	1 1	1 1
Heure du pic		19-20h	19-20h	12-13h 11-12h à 13-14h	12-13h 11-12h à 13-14h	11-12h 11-12h à 13-14h	12-13h 11-12h à 14-15h
Plage d'émission (h)		4	4	6,0 ± 0,3 5 à 7	4,8 ± 0,3 4 à 6	5,3 ± 0,4 3 à 7	4,9 ± 0,6 3 à 8
Intensité de la plage (%)		52,0	92,9	79,7 ± 3,6 70,6 à 89,4	67,8 ± 3,9 50 à 80,5	86,6 <sup>b</sup> ± 3,4 68,4 à 97,7	74,9 ± 7,6 38,9 à 95,2
Asymétrie		+1,5	+1,8	+0,5 ± 0,4 -0,6 à +1,8	0 ± 0,3 -0,9 à +1,6	+0,6 ± 0,2 -0,4 à +1,4	+0,9 ± 0,6 -2,1 à +2,5
Aplatissement		+2,7	+3,2	+0,2 ± 0,9 -1,7 à +3,9	-0,6 ± 0,7 -2,4 à +2,8	-0,4 ± 0,9 (N=7) -3,0 à +2,9	+2,6 ± 1,3 (N=7) -3,3 à +6,7
Cercaires la nuit subjective (%)		58,7	96,8	11,5 ± 3,2 4,7 à 21,3	18,5 ± 2,9 7,5 à 33,3	5,7 ± 3,2 0 à 25,3	15,2 ± 6,2 0 à 41,7
Nombre total de cercaires	7,3 ± 4,8 0 à 19	196	156	1940,3 ± 481,1 856 à 3755	151,8 <sup>a</sup> ± 49 26 à 324	227,8 ± 66 38 à 519	444,9 ± 109 166 à 948
N° mollusque		N°2	N°6				

ES : Ecart-standard ; L-O : Lumière-Obscurité (12-12) ; O-L : Obscurité-Lumière (12-12) ; Tests Wilcoxon Mann Whitney significatifs ( $p$ -value<0,01) pour les effectifs supérieurs à 4 individus : Comparaison avec le jour précédent : <sup>a</sup>W = 48,  $p$  = 0,0007, <sup>b</sup> W =5,  $p$  = 0,003.

## Partie I

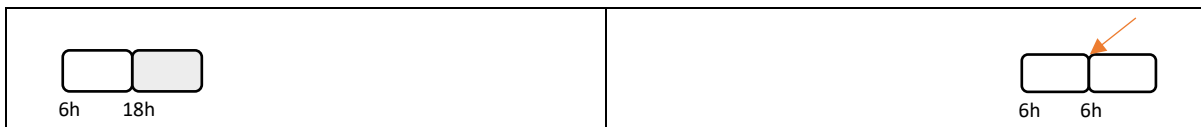
Chacun des paramètres d'un jour a été comparé avec le même paramètre du jour précédent et les résultats sont présentés ci-dessous.

### Jour 1 versus Jour 2



Aucun des paramètres de description n'est significativement différent entre le profil du jour 1 et le profil du jour 2 ( $p > 0,01$ ). Ces profils ont les caractéristiques du chronotype diurne avec la présence d'un seul pic à 12-13h et une plage entre 4h et 5h avec une intensité moyenne de la plage supérieure à 90%. Ils partagent aussi un coefficient d'asymétrie modéré (positif ou négatif) ainsi qu'une émission nocturne quasi inexistante. Le nombre total de cercaires émises est d'environ 700.

### Jour 2 versus 1<sup>er</sup> jour de lumière continue



Le régime de lumière continue débute à la seconde moitié de la 1<sup>ère</sup> journée de lumière continue. Les 12 premières heures du 1<sup>er</sup> jour de lumière continue ne sont pas affectées par le changement de photopériode car elles correspondent au début de jour classique. Aucune émission significative n'est observée dans la seconde moitié du 1<sup>er</sup> jour de lumière continue. Ainsi, aucun des paramètres de description n'est significativement différent entre le profil du jour 2 et le profil du 1<sup>er</sup> jour de la lumière continue ( $p > 0,01$ ). La différence d'une heure pour l'heure du pic n'est pas significative entre les profils du jour 2 et du 1<sup>er</sup> jour de lumière continue puisqu'elle est comprise dans leur variabilité individuelle ( $p > 0,01$ ).

### 1<sup>er</sup> jour de lumière continue versus 2<sup>ème</sup> jour de lumière continue



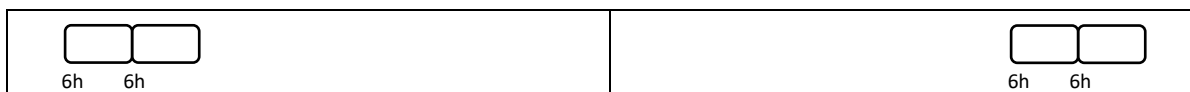
Après 24h de lumière continue, il y a une perte totale de rythme et une émission quasi inexistante. Sur les 6 profils du 2<sup>ème</sup> jour de lumière continue, aucun n'a émis plus de 6 cercaires. Ce changement de comportement est d'autant plus important puisqu'en 24h de lumière continue, le nombre total moyen de cercaires émises est passé de 830 à 3.

### 2<sup>ème</sup> jour de lumière continue versus 3<sup>ème</sup> jour de lumière continue



Aucun des 6 profils du 3<sup>ème</sup> jour de lumière continue n'a émis plus de 5 cercaires. Le nombre total de cercaires émises n'est pas significativement différent entre le 2<sup>ème</sup> jour de lumière continue et le 3<sup>ème</sup> jour de lumière continue ( $p > 0,01$ ).

### 3<sup>ème</sup> jour de lumière continue versus 7<sup>ème</sup> jour de lumière continue



A partir du 7<sup>ème</sup> jour, deux comportements se distinguent entre les 6 profils. Ceux qui ont émis moins de 20 cercaires (4/6) et ceux qui émettent plus de 20 cercaires (2/6). Cette différence de



## Partie I

comportement demeure jusqu'au dernier jour de la lumière continue chez les mêmes individus. Les 4 profils du 7<sup>ème</sup> jour de lumière continue qui ont émis moins de 20 cercaires sont les mêmes individus que ceux du 13<sup>ème</sup> jour de lumière continue qui ont émis moins de 20 cercaires. Le nombre moyen total de cercaires émises pour ces 4 profils du 7<sup>ème</sup> jour de lumière continue n'est pas significativement différent de celui du profil du 3<sup>ème</sup> jour de lumière continue ( $p > 0,01$ ).

Les deux profils restants du 7<sup>ème</sup> jour de lumière continue ont émis plus de 20 cercaires. Ils sont issus des mollusques n°2 et n°6. La majorité de leur émission se concentre sur 2 heures avec toutefois 9 heures d'écart entre leur « pic » d'émission respectifs. La majorité de leur émission se fait dans la deuxième moitié du 7<sup>ème</sup> jour de lumière continue. Chez le mollusque n°2, la majorité des cercaires sont émises dans les deux dernières heures du 7<sup>ème</sup> jour de lumière continue tandis que chez le mollusque n°6 la majorité des cercaires sont émises entre 8 et 10h de la seconde moitié du 7<sup>ème</sup> jour de lumière continue. A noter que 9 fois plus de cercaires ont été émises chez le mollusque n°2.

### 7<sup>ème</sup> jour de lumière continue *versus* 13<sup>ème</sup> jour de lumière continue (4/6)



Du 2<sup>ème</sup> jour de lumière continue jusqu'au dernier, les mollusques n°1, 3, 4, et 5 ont émis moins de 20 cercaires avec une moyenne globale de 6 cercaires par jour. Le nombre total moyen de cercaires émises n'est pas significativement différent entre les jours de lumière continue. Aucune période de libre cours n'a pu être mise en évidence.

### 7<sup>ème</sup> jour de lumière continue *versus* 13<sup>ème</sup> jour de lumière continue (1/6)



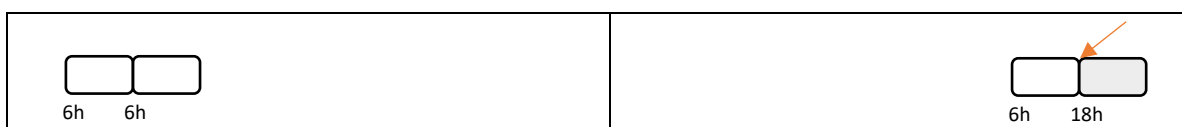
Le mollusque n°2 a émis plus de 20 cercaires les 7<sup>ème</sup> jour, 8<sup>ème</sup> jour, 12<sup>ème</sup> jour et 13<sup>ème</sup> jour de lumière continue, avec en moyenne une émission de 400 cercaires. Aucune période de libre cours comprise entre 20h et 28h n'a pu être mise en évidence.

### 7<sup>ème</sup> jour de lumière continue *versus* 13<sup>ème</sup> jour de lumière continue (1/6)



Le mollusque n°6 a émis plus de 20 cercaires les 7<sup>ème</sup> jour, 8<sup>ème</sup> jour, 12<sup>ème</sup> jour et 13<sup>ème</sup> jour de lumière continue, avec en moyenne une émission de 107 cercaires. Aucune période de libre cours comprise entre 20h et 28h n'a pu être mise en évidence.

### 13<sup>ème</sup> jour de lumière continue *versus* 1<sup>er</sup> jour du retour L-O



Le retour de l'alternance lumière-obscurité débute avec l'obscurité à 18h. Les deux comportements distinctifs observés lors du 13<sup>ème</sup> jour de lumière continue restent inchangés au 1<sup>er</sup> jour du retour de la photopériode équilibrée (L-O). Les mêmes 4 individus ont des profils avec moins de 20 cercaires. Les mollusques n°2 et 6 ont émis plus de 20 cercaires. La majorité de l'émission cercarienne se fait au retour de l'obscurité, avec respectivement pour les mollusques n°2 et 6, 58,7 et 96,8% de cercaires nocturnes. Le pic de cette émission nocturne a lieu une heure après le retour de l'obscurité (19-20h).

## Partie I

### 1<sup>er</sup> jour du retour L-O versus 2<sup>ème</sup> jour du retour L-O



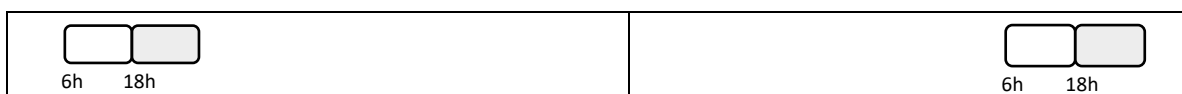
Au 2<sup>ème</sup> jour du retour de la photopériode équilibrée (L-O), les 6 profils ont émis plus de 20 cercaires. Contrairement au jour précédent, ces profils ont les caractéristiques du chronotype diurne (1<sup>er</sup> jour, 2<sup>ème</sup> jour et 1<sup>er</sup> jour Lumière continue) avec la présence d'un seul pic à 12-13h et le pourcentage de cercaire nocturne moyen qui est descendu à 11%.

### 2<sup>ème</sup> jour du retour L-O versus 3<sup>ème</sup> jour du retour L-O



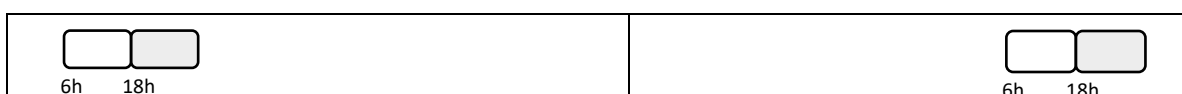
Les huit profils du 3<sup>ème</sup> jour du retour L-O ont émis plus de 20 cercaires. A l'exception du nombre total de cercaires, aucun des autres paramètres de description n'est significativement différent entre le profil du 2<sup>ème</sup> jour du retour de L-O et le 3<sup>ème</sup> jour du retour L-O. Le nombre total moyen de cercaires du 3<sup>ème</sup> jour du retour L-O est significativement plus petit que celui du 2<sup>ème</sup> jour du retour L-O (respectivement 1940 et 152 cercaires). Même si la majorité de l'émission cercarienne du 3<sup>ème</sup> jour du retour L-O est diurne, le pourcentage moyen de cercaires nocturnes reste élevé (18,5%) comparé au pourcentage classique observé chez le chronotype diurne (~1-2%).

### 3<sup>ème</sup> jour du retour L-O versus 7<sup>ème</sup> jour du retour L-O



Les huit profils du 7<sup>ème</sup> jour du retour L-O ont émis plus de 20 cercaires. A l'exception de l'intensité de la plage d'émission, aucun des paramètres de description n'est significativement différent entre le profil du 3<sup>ème</sup> jour du retour de L-O et le profil du 7<sup>ème</sup> jour du retour L-O. L'intensité de la plage du 7<sup>ème</sup> jour du retour L-O est significativement plus élevée que celle du 3<sup>ème</sup> jour du retour L-O (respectivement 86,6% et 67,8%). Même si la majorité de l'émission cercarienne du 7<sup>ème</sup> jour du retour L-O est diurne, le pourcentage moyen de cercaires nocturnes reste élevé (5,7%) comparé au pourcentage classique observé chez le chronotype diurne (~1-2%).

### 7<sup>ème</sup> jour du retour L-O versus 8<sup>ème</sup> jour du retour L-O



Les huit profils du 8<sup>ème</sup> jour du retour L-O ont émis plus de 20 cercaires. Aucun des paramètres de description n'est significativement différent entre le profil du 3<sup>ème</sup> jour du retour de L-O et le 7<sup>ème</sup> jour du retour L-O. Même si la majorité de l'émission cercarienne du 7<sup>ème</sup> jour du retour L-O est diurne, le pourcentage moyen de cercaires nocturnes reste élevé (15,2%) comparé au pourcentage classique observé chez le chronotype diurne (~1-2%).

#### I.2.2.1.3. Discussion et conclusion

Les jours 1 et 2 ainsi que le 1<sup>er</sup> jour de lumière continue (L-L) représentent le chronotype classique de *S. mansoni*. Le changement de photopériode s'étant opéré au milieu de la journée (L-L à 18h) après le pic, les 12 premières heures correspondent à une émission classique. De plus aucune émission significative n'a été observée durant les douze heures suivantes du 1<sup>er</sup> jour de lumière continue. Les

## Partie I

2<sup>ème</sup> et 3<sup>ème</sup> jours de lumière continue marquent un arrêt brusque dans l'émission des cercaires puisqu'aucun profil ne dépasse les six cercaires. Cet arrêt d'émission est d'autant plus impressionnant que l'on passe d'un nombre total moyen de cercaires de 753 (Jour 1, jour 2 et 1<sup>er</sup> jour L-L) à 2 (2<sup>ème</sup> jour et 3<sup>ème</sup> jour de L-L).

C'est à partir du 7<sup>ème</sup> jour de lumière continue qu'un changement de comportement est observé. Quatre individus vont rester sous la barre des 20 cercaires avec un nombre total moyen de 5 cercaires jusqu'au dernier jour de lumière continue (13<sup>ème</sup> jour de L-L). Les deux individus restants vont émettre plus de 20 cercaires pendant les 7<sup>ème</sup> jour, 8<sup>ème</sup> jour, 12<sup>ème</sup> jour et 13<sup>ème</sup> jour de lumière continue avec un nombre total moyen de cercaires de 600 (mollusque n°2) et 107 (mollusque n°6). En plus du nombre moyen de cercaires émises, les deux individus se distinguent par la désynchronisation de leur émission cercarienne. Même si des cercaires ont été émises, aucune période de libre cours n'a pu être mise en évidence.

Le 1<sup>er</sup> jour de retour à la photopériode équilibrée s'opère avec le retour de l'obscurité à 18h. Pour les deux mollusques qui ont émis plus de 20 cercaires, la majorité de l'émission est nocturne s'opérant une heure après le début du retour de la nuit. Après douze heures d'obscurité, le 2<sup>ème</sup> jour du retour de la photopériode équilibrée, tous les mollusques émettent de nouveau plus de 20 cercaires avec un pic diurne classique à 12-13h et une explosion du nombre total de cercaires émises (1940 cercaires en moyenne). Le pourcentage de cercaires nocturnes et la plage d'émission sont significativement plus grands que ceux observés au début de l'expérience ( $p < 0,01$ ). Une émission plus longue avec plus de cercaires nocturnes illustre une réadaptation à la photopériode équilibrée. En effet, au 8<sup>ème</sup> jour du retour de la photopériode équilibrée, aucun des paramètres de description ne sont significativement différents avec les Jour 1, Jour 2 et le 1<sup>er</sup> jour de lumière continue ( $p > 0,01$ ).

L'absence d'émission cercarienne ou la présence d'une émission erratique en lumière continue, couplées à un retour immédiat à la normale dès le recommencement d'une alternance jour/nuit sont des observations en faveur d'un rythme de nature exogène.

En conclusion, en présence de lumière continue le rythme d'émission des cercaires n'est pas maintenu. Il n'y a donc pas d'évidence de rythme circadien. Cette expérience montre aussi que la lumière seule n'est pas suffisante pour déclencher une émission rythmée et qu'il faut une alternance jour/nuit. La variation impressionnante du nombre de cercaires observées sous les différentes photopériodes nous interroge sur le phénomène au niveau intra-mollusque.

### 1.2.2.1.4. Expérience d'histologie

Si l'analyse de l'expérience de lumière continue est faite d'un point de vue production cercarienne, l'arrêt brusque de l'émission pendant 12 jours de lumière continue et l'explosion de l'émission cercarienne au 2<sup>ème</sup> jour du retour de la photopériode équilibrée nous interroge sur le devenir des cercaires pendant ce laps de temps. Bien qu'important, le nombre de cercaires émises le 2<sup>ème</sup> jour du retour L-O ne correspond pas à l'addition des douze jours d'absence d'émission. Pour comprendre où sont passées les cercaires non-émises, deux hypothèses sont envisagées. La première est que les cercaires sont sorties des sporocystes mais sont en attente des conditions propices de photopériode pour sortir du mollusque. La deuxième hypothèse est que les cercaires sont soit retenues soit en attente à l'intérieur des sporocystes.

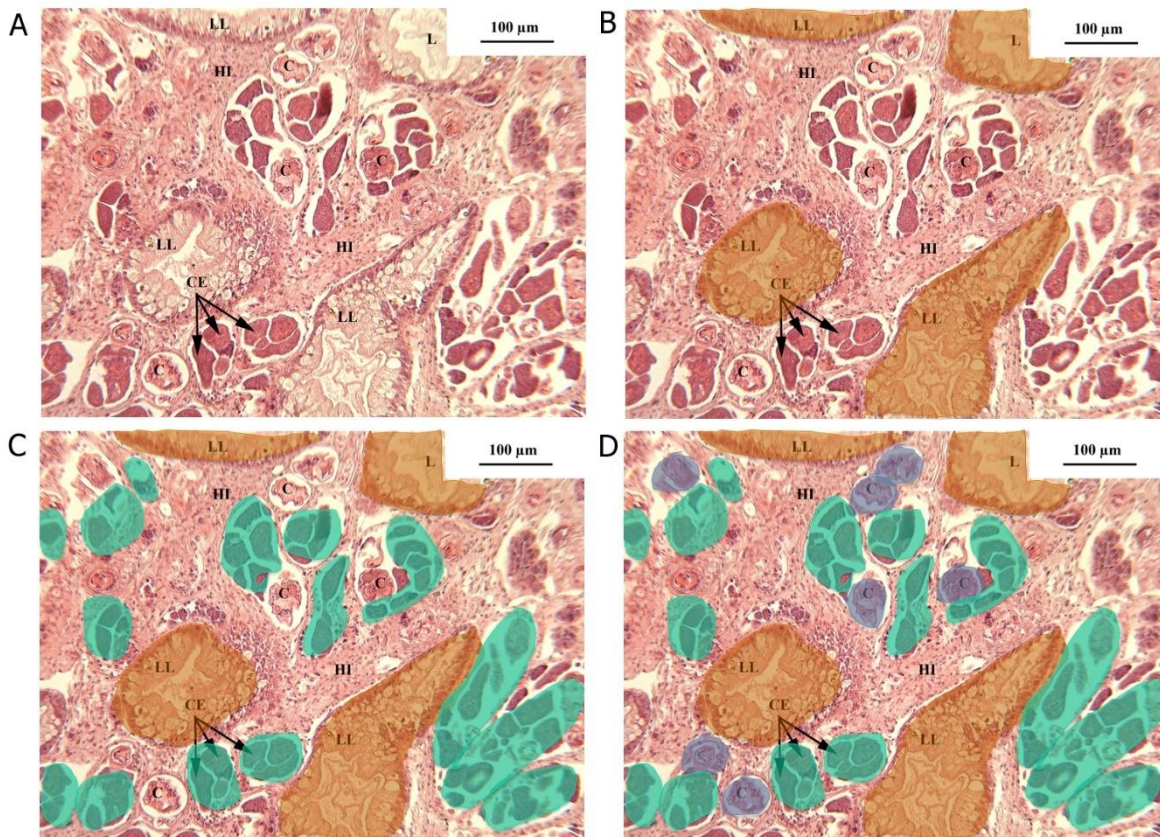
Pour répondre à cette question, une étude histologique a été réalisée sur des *B. glabrata* infestés avec le chronotype diurne de *S. mansoni* sous deux conditions expérimentales différentes ; soit après 13 jours de photopériode équilibrée (L-O 12-12) soit après 13 jours de lumière continue (L-L 12-12). Les coupes sont réalisées sur l'hépatopancreas et l'ovotestis car c'est dans ces zones que se concentrent

## Partie I

les parasites. La coloration utilisée est l'Hématoxyline Eosine Safran (HES) ; elle met en évidence les noyaux en violet (hématoxyline), les cytoplasmes en rose (éosine) et les fibres de collagène en jaune (Safran).

### I.2.2.1.5. Résultats histologie après 13 jours de photopériode équilibrée

Après 13 jours de photopériode équilibrée quelques cellules hôtes sont observées (Figure\_28\_A et B). Il y a beaucoup plus d'embryons cercariens que de cercaires matures (Figure\_28\_C et D). Cela s'explique par le fait que, toutes les 24h, les cercaires matures sont émises et que la cercariogenèse suit son cours pour les remplacer.



**Figure 28 : Coupe histologique d'hépatopancreas de *B. glabrata* infesté avec le chronotype diurne de *S. mansoni* (Oman) après 13 jours de photopériode équilibrée (L-O)** A) Coupe histologique avec coloration HES, image d'origine. B) Coupe histologique avec mise en évidence des cellules de l'hôte en orange. C) Coupe histologique avec mise en évidence des embryons cercariens en turquoise. D) Coupe histologique avec mise en évidence des cercaires matures en bleu foncé. Abréviations : C : Cercaria (cercaire), CE : Cercarial Embryo (embryon cercarien), HI : Hemocyte Invasion (invasion hémocytaire), L : Lumen (lumière), LL : Lobular Liver (lobule digestif).

### I.2.2.1.6. Résultats histologie après 13 jours de lumière continue

Après 13 jours de lumière continue, très peu de cellules hôtes sont observées (Figure\_29\_A et B). On peut aussi voir quelques embryons cercariens (Figure\_29\_C). Toutefois, la majorité du tissu est rempli par des cercaires matures (Figure\_29\_D). Ces cercaires sont contenues à l'intérieur des sporocystes (espaces blancs autour des cercaires). C'est donc l'hypothèse des cercaires attendant à l'intérieur des sporocystes qui est validée. De plus le ratio de cercaires matures et d'embryons cercariens illustre un arrêt ou un très fort ralentissement de la cercariogenèse qui explique pourquoi après le retour de la

## Partie I

photopériode équilibrée, le nombre de cercaires observées n'est pas l'addition des 12 jours d'arrêt d'émission.

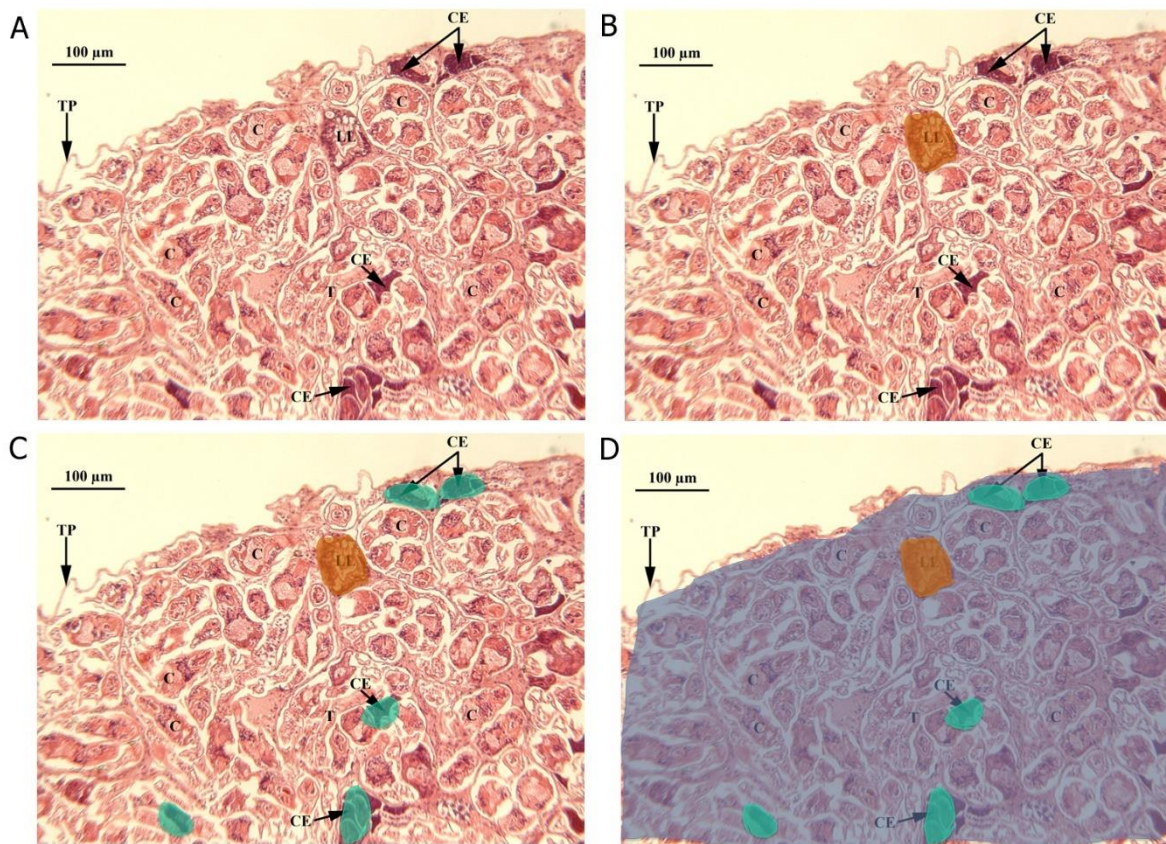


Figure 29 : Coupe histologique d'hépatopancréas de *B. glabrata* infesté avec le chronotype diurne de *S. mansoni* (Oman) après 13 jours de lumière continue A) Coupe histologique avec coloration HES, image d'origine B) Coupe histologique avec mise en évidence des cellules de l'hôte en orange C) Coupe histologique avec mise en évidence des embryons cercariens en turquoise D) Coupe histologique avec mise en évidence des cercaires matures en bleu foncé. Abréviations : C : Cercaria (cercaire), CE : Cercarial Embryo (embryon cercarien), LL : Lobular Liver (lobule digestif), T : Tail (queue de la cercaire) et TP : Tunica propria (tégument du mollusque).

### I.2.2.1.7. Discussion expérience lumière continue et histologie

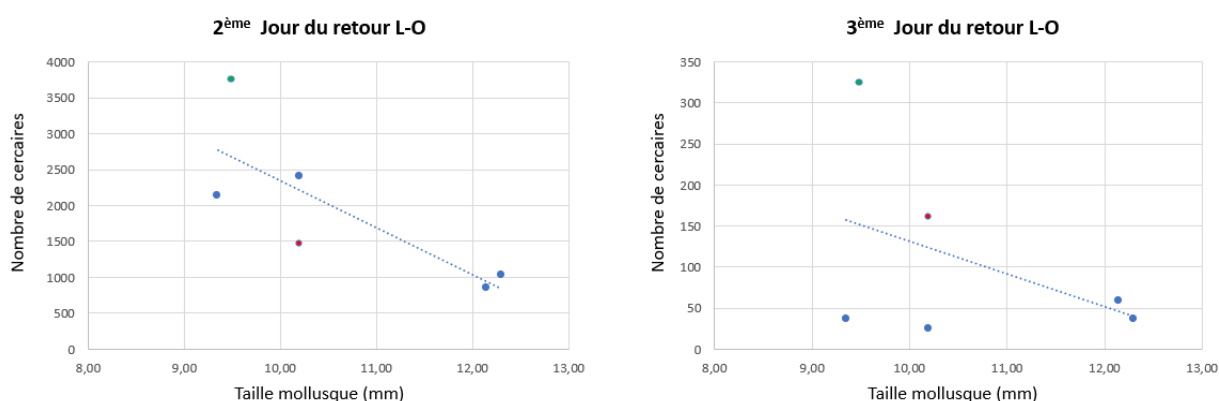
Les observations des coupes histologiques apportent un éclaircissement supplémentaire sur les résultats de l'expérience de lumière continue. Pour la majorité des individus, l'émission des cercaires cesse avec la lumière continue. Les observations histologiques montrent que les cercaires matures s'accumulent à l'intérieur des sporocystes et que la cercariogenèse est arrêtée.

Au 2<sup>ème</sup> jour du retour de la photopériode équilibrée, il y a une explosion du nombre de cercaires émises allant de 856 à 3755 cercaires (Tableau\_20). Il est intéressant de noter que le plus grand nombre de cercaires a été émis par l'individu n°6 qui pourtant a émis des cercaires en période de lumière continue (Tableau\_21).

**Tableau 21 : Corrélation entre la taille des mollusques et la production cercarienne du jour 1 et 2 de la photopériode équilibrée, du 1<sup>er</sup> jour de la lumière continue et des 2<sup>ème</sup> jour et 3<sup>ème</sup> jour de la photopériode équilibrée après 15 jours de lumière continue**

Mollusque	Taille des mollusques au début de l'expérience de lumière continue (mm)	Production cercarienne au jour 1 de la photopériode équilibrée (L-O)	Production cercarienne au jour 2 de la photopériode équilibrée (L-O)	Production cercarienne au 1er jour de lumière continue (L-L)	Production cercarienne au 2 <sup>ème</sup> jour du retour de la photopériode équilibrée (L-O)	Production cercarienne au 3 <sup>ème</sup> jour du retour de la photopériode équilibrée (L-O)
1	10,20	1036	986	1313	2409	26
2	10,20	1589	901	804	1463	161
3	12,15	274	464	680	856	59
4	13,00	575	775	835		
5	9,30	476	477	671		
6	9,50	546	474	678	3755	324
8	9,35				2132	37
9	12,30				1027	36
Coefficient de corrélation		-0,28	0,10	-0,02	-0,78	-0,44

Il n’y a pas de corrélation évidente entre la taille du mollusque et le nombre de cercaires émises pour les jours 1 et 2 de la photopériode équilibrée (L-O), ni pour le 1<sup>er</sup> jour de la lumière continue (L-L) dont les douze premières heures correspondent à une émission classique. Par contre, il existe une corrélation négative entre la taille des mollusques et le nombre de cercaires émises les 2<sup>ème</sup> et 3<sup>ème</sup> jours de photopériode équilibrée après 15 jours de lumière continue (Figure\_30 et Tableau\_21).

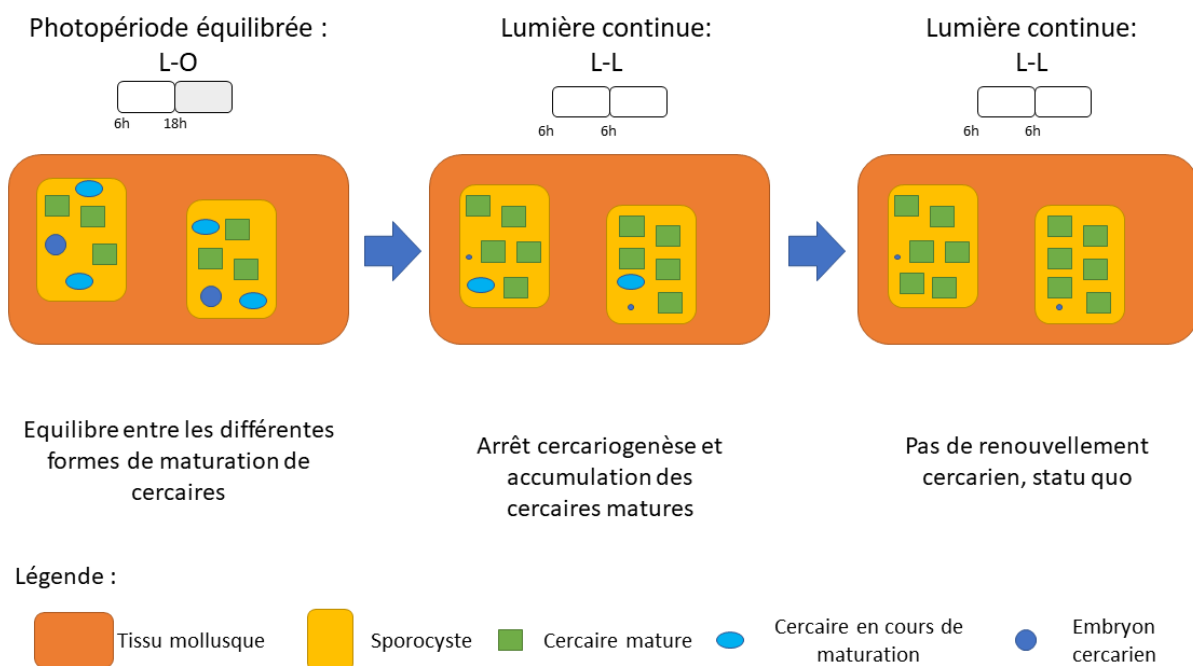


**Figure 30 : Corrélation entre la taille des mollusques et la production cercarienne aux 2<sup>ème</sup> et 3<sup>ème</sup> jours du retour de la photopériode équilibrée.** En abscisse la taille des mollusques et en ordonnée le nombre de cercaires. Les coefficients de corrélation sont respectivement de -0,78 pour le 2<sup>ème</sup> jour du retour de la photopériode équilibrée et de -0,44 pour le 3<sup>ème</sup> jour du retour de la photopériode équilibrée. Les points rouges indiquent le mollusque n°2 et les points verts le mollusque n°6.

## Partie I

En période de photopériode classique L-O, la production cercarienne dépend de plusieurs facteurs, dont la taille du mollusque, la charge parasitaire, la capacité de développement du parasite et la permissivité de l'hôte. Dans notre cas, la charge parasitaire n'entre pas en compte car toutes les infestations sont faites avec un seul miracidium. Si la corrélation entre taille du mollusque et production cercarienne a été prouvée chez *S. mansoni* (Porto Rico) chez *B. glabrata* (Niemann & Lewis, 1990) il n'y pas de corrélation évidente dans cette expérience (Tableau\_21).

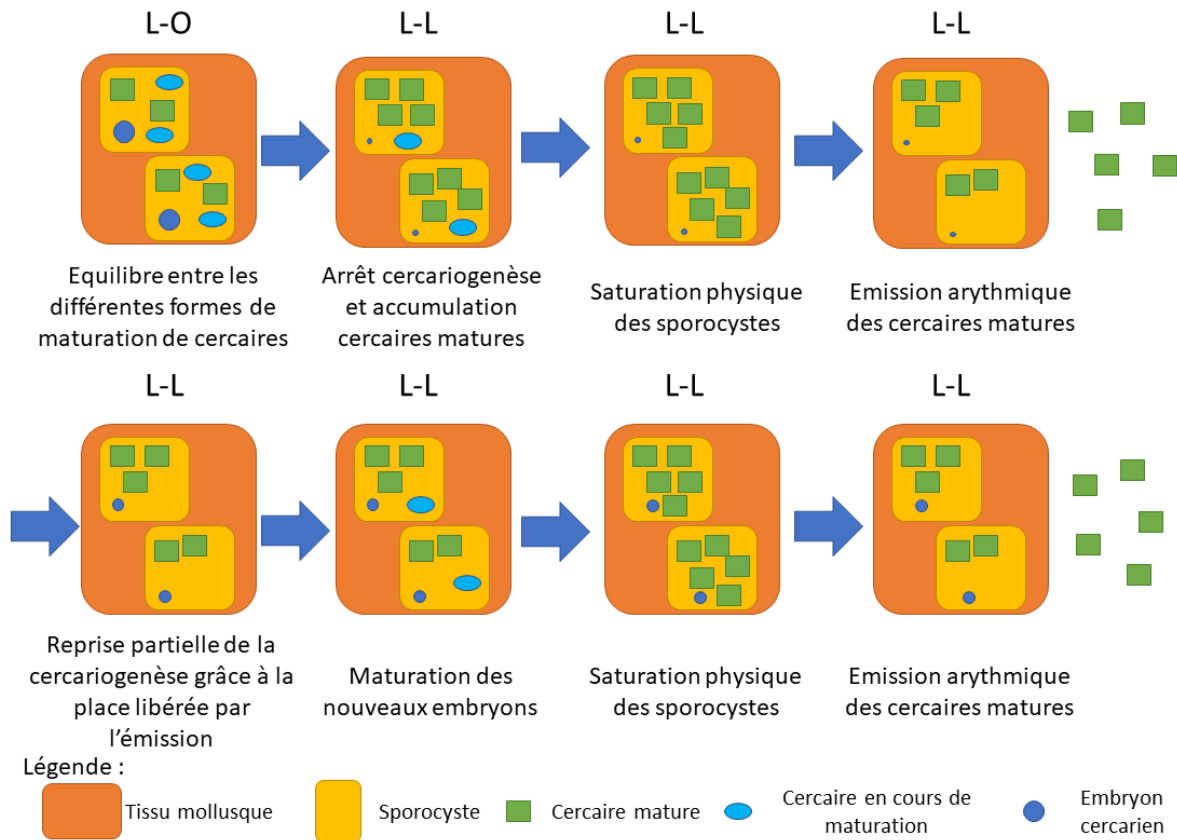
Les observations histologiques ont montré un arrêt de la cercariogénèse et une accumulation des cercaires matures dans les sporocystes. Nous pouvons émettre l'hypothèse que plus le mollusque est grand plus les sporocystes pleins de cercaires ont de la place dans les tissus les uns par rapport aux autres, et donc peuvent tenir plus longtemps sans saturer les tissus mollusques. Nous pouvons alors penser que, de par l'encombrement intrasporocyste, la cercariogénèse est complètement arrêtée, et que le sporocyste plein reste dans un état de statu quo Figure\_31.



**Figure 31 : Schéma de l'hypothèse du devenir des cercaires en lumière continue dans les mollusques de grande taille**

A l'inverse, dans les petits mollusques, la saturation limite des tissus arrive plus rapidement et oblige une sortie massive de cercaires. Nous pouvons alors penser que, dans ce cas, la cercariogénèse est très fortement ralentie mais pas totalement arrêtée (Figure\_32). Les cercaires émises sont lentement remplacées, ce qui explique la succession de jours avec et sans émissions significatives pour les mollusques 2 et 6 (les plus petits du groupe).

Partie I



**Figure 32 : Schéma de l'hypothèse du devenir des cercaires en lumière continue dans les mollusques de petite taille**

Cette hypothèse reste valable même si le mollusque n°2 (10,20 mm), plus grand que le mollusque n°6 (9,50 mm), émet plus de cercaires. Lors des premiers jours de l'expérience en photopériode équilibrée (L-O), le mollusque n° 2 a émis en moyenne deux fois plus de cercaires que le mollusque n°6. On peut donc dire que le mollusque n°6, plus petit, pourrait avoir une concentration parasitaire moins forte et donc émettre moins de cercaires ce qui signifierait que la différence observée n'est pas uniquement due à la taille. Dans cette expérience, seule la taille en début d'expérience a été mesurée. L'évolution de la taille des mollusques et donc de la capacité de croissance de l'hôte dépend de sa réponse à la parasitose. Deux mollusques de même taille, peuvent avoir une réponse différente à la présence du parasite qui peut provoquer une différence de croissance. Cela pourrait notamment expliquer la différence de comportement entre le mollusque n°1 qui a toujours émis moins de 20 cercaires en lumière continue et le mollusque n°2 avec lequel il partageait la même taille en début d'expérience. Pour approfondir ces résultats, il faudrait mesurer les mollusques en début et fin d'expérience, et aussi réaliser l'expérience des coupes histologiques avec des mollusques de tailles différentes. En effet, la taille du mollusque pour la coupe histologique en lumière continue mesurait 13,55mm, ce qui fait partie des tailles les plus grandes.

Indépendamment du rythme de l'émission des cercaires, il existe un rythme de la production cercarienne (nombre de cercaires) ainsi qu'un rythme de développement des cercaires (abondance relative entre les stades embryonnaires et les stades cercaires en fin de maturation) (Théron, 1982). D'un point de vue histologique, on peut interpréter les résultats en lumière continue par le fait que l'arrêt de l'alternance jour/nuit induit un arrêt de synchronisme dans le développement des cercaires. Sans repère temporel, le renouvellement de génération de cercaires s'arrête. L'agent de ce renouvellement est inconnu. Il peut soit être un stock initial de cellules germinales présentes dans le



## Partie I

sporocyste qui entrerait en maturation de façon échelonnée, soit être une production rythmique de cellules germinales générées à la fin de chaque cycle de maturation (génération). Dans les deux cas, la lumière continue induit leur arrêt de maturation. En parallèle, les générations de cercaires déjà en cours de maturation continuent leur développement jusqu'au stade mature. Il se crée alors l'accumulation de cercaires matures dans les sporocystes observés après 13 jours de lumière continue. Pour les parasites présents dans les plus petits mollusques, cette accumulation peut se transformer en encombrement physique du sporocyste qui, faute de place, voit les cercaires les plus matures sortir. Cette sortie des cercaires peut, soit être un processus issu du sporocyste qui va réguler son encombrement, soit un processus des cercaires qui perçoivent leur densité intra-sporocyste, soit plus simplement une pression physique qui fait sortir les cercaires.

### I.2.2.2. Chronotype nocturne (expérience d'obscurité continue)

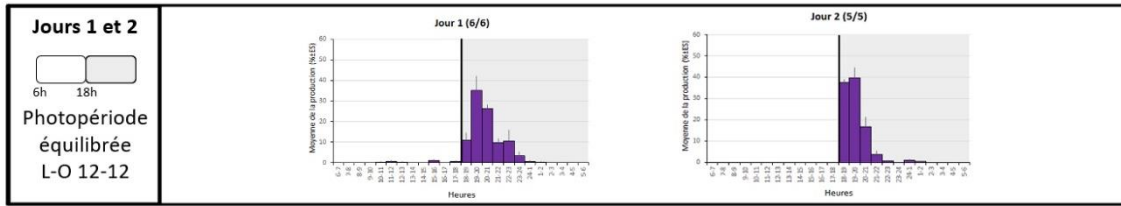
#### I.2.2.2.1. Protocole obscurité continue

L'expérience de chronobiologie en obscurité continue a été effectuée sur une période de 20 jours et réalisée sur 6 *Biomphalaria pfeifferi* infestés par le chronotype nocturne de *S. mansoni*. Ces jours se divisent en trois phases d'expérimentation. Les jours 1 et 2 représentent les conditions classiques, avec une photopériode équilibrée L-O 12-12 ; la lumière s'allumant à 6h du matin et s'éteignant à 18h le soir. A partir du jour 3, la lumière reste éteinte à 6h du matin. Le jour 3 correspond au 1<sup>er</sup> jour d'obscurité continue et le jour 14 au 12<sup>ème</sup> et dernier jour d'obscurité continue O-O 12-12. A partir du jour 15 et jusqu'à la fin de l'expérimentation, la lumière est rallumée à 6h pour un retour de la photopériode équilibrée L-O 12-12 identique aux conditions classiques.

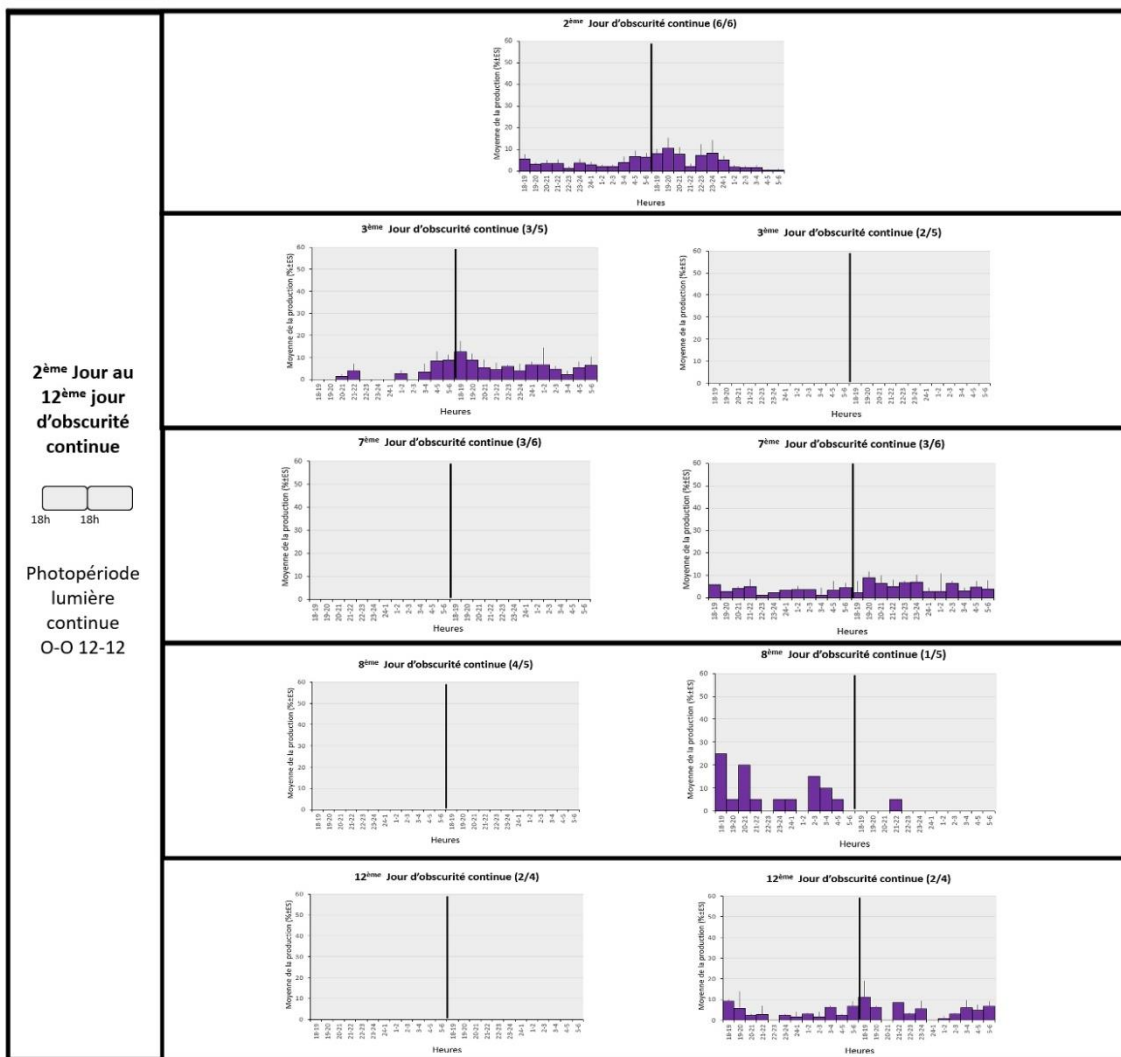
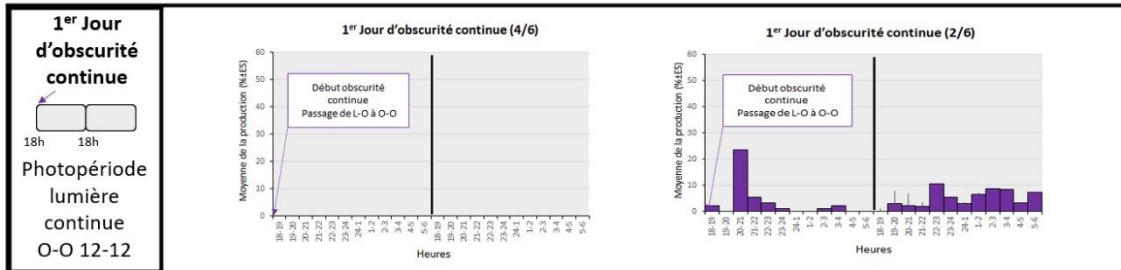
#### I.2.2.2.2. Expérience d'obscurité continue

Les résultats moyens, accompagnés des variations individuelles, sont présentés dans la Figure\_33 et le Tableau\_22 pour la partie obscurité continue ainsi que dans la Figure\_34 et le Tableau\_23 pour la partie retour de la photopériode équilibrée. Pour cette expérience, le paramètre du nombre total de cercaires émises n'est pas pris en compte, mais reste présent dans les données. Pour les jours d'obscurité continue (O-O), le pourcentage de cercaires nocturnes correspondra au pourcentage de cercaires émises durant les douze dernières heures de la journée, c'est-à-dire pendant la nuit subjective.

Partie I



## Obscurité continue



**Figure 33 : Histogramme de l'expérience de chronobiologie d'Obscurité continue pour le chronotype nocturne (1/2), phase obscurité continue.** L'expérimentation voit se succéder trois conditions : Photopériode équilibrée L-O 12-12 (jours 1 et 2), obscurité continue (1er au 12ème jour) et le retour à la photopériode équilibrée en figure 32 (1er au 6ème jour). L'obscurité continue commence à 6h. En blanc, période de lumière (indiquée de 6h à 18h) ; en gris, période d'obscurité (indiquée de 18h à 6h).

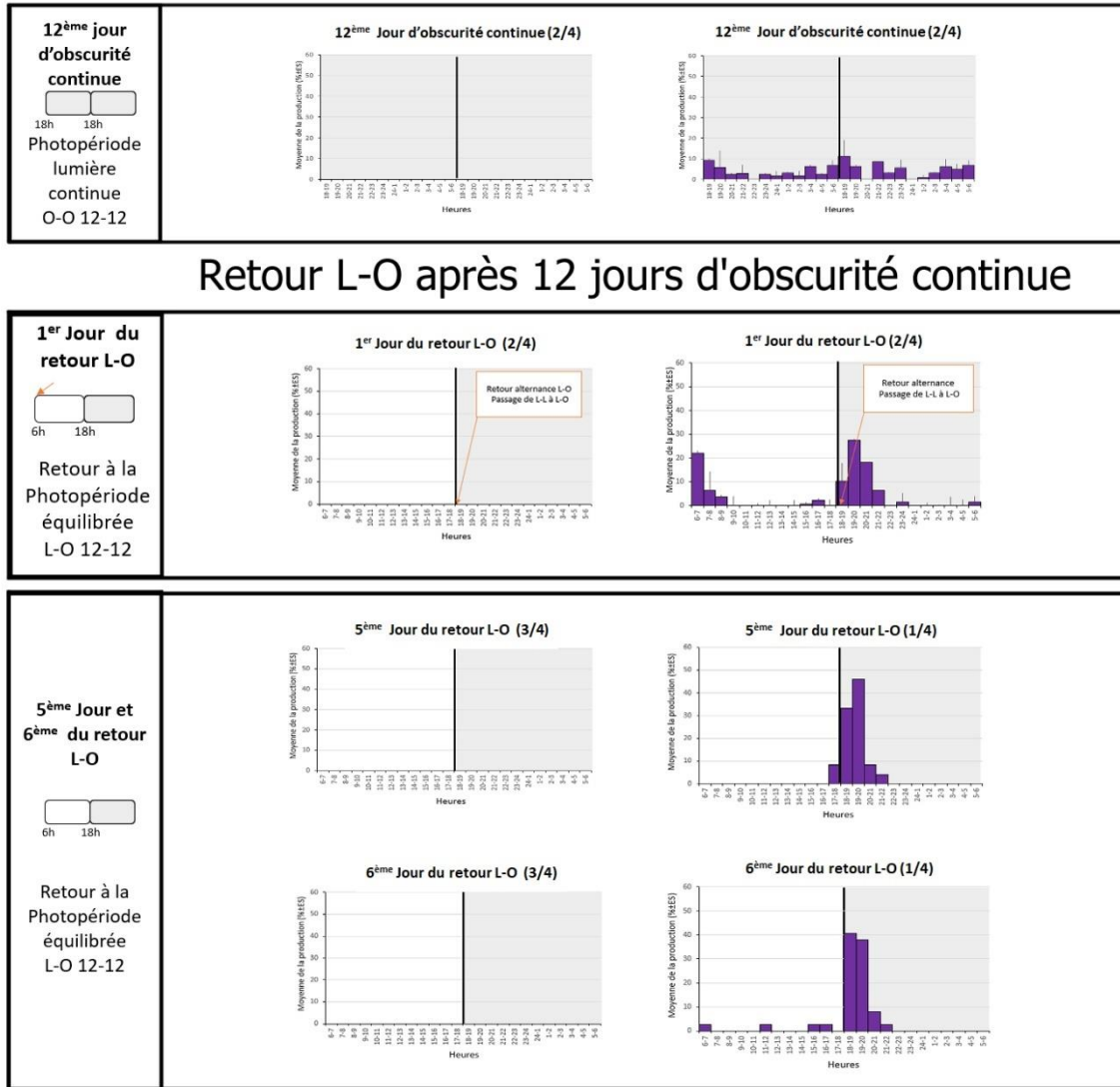
Partie I

**Tableau 22 : Paramètres des profils moyens d'émission, accompagnés des variations individuelles, pour les parasites du chronotype nocturne de *S. mansoni* de l'expérience d'obscurité continue (1/2)**

Paramètres	Moyenne ± ES Variations individuelles												
	Jour 1	Jour 2	1 <sup>er</sup> jour de O-O	1 <sup>er</sup> jour de O-O	2 <sup>ème</sup> Jour de O-O	3 <sup>ème</sup> Jour de O-O	3 <sup>ème</sup> Jour de O-O	7 <sup>ème</sup> jour de O-O	7 <sup>ème</sup> jour de O-O	8 <sup>ème</sup> Jour de O-O	8 <sup>ème</sup> Jour de O-O	12 <sup>ème</sup> Jour de O-O	12 <sup>ème</sup> Jour de O-O
Effectif	6/6	5/5	4/6	2/6	6/6	3/5	2/5	3/6	3/6	4/5	1/5	2/4	2/4
Photopériode	L-O	L-O	O-O	O-O	O-O	O-O	O-O	O-O	O-O	O-O	O-O	O-O	O-O
Nombre de pics	1 1	1 1		1 1									
Heure du pic	19-20h 18-19h à 22-23h	19-20h 18-19h à 19-20h		20-21h (am) à 22- 23h (pm)	3-4h (am) à 23-24h (pm)	18-19h (pm) à 1- 2h (pm)			19-20h (pm) à 23- 24h (pm)		18-19h (am)		19-20h (am) à 18- 19h (pm)
Plage d'émission (h)	4,3 ± 0,4 3 à 5	4,2 ± 0,4 3 à 5		2,5 ± 0,7 2 à 3	3,5 ± 0,4 2 à 4	4,3 ± 1,1 3 à 6			2,7 ± 0,8 2 à 4		4		2 ± 1,4 1 à 3
Intensité de la plage (%)	91,9 ± 1,6 89,1 à 98,8	97,8 ± 0,5 96,2 à 98,7		39,5 ± 11,7 31,3 à 47,7	50,8 <sup>a</sup> ± 10,8 26,8 à 86,6	48,4 ± 11,6 34,8 à 66,7			28,7 ± 12,2 16,0 à 48,3		55		33,2 ± 30,8 11,4 à 55
Asymétrie	+0,6 ± 0,5 -1,3 à +1,7	-0,4 ± 0,5 -1,8 à +0,4		+2 (N=1)	+1,2 ± 0,4 (N=5) 0 à +1,8	+0,3 ± 0,9 -0,9 à +1,5			-2 (N=1)		+0,2		+0,2 (N=1)
Aplatissement	-0,3 ± 1,5 (N=5) -3,1 à +3,2	-1,5 ± 1,9 (N=4) -4,0 à +3,3		(N=0)	+0,9 ± 2,1 (N=4) -4,4 à +3,3	-1 ± 0,4 (N=2) -1,5 à -0,6			4 (N=1)		-4,9		-4,9 (N=1)
Cercaires la nuit (%)	97,1 ± 1,8 91,3 à 100	100 100		61,0 ± 31,6 38,6 à 83,3	54,9 ± 12,2 19,7 à 99,5	71,1 ± 12,4 51,5 à 82,6			60 ± 6,3 51 à 69		5		56,0 ± 10,4 48,6 à 63,3
Nombre total de cercaires	63,3 ± 5,5 46 à 80	84,0 ± 15 52 à 122	10,8 ± 3,7 3 à 17	46,0 ± 2,8 44 à 48	123 ± 55,1 34 à 340	56 ± 20,7 33 à 89	13 ± 8,5 7 à 19	11 ± 6,5 1 à 19	58,3 ± 24,2 29 à 96	8,5 ± 4,1 3 à 19	20	9 ± 2,8 7 à 11	47,5 ± 17,7 35 à 60

ES : Ecart-standard ; L-O : Lumière-Obscurité (12-12) ; O-L : Obscurité-Lumière (12-12)

Tests Wilcoxon Mann Whitney significatifs ( $p$ -value<0,01) pour les effectifs supérieurs à 4 individus : Comparaison avec le jour précédent



**Figure 34 : Histogramme de l'expérience de chronobiologie d'Obscurité continue pour le chronotype nocturne (2/2), phase retour à la photopériode équilibrée.** L'expérimentation voit se succéder trois conditions : Photopériode équilibrée L-O 12-12 en figure 31 (jours 1 et 2), obscurité continue en figure 31 (1<sup>er</sup> au 12<sup>ème</sup> jour) et le retour à la photopériode équilibrée (1<sup>er</sup> au 6<sup>ème</sup> jour). Le retour à la photopériode continue se fait avec le retour de la lumière à 6h. En blanc, période de lumière (indiquée de 6h à 18h) ; en gris, période d'obscurité (indiquée de 18h à 6h).

Partie I

**Tableau 23 : Paramètres des profils moyens d'émission, accompagnés des variations individuelles, pour les parasites du chronotype nocturne de *S. mansoni* de l'expérience d'obscurité continue (2/2)**

Paramètres	Moyenne ± ES <i>Variations individuelles</i>					
	1 <sup>er</sup> jour du retour L-O	1 <sup>er</sup> jour du retour L-O	5 <sup>ème</sup> jour du retour L-O	5 <sup>ème</sup> jour du retour L-O	6 <sup>ème</sup> jour du retour L-O	6 <sup>ème</sup> jour du retour L-O
Effectif	2/4	2/4	3/4	1/4	1/4	3/4
Photopériode	L-O	L-O	L-O	L-O	L-O	L-O
Nombre de pics	1 1			1	1	
Heure du pic	19-20h			19-20h	18-19h	
Plage d'émission (h)	4			4	4	
Intensité de la plage (%)	62,1 ± 16,1 50,7 à 73,5			95,8	89,2	
Asymétrie	+0,5 ± 1,2 -0,3 à +1,4			+0,4	-0,1	
Aplatissement	0 ± 2,2 -1,6 à +1,5			-3,9	-5,2	
Cercaires la nuit (%)	65,1 ± 20,3 50,7 à 79,44			91,7	89,2	
Nombre total de cercaires	52,5 ± 26,2 34 à 71	15,0 ± 1,4 14 à 16	8,3 ± 2,9 7 à 13	24	37	4,0 ± 1,2 2 à 5

ES : Ecart-standard ; L-O : Lumière-Obscurité (12-12)

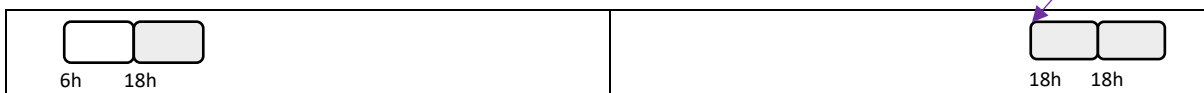
Chacun des paramètres d'un jour a été comparé avec le même paramètre du jour précédent et les résultats sont présentés ci-dessous.

### Jour 1 versus Jour 2



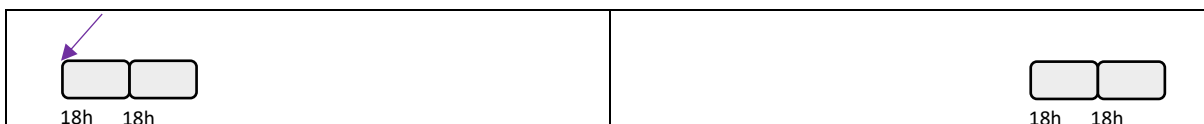
Aucun des paramètres de description n'est significativement différent entre le profil du jour 1 et le profil du jour 2 ( $p > 0,01$ ). Ces profils ont les caractéristiques du chronotype nocturne avec la présence d'un seul pic à 19-20h et une plage d'environ 4h avec une intensité moyenne de la plage supérieure à 90%. Ils partagent aussi un coefficient d'asymétrie modéré (positif ou négatif) ainsi qu'une émission nocturne supérieure à 97%. Le nombre total de cercaires émises est en moyenne de 73.

### Jour 2 versus 1<sup>er</sup> jour d'obscurité continue



Le régime d'obscurité continue débute dès le commencement de la 1<sup>ère</sup> journée d'obscurité continue, avec la lumière laissée éteinte à 6h. Trois comportements se distinguent parmi les six profils observés. Quatre profils sur six ont émis moins de 20 cercaires. Un profil issu du mollusque n°2 a eu son émission la plus élevée 14 heures après son pic du jour précédent. Le dernier profil issu du mollusque n°3 a eu son émission la plus élevée 27 heures après son pic du jour précédent. Pour les deux profils qui ont émis plus de 20 cercaires, l'intensité de la plage principale d'émission a baissé à 39,5% contre les 97,8% du jour 2 de la photopériode équilibrée. Le pourcentage moyen de cercaires émises dans la nuit subjective est descendu à 61% contre les 100% du jour 2 de la photopériode équilibrée.

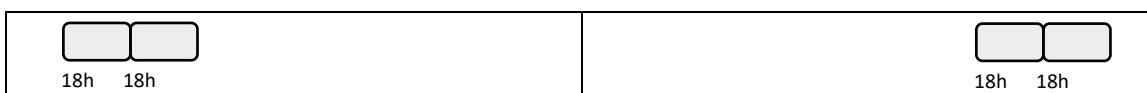
### 1<sup>er</sup> jour d'obscurité continue versus 2<sup>ème</sup> jour d'obscurité continue



Après 24h d'obscurité continue, les 6 profils ont émis plus de 20 cercaires. La plage principale moyenne d'émission et le pourcentage de cercaires émises pendant la période de nuit subjective sont significativement plus faibles que ceux du jour 2 de la photopériode équilibrée ( $w = 30, p = 0,004$  et  $W = 30, p = 0,006$ ). Il n'y a pas une heure consensus pour l'émission la plus haute entre les six profils.

Le profil issu du mollusque n°2 a eu son émission la plus élevée 35h après son émission la plus élevée du 1<sup>er</sup> jour d'obscurité. Le profil issu du mollusque n°3 a eu son émission la plus élevée 22h après son émission la plus élevée du 1<sup>er</sup> jour d'obscurité.

### 2<sup>ème</sup> jour d'obscurité continue versus 3<sup>ème</sup> jour d'obscurité continue

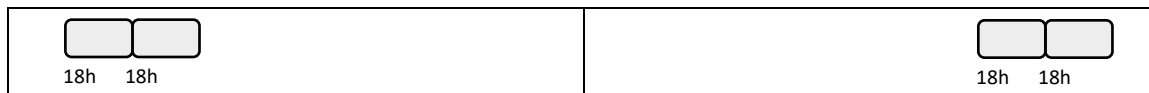


Deux profils sur les 5 observés le 3<sup>ème</sup> jour d'obscurité continue ont émis moins de 20 cercaires. Parmi les trois profils qui ont émis plus de 20 cercaires, on retrouve ceux issus des mollusques n°2 et n°3. Le profil issu du mollusque n°2 a eu son émission la plus élevée 23 heures après son émission la plus élevée du 2<sup>ème</sup> jour d'obscurité continue. Le profil issu du mollusque n°3 a eu son émission la plus élevée 22 heures après son pic du jour précédent.

Le troisième profil qui a émis plus de 20 cercaires est issu du mollusque n°9 qui a été intégré à l'expérimentation pour remplacer un mollusque décédé. Il n'y a pas de donnée de comptage pour cet individu avant le 3<sup>ème</sup> jour d'obscurité continue.

Pour ces trois profils qui ont émis plus de 20 cercaires, l'intensité de la plage principale d'émission reste faible par rapport à celle du jour 2 de la photopériode équilibrée mais équivalente à l'intensité moyenne des jours en obscurité continue. Cette observation est aussi valable pour le nombre moyen de cercaires émises en nuit subjective.

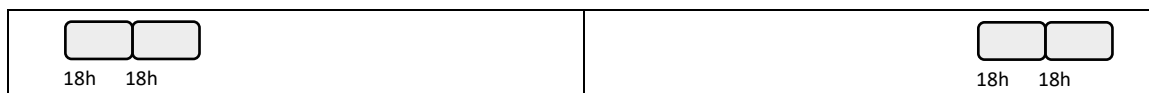
### 3<sup>ème</sup> jour d'obscurité continue versus 7<sup>ème</sup> jour d'obscurité continue



A partir du 7<sup>ème</sup> jour d'obscurité, tous les individus mollusques ont été remplacés par des nouveaux. Aucune période de libre cours n'a pu être mise en évidence pour les individus décédés.

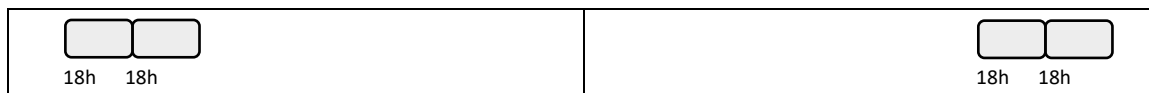
Au 7<sup>ème</sup> jour d'obscurité continue, trois profils sur six ont émis moins de 20 cercaires. Les trois profils restants sont issus des individus mollusques n°12, 13 et 14. Pour ces trois profils, l'intensité moyenne de la plage d'émission est la plus faible depuis le début de l'expérimentation, avec 28,7%. La majorité de l'émission se fait pendant la nuit subjective, avec 60% de cercaires.

### 7<sup>ème</sup> jour d'obscurité continue versus 8<sup>ème</sup> jour d'obscurité continue



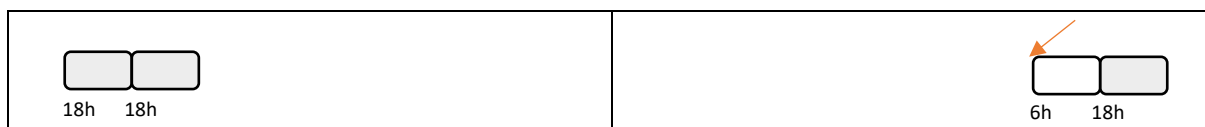
Au 8<sup>ème</sup> jour d'obscurité continue, quatre profils sur cinq ont émis moins de 20 cercaires. Le profil restant est issu du mollusque n°13. Le profil issu du mollusque n°13 a eu son émission la plus élevée 11 heures après son émission la plus élevée du 7<sup>ème</sup> jour d'obscurité continue. La majorité de son émission s'est faite pendant le jour subjectif (5% de cercaires émises en nuit subjective).

### 8<sup>ème</sup> jour d'obscurité continue versus 12<sup>ème</sup> jour d'obscurité continue



Au 12<sup>ème</sup> jour d'obscurité continue, deux profils sur les quatre observés ont émis moins de 20 cercaires. Les deux profils restants sont issus des mollusques n°12 et 13. Il n'y a pas de consensus dans l'heure de l'émission la plus élevée. L'intensité moyenne de la plage d'émission principale reste faible, avec 33,2%.

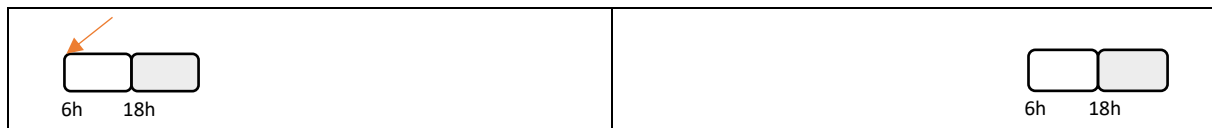
### 12<sup>ème</sup> jour d'obscurité continue versus 1<sup>er</sup> jour du retour L-O



Le retour de l'alternance lumière-obscurité débute avec la lumière à 6h. Les deux comportements distinctifs observés lors du 12<sup>ème</sup> jour d'obscurité continue restent inchangés au 1<sup>er</sup> jour du retour de la photopériode équilibrée (L-O). Les mêmes 2 individus ont des profils avec moins de 20 cercaires. Les mollusques n°12 et n°13 ont émis plus de 20 cercaires. Une émission cercarienne conséquente a eu lieu à 6h du matin dès le retour de la lumière. Toutefois, elle ne peut pas être apparentée à un pic mais plutôt à une réponse au retour de la lumière puisque le chronotype est nocturne et que le pic classique a lieu à 19-20h le même jour. Si le pic a lieu à la bonne heure pour les deux profils, il faut cependant

noter que l'intensité et le pourcentage de cercaires nocturnes sont plus faibles que ceux des jours 1 et jour 2 de photopériode équilibrée.

### Du 2<sup>ème</sup> jour du retour L-O jusqu'au 6<sup>ème</sup> jour du retour L-O



A partir du 2<sup>ème</sup> jour du retour L-O et jusqu'à la fin de l'expérience (6<sup>ème</sup> jour retour L-O), le même comportement est observé. Trois mollusques sur quatre ont des profils d'émission cercarien avec moins de 20 cercaires durant cette période. Seuls les profils du mollusque n° 13 ont émis plus de 20 cercaires. Ces profils ont tous les caractéristiques d'un profil classique du chronotype nocturne, et illustrent ainsi un retour à la normale avec un seul pic à 18-19h ou 19-20h qui a une intensité aux alentours de 90% sur une plage de 4h avec un pourcentage de cercaires nocturnes supérieur à 90%.

#### I.2.2.2.3. Discussion et conclusion

Les jours 1 et 2 de photopériode équilibrée représentent le chronotype nocturne classique de *S. mansoni*. Le 1<sup>er</sup> jour d'obscurité continue marque un arrêt de l'émission des cercaires pour la majorité des profils. Pour la minorité des profils qui émettent des cercaires, il y a une désynchronisation de leur émission cercarienne, avec d'un côté un profil qui fait son émission la plus élevée 14 heures après le pic de la veille et l'autre profil après 27 heures. Pour ces deux individus, les jours suivants, les écarts entre les émissions les plus élevées sont disparates (respectivement 35 heures et 22h heures). Contrairement à l'expérience faite avec le chronotype diurne, le suivi individuel des profils n'a pas pu être mené jusqu'à la fin de l'expérimentation à cause de la mortalité des mollusques. Toutefois pour les quelques jours de suivi individuels qui ont pu être faits, aucune période de libre cours n'a pu être mise en évidence.

De façon globale, la période d'obscurité continue voit l'alternance de deux comportements (plus ou moins de 20 cercaires) en intra et inter-individuel. Il est intéressant de voir que les profils qui émettent le plus de cercaires sont issus des mêmes mollusques. Contrairement au chronotype diurne, le nombre total de cercaires ne s'est pas révélé être un paramètre pertinent. Cela s'explique par le fait que la production cercarienne des *B. pfeifferi* est beaucoup moins élevée que celle des *B. glabrata* utilisés pour le chronotype diurne.

Le 1<sup>er</sup> jour de retour à la photopériode équilibrée s'opère avec le retour de la lumière à 6h. Pour les mollusques qui ont émis des cercaires, le pic nocturne classique est immédiatement de retour. Il se fait en parallèle d'une émission conséquente qui a eu lieu à 6h et s'apparente à une réponse physique à la lumière. Ce phénomène physique de réponse à la photopériode opposée a aussi été observé avec le chronotype diurne (émission nocturne avec le retour de la nuit lors du 1<sup>er</sup> jour du retour de la photopériode équilibrée).

Il est important de voir que la moitié des individus sont restés sous la barre des 20 cercaires malgré le retour de la photopériode équilibrée. Cela peut s'expliquer par le fait que plus de 12 jours d'obscurité continue sont éreintants pour les organismes et que les individus mollusques étaient sur une dynamique de mortalité. On est passé de 6 individus au début de l'expérimentation à 4 car tous les remplaçants étaient soit utilisés soit morts. Pour les cinq derniers jours de l'expérimentation seul un individu a émis plus de 20 cercaires avec un retour à un rythme d'émission classique.



La forte diminution de l'émission cercarienne ou d'une émission erratique en obscurité continue, couplée à un retour immédiat à la normale dès le recommencement d'une alternance jour/nuit sont des observations en faveur d'un rythme de nature exogène.

En conclusion, en présence d'obscurité continue, le rythme d'émission des cercaires n'est pas maintenu. Il n'y a donc pas d'évidence de rythme circadien. Cette expérience montre aussi que l'obscurité seule n'est pas suffisante pour déclencher une émission rythmée et qu'il faut une alternance jour/nuit.

### I.3. Discussion

Pour déterminer si le rythme d'émission des cercaires était circadien, l'adaptabilité et la constance du rythme ont été étudiées. L'adaptabilité du rythme aux changements du zeitgeber a été testée par l'inversion de photopériode. Nos résultats chez le chronotype diurne, que ce soit par 24h de nuit ou 24h de jour, montrent une adaptation directe et totale. Ces résultats sont en accord avec ceux obtenus par Luttermoser (1955), Asch (1972) et Glaudel & Edges (1973) avec la souche Porto Rico. Malheureusement, le comptage des cercaires dans ces études s'est fait toutes les 6 (Asch, 1972 ; Glaudel and Edges, 1973) ou 12 heures (Luttermoser, 1955) et ne peut pas être finement comparé avec les nôtres, notamment sur la question de l'émission subsidiaire aux transitions jour/nuit. Le chronotype nocturne, bien que la majorité de l'émission s'adapte dès 12 heures de nuit, passe par une étape intermédiaire bimodale précédée suivant les cas d'une chute de l'émission cercarienne (émission inférieure à 20 cercaires). L'émission à l'ancienne heure de pic pourrait s'apparenter à une perturbation causée par les 12 heures de décalage qui induiraient une sensibilité accrue à la transition jour/nuit. Toutefois, le pic nocturne se produisant à la transition jour/nuit complexifie la compréhension de cette émission secondaire. Pour peut-être éliminer ce pic subsidiaire à la transition jour/nuit, une expérience de chronobiologie avec un décalage plus léger pourrait être faite. Cette émission chez le chronotype nocturne, dès 12h de nuit supplémentaire lors du 1<sup>er</sup> jour de photopériode inversée pourrait être en faveur d'un rythme plutôt exogène. Peu importe le chronotype, le rythme d'émission suit le changement de photopériode.

La constance du rythme d'émission a été testée avec la photopériode où le pic a lieu, c'est-à-dire lumière continue pour le chronotype diurne et obscurité continue pour le chronotype nocturne. Nos résultats chez le chronotype diurne montrent, qu'à la lumière continue, il y a un arrêt de l'émission cercarienne puis une émission disparate pour quelques individus, sans la possibilité de mettre en évidence une période de libre cours (période régulière entre deux pics comprise entre 20 et 28 heures). Ces résultats sont en accord avec les travaux de Valle, Pellegrino & Alvarenga (1973) (lumière et obscurité continues) et Williams, Wessels & Gilbertson (1984) (lumière continue) faits sur la souche Porto Rico. Toutefois, ces deux études ne peuvent pas être finement comparées aux nôtres, avec, d'un côté, seulement 20% des cercaires comptées toutes les 12h et de l'autre 3 aliquots de 1 $\mu$ L toutes les deux heures, respectivement. Ceci est d'autant plus vrai pour les questionnements portant sur la variabilité des comportements individuels. Tout comme le chronotype diurne, le chronotype nocturne en condition d'obscurité continue, a soit une émission inférieure à 20 cercaires soit une émission disparate sans période de libre cours. Dans les deux chronotypes, le 1<sup>er</sup> jour du retour de la photopériode équilibrée voit un pic d'émission dans l'élément inverse (un pic nocturne pour le chronotype diurne et un pic diurne pour le chronotype nocturne) suivi d'un retour à la normale dès le lendemain. Ce pic dans l'élément inverse (nuit pour le chronotype diurne et jour pour le chronotype nocturne) est une réponse « physique » suite au brusque changement après au moins 10 jours de photopériode continue. Ces résultats montrent l'importance de l'alternance jour/nuit pour déclencher

une émission rythmique des cercaires. Les expériences de Nojima & Sato (1982) sur la souche Porto Rico ont montré qu'il fallait entre 10 et 14 heures d'obscurité pour maintenir un rythme d'émission normale et qu'en dessous de 8 heures d'obscurité le rythme était perdu. Au-dessus de 16 heures d'obscurité, un « pic » supplémentaire apparaît dès le retour de la lumière en plus du pic classique (Nojima & Sato, 1982). Ce « pic » qui ne dure en fait qu'une heure, ressemble au « pic » subsidiaire observé chez le chronotype nocturne lors de nos expériences d'inversion et d'obscurité continue. L'hypothèse qui peut résulter de ces observations est qu'une modification importante de la photopériode (comme une inversion de 12h ou une obscurité trop prolongée) induit un stress chez le parasite qui va émettre des cercaires mécaniquement dès le retour de l'élément opposé (nuit pour le chronotype diurne et jour pour le chronotype nocturne) sur une période très courte pour ensuite faire son pic normal. Cette hypothèse peut être appuyée par nos résultats d'histologie qui ont montré qu'en condition de lumière continue il y a une accumulation des cercaires matures. La modification importante de la photopériode pourrait impliquer une accumulation de cercaires matures dont une petite partie serait expulsée dès le retour de l'élément opposé (nuit pour le chronotype diurne et jour pour le chronotype nocturne). Cette expulsion pourrait être soit une réaction purement physique soit une réaction du parasite afin de faire de la place dans les sporocystes pour retrouver un rythme normal dans l'émission mais aussi dans le développement cercarien.

L'utilisation des *B. glabrata* pour l'expérience de lumière continue a mis en exergue l'arrêt de production et de différenciation cercarienne en photopériode continue. Les travaux de Théron en 1982 ont montré l'existence d'un rythme mensuel de la production des cercaires et des indices d'abondance des différentes formes de maturation cercarienne. Cette accumulation des cercaires matures à l'intérieur des sporocystes nous permet de déduire que l'alternance jour/nuit est le zeitgeber principal qui contrôle ces deux autres rythmes ; et par extension que l'utilisation de la photopériode continue, impactant négativement la production et le développement cercarien, a peut-être perturbé l'observation du rythme journalier d'émission des cercaires. Cette déduction amène à penser qu'une expérience de photopériode continue est une condition trop stricte pour pouvoir mettre en évidence un rythme circadien de l'émission des cercaires chez *S. mansoni*. En effet, le test de la maintenance du rythme en photopériode continue est l'expérience la plus simple mais pas forcément la plus adaptée en fonction de la nature fonctionnelle du rythme et du modèle biologique. C'est le cas pour *Ostreococcus tauri* dont la régulation du cycle cellulaire sous le contrôle d'une horloge circadienne est dépendant de la lumière (Moulager, Monnier et al., 2007) ; de ce fait, une expérience d'obscurité continue ne permet pas de révéler le rythme circadien du cycle cellulaire chez *Ostreococcus tauri*.

Une expérience alternative qui pourrait mettre en évidence un rythme circadien chez *S. mansoni* serait l'utilisation de lumière pulsée. Aschoff explique que, lors d'une expérience de condition continue, « l'arrêt soudain ou l'amortissement du rythme observé n'est pas une preuve convaincante contre le rythme endogène ou pour le rythme exogène » (Aschoff, 1960).

Nos résultats des expériences de photopériode continue peuvent s'expliquer par deux modèles de régulation des rythmes des horloges endogènes. Il y a d'un côté le modèle du sablier (ou « Hourglass ») (Queiroz 1979 ; Rensing, Meyer-Grahlé & Ruoff, 2001) qui nécessite un unique signal du zeitgeber pour enclencher un cycle. Si le signal n'est pas renouvelé alors le rythme disparaît. En d'autres termes, si l'horloge endogène est symbolisée par un sablier et le rythme par le sable qui s'écoule, le zeitgeber enclenche le retournement du sablier qui entraîne l'écoulement du sable jusqu'à épuisement (Figure\_35\_A). S'il n'y pas de nouveau signal du zeitgeber pour retourner le sablier, le sable ne s'écoule pas et il n'y a plus de rythme (Figure\_35\_B). Un exemple moins trivial issu de Queiroz (1979) prend comme signal de zeitgeber « le passage jour/nuit qui déclenche le changement de concentration d'un

effecteur X qui doit atteindre un seuil permettant le démarrage des mécanismes conduisant à l'induction » du rythme.

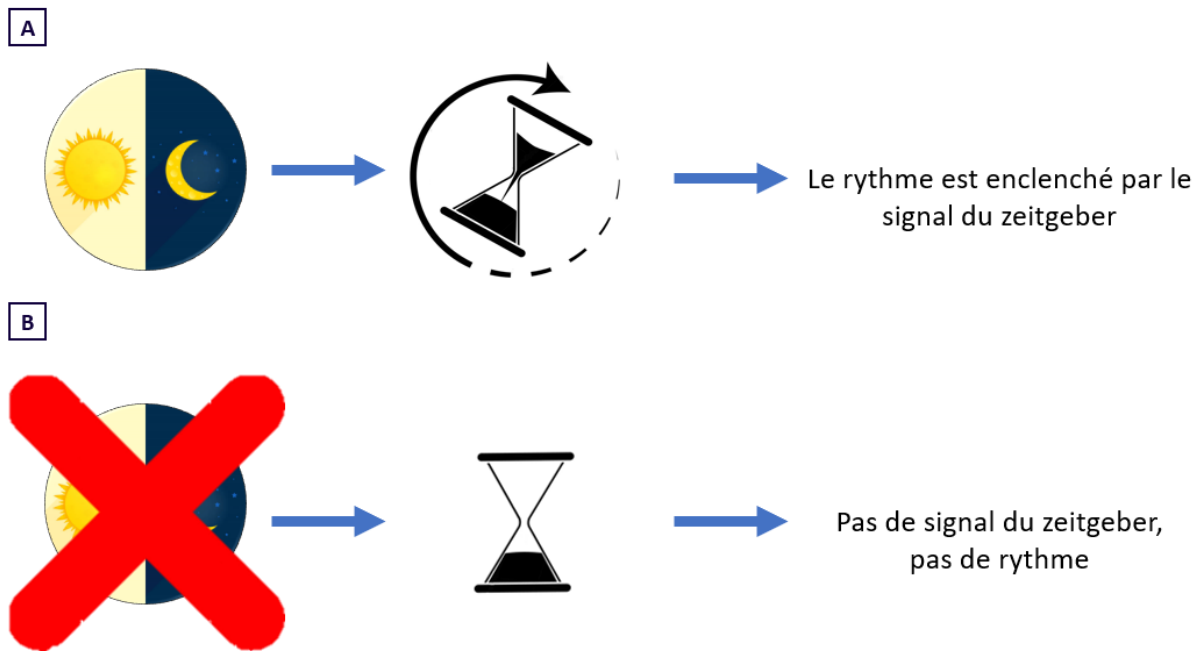
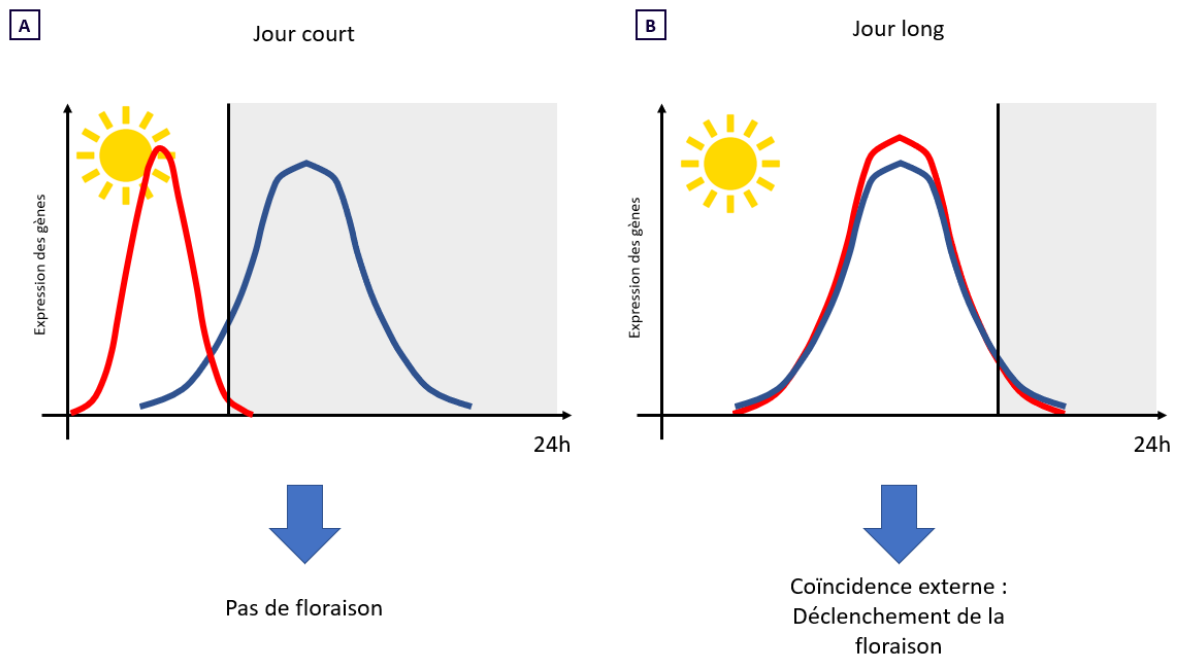


Figure 35 : Schéma du modèle du sablier (« hourglass »). A) Cas où le zeitgeber est présent. B) Cas où le zeitgeber est absent.

Le second modèle est celui de la coïncidence externe, théorisée par Bünning (*in* Queiroz 1979). Deux composantes doivent être intégrées par l'organisme pour que le rythme se déclenche. Une information externe issue du zeitgeber et une information interne provenant de l'horloge circadienne (signal moléculaire à un moment spécifique de la journée). Le rythme s'obtient lorsque les deux signaux coïncident. C'est le cas de la floraison chez *Arabidopsis thaliana* où l'expression du gène FKF1 (Flavin-Binding, Kelch Repeat, F-Box) contrôlée par l'horloge circadienne doit coïncider avec l'expression de GI (Gigantea) pour enclencher une cascade de régulation qui entraînera la floraison (Sawa, Nusinow et al., 2007). Cette coïncidence des deux expressions se déroule lors des jours longs (Figure\_36).



**Figure 36 : Schéma du modèle de la coïncidence externe à travers l'exemple simplifié de la floraison d'*Arabidopsis thaliana*.** A) Cas où les journées ont une période lumineuse courte. B) Cas où les journées ont une période lumineuse longue. Pour chaque diagramme, il y a en rouge l'expression du gène sous contrôle de l'horloge circadienne (FKF1) et en bleu l'expression du gène qui n'est pas régulée par l'horloge circadienne (GI). Les abscisses représentent les 24h d'une journée et les ordonnées le taux d'expression des gènes. La partie blanche correspond à la période de lumière et la partie grisée correspond à la période d'obscurité.

Pour discriminer les deux modèles et la nature endogène ou exogène du rythme d'émission des cercaires, il faudrait réaliser une expérience lumière pulsée (une à deux heures). Pour cela il faudrait tester, en parallèle, plusieurs groupes de mollusques parasités où chaque groupe recevrait un seul pulse de lumière à une heure différente de la journée. Si le rythme d'émission des cercaires s'enclenche après chaque pulse peu importe le groupe (6h après le début de la lumière pour le chronotype diurne et 12h pour le nocturne), alors le rythme est circadien et suit le modèle du sablier. Si le rythme d'émission se déclenche quand le pulse a été émis à une heure précise et pas dans les autres, alors le rythme est circadien et suit le modèle de la coïncidence externe. Si le rythme n'a pas été déclenché dans un seul des groupes, alors le rythme est exogène.

Une complexité supplémentaire à la compréhension du rythme d'émission des cercaires est que l'on ignore l'échelle du rythme observé. C'est-à-dire si le rythme d'émission des cercaires provient de la population des cercaires, d'un rythme propre au sporocyste ou d'une action conjointe des deux. La position d'attente des cercaires matures à l'intérieur des sporocystes en condition de photopériode continue ne permet pas d'exclure une des hypothèses. Ceci se rajoute au fait, que l'on ignore le mécanisme exact de sortie des cercaires des sporocystes. Chez *S. mansoni*, le sporocyste a un pore de sortie « birth pore » à une de ses extrémités qui s'ouvre et se ferme activement (Théron, 2015). Les observations histologiques issues de Théron (2015) ont pu mettre en évidence des cercaires qui sortent par ce pore ; ce qui serait en faveur soit d'un contrôle du sporocyste soit d'une action conjointe sporocyste-cercaire. Toutefois, on peut mettre en balance ces observations avec le fait que ce pore n'a pas été observé chez *S. bovis*, *S. haematobium* et *S. intercalatum* qui appartiennent au groupe des

œufs avec éperon terminal (Théron, 2015). L'hypothèse explicitée par cet auteur est que les cercaires de ces trois espèces sortiraient en déchirant la paroi des sporocystes allant dans le sens d'une action de la cercaire ou d'une action conjointe cercaires-sporocystes.

Pour conclure, cette première partie a permis de poser les bases du phénotype du rythme d'émission des cercaires pour les deux chronotypes de *S. mansoni* Oman (Figure\_37). En condition de photopériode équilibrée (L-O), les deux chronotypes partagent une émission unimodale, où ni l'espèce de l'hôte mollusque, ni le sexe du parasite n'ont d'influence sur le rythme. Ces observations minimisent le rôle que pourrait avoir l'hôte en tant que zeitgeber et vont plutôt dans le sens d'une perception directe de l'environnement par le parasite. Les résultats de chronobiologie pour les deux chronotypes n'ont pas mis en évidence une nature circadienne pour le rythme d'émission des cercaires. Un rythme circadien repose sur l'entraînabilité du rythme aux variations de zeitgeber, la constance du rythme et la compensation du rythme face à des changements de température. Si nos résultats montrent que les deux chronotypes s'adaptent à la variation du zeitgeber (directement ou indirectement), il n'y a pas de constance du rythme en condition continue (L-L ou O-O). Toutefois, le rythme est retrouvé dès le retour de la photopériode équilibrée. La dernière caractéristique qui pourrait trancher est la compensation du rythme face à la température mais elle n'a pas été testée dans cette thèse. Pour répondre à ce point, il faudrait observer l'influence de la température sur le rythme. C'est-à-dire regarder si le rythme d'émission des cercaires reste inchangé quand le mollusque est par exemple à 20°C, 26°C (condition contrôle) ou 30°C en photopériode équilibrée (L-O). D'autres expériences de chronobiologie peuvent être faites pour aiguiller la compréhension moléculaire du phénomène en nous interrogeant sur la perception du signal lumineux par le parasite. Dans l'environnement, les parasites ont accès à l'ensemble du spectre lumineux, or un photorécepteur capte une gamme précise du spectre lumineux. Pour déterminer quelle est la nature des photorécepteurs impliqués, il serait intéressant de remplacer la lumière blanche par une lumière colorée (rouge, bleue ou verte). Les photorécepteurs qui absorbent la lumière dans le rouge (type phytochrome) peuvent déjà être éliminés de l'expérimentation ; en effet, l'utilisation de la lumière pour nous éclairer lors des transferts de mollusques dans l'obscurité ne perturbe pas le rythme d'émission des cercaires.

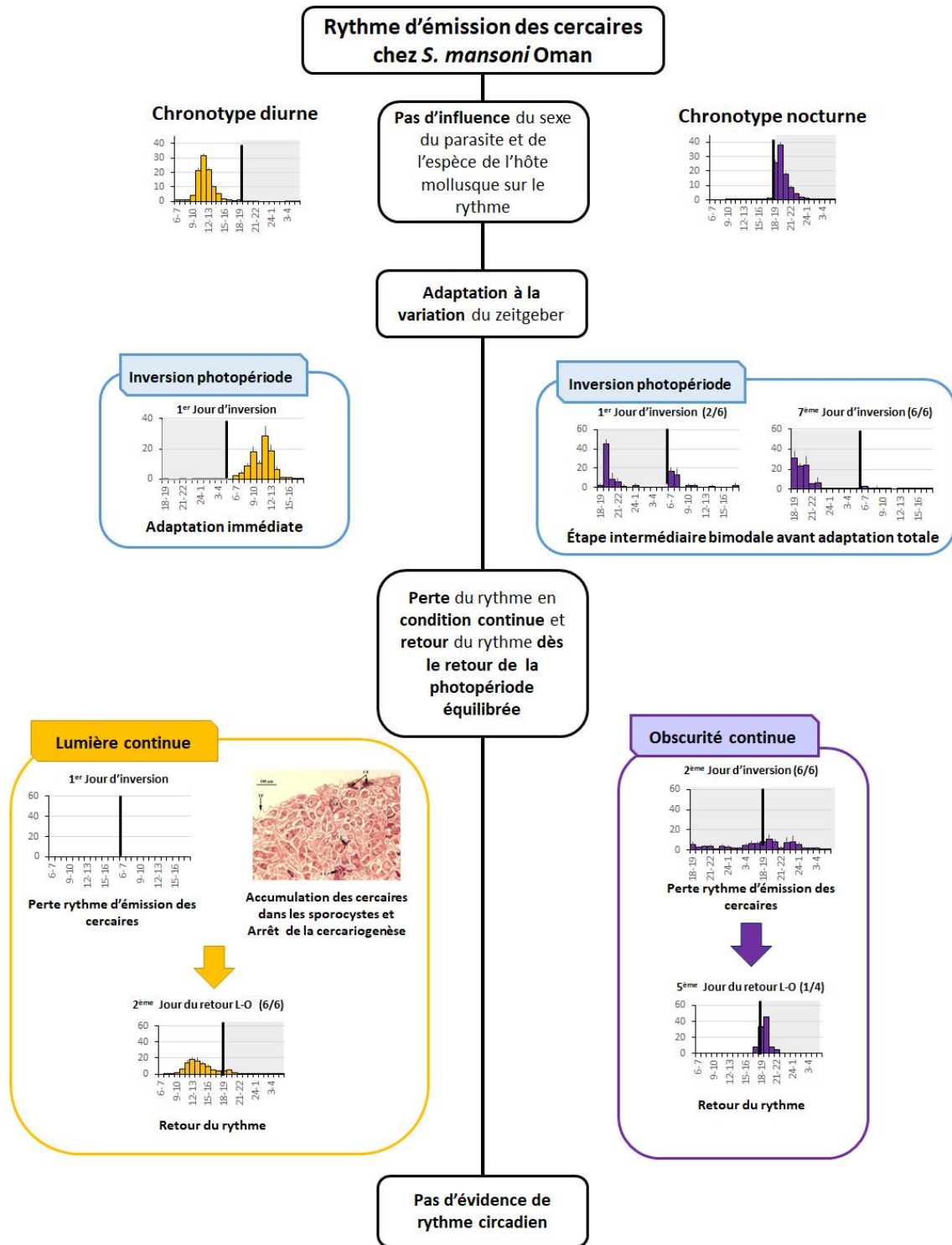


Figure 37 : Résumé des principaux résultats de la première partie.



## Partie II : Déterminants moléculaires du rythme d'émission des cercaires chez *Schistosoma mansoni*

Le caractère gonochorique des schistosomes permet de réaliser diverses expériences de croisements intra et interspécifiques. Ainsi, le déterminisme génétique du rythme d'émission des cercaires chez les schistosomes, et plus particulièrement chez *S. mansoni*, (Théron & Combes, 1983 ; Théron & Combes, 1988 ; Théron, 1989) a pu être démontré. Cette deuxième partie de thèse a pour objectif d'identifier les déterminants moléculaires du rythme d'émission des cercaires. Cet objectif se réalisera autour de trois approches complémentaires qui sont la génétique, l'épigénétique et la transcriptomique. L'intérêt de cette transversalité peut se comprendre au travers d'un exemple musical (Figure\_38).

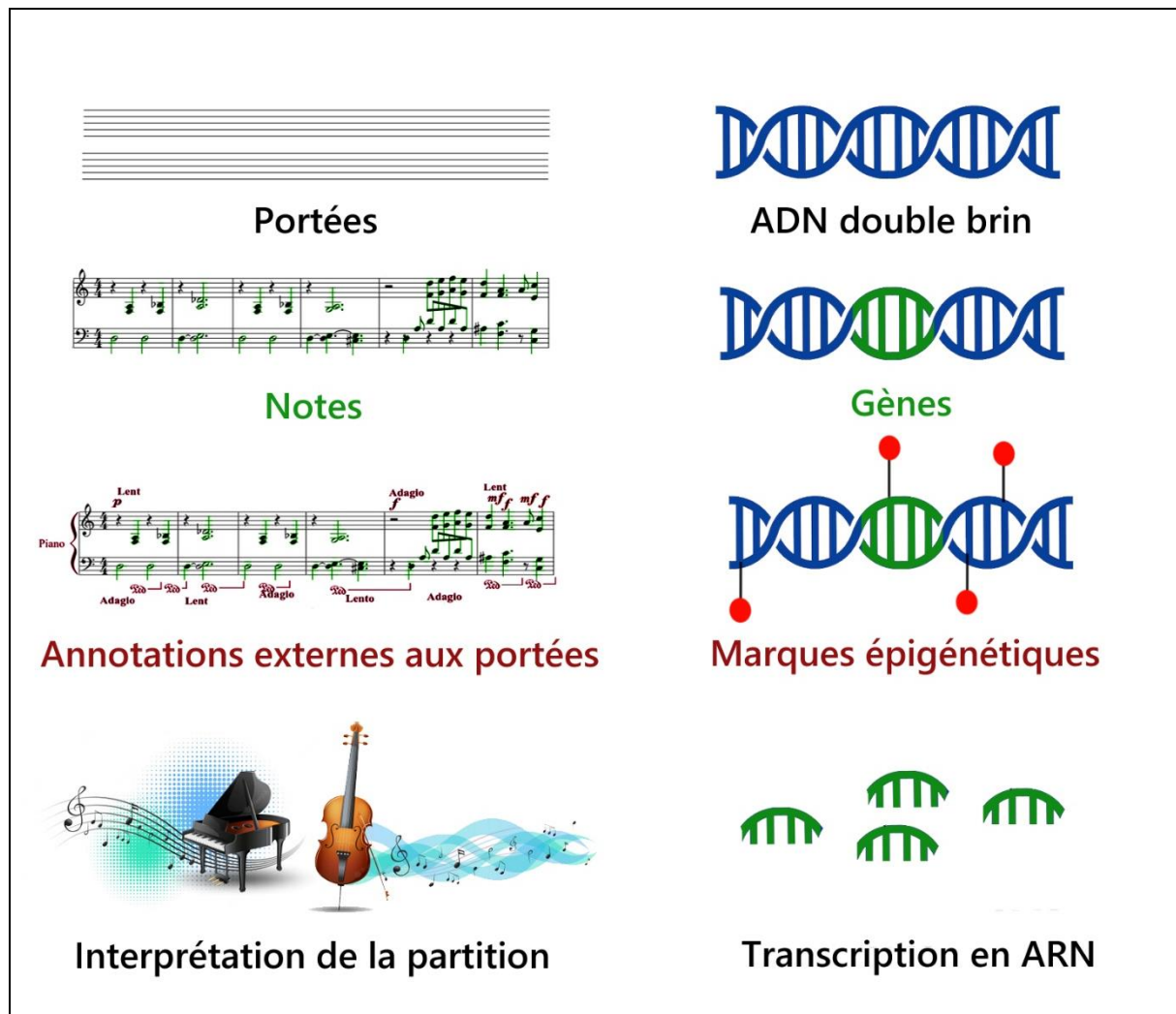


Figure 38 : Parallèle entre la partition musicale et l'approche moléculaire tri-disciplinaire (Génétique, Epigénétique et Transcriptomique). Explication ci-dessous.



Si un morceau de musique symbolise le phénotype, la partition musicale représente l'ADN où sont présentes toutes les informations nécessaires pour jouer le morceau de musique. Les portées représentent l'ADN double brin et les notes de musique les gènes. Toutes les annotations autour des portées qui indiquent les nuances à apporter à l'interprétation musicale, sans pour autant modifier l'ordre des notes, correspondent aux marques épigénétiques (méthylation de l'ADN dans l'exemple). Enfin, la transcriptomique est représentée par la mélodie jouée, qui dépend de la façon dont la partition est retranscrite, en prenant en compte ou non les annotations. Pour comprendre les subtilités de la musique, il est donc essentiel de ne pas uniquement s'intéresser aux notes (gènes) mais aussi de prendre en compte toutes les nuances de jeu apportées par les annotations (épigénétique) et les choix qui sont faits lors de l'interprétation de la partition (transcriptomique).

Le premier chapitre se concentrera sur l'aspect génétique avec la technique de linkage mapping pour identifier les gènes responsables du phénotype du rythme d'émission cercarienne.

Le deuxième chapitre portera sur l'étude des marqueurs moléculaires rythmiques avec une approche épigénétique et une approche transcriptomique partageant la même expérience d'échantillonnage biologique. L'épigénétique sera abordée au travers de la technique de ChIPmentation et la transcriptomique au travers de la technique de RNA-seq. Dans les deux cas, le but sera de déterminer s'il y a un rythme d'acétylation et méthylation des histones/expression des gènes pour mettre en lumière d'autres potentiels gènes candidats. Enfin, l'approche transcriptomique permettra de faire le lien entre les différentes approches moléculaires et d'affiner la liste des gènes candidats.

### II.1. : Linkage mapping

De façon générale, la génétique étudie la transmission de caractères héréditaires. Ces caractères héréditaires reposent sur l'ADN et plus particulièrement sur les gènes. L'étude d'un phénotype à travers le prisme génétique consiste donc à s'intéresser aux gènes et à la façon dont ils sont transmis aux générations suivantes. L'ADN est regroupé en chromosomes et chaque chromosome est présent en deux exemplaires, chaque exemplaire provenant de l'un des deux parents dans le cas d'une reproduction sexuée (méiose). Cette paire de chromosomes contient les mêmes gènes mais pas forcément les mêmes versions, on parle alors d'allèles pour les différentes versions d'un même gène présentes sur des chromosomes homologues. Le génotype est alors la combinaison des deux allèles d'un même gène dont l'expression va produire un phénotype. Pour un gène donné, un individu qui porte deux allèles identiques est appelé homozygote. Si les deux allèles sont différents, l'individu est alors hétérozygote. L'expression du génotype dépend de l'interaction entre les allèles. On parle d'allèle dominant si la présence d'un seul exemplaire est suffisante pour s'exprimer et il est à l'inverse récessif si la présence des deux exemplaires est nécessaire à son expression.

Le rythme d'émission des cercaires est un caractère héréditaire chez les schistosomes. Il a été prouvé chez *S. mansoni* au travers de croisements intraspécifiques et interspécifiques.

- Lors du croisement intraspécifique entre ***S. mansoni* Guadeloupe (chronotype précoce)** et ***S. mansoni* Brésil (chronotype tardif)** (Théron & Combes, 1983), les progénitures présentent un profil d'émission **bimodal équilibré** entre les pics précoce et tardif parentaux, ce qui illustre la **codominance** des gènes responsables du rythme de l'émission des cercaires pour les deux chronotypes.
- Le croisement intraspécifique entre les **chronotypes précoce et tardif de *S. mansoni* Guadeloupe** (Théron & Combes, 1988), donne des progénitures avec un **profil unimodal au**

**milieu des deux pics parentaux**, ce qui indique que les deux chronotypes sont dus à une variation allélique du même gène. Par contre, l'utilisation du chronotype tardif d'une autre souche (Brésil) a montré que les deux chronotypes sont dus à deux gènes différents (Théron & Combes, 1988).

- Le croisement interspécifique entre *S. mansoni* Guadeloupe précoce et *S. rodhaini* (nocturne) (Théron, 1989) donne des progénitures avec un **profil bimodal** des deux pics parentaux, dont **le pic le plus important correspond au chronotype précoce**.
- A l'inverse, le croisement interspécifique entre *S. mansoni* Guadeloupe tardif et *S. rodhaini* (nocturne) (Théron, 1989) donne des progénitures avec un **profil bimodal** des deux pics parentaux, dont **le pic le plus important correspond au chronotype nocturne**. Ceci suggère que le rythme d'émission des cercaires est polygénique (Théron, 1989).

Ce déterminisme génétique du rythme d'émission des cercaires, couplé à la facilité de croiser les espèces grâce à leur caractère gonochorique ainsi que la bonne qualité d'assemblage du génome de *S. mansoni* version 7 sont les conditions idéales pour faire du linkage mapping (Anderson, LoVerde et al., 2018). Le linkage mapping, et plus particulièrement la technique de QTL (Quantitative Trait Loci) mapping, est le fait d'expliquer la variation phénotypique d'une population par sa variation génotypique (Broman & Sen, 2009). Cette technique permet donc de détecter des QTLs qui contribuent à la variation d'un trait quantitatif (phénotype). L'identification de ces loci se fait grâce à des expériences de deux croisements consécutifs (rétrocroisement ou croisement test). Le rétrocroisement ou « back-cross » est le croisement entre un individu progéniture F1 et un de ses parents F0 (Figure\_39\_A), il permet de détecter les QTL pour un caractère récessif. Le croisement test ou « intercross » est le croisement entre deux progénitures F1, il permet de détecter les QTLs pour un caractère dominant (Figure\_39\_B).

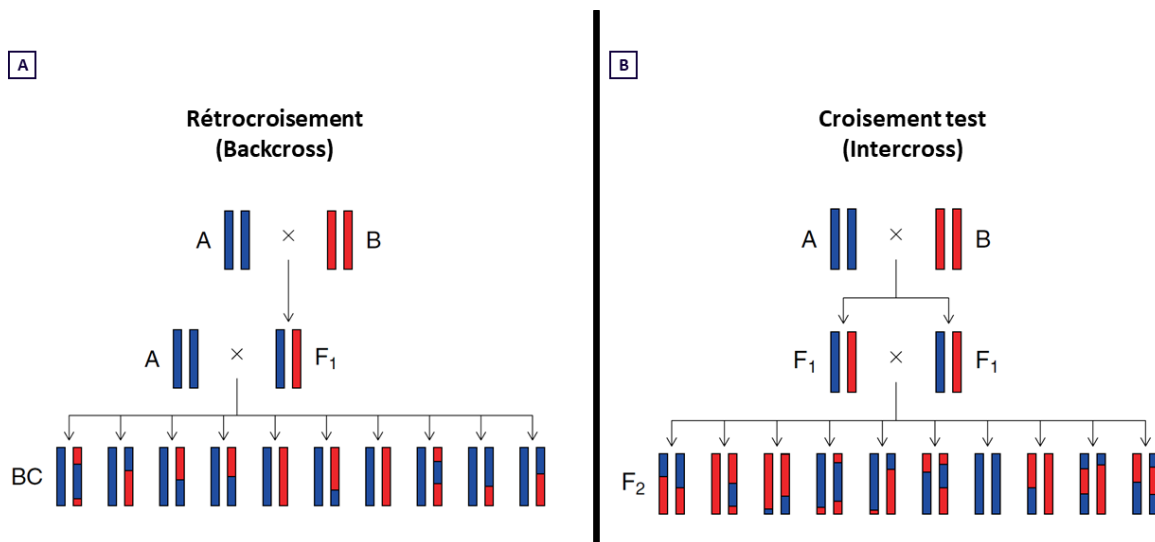


Figure 39 : Schéma de la ségrégation de chromosomes autosomiques dans le cas d'un rétrocroisement et d'un croisement test. (Figure modifiée à partir de Broman & Sen (2009)). A et B sont deux souches aux phénotypes distincts dont les chromosomes sont respectivement représentés en bleu et rouge A) Rétrocroisement : Le phénotype étudié est récessif chez la souche A. B) Croisement test : Le phénotype étudié est dominant chez la souche A et/ou la souche B.

Dans les deux cas, le croisement initial, dit F0, s'opère entre deux individus homozygotes aux phénotypes distincts (ex : les chronotypes diurne et nocturne) (Figure\_39). Ce croisement F0 donne une génération F1 hétérozygote qui a un génotype à 50% maternel et à 50% paternel. Le troisième croisement donnera soit une génération Backcross BC, soit une génération F2 (en fonction du type de croisement) où les individus de cette troisième génération présentent une mosaïque de variation de pourcentage des génotypes parentaux due à la recombinaison lors de la méiose (Figure\_39). La détection des QTLs se fait sur cette troisième génération (F2 ou BC) au travers d'une régression statistique entre la variation génotypique de chaque individu F2 et la variation de leur phénotype. Plus le nombre d'individus F2 est grand, meilleure est la précision de la détection.

L'objectif de ce chapitre est de déterminer quels sont les loci impliqués dans le rythme d'émission des cercaires chez *Schistosoma mansoni*. Le rythme d'émission des cercaires étant un caractère co-dominant entre les différents chronotypes, deux croisements tests ont été réalisés pour répondre à cet objectif : un croisement test (intercross) intraspécifique entre les deux chronotypes (diurne et nocturne) de *S. mansoni* et un autre croisement test (intercross) interspécifique entre le chronotype diurne de *S. mansoni* Oman et un schistosome nocturne *S. rodhaini*. L'intérêt de ce second croisement est de voir si les loci impliqués dans le rythme d'émission du chronotype nocturne de *S. mansoni* sont les mêmes que ceux de *S. rodhaini*.

Le premier point de ce chapitre sera consacré à une présentation de la méthode utilisée, le second point présentera les résultats du QTL mapping pour le croisement intraspécifique et le troisième point pour le croisement interspécifique.

### II.1.1. Méthode du linkage QTL mapping

La méthode de QTL mapping se divise en quatre grandes étapes, les croisements et les phénotypages, la préparation moléculaire des échantillons, le génotypage et enfin la détection des QTLs. Ces expériences sont issues d'une collaboration avec l'équipe de Tim Anderson du Texas Biomedical Research Institute. Les croisements et le phénotypage sont réalisés à Perpignan tandis que la préparation moléculaire et le génotypage sont faits à San Antonio. L'analyse QTL est faite entre les deux sites.

Les croisements consistent à générer, phénotyper et échantillonner les générations F0, F1 et F2. La production des parasites a été détaillée dans les méthodes de l'entretien du cycle ; ils sont obtenus à partir d'une infestation mono-miracidiale de *B. pfeifferi*. Chaque mollusque contient le génome d'un seul individu parasite. Pour la génération F0 et F1, une étape supplémentaire de sexage a été faite pour faire les croisements (cf 4. Méthode du sexage des parasites au stade cercaire). Le phénotype utilisé dans nos expériences est la moyenne du pourcentage de cercaires nocturnes (trait quantitatif). Le phénotypage se fait au travers d'une expérience de chronobiologie de trois jours consécutifs en photopériode équilibrée (L-O 12-12) (cf 3. Méthode de chronobiologie de l'émission cercarienne). Afin d'éviter tout biais sexuel dans les résultats, les croisements et leurs réciproques sont réalisés à chaque expérience.

La préparation moléculaire se répartit entre l'extraction de l'ADN, la préparation des banques et le séquençage. Pour le cas du croisement intraspécifique, seuls les exomes (partie exprimée des gènes) générés à partir de vers adultes conservés dans l'alcool à 90° ont été utilisés pour générer les banques (Whole Exome Sequencing). Pour le croisement interspécifique, tout le génome, généré à partir de cercaires plongées dans l'azote liquide et conservées à -80°C, a été utilisé (Whole Genome Sequencing). Pour éviter un effet de lot (batch effect), l'extraction d'ADN des échantillons a été réalisée dans un ordre aléatoire généré par ARTS (Automated Randomization of multiple Traits for Study design) (Maienschein-Cline, Lei et al., 2014).

Le génotypage consiste dans un premier temps à faire des alignements contre la version 7 du génome de *S. mansoni*. On crée un fichier VCF (Variant Call file) qui répertorie les variations des séquences génétiques, et plus particulièrement dans notre cas les SNPs (Single Nucleotide Polymorphism). Grâce à VCFtool, on sélectionne les SNPs qui ont une ségrégation de type mendélienne c'est-à-dire une répartition dans la F2 de ~25% homozygotes pour le parent F0 diurne (DD) , ~25% homozygotes pour le parent F0 nocturne (NN) et ~50% hétérozygotes (DN et ND). On crée ensuite un fichier génotype (csvs) qui contient tous les individus F2, avec, pour chaque marqueur sélectionné, la position du marqueur et le génotype correspondant (DD, NN ou ND/DN) ou NA quand il est impossible de statuer. Pour l'analyse QTL, seuls les échantillons dont le génotype contient moins de 20% de données manquantes (NA) sont utilisés.

La détection des QTLs se fait avec la version 1.47-9 de R/qtl (Broman & Sen 2009). Elle se divise en plusieurs étapes, la vérification de la distribution des phénotypes, la vérification des marqueurs génotypiques, l'analyse QTL et la détermination de la transmission allélique (Figure\_40). La vérification des phénotypes consiste à tester la normalité des échantillons pour déterminer quel modèle utiliser. La vérification des marqueurs génotypiques consiste à repérer les potentielles erreurs. L'analyse QTL consiste à scanner la présence d'un ou de plusieurs QTLs en déterminant un LOD score QTL (logarithm of odds). Pour cela, un seuil de détection avec un risque de 5% est déterminé avec 1000 permutations. Si plusieurs QTLs sont détectés (supérieur au seuil), il faut alors déterminer le type d'interaction qu'il y a entre eux (épistasie ou effet additif). Enfin, la détermination de la transmission allélique sert à comprendre comment les allèles parentaux F0 diurne et nocturne interagissent (dominance, récessivité). Cela se fait en regardant la répartition des génotypes (DD, ND/DN et NN) en fonction de la distribution du phénotype pour le marqueur QTL qui a le LOD score le plus haut. Si les répartitions des génotypes sont significativement différentes entre les trois groupes (DD, ND/DN et NN), les allèles parentaux D et N sont co-dominants. Si un groupe homozygote (ex : NN) est significativement différent des deux autres (DD et DN/ND) et que ces derniers ne sont pas significativement différents entre eux, alors l'allèle N est récessif et l'allèle D est dominant.

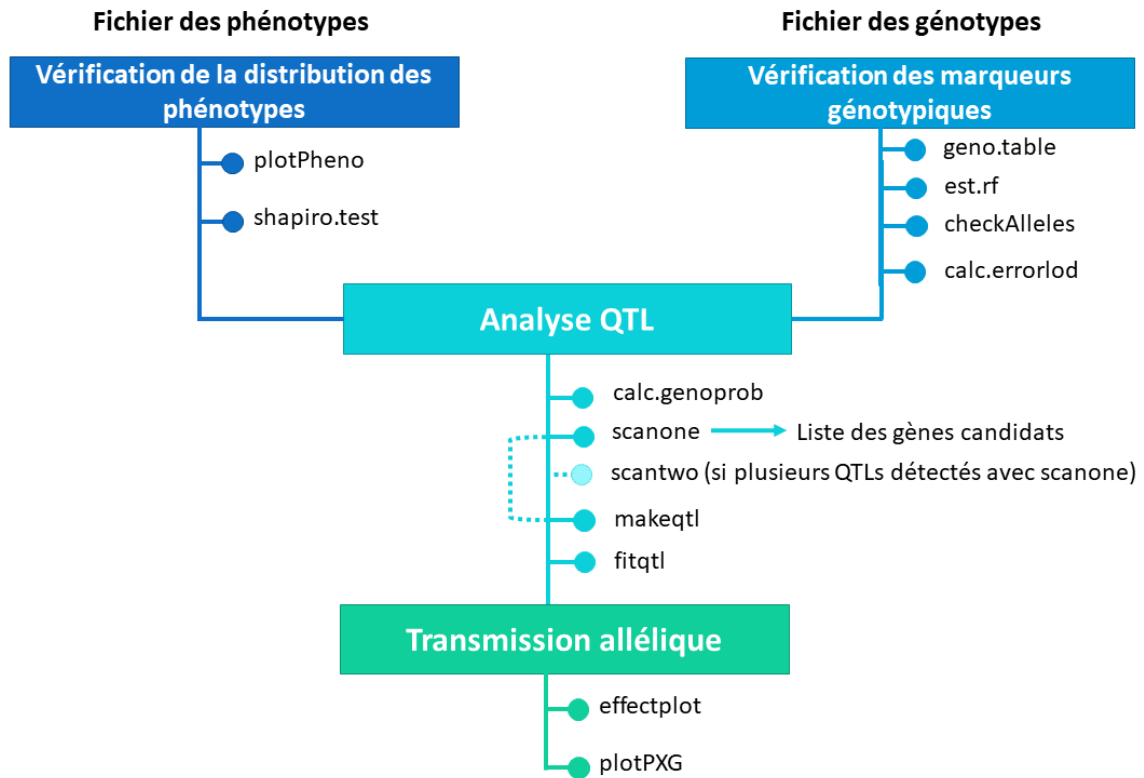


Figure 40 : Schéma de l'analyse QTL avec les fonctions utilisées dans R/qtl. Détail des fonctions (Broman & Sen 2009) : *PlotPheno* affiche le graphique des phénotypes et *shapiro.test* évalue la normalité des données. *Geno.table* crée une table des génotypes pour chaque marqueur et permet de déterminer si la répartition des marqueurs suit une ségrégation mendélienne. *Est.rf* estime la fraction de recombinaison par paire entre les marqueurs. *checkAllele* identifie les marqueurs avec des allèles potentiellement inversés. *calc.errorlod* calcule les LOD score des erreurs de génotypage. *Calc.genoprob* calcule les probabilités des génotypes conditionnels. *Scanone* analyse le génome avec un modèle de QTL unique. *Scantwo* fait une analyse bidimensionnelle du génome avec un modèle à deux QTLs et permet d'identifier la nature des interactions quand plusieurs QTLs sont détectés. *Makeqtl* crée un objet QTL pour utiliser la fonction *fitqtl* qui fait une analyse multidimensionnelle du génome. *Fitqtl* détermine ainsi le pourcentage de variation phénotypique expliquée par le QTL. *Effectplot* trace les moyennes phénotypiques de groupes de génotype définis par un marqueur QTL. *plotPXG* a la même fonction que *Effectplot* mais génère à la place un dot plot.

## II.1.2. Croisement intraspécifique : *Schistosoma mansoni* diurne X *Schistosoma mansoni* nocturne

Les croisements *Schistosoma mansoni* diurne avec *S. mansoni* nocturne ont été réalisés avant cette thèse. Les croisements et le phénotypage ont été réalisés par Hélène Moné et Gabriel Mouahid. La préparation moléculaire des échantillons jusqu'à l'analyse a été réalisée par l'équipe de Tim Anderson (Texas Biomedical Research Institute). Les premières analyses QTL ont été réalisées en utilisant la version 5 du génome de *S. mansoni* (Protasio, Tsai et al., 2012). Une nouvelle analyse QTL a été réalisée par Frédéric Chevalier (Texas Biomedical Research Institute) avec la version 7 du génome de *S. mansoni* (GCA\_000237925.3). Ce sont les résultats de cette nouvelle analyse et plus particulièrement la nouvelle liste de gènes qui vont être présentés ci-dessous.

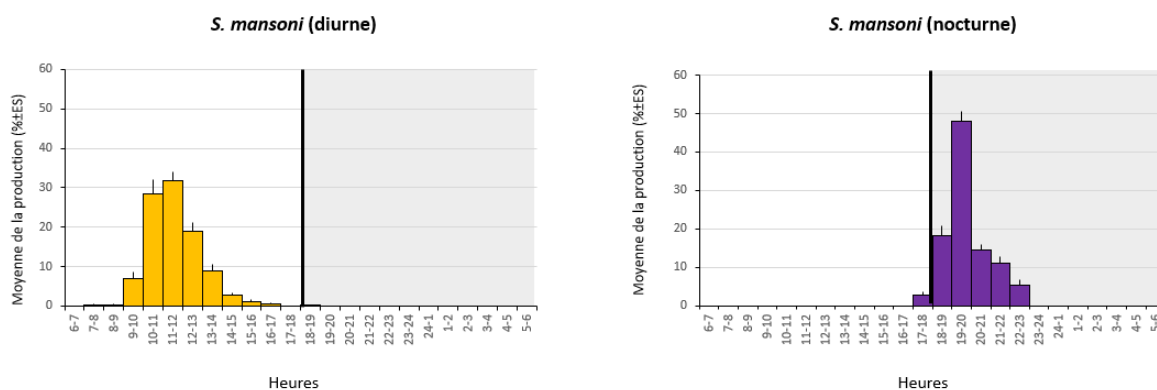
### II.1.2.1. Protocole

Les parasites F0 utilisés pour ce croisement sont les chronotypes diurne et nocturne de *S. mansoni* (Oman). Les hôtes intermédiaires mollusques sont des *Biomphalaria pfeifferi* originaires de la région du Dhofar à Oman. Les croisements réciproques se nomment croisement A (F0 mâle *S. mansoni* diurne X F0 femelle *S. mansoni* nocturne) et croisement B (F0 mâle *S. mansoni* nocturne X F0 femelle *S. mansoni* diurne). Pour cette nouvelle analyse, les individus F2 des croisements A et B ont été utilisés en une seule analyse QTL.

### II.1.2.2. Présentation des phénotypes

#### II.1.2.2.1. Les phénotypes F0

Les résultats moyens, accompagnés des variations individuelles, sont présentés dans la Figure\_41 et le Tableau\_24. Pour *S. mansoni* diurne, 18 profils ont émis plus de 20 cercaires et 15 profils ont émis plus de 20 cercaires pour le chronotype nocturne.



**Figure 41 : Chronobiologie des chronotypes diurne (orange) et nocturne (violet) de *S. mansoni* pour l'expérience de QTL mapping.** Chaque histogramme a, en ordonnées, la moyenne du pourcentage des cercaires émises ± écart standard. L'axe des abscisses représente les 24h de la journée, débutant à 6h, début du jour, et finissant à 6h, fin de la nuit. L'arrière-plan gris correspond à l'obscurité. La ligne noire marque le début de l'obscurité à 18h.

**Tableau 24 : Paramètres des profils moyens d'émission, accompagnés des variations individuelles, pour les parasites diurne et nocturne de *S. mansoni* pour l'expérience de QTL mapping**

Paramètres	Moyenne $\pm$ ES Variations individuelles		Test de Wilcoxon- Mann-Whitney
	<i>S. mansoni</i> diurne (F0)	<i>S. mansoni</i> nocturne (F0)	
Effectif	18	15	
Photopériode	L-O	L-O	
Nombre de pics	1 1	1 1	
Heure du pic	11-12h 10-11h à 12-13h	19-20h 19-20h	$\neq$
Plage d'émission (h)	4,4 $\pm$ 0,2 3 à 7	4,2 $\pm$ 0,2 3 à 6	W =148, $p = 0,6$
Intensité de la plage (%)	95,1 $\pm$ 0,8 88,2 à 99,1	95,4 $\pm$ 1,1 86,6 à 100	W =106, $p = 0,3$
Asymétrie	+0,3 $\pm$ 0,2 -1,5 à +1,9	+1 $\pm$ 0,3 -1,6 à +2,1	W = 77, $p = 0,04$
Aplatissement	-0,9 $\pm$ 0,6 (N=17) -5,9 à +3,8	+2 $\pm$ 0,7 (N=12) -4,1 à +4,5	W = 35, $p = 0,002$
Cercaires la nuit (%)	0,2 $\pm$ 0,1 0,0 à 1,6	97,2 $\pm$ 1 87,7 à 100	W =0, $p = 3,2e-07$
Nombre total de cercaires	202,1 $\pm$ 28,9 55 à 497	144 $\pm$ 20,8 75 à 328	W =177, $p = 0,1$

Le chronotype diurne a une émission unimodale dont l'heure du pic moyen est à 11-12h. La plage horaire moyenne d'émission est d'environ 4h avec une intensité de 95%. La plage d'émission moyenne est symétrique avec une distribution aplatie. Il n'y a quasiment pas d'émission de cercaires nocturne.

Le chronotype nocturne a une émission unimodale dont l'heure du pic moyen est à 19-20h. La plage horaire moyenne d'émission est d'environ 4h avec une intensité de 95%. La plage d'émission moyenne a une asymétrie sur la droite avec une distribution centrée. L'émission nocturne est majoritaire avec en moyenne 97,2% des cercaires émises la nuit.

Les comparaisons sont présentées dans le Tableau\_24. Les profils moyens des deux chronotypes se différencient par l'heure du pic (11-12h pour les diurnes et 19-20h pour les nocturnes), le pourcentage de cercaires émises la nuit (0% pour les diurnes et 98% pour les nocturnes ;  $p < 0,01$ ) et le coefficient d'aplatissement (-0,9 pour les diurnes et +2 pour les nocturnes ;  $p < 0,01$ ). La plage d'émission du chronotype nocturne est plus centrée autour de l'heure du pic que le chronotype diurne. Les deux chronotypes partagent la présence d'un seul pic. De plus, les plages d'émission, l'intensité d'émission, leurs coefficients d'asymétrie et le nombre total de cercaires ne sont pas significativement différents entre les diurnes et les nocturnes ( $p > 0,01$ ).

#### II.1.2.2.2. Le croisement A (Figure\_42)

Le croisement intraspécifique A est un croisement entre un mâle F0 *S. mansoni* diurne et une femelle F0 *S. mansoni* nocturne. La génération F1 issue de ce croisement a un profil bimodal, avec un pic principal vers 13-14h et un pic secondaire vers 19-20h. Le pourcentage de cercaires nocturne moyen est de 23,57%. La génération F2, issue du croisement entre un mâle F1 et une femelle F1, contient 113 individus. Le pourcentage de cercaires nocturne moyen est de 24,29% avec des valeurs individuelles allant de 0% à 87,99%. La répartition des profils de chronobiologie de la F2 est de 25,7% ressemblant à la F0 diurne (1ère ligne des F2 ; Figure\_42), 13,3% ressemblant à la F0 nocturne (dernière ligne des

Partie II : Linkage mapping

F2 ; Figure\_42) et 61,1% hybride ou inclassable (les 2<sup>ème</sup> et 3<sup>ème</sup> lignes des F2 ; Figure\_42). L'ensemble des profils F2 sont à retrouver en Annexe 3.

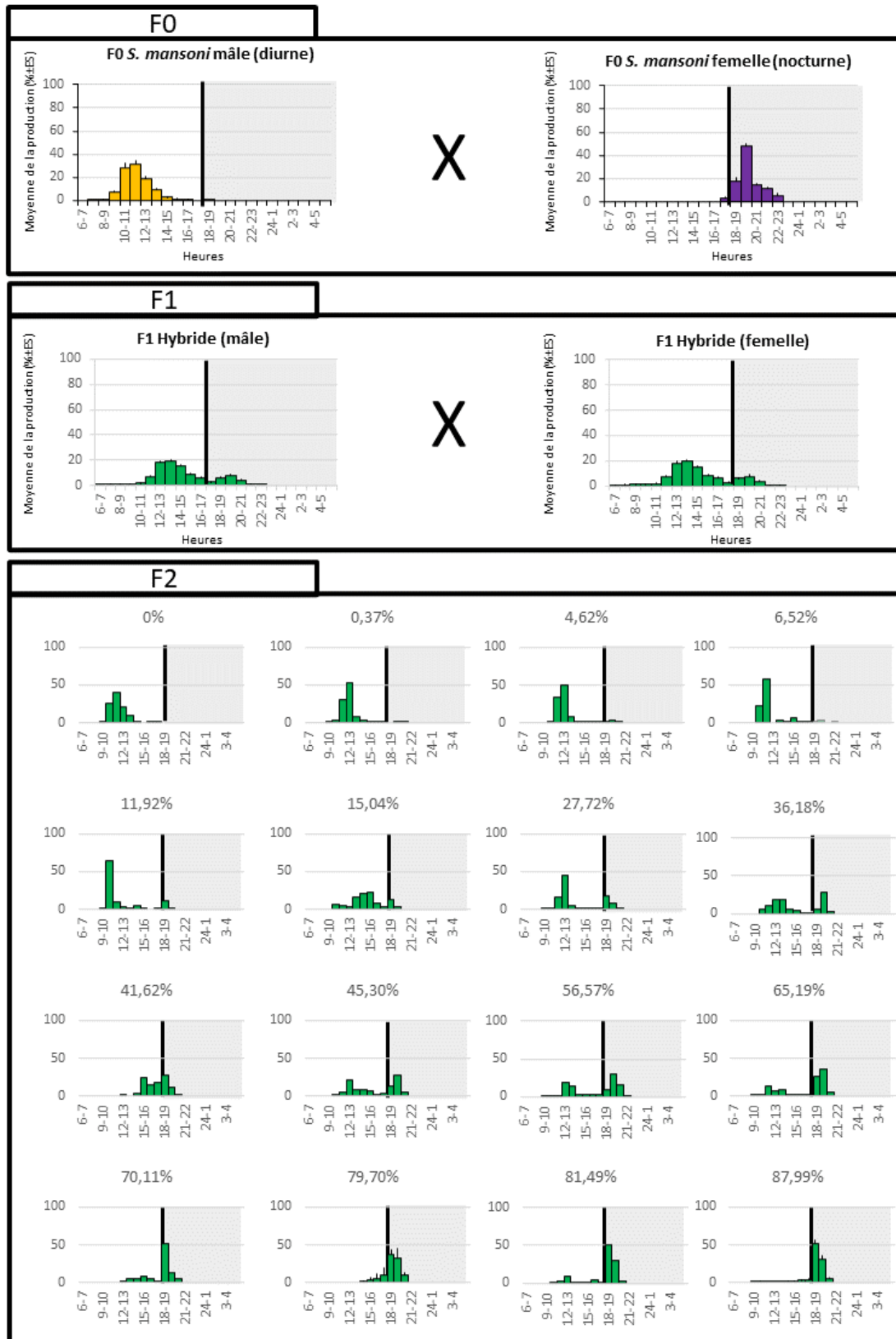


Figure 42 : Chronobiologie du croisement A intraspécifique, avec F0 diurne (orange), F0 nocturne (violet) et une sélection d'individus F2 hybrides (vert).



Chaque histogramme a, en ordonnées, la moyenne du pourcentage des cercaires émises  $\pm$  écart standard ; au-dessus de chacun des individus F2 est indiqué le pourcentage total des cercaires émises la nuit. L'axe des abscisses représente les 24h de la journée, débutant à 6h, début du jour, et finissant à 6h, fin de la nuit. L'arrière-plan gris correspond à l'obscurité. La ligne noire marque le début de l'obscurité à 18h. Les individus F2 sont une sélection des divers pourcentages de cercaires nocturnes. Ils sont classés du plus faible pourcentage de cercaires nocturne au plus élevé. La première ligne des F2 représente une sélection d'individus proche du phénotype F0 diurne. Les 2<sup>ème</sup> et 3<sup>ème</sup> lignes représentent une sélection d'individus au phénotype hybride. La 4<sup>ème</sup> ligne représente une sélection d'individus proches du phénotype F0 nocturne.

### II.1.2.2.3. Le croisement B (Figure\_43)

Le croisement intraspécifique B est un croisement entre un mâle F0 *S. mansoni* nocturne et une femelle F0 *S. mansoni* diurne. La génération F1 issue de ce croisement a un profil bimodal, avec un pic principal vers 13-14h et un pic secondaire vers 18-19h. Le pourcentage de cercaires nocturne moyen est de 14,13%. La génération F2, issue du croisement entre un mâle et une femelle F1, contient 102 individus. Le pourcentage de cercaires nocturne moyen est de 26,36% avec des valeurs individuelles allant de 0% à 90,16%. La répartition des profils de chronobiologie de la F2 est de 24,5% ressemblant à la F0 diurne (1<sup>ère</sup> ligne des F2 ; Figure\_43), 14,7% ressemblant à la F0 nocturne (dernière ligne des F2 ; Figure\_43) et 60,8% hybride ou inclassable (les 2<sup>ème</sup> et 3<sup>ème</sup> lignes des F2 ; Figure\_43). L'ensemble des profils F2 sont à retrouver en Annexe 3.

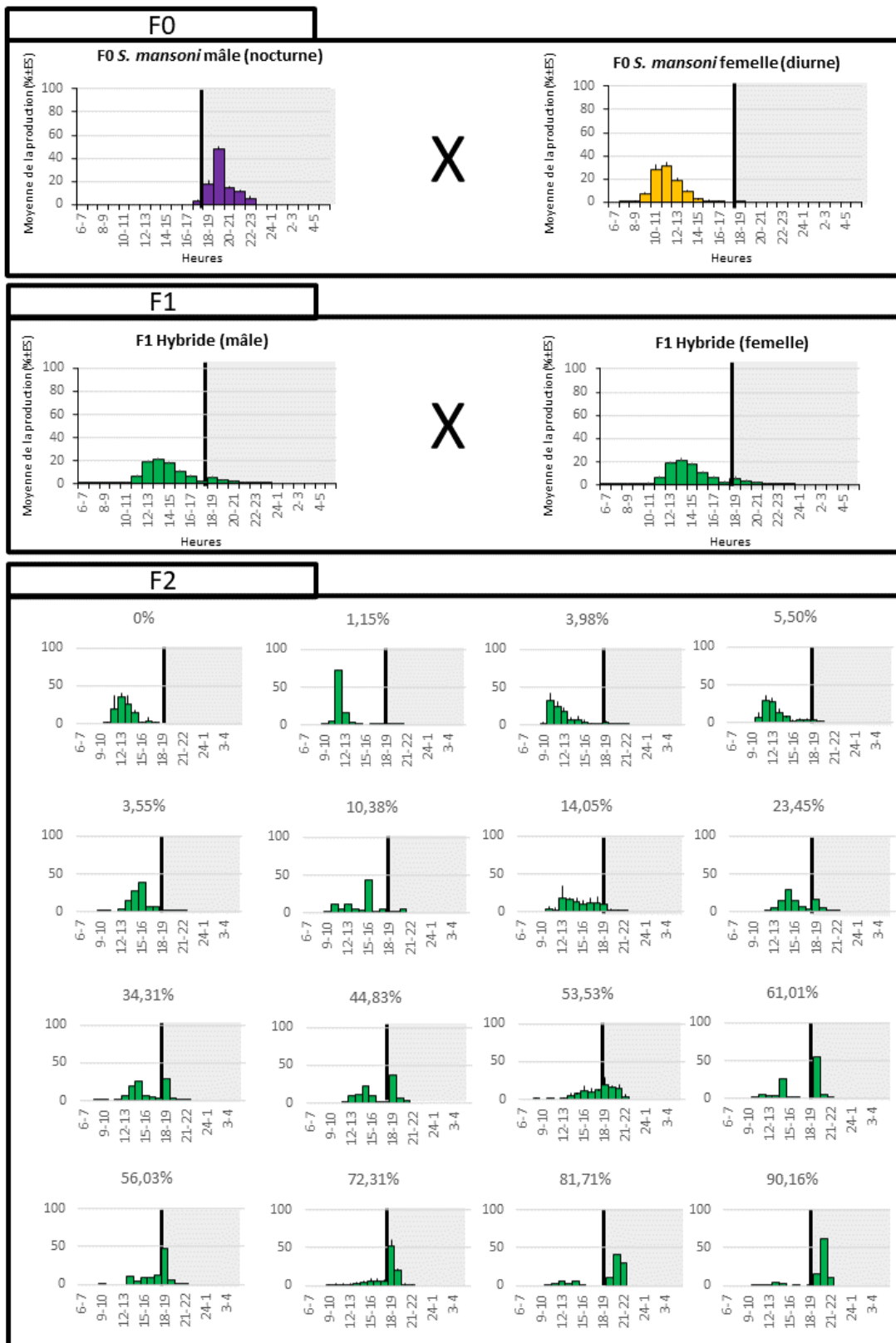


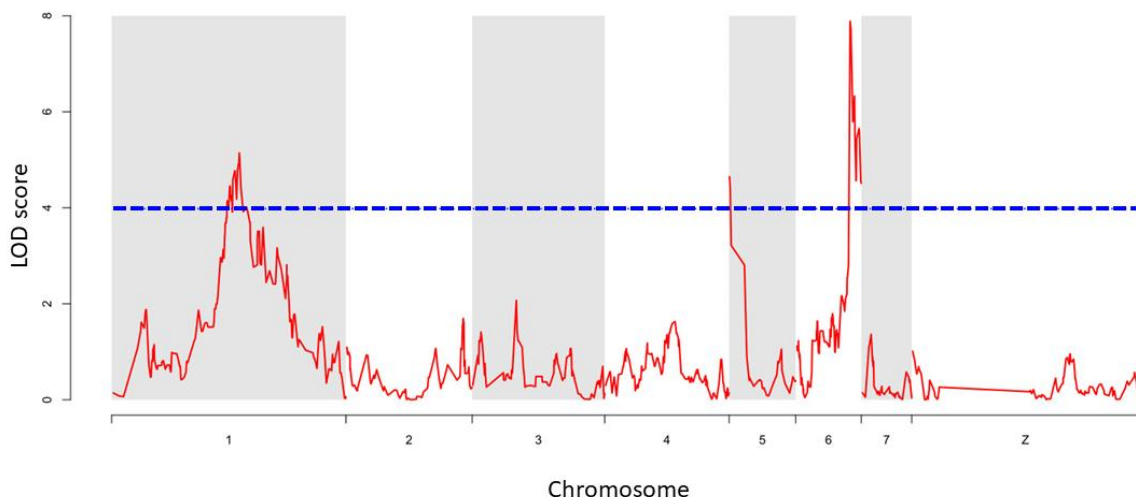
Figure 43 : Chronobiologie du croisement B intraspécifique, avec F0 diurne (orange), F0 nocturne (violet), F1 et une sélection d'individus F2 hybrides (vert).

Chaque histogramme a, en ordonnées, la moyenne du pourcentage des cercaires émises  $\pm$  écart standard ; au-dessus de chacun des individus F2 est indiqué le pourcentage total des cercaires émises la nuit. L'axe des abscisses représente les 24h de la journée, débutant à 6h, début du jour, et finissant à 6h, fin de la nuit. L'arrière-plan gris correspond à l'obscurité. La ligne noire marque le début de l'obscurité à 18h. Les individus F2 sont une sélection des divers pourcentages de cercaires nocturnes. Ils sont classés du plus faible pourcentage de cercaires nocturne au plus élevé. La première ligne des F2 représente une sélection d'individus proches du phénotype F0 diurne. Les 2ème et 3ème lignes représentent une sélection d'individus au phénotype hybride. La 4ème ligne représente une sélection d'individus proches du phénotype F0 nocturne.

Il n'y a pas de différence significative entre les profils d'émission des générations F1 des croisements intraspécifiques A et B. De plus, le pourcentage de cercaires nocturne moyen n'est pas significativement différent entre les générations F1 et F2 des deux croisements intraspécifiques ( $p > 0,01$ ). Ces résultats sont en accord avec ceux de la première partie qui montraient que le sexe, et donc ici le sens du croisement, n'a pas d'influence sur le rythme d'émission des cercaires.

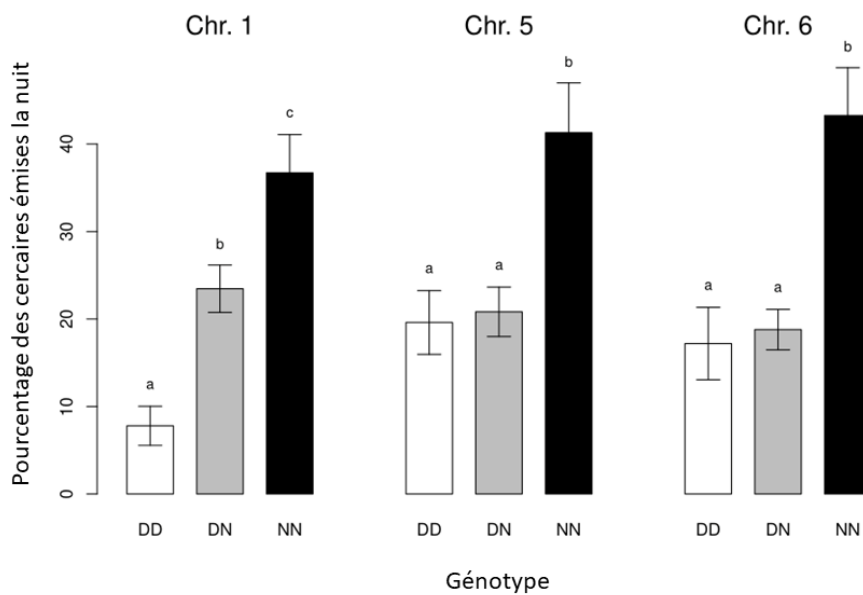
### II.1.2.3. QTL mapping

Cette nouvelle analyse QTL a été réalisée en utilisant conjointement les individus F2 des croisements A et B. La préparation moléculaire des échantillons s'est faite en récupérant uniquement l'exome. Le phénotype utilisé est la moyenne du pourcentage de cercaires nocturne. Elle a été réalisée avec l'algorithme EM (Expectation Maximisation pour maximisation de l'espérance) de la fonction scanone du package R/qtl avec un seuil de détection à 5% déterminé avec 1000 permutations. Cette nouvelle analyse révèle 3 QTLs sur les chromosomes 1, 5 et 6 (Figure\_44), c'est-à-dire les chromosomes qui ont un LOD score supérieur au seuil fixé à 5% (ligne pointillée).



**Figure 44 : LOD score pour les croisements intraspécifiques de *Schistosoma mansoni* obtenu avec l'algorithme EM (Espérance-Maximisation) de la fonction scanone.** En abscisse, les chromosomes, et en ordonnée les LOD scores. Les LOD score significatifs sont ceux qui sont supérieurs au seuil de 5% symbolisé par la ligne en pointillé bleue. Les chromosomes impairs sont grisés.

Ces 3 QTLs sont additifs et expliquent 28% de la variation phénotypique. La transmission allélique pour les QTLs sur les chromosomes 1, 5 et 6 est illustrée en Figure\_45. Pour le QTL du chromosome 1, les trois génotypes sont significativement différents les uns des autres ( $p < 0,01$ ), indiquant que les allèles D (Diurne) et N (Nocturne) sont codominants. Pour les QTLs sur les chromosomes 5 et 6, la moyenne du pourcentage de cercaires nocturne n'est pas significativement différente entre les individus avec un génotype homozygote diurne (DD) et un génotype hétérozygote (ND/DN) ( $p > 0,01$ ). Seul le groupe homozygote nocturne (NN) a une moyenne de cercaires nocturne significativement plus élevée que les deux autres génotypes. Ceci indique que l'allèle N (nocturne) est récessif par rapport à l'allèle D (diurne). En effet, il faut deux allèles N pour que la moyenne de cercaires nocturne soit plus élevée.



**Figure 45 : Transmission allélique pour les QTLs identifiés pour le croisement intraspécifique entre les deux chronotypes de *Schistosoma mansoni*.** En abscisse, les génotypes homozygotes diurnes (DD) en blanc, hétérozygotes (DN) en gris et homozygotes nocturne (NN) en noir. En ordonnée, le pourcentage des cercaires émises la nuit. Les lettres « a », « b » et « c » désignent les  $p$ -values significativement différentes. Ainsi, s'il y a « a, b et c » les trois génotypes sont significativement différents. S'il y a uniquement « a » et « b », alors un seul génotype est significativement différent des deux autres. Pour le chromosome 1, les trois génotypes sont significativement différents les uns des autres ( $p < 0,01$ ). Pour les chromosomes 5 et 6, seul le génotype homozygote nocturne (NN) est significativement différent des deux autres génotypes ( $p < 0,01$ ). Les hétérozygotes DN regroupent les génotypes DN et ND.

#### II.1.2.4. Gènes candidats

Pour ces 3 QTLs, 462 gènes ont été identifiés dans le chromosome 1, 63 gènes dans le chromosome 5 et 169 gènes dans le chromosome 6 (liste des gènes en Annexe 4). Ces gènes ont été annotés dans la version WBPS16 du génome de *Schistosoma mansoni* (annotation GFF) (Howe, Bolt et al., 2016 ; Howe, Bolt et al., 2017). Pour compléter l'annotation officielle sur ces gènes, une annotation HHsearch a été réalisée (Le Clec'h, Chevalier et al., 2021). Elle est basée sur la génération de modèles Hidden Markov (HMM Hidden Markov model) à partir d'un alignement multiple des séquences et de la comparaison des modèles HMM pour reconnaître des protéines homologues (Steinegger, Meier et al., 2019).

Une première analyse basée sur les annotations génomiques et les résultats HHsearch a été réalisée pour faire une sélection subjective basée sur les gènes de l'horloge (rythme circadien), les gènes impliqués dans la phototransduction (détection de la photopériode) et les gènes impliqués dans l'acétylation et la méthylation des histones (pour trouver un lien avec les travaux en épigénétique).

Pour discerner les implications fonctionnelles et les gènes candidats, deux analyses d'enrichissement GO (Gene Ontology) et Kegg pathway ont été réalisées. L'enrichissement GO a été fait avec Blast2GO version 5.2 en utilisant le Fisher's Exact Test ( $p < 0,05$ ). L'analyse d'enrichissement Kegg pathway a été faite sur le serveur KAAS en utilisant le programme Blast avec la méthode SBH (Single-directional Best Hit).

#### II.1.2.4.1. Les gènes de l'horloge et le domaine Basic Helix-Loop-Helix

Parmi les 694 gènes QTLs, il n'y a aucun gène canonique de l'horloge. Toutefois, on retrouve 5 gènes qui présentent un domaine BHLH (Basic Helix-Loop-Helix) pour hélice-boucle-hélice basique. Ce domaine est présent dans les gènes canoniques de l'horloge comme CLOCK, BMAL1 et PER-1 et permet la liaison à l'ADN et plus particulièrement aux E-box (enhancer box) important dans leur rôle de facteur de transcription (Delauney & Laudet, 1998). Toutefois, l'annotation génomique de ces 5 gènes (Tableau\_25) définit ce domaine comme étant de type Myc, où Myc est un facteur de transcription proto-oncogène qui recrute des histones acétyltransférases. Si l'on s'intéresse aux annotations HHsearch, on trouve une ARNT (Aryl hydrocarbon receptor nuclear translocator-like), le pendant non circadien de BALM1, mais aussi le facteur de transcription SREBP-1a (Sterol Regulatory Element-Binding Proteins 1) et le facteur de transcription HES-1 (hairy and enhancer of split-1). Chez la souris, l'expression du gène SREBP est sous le contrôle de l'horloge circadienne (Matsumoto, Ishihare et al., 2010). Enfin, HES-1 est impliqué dans la différenciation et la prolifération cellulaire dans l'embryogénèse. Ce dernier gène est un oscillateur moléculaire avec une période de 2h, dont sa propre protéine va inhiber son expression (Kageyama, Ohtsuka & Kobayashi, 2007).

Tableau 25 : Liste des gènes QTLs avec un motif BHLH (Basic Helix Loop Helix)

Intérêt	Chr.	Identifiant (ID) du gène	Annotation GFF	Annotation HHsearch
Motif BHLH	1	Smp_132810	Myc-type, Basic Helix-Loop-Helix (BHLH) domain	Myc proto-oncogene protein {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
		Smp_132800		Aryl hydrocarbon receptor nuclear translocator-like
		Smp_019460		SREBP-1a {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
		Smp_132790		Transcription factor HES-1
		Smp_343020		DNA-binding protein inhibitor ID-1

Chr. : chromosome

#### II.1.2.4.2. Les rhodopsines et autres gènes impliqués dans la phototransduction

On retrouve aussi, comme gènes remarquables, une rhodopsine (Smp\_146470) et 16 autres gènes rhodopsine-like (Tableau\_26). La rhodopsine est un photorécepteur impliqué dans la phototransduction rétinienne (Chabre, 1998). La phototransduction est la conversion d'un signal lumineux en signal compréhensible par le système nerveux.

Tableau 26 : Liste des gènes QTLs rhodospine et rhodospine\_like

Intérêt	Chr.	Identifiant (ID) du gène	Annotation GFF	Annotation HHsearch
Rhodopsine	1	Smp_146470	NA	Rhodopsin
Rhodopsine-like	1	Smp_180140	G protein-coupled receptor, rhodopsin-like	P2Y purinoceptor 1, Rubredoxin
		Smp_316850		C-C chemokine receptor type 2
		Smp_204580		Neurotensin receptor type 1, Neurotensin
		Smp_118040		Chimera of Proteinase-activated receptor 1
		Smp_137320		Srg
		Smp_027940		ADENOSINE RECEPTOR A2A
		Smp_137310		7TM_GPCR_Srw
		Smp_137300		7TM_GPCR_Srw
		Smp_319310	NA	7TM_GPCR_Srab
		Smp_178420	GPCR, rhodopsin-like, 7TM	ADENOSINE RECEPTOR A2A
		Smp_132730	G protein-coupled receptor, rhodopsin-like	Neurotensin receptor type 1, Endolysin, NEUROTENSIN RECEPTOR
		Smp_127310		Chimera of Proteinase-activated receptor 1
		Smp_140250		Platelet-activating factor receptor, Flavodoxin
		Smp_193610		Lysozyme C (E.C.3.2.1.17)
Smp_126730	7tm_1			
Smp_007070	M3-mT4L, Tiotropium			

Chr. : chromosome

Un schéma de la transmission, de l'amplification et de la régulation du signal lors de la phototransduction avec la rhodopsine est présenté en Figure\_46. Une cellule photoréceptrice rétinienne de type bâtonnet est composée d'un segment externe qui contient des disques, d'un segment interne qui contient le noyau et les mitochondries et enfin d'une terminaison synaptique. Les disques présents dans le cytoplasme du segment externe sont des éléments cellulaires contenant les photopigments (ici rhodopsine). C'est donc dans les disques que commence la phototransduction.

- De façon simplifiée, **quand il fait nuit**, les canaux GMPc (Guanine Monophosphate cyclique) dépendant des cellules photoréceptrices sont ouverts et laissent entrer dans le milieu cellulaire des ions  $\text{Na}^+$  et  $\text{Ca}^{2+}$ . En parallèle, d'autres canaux laissent sortir des ions  $\text{K}^+$ , ce qui entraîne un état dépolarisé de la cellule photoréceptrice. Cette dépoliarisation de la cellule entraîne alors la libération de neurotransmetteur dans les synapses (glutamate).
- **A la lumière**, la rhodopsine (photorécepteur) est activée. La rhodopsine activée catalyse l'activation par le GTP de la transducine (protéine G). L'activation de cette protéine hétérotrimérique ( $\text{T}\alpha\text{GDP-T}\beta\gamma$ ) entraîne la séparation des deux sous-unités, avec d'un côté  $\text{T}\alpha\text{GTP}$  et de l'autre  $\text{T}\beta\gamma$ . La forme active de la transducine ( $\text{T}\alpha\text{GTP}$ ) active l'enzyme phosphodiesterase qui hydrolyse les GMPc en 5'-GMP. La concentration en GMPc baisse dans la cellule photoréceptrice ce qui entraîne la fermeture des canaux ioniques GMPc dépendants. L'arrêt de l'entrée des ions  $\text{Na}^+$  et  $\text{Ca}^{2+}$  couplé à la sortie des ions  $\text{K}^+$  qui n'est pas perturbée, entraîne une hyperpolarisation de la cellule et ainsi un arrêt de la libération des neurotransmetteurs (glutamate). Cet arrêt de l'émission du neurotransmetteur depuis la cellule photoréceptrice est perçu comme un signal par les autres cellules.

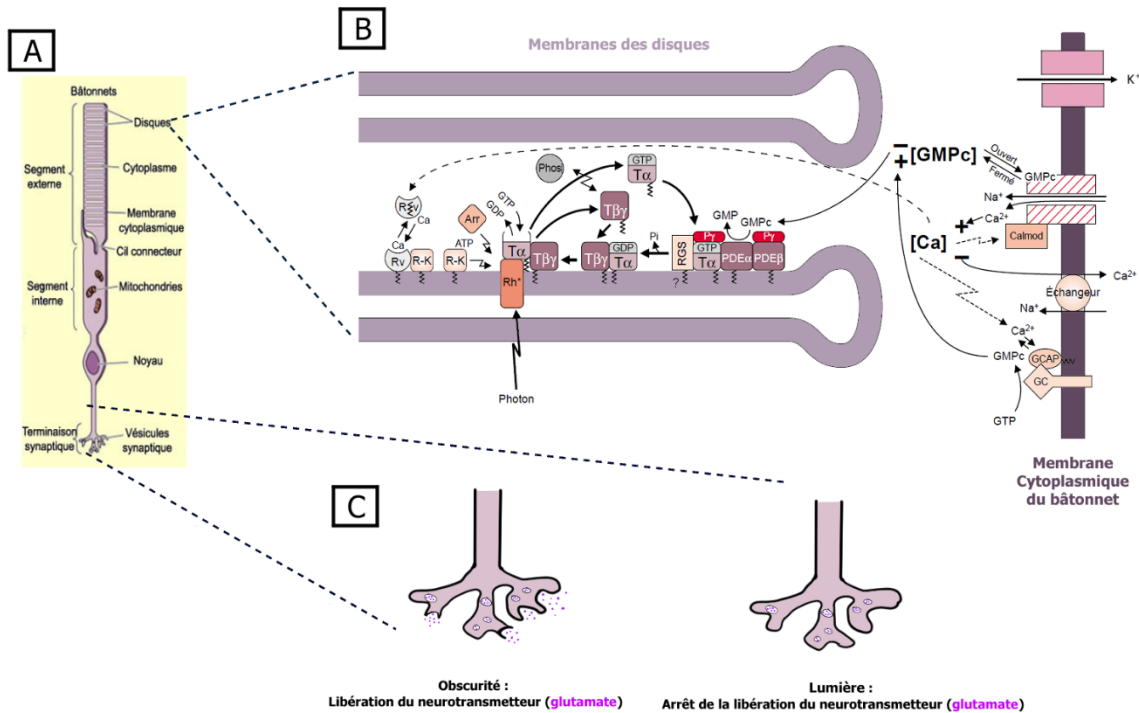


Figure 46 : **Schéma de la phototransduction du signal dans les bâtonnets rétiens.** A) Une cellule photoréceptrice complète (bâtonnet rétinien) (figure issue du site <https://lecerveau.mcgill.ca>) B) Disques qui contiennent les photorécepteurs (Rhodopsine) et la membrane plasmique de la cellule photoréceptrice avec le détail de la transmission, de l'amplification et de la régulation du signal (figure modifiée à partir de Chabre (1998)). La membrane des disques est colorée en violet clair et la membrane cytoplasmique de la cellule photoréceptrice est colorée en violet foncé. Arr : arrestine ; Calmod: calmoduline ; RGS: GTPase activating protein ; GC: guanylyl cyclase ; GCAP: calciprotéine activatrice de la guanylyl cyclase ; PDE $\alpha\beta$  : sous-unité catalytique de la phosphodiésterase ; Py : sous-unité inhibitrice; Phos : phosducine ; Rh\* : rhodopsine photoactivée ; R-K : rhodopsine kinase ; Rv : recoverine ; T $\alpha$ , T $\beta\gamma$  : sous-unités de la transducine. C) Terminaison synaptique de la cellule photoréceptrice à l'obscurité et à la lumière.

La rhodopsine qui a été activée par la lumière est ensuite phosphorylée par une rhodopsine kinase (spécifique sérine/thréonine kinase), ce qui permet à l'arrestine de se fixer à la rhodopsine phosphorylée. Ce complexe rhodopsine-arrestine met un terme à toute nouvelle interaction entre la transducine et la rhodopsine.

- **Avec le retour de la nuit**, un guanylate cyclase va produire du GMPc, ce qui va entraîner une augmentation de la concentration en GMPc dans le cytoplasme et ainsi l'ouverture des canaux GMPc dépendant. Cette ouverture des canaux provoque une augmentation du Ca<sup>2+</sup> cellulaire et la capture de la rhodopsine kinase par la recoverine. La recoverine est un inhibiteur de la rhodopsine kinase qui est calcium-dépendant.
- Lorsque la concentration de Ca<sup>2+</sup> est basse (**jour**), la recoverine devient soluble et relâche la rhodopsine kinase qui va alors phosphoryler la rhodopsine. A concentration élevée de Ca<sup>2+</sup> (**obscurité**), la recoverine est liée à la membrane du disque et à la rhodopsine kinase empêchant ainsi la phosphorylation de la rhodopsine. La rhodopsine phosphorylée couplée à l'arrestine est alors déphosphorylée par une protéine phosphatase 2A (PP2A) et séparée de l'arrestine. Dans cet état, la rhodopsine est de nouveau prête à être activée par la lumière et à recommencer un cycle de phototransduction. En parallèle du phénomène de déphosphorylation de la rhodopsine, la sous-unité T $\beta\gamma$  de la transducine est capturée par une

phosducine et la sous-unité  $\alpha$ GTP est hydrolysée pour devenir  $\alpha$ GDP et reformer la transducine hétérotrimérique inactive ( $\alpha$ GDP- $\beta\gamma$ ) prête elle aussi à réenclencher la transduction du signal.

Parmi tous les acteurs de la phototransduction cités, on retrouve, dans les gènes QTL (Tableau\_26 et Tableau\_27) : la rhodopsine, la protéine G, la phosphodiesterase, la sérine/thréonine kinase, l'arrestine, la recoverine, la protéine phosphatase 2A (PP2A), la GTPase activating protein et la guanylyl cyclase.

Tableau\_27 : Liste des gènes QTLs en lien avec la phototransduction

Intérêt	Chr.	Identifiant (ID) du gène	Annotation GFF	Annotation HHsearch
Protéine G (Transducine)	1	Smp_132750	G-protein alpha subunit, group Q	G-protein alpha subunit Galpha7
		Smp_010230	G-protein, beta subunit	beta1-subunit of the signal-transducing G protein heterotrimer {Cow ( <i>Bos taurus</i> ) [TaxId: 9913]}
	6	Smp_016630	Guanine nucleotide binding protein (G-protein), alpha subunit	G-protein alpha subunit Galpha7
Phosphodiesterase	1	Smp_105040	Endonuclease/exonuclease/phosphatase	2',5'-phosphodiesterase 12 (E.C.3.1.4.-,3.1.13.4)
GPCR kinase, sérine/thréonine kinase (rhodospine kinase)		Smp_210820	GPCR kinase	Beta-adrenergic receptor kinase 1 (E.C.2.7.11.15)
		Smp_068060	Protein kinase domain	Serine/threonine protein kinase TAO2 {Rat ( <i>Rattus norvegicus</i> ) [TaxId: 10116]}
		Smp_005190	Protein kinase domain	STE20-like serine/threonine-protein kinase, SLK {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
		Smp_009800	Protein kinase domain	Serine/threonine-protein kinase Nek2 {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
	5	Smp_158950	Protein kinase-like domain superfamily	SERINE/THREONINE-PROTEIN KINASE HASPIN (E.C.2.7.11.1)
Arrestine	1	Smp_126080	Arrestin	Bovine arrestin-2 (full length)
Recoverine et domaine EF-hand	5	Smp_158980	EF-hand domain	Recoverin {Cow ( <i>Bos taurus</i> ) [TaxId: 9913]}
		Smp_158990	EF-hand domain	Frequenin (neuronal calcium sensor 1) {Baker's yeast ( <i>Saccharomyces cerevisiae</i> ) [TaxId: 4932]}
	6	Smp_156180	EF-hand domain	DUF5580
		Smp_136640	EF-hand domain	Grancalcin {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
	1	Smp_083740	EF-hand domain	Myosin Essential Chain {Bay scallop ( <i>Aequipecten irradians</i> ) [TaxId: 31199]}



		Smp_005230	EF-hand domain	Calcium-dependent protein kinase sk5 CLD {Soybean ( <i>Glycine max</i> ) [TaxId: 3847]}
		Smp_137410	EF-hand domain pair	M-CALPAIN LARGE AND SMALL SUBUNITS
		Smp_132670	EF-hand domain	Myosin Essential Chain {Bay scallop ( <i>Aequipecten irradians</i> ) [TaxId: 31199]}
protéine phosphatase 2A (PP2A)	5	Smp_165490	Calcineurin-like phosphoesterase domain, ApaH type	Protein phosphatase 2A catalytic subunit alpha isoform, PP2A-alpha {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
GTPase activating protein	1	Smp_094890	Rho GTPase-activating protein domain	Beta-chimaerin, C-terminal domain {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
		Smp_028030	Rho GTPase-activating protein domain	SLIT-ROBO Rho GTPase-activating protein 2
		Smp_180150	Rho GTPase-activating protein domain	Beta-chimaerin, C-terminal domain {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
		Smp_180160	Rho GTPase-activating protein domain	Graf {Chicken ( <i>Gallus gallus</i> ) [TaxId: 9031]}
		Smp_319130	FCH domain	SLIT-ROBO Rho GTPase-activating protein 2
		Smp_007210	Rab-GTPase-TBC domain	Rab GTPase-activating protein 1
	5	Smp_165530	Rho GTPase-activating protein domain	p50 RhoGAP domain {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
guanylyl cyclase	1	Smp_211230	Adenylyl cyclase class-3/4/guanylyl cyclase	Soluble guanylyl cyclase alpha-1 subunit
		Smp_097470	Adenylyl cyclase class-3/4/guanylyl cyclase	Soluble guanylyl cyclase alpha-1 subunit
GTPase		Smp_067140	Small GTPase	RhoA {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
		Smp_025740	Small GTPase	RhoE (RND3) {Mouse ( <i>Mus musculus</i> ) [TaxId: 10090]}

Chr. : chromosome

#### II.1.2.4.3. Les gènes impliqués dans l'acétylation des histones

Un autre point non exhaustif sur les gènes QTLs du croisement intraspécifique est la présence de gènes qui codent pour les histones H2A, H3, H4, mais aussi des gènes qui codent pour des histones acétyl transférase (« writer ») et des histones désacétylases (« eraser ») (Tableau\_28). La présence des « writer » (enzyme qui fait les modifications des queues d'histones) et d' « eraser » (enzyme qui enlève les modifications des queues d'histones) sont encourageantes pour l'implication de l'acétylation des histones dans le phénomène du rythme d'émission des cercaires.

Tableau 28 : Liste des gènes QTLs en lien avec les histones et l'acétylation/désacétylation des histones

Intérêt	Chr.	Identifiant (ID) du gène	Annotation GFF	Annotation HHsearch
Histones	1	Smp_246990	Transcription factor CBF/NF-Y/archaeal histone domain	Histone H3.1, Histone H4, Histone
		Smp_036220	Histone H2B	Histone H2B {Chicken ( <i>Gallus gallus</i> ), erythrocytes [TaxId: 9031]}
		Smp_124840	Zinc finger, UBR-type	Histone H3.2, Histone H4, Histone
		Smp_318950	Histone H3/CENP-A	Leishmania histone H3, Histone H4
Histone acétylase	6	Smp_210630	Histone H2A	Histone H2A {Chicken ( <i>Gallus gallus</i> ), erythrocytes [TaxId: 9031]}
		Smp_131320	Zinc finger, PHD-type	Probable histone acetyltransferase MYST1 {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
Histone désacétylase	1	Smp_005210	Histone deacetylase family	METASTASIS-ASSOCIATED PROTEIN MTA1, HISTONE DEACETYLASE
	5	Smp_330000	Sin3 associated polypeptide p18	HISTONE DEACETYLASE COMPLEX SUBUNIT SAP18

Chr. : chromosome

#### II.1.2.4.4. Enrichissement GO : Processus biologiques

Les analyses d'enrichissement GO des gènes QTLs n'ont pas donné de résultats significatifs en utilisant la limite usuelle FDR < 0,05 (False Discovery Rate). Les résultats qui suivent ont été obtenus en utilisant la limite de  $p < 0,05$ .

On retrouve, parmi les processus biologiques mis en évidence par les analyses d'enrichissement GO, le rythme circadien, le sommeil, le transport de neurotransmetteur, la signalisation cellule à cellule et la maturation développementale (Figure 47). L'annotation GO « rythme circadien » ne correspond pas à un gène de l'horloge (clock) mais au gène **Smp\_105040** qui est décrit comme la nocturnine dans le génome de *S. mansoni*. La nocturnine contient un domaine Endonuclease/exonuclease/phosphatase qui est commun avec la phosphodiesterase. Ce gène est connu pour avoir une expression rythmique avec un pic nocturne et être régulé par les gènes de l'horloge (Green & Besharse, 1996). L'annotation GO « sommeil » correspond au gène **Smp\_345020**, le gène quiver (QVR) chez la drosophile qui est impliqué dans la régulation positive des cycles veille/sommeil circadiens mais aussi dans la modulation de la transmission synaptique chimique et la régulation négative de potentiel membranaire (<https://www.ncbi.nlm.nih.gov/gene?Db=gene&Cmd=DetailsSearch&Term=2768718>). L'annotation GO « signalisation cellule à cellule » correspond, entre autres, au gène **Smp\_126730**, un récepteur 5HT-1 (5-hydroxytryptamine). Le récepteur 5HT-1 est en fait un récepteur sérotoninergique de type rhodopsine-like. Chez la forme adulte de *S. mansoni*, il a été montré que **Smp\_126730** est un des récepteurs responsables de l'effet moteur provoqué par la sérotonine qui est un puissant myo-excitateur (Patocha, Sharma et al. 2014). De plus, chez la forme sporocyste, il a été montré que les récepteurs 5HT sont importants dans la production de sporocystes filles (secondaires) (Boyle & Yoshino, 2005). Parmi les gènes qui portent l'annotation GO « maturation développementale », on retrouve le gène **Smp\_009630**, un facteur de transcription putatif six/sine homeobox qui, chez la drosophile, est directement responsable de la formation de l'œil (Weasner, Salzer & Kumar, 2007).

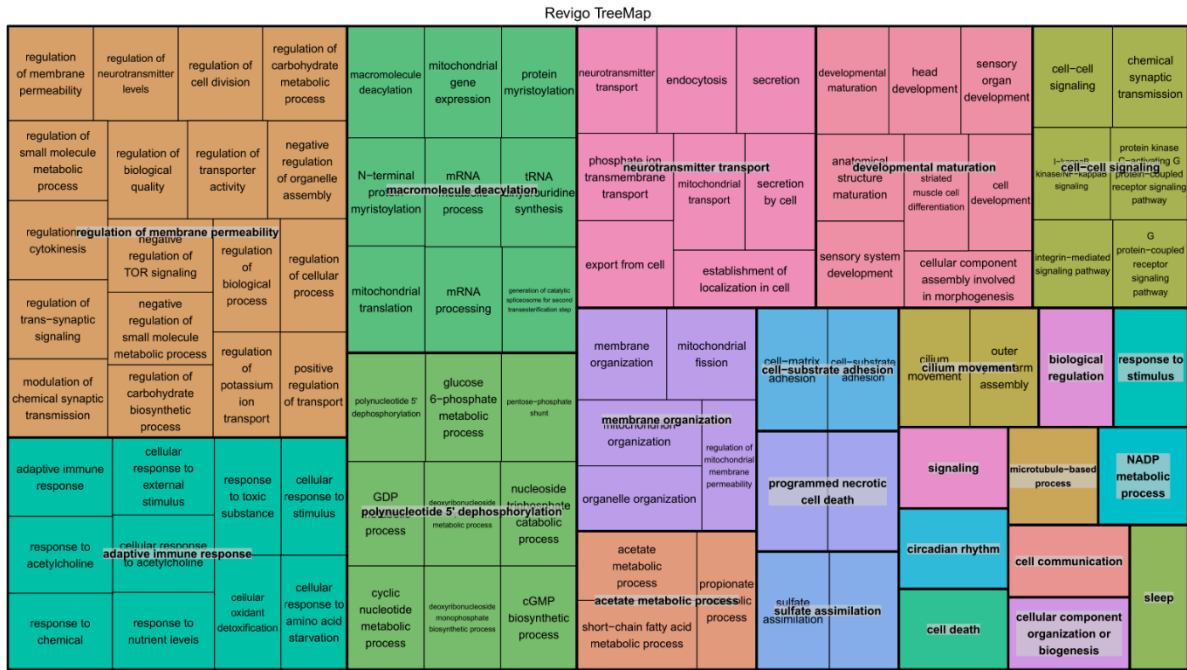


Figure 47 : Treemap des résultats de l'analyse d'enrichissement GO des processus biologiques obtenus pour les gènes QTLs du croisement intraspécifique. Chaque couleur représente un grand ensemble de processus biologiques. Il y a, de gauche à droite : Régulation de la perméabilité membranaire, déacylation des macromolécules, transports des neurotransmetteurs, maturation du développement, signalisation cellule à cellule, réponse de l'immunité adaptative, déphosphorylation 5' polynucléotides, organisation de la membrane, adhésion cellule-substrat, mouvement des cils, régulation biologique, réponse à des stimulus, mort cellulaire programmée, signalisation, processus basé sur les microtubules, processus métabolique NADP, processus métabolique acétate, assimilation du sulfate, rythme circadien, communication cellulaire, mort cellulaire, organisation ou biogénèse des composants cellulaires et sommeil.

II.1.2.4.5. Kegg pathway

Les analyses Kegg pathway montrent que les gènes QTLs sont potentiellement impliqués dans 266 voies moléculaires différentes. Parmi toutes ces voies, cinq retiennent notre attention : le rythme circadien, l'entraînement circadien, la phototransduction, la phototransduction chez la drosophile et la transduction olfactive (Tableau\_29).

Tableau 29 : Liste des gènes QTLs correspondant aux Kegg pathways sélectionnés

Kegg pathway	KO annotation	Chr.	Identifiant (ID) du gène	Annotation GFF	Annotation HHsearch
Rythme circadien	K07200	1	Smp_315790	CBS domain	5'-AMP-activated protein kinase catalytic subunit
Entraînement du rythme circadien	K04630	6	Smp_016630	Guanine nucleotide binding protein (G-protein), alpha subunit	G-protein alpha subunit Galpha
	K04634	1	Smp_132750	G-protein alpha subunit, group Q	
	K04536		Smp_010230 *	G-protein, beta subunit	beta1-subunit of the signal-transducing G protein

Photo-transduction	K04536		Smp_010230 *		heterotrimer {Cow ( <i>Bos taurus</i> ) [TaxId: 9913]}
Photo-transduction chez la drosophile	K00910		Smp_210820 *	GPCR kinase	Beta-adrenergic receptor kinase 1 (E.C.2.7.11.15)
	K04536		Smp_010230 *	G-protein, beta subunit	beta1-subunit of the signal-transducing G protein heterotrimer {Cow ( <i>Bos taurus</i> ) [TaxId: 9913]}
Transduction olfactive	K00910		Smp_210820 *	GPCR kinase	Beta-adrenergic receptor kinase 1 (E.C.2.7.11.15)
	K04439		Smp_126080	Arrestin	Bovine arrestin-2 (full length)
	K04536		Smp_010230 *	G-protein, beta subunit	beta1-subunit of the signal-transducing G protein heterotrimer {Cow ( <i>Bos taurus</i> ) [TaxId: 9913]}
	K05849		Smp_152830	Na-Ca exchanger/ integrin-beta4	Na/Ca exchange protein

Chr. : chromosome ; les gènes marqués d'un astérisque (\*) sont des gènes correspondant à plusieurs Kegg pathways sélectionnés.

On retrouve un seul gène dans le pathway rythme circadien, qui est **Smp\_315790** une sous-unité gamma de la kinase 5'AMP-activatrice de protéine (AMPK). La phosphorylation de la protéine CRY (issue du gène clock Cryptochrome) est médiée par des AMPK ce qui induit la dégradation de CRY et ainsi régule la boucle de rétro contrôle négative de l'horloge circadienne (Lee & Kim, 2013).

Il y a trois gènes dans la voie de l'entraînement circadien, c'est-à-dire les processus de synchronisation de l'horloge interne avec les signaux environnementaux. Le gène **Smp\_016630** est une putative sous unité alpha de la protéine G. Le gène **Smp\_010230**, retrouvé dans trois voies différentes qui sont l'entraînement circadien, la phototransduction et la transduction olfactive, est la sous-unité bêta de la protéine G/transducine. Le gène **Smp\_132750** est une sous unité alpha du groupe Q de la protéine G. Ce gène est aussi retrouvé dans la voie de la phototransduction de la drosophile.

Le gène **Smp\_210820** est le second gène qui est présent dans la voie de la phototransduction chez la drosophile. Ce gène est une kinase des récepteurs couplés aux protéines G (GPCR kinase) et potentiellement une rhodopsine kinase. Il est aussi présent dans la voie de la transduction olfactive.

La voie de la transduction olfactive contient deux autres gènes, en plus de ceux précédemment cités : le gène **Smp\_126080** qui est l'arrestine et le gène **Smp\_152830** qui est un échangeur d'ion sodium calcium.

### II.1.2.5. Conclusion

Le QTL mapping issu des croisements intraspécifiques de *S. mansoni* a mis en évidence 3 QTLs, sur les chromosomes 1, 5 et 6. Parmi tous les gènes candidats, aucun gène canonique de l'horloge n'est présent, ce qui est en accord avec Rawlinson, Reid et al. (2021) qui n'ont pas trouvé de gène canonique de l'horloge dans le génome de *S. mansoni*. On peut toutefois noter la présence de gènes avec un domaine BHLH, qui est présent dans certains gènes canoniques de l'horloge. La sélection subjective, basée sur les annotations GFF et HHsearch, a mis en évidence des gènes impliqués dans la phototransduction et plus largement dans la transduction des signaux environnementaux, avec la rhodopsine comme probable photorécepteur. Ceci est en accord avec la découverte de plusieurs rhodopsines et GPCR (récepteurs couplés aux protéines G) dont certaines sont impliquées dans la

reconnaissance par le miracidium de certains peptides sécrétés par le mollusque dans l'eau (Hoffmann, Davis et al., 2001 ; El-Shehabi, Taman et al., 2012 ; Liang, Zao et al., 2016; Phan, Liang et al., 2022). La sélection subjective a aussi mis en évidence des gènes impliqués dans l'acétylation des histones. La présence de gènes avec un domaine BHLH de type Myc qui recrute les histones acétylases ainsi que la présence de gènes histone acétylase et désacétylase sont encourageants pour l'implication de l'acétylation des histones dans le rythme d'émission des cercaires.

Enfin, les analyses d'enrichissement GO et de Kegg pathway ont confirmé l'implication fonctionnelle de certains gènes dans les voies moléculaires touchant à la régulation du rythme circadien et à la phototransduction, et plus largement dans la transduction des signaux environnementaux.

### II.1.3. Croisement interspécifique : *Schistosoma mansoni* (diurne) X *Schistosoma rodhaini* (nocturne)

Les croisements F0 et F1 de *Schistosoma mansoni* diurne avec *Schistosoma rodhaini* nocturne ont été réalisés avant cette thèse par Hélène Moné et Gabriel Mouahid. La préparation moléculaire des échantillons a été faite dans le cadre de mon stage d'été de 3 mois au Texas Biomedical Research Institute dans l'équipe de Tim Anderson, avec l'aide de Robbie Diaz. Le génotypage a été réalisé sur la version 7 du génome de *S. mansoni* avec le pipeline du génotypage multiplex shotgun (Andolfatto, Davison et al., 2011) adapté par Frédéric Chevalier. L'analyse QTL a été réalisée à Perpignan.

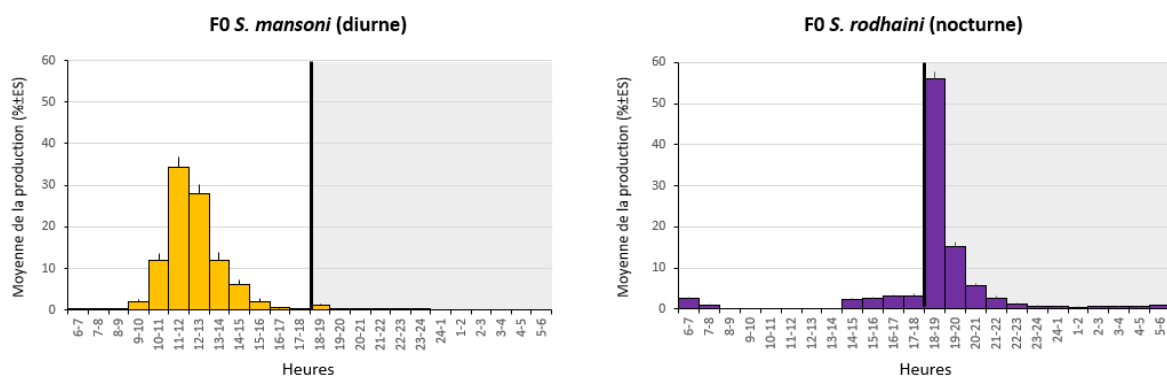
#### II.1.3.1. Protocole

Les parasites F0 utilisés pour ce croisement sont le chronotype diurne de *S. mansoni* (Oman) et le parasite *S. rodhaini*. Les hôtes intermédiaires mollusques *Biomphalaria pfeifferi* ont été prélevés sur le terrain dans la région du Dhofar à Oman. Les croisements réciproques se nomment croisement A (F0 mâle *S. mansoni* diurne X F0 femelle *S. rodhaini* nocturne) et croisement B (F0 mâle *S. rodhaini* nocturne X F0 femelle *S. mansoni* diurne). L'analyse QTL a été faite sur chacun des deux croisements interspécifiques.

#### II.1.3.2. Présentation des phénotypes

##### II.1.3.2.1. Les phénotypes F0

Les résultats moyens, accompagnés des variations individuelles, sont présentés dans la Figure\_48 et le Tableau\_30. Pour *S. mansoni* diurne, 24 profils ont émis plus de 20 cercaires et 41 profils ont émis plus de 20 cercaires pour *S. rodhaini*.



**Figure 48 : Chronobiologie du chronotype diurne de *S. mansoni* (orange) et de *S. rodhaini* nocturne (violet) pour l'expérience de QTL mapping.** Chaque histogramme a, en ordonnées, la moyenne du pourcentage des cercaires émises ± écart standard. L'axe des abscisses représente les 24h de la journée, débutant à 6h, début du jour, et finissant à 6h, fin de la nuit. L'arrière-plan gris correspond à l'obscurité. La ligne noire marque le début de l'obscurité à 18h.

**Tableau 30 : Paramètres des profils moyens d'émission, accompagnés des variations individuelles, pour les parasites diurne *S. mansoni* et nocturne *S. rodhaini* pour l'expérience de QTL mapping**

Paramètres	Moyenne ± ES Variations individuelles		Test de Wilcoxon- Mann-Whitney
	<i>S. mansoni</i> diurne (F0)	<i>S. rodhaini</i> nocturne (F0)	
Effectif	24	41	
Photopériode	L-O	L-O	
Nombre de pics	1 1	1 1	
Heure du pic	11-12h 10-11h à 13-14h	18-19h 18-19h à 19-20h	≠
Plage d'émission (h)	4,5 ± 0,2 3 à 6	3,7 ± 0,1 3 à 6	W = 727, p = 0,0007
Intensité de la plage (%)	92 ± 1,4 65,6 à 98,2	81,1 ± 1,1 66,7 à 95	W = 883, p = 1,1e-07
Asymétrie	+0,4 ± 0,1 -1,3 à +1,6	+1,5 ± 0,1 -0,5 à +2,3	W = 104, p = 1,4e-07
Aplatissement	-0,9 ± 0,5 (N=22) -5,8 à +2,8	+2,6 ± 0,4 (N=22) -2,9 à +5,6	W = 47, p = 6,7e-07
Cercaires la nuit (%)	1,8 ± 0,5 0,0 à 10,1	85,3 ± 1,2 63,5 à 96,6	W = 0, p = 2,4e-11
Nombre total de cercaires	355,4 ± 41,5 58 à 811	559,4 ± 51,1 63 à 1269	W = 302, p = 0,00999

*S. mansoni* diurne a une émission unimodale dont l'heure du pic moyen est à 11-12h. La plage horaire moyenne d'émission est d'environ 4h avec une intensité de 92%. La plage d'émission moyenne est symétrique, avec une distribution aplatie. Il n'y a quasiment pas d'émission de cercaires nocturne.

*S. rodhaini* nocturne a une émission unimodale dont l'heure du pic moyen est à 18-19h. La plage horaire moyenne d'émission est un peu moins de 4h avec une intensité de 81%. La plage d'émission moyenne a une asymétrie élevée sur la droite, avec une distribution centrée. L'émission nocturne est majoritaire, avec en moyenne 85% des cercaires émises la nuit.

Les comparaisons sont présentées dans le Tableau\_30. Les profils moyens des deux schistosomes se différencient par tous les paramètres de comparaison ( $p < 0,01$ ), à l'exception du pic unimodal. Comparé à *S. mansoni*, le parasite *S. rodhaini* a une plage d'émission plus courte et plus centrée autour de l'heure du pic mais avec une intensité plus faible. Ces résultats illustrent bien l'unicité des deux phénotypes F0.

#### II.1.3.2.2. Le croisement A

Le croisement interspécifique A est un croisement entre un mâle F0 *S. mansoni* diurne et une femelle F0 *S. rodhaini* nocturne. La génération F1 issue de ce croisement a un profil tri-modal, avec deux pics majoritaires à 6-7h et 10-11h et un pic secondaire à 18-19h. Le pourcentage de cercaires nocturne moyen est de 11,19%. La génération F2, issue du croisement entre un mâle et une femelle F1, contient 94 individus. Le pourcentage de cercaires nocturne moyen est de 25,25% avec des valeurs individuelles allant de 0% à 94,46%. La répartition des profils de chronobiologie de la F2 est de 17% ressemblant à la F0 diurne (1ère ligne des F2 ; Figure\_49), 12,8% ressemblant à la F0 nocturne (dernière ligne des F2 ; Figure\_44) et 70,2% hybride ou inclassable (les 2<sup>ème</sup> et 3<sup>ème</sup> lignes des F2 ; Figure\_49). L'ensemble des profils F2 sont à retrouver en Annexe 5.

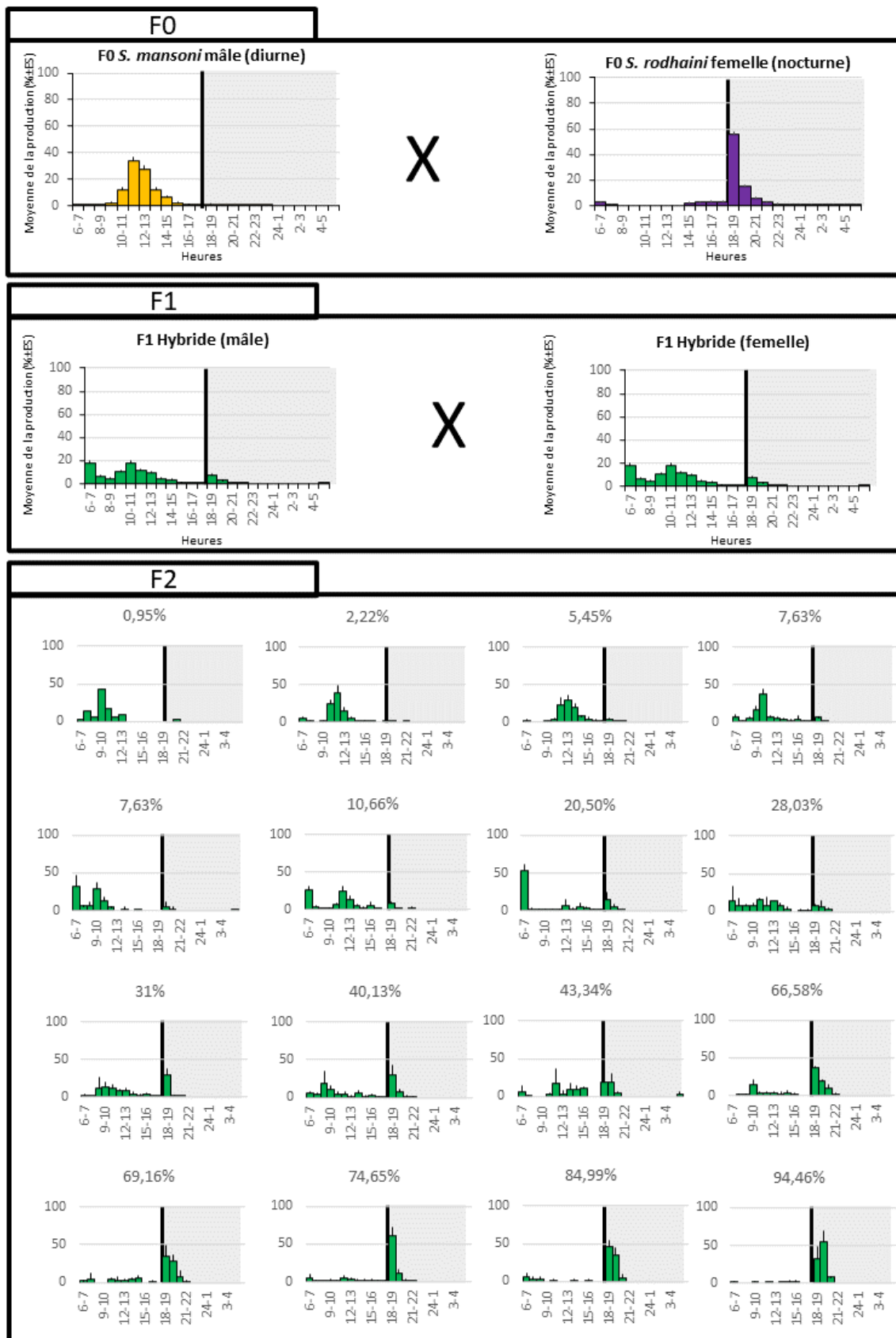


Figure 49 : Chronobiologie du croisement A interspécifique, avec F0 diurne (orange), F0 nocturne (violet) et F1 et une sélection d'individus F2 hybrides (vert).



Chaque histogramme a, en ordonnées, la moyenne du pourcentage des cercaires émises  $\pm$  écart standard. L'axe des abscisses représente les 24h de la journée, débutant à 6h, début du jour, et finissant à 6h, fin de la nuit. L'arrière-plan gris correspond à l'obscurité. La ligne noire marque le début de l'obscurité à 18h. Les individus F2 sont une sélection des divers pourcentages de cercaires nocturnes. Ils sont classés du plus faible pourcentage de cercaires nocturne au plus élevé. La première ligne des F2 représente une sélection d'individus proches du phénotype F0 diurne. Les 2<sup>ème</sup> et 3<sup>ème</sup> lignes représentent une sélection d'individus au phénotype hybride. La 4<sup>ème</sup> ligne représente une sélection d'individus proches du phénotype F0 nocturne.

### II.1.3.2.3. Le croisement B

Le croisement interspécifique B est un croisement entre un mâle F0 *S. rodhaini* nocturne et une femelle F0 *S. mansoni* diurne. La génération F1 issue de ce croisement a un profil tri-modal, avec deux pics majoritaires à 6-7h et 10-11h et un pic secondaire à 18-19h. Le pourcentage de cercaires nocturne moyen est de 8,75%. La génération F2, issue du croisement entre un mâle F1 et une femelle F1, contient 115 individus. Le pourcentage de cercaires nocturne moyen est de 27,53% avec des valeurs individuelles allant de 0% à 97,16%. La répartition des profils de chronobiologie de la F2 est de 25,2% ressemblant à la F0 diurne (1<sup>ère</sup> ligne des F2 ; Figure\_50), 13,9% ressemblant à la F0 nocturne (dernière ligne des F2 ; Figure\_50) et 60,9% hybride ou inclassable (les 2<sup>ème</sup> et 3<sup>ème</sup> lignes des F2 ; Figure\_50). L'ensemble des profils F2 sont à retrouver en Annexe 5.

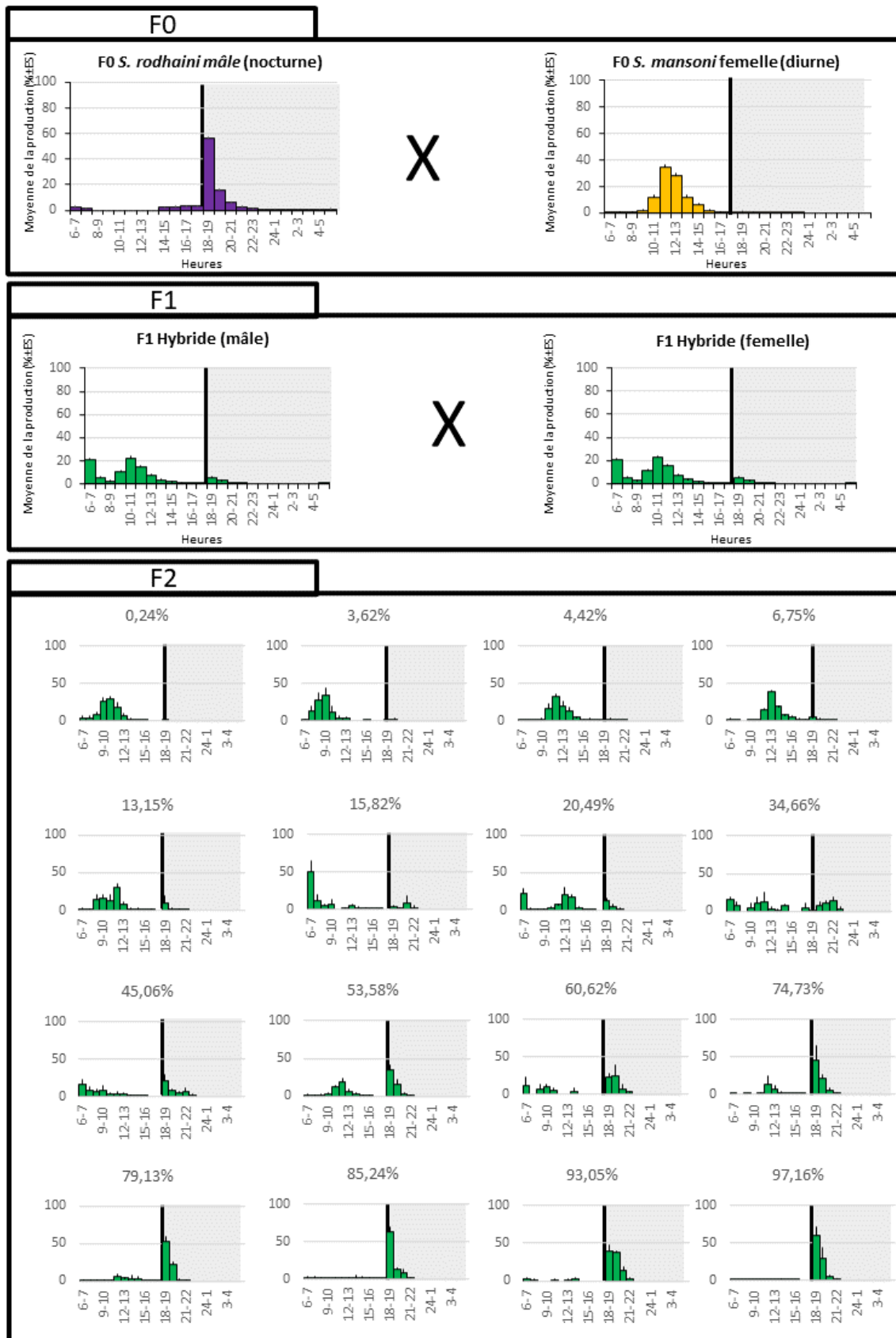


Figure 50 : Chronobiologie du croisement B interspécifique, avec F0 diurne (orange), F0 nocturne (violet) et F1 et une sélection d'individus F2 hybrides (vert).

Chaque histogramme a, en ordonnées, la moyenne du pourcentage des cercaires émises  $\pm$  écart standard. L'axe des abscisses représente les 24h de la journée, débutant à 6h, début du jour, et finissant à 6h, fin de la nuit. L'arrière-plan gris correspond à l'obscurité. La ligne noire marque le début de l'obscurité à 18h. Les individus F2 sont une sélection des divers pourcentages de cercaires nocturnes. Ils sont classés du plus faible pourcentage de cercaires nocturne au plus élevé. La première ligne des F2 représente une sélection d'individus proches du phénotype F0 diurne. Les 2<sup>ème</sup> et 3<sup>ème</sup> lignes représentent une sélection d'individus au phénotype hybride. La 4<sup>ème</sup> ligne représente une sélection d'individus proches du phénotype F0 nocturne.

Il n'y a pas de différence significative entre les profils d'émission des générations F1 des croisements interspécifiques A et B. De plus, le pourcentage de cercaires nocturne moyen n'est pas significativement différent entre les générations F1 et F2 des deux croisements intraspécifiques ( $p > 0,01$ ). Ces résultats sont en accord avec ceux de la première partie qui montraient que le sexe, et donc ici le sens du croisement, n'a pas d'influence sur le rythme d'émission des cercaires.

### II.1.3.3. QTL mapping

Les analyses QTL ont été faites en parallèle sur les deux croisements interspécifiques. La préparation moléculaire des échantillons s'est faite en utilisant tout le génome. Le génotypage regroupe 400 marqueurs répartis sur les 8 chromosomes du génome de *S. mansoni* (7 autosomes et 1 sexuel). Pour le croisement A, 87 individus F2 ont servis pour l'analyse QTL et 97 individus F2 pour le croisement B. Le phénotype utilisé est la moyenne du pourcentage de cercaires nocturne. La détection des QTLs a été faite avec la méthode de l'algorithme EM de la fonction scanone avec un seuil de détection à 5% déterminé avec 1000 permutations.

#### II.1.3.3.1. Croisement A

L'analyse QTL du croisement A révèle un QTL dans le chromosome 3 avec un LOD score de 5,04 pour un seuil 5% de 4,0 (Figure\_51).

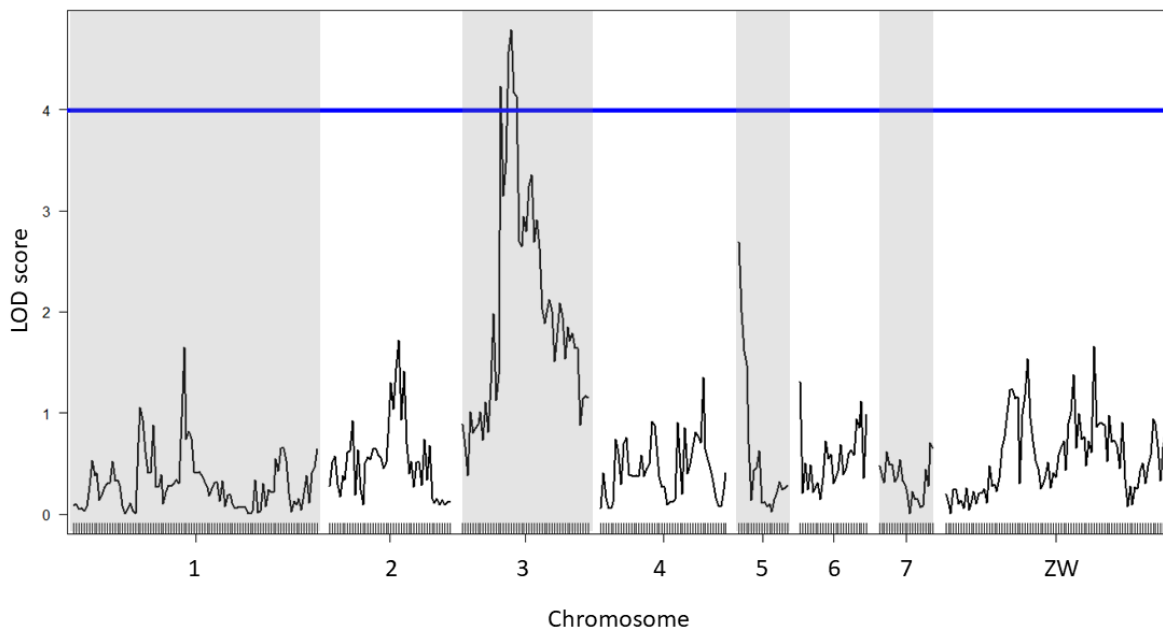
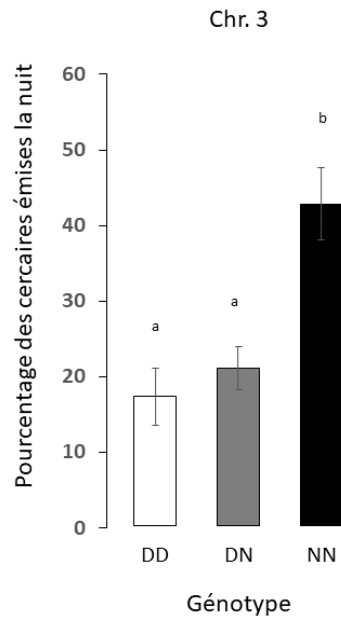


Figure 51 : LOD score QTL pour le croisement A interspécifique entre *Schistosoma mansoni* et *Schistosoma rodhaini* obtenu avec l'algorithme EM de la fonction scanone. En abscisse, les chromosomes, et en ordonnée les LOD scores. Les LOD score significatifs sont ceux qui sont supérieurs au seuil de 5% symbolisé par la ligne en bleue. Les chromosomes impairs sont grisés.

Ce QTL sur le chromosome 3 explique 22,38% de la variation phénotypique ( $P$ -value 2,38 e-05). La répartition géotypique en fonction de la variation phénotypique pour ce QTL sur le chromosome 3 est illustrée en Figure\_52. La moyenne du pourcentage de cercaires nocturne n'est pas significativement différente entre les individus avec un géotype homozygote diurne (DD) et un géotype hétérozygote (ND/DN) pour le QTL du chromosome 3 ( $p > 0,01$ ). Seul le groupe homozygote nocturne (NN) a une moyenne de cercaires nocturne significativement plus élevée que les deux autres géotypes. Ceci indique que l'allèle N (nocturne) est récessif par rapport à l'allèle D (diurne). En effet, il faut deux allèles N pour que la moyenne des cercaires nocturne soit plus élevée pour le QTL du chromosome 3.



**Figure 52 : Effet plot pour le QTL du chromosome 3 pour le croisement A interspécifique entre *Schistosoma mansoni* et *Schistosoma rodhaini* obtenu avec l'algorithme EM de la fonction scanone.** En abscisse, les géotypes homozygotes diurnes (DD) en blanc, hétérozygotes (DN) en gris et homozygotes nocturne (NN) en noir. En ordonnée, le pourcentage des cercaires émises la nuit. Les lettres « a », « b » et « c » désignent les  $p$ -values significativement différentes. Ainsi, s'il y a « a », « b » et « c », les trois géotypes sont significativement différents. S'il y a uniquement « a » et « b », alors un seul géotype est significativement différent des deux autres. Pour le chromosome 3, seul le géotype homozygote nocturne (NN) est significativement différent des deux autres géotypes ( $W = 264$ ,  $p = 0,0008$  pour ND et  $W = 331$ ,  $p = 0,0008$  pour DD). Les hétérozygotes DN regroupent les géotypes DN et ND.

## II.1.3.3.2. Croisement B

L'analyse QTL du croisement B révèle deux QTLs pour un seuil 5% de 4,08 (Figure\_53). Le premier QTL est dans le chromosome 3 avec un LOD score de 4,52 et le second se situe dans le chromosome 4 avec un LOD score de 4,34.

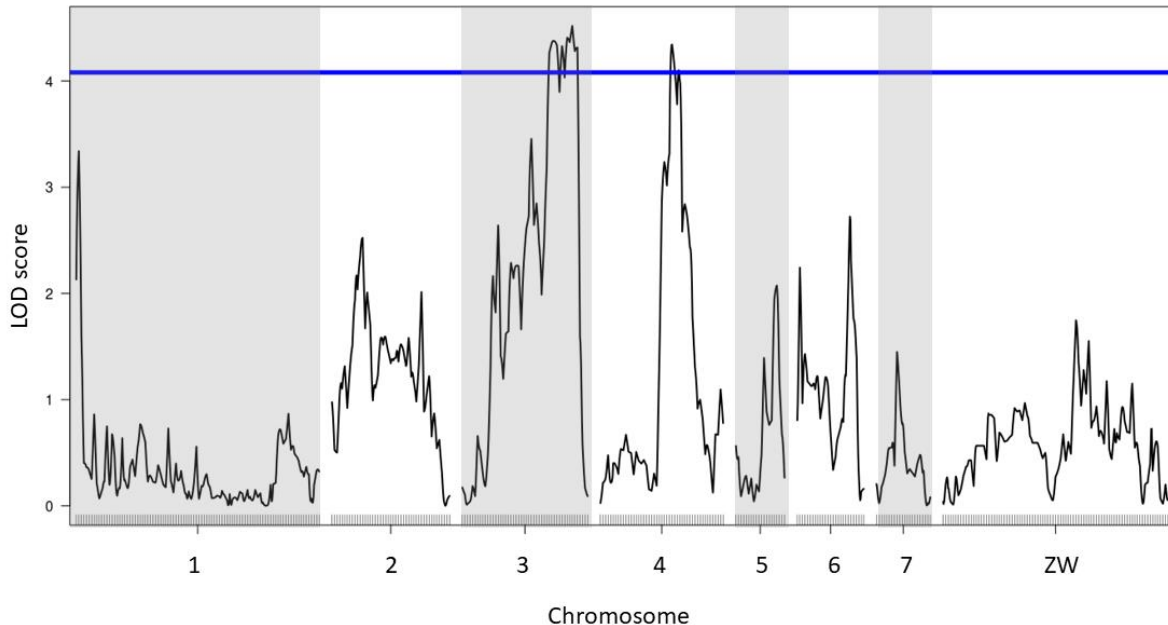


Figure 53 : LOD score QTL pour le croisement B interspécifique de *Schistosoma mansoni* et *Schistosoma rodhaini* obtenu avec l'algorithme EM de la fonction scanone. En abscisse, les chromosomes, et en ordonnée les LOD scores. Les LOD score significatifs sont ceux qui sont supérieurs au seuil de 5% symbolisé par la ligne en bleu. Les chromosomes impairs sont grisés.

Pour déterminer la nature des interactions entre les deux QTLs, on dresse de façon simplifiée, un modèle théorique d'interaction épistatique et un modèle théorique d'interaction d'effet additif. Un test statistique est ensuite réalisé pour déterminer lequel de ces deux modèles théoriques explique le mieux la nature de l'interaction des QTLs que nous avons détectée. Pour cela, le seuil est fixé à 5% et est a été déterminé avec 500 permutations donnant respectivement un seuil de 12,3 pour l'épistasie et de 7,67 pour l'effet additif.

Seul le LOD score de l'interaction additive des QTLs sur les chromosomes 3 et 4 s'est révélé significatif (Tableau\_31).

Tableau 31 : Tableau récapitulatif du scantwo pour déterminer l'interaction (épistasie et l'effet additif) des QTLs pour le croisement B interspécifique

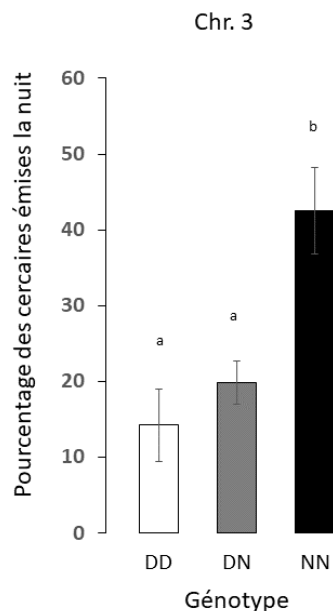
Paire des locus	Position du locus 1 (chr3)	Position du locus 2 (chr4)	LOD Épistasie	LOD Addition
Seuil 5%			12,3	7,67
Chr3 :Chr4	215	157	11,12	8,71

Cette interaction additive entre les QTLs entre les chromosomes 3 et 4 explique 32,78% de la variation phénotypique (Tableau\_32). Pris individuellement, le QTL du chromosome 3 explique 14,17% de la variation phénotypique et le QTL du chromosome 4, 13,47%, ce qui illustre que les deux QTLs sont importants.

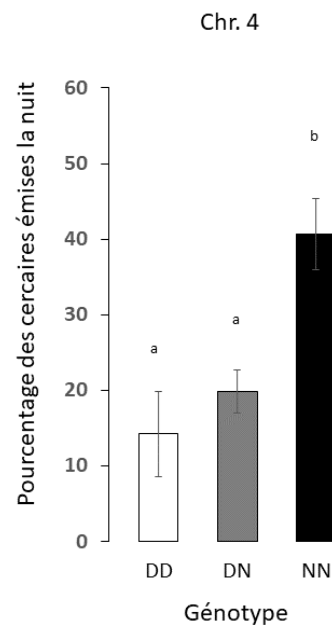
**Tableau 32 : Explication de la variance du phénotype (moyenne du pourcentage de cercaires nocturne) par le modèle additif des QTLs et les QTLs séparés pour le croisement B interspécifique**

Loci	LOD score (QTL)	% variance	P-value
<b>Modèle additif Chr3 + Chr4</b>	8,37	32,78	1.87e-07
<b>Chr3</b>	4,02	14,17	0,0002
<b>Chr4</b>	3,85	13,47	0,0002

La répartition génotypique en fonction de la variation phénotypique pour les QTLs sur les chromosomes 3 et 4 est illustrée en Figure\_54 et Figure\_55. Pour les deux QTLs, la moyenne du pourcentage de cercaires nocturne n'est pas significativement différente entre les individus avec un génotype homozygote diurne (DD) et un génotype hétérozygote (ND/DN) pour le QTL du chromosome 3 ( $p > 0,01$ ). Seul le groupe homozygote nocturne (NN) a une moyenne de cercaires nocturne significativement plus élevée que les deux autres génotypes. Ceci indique que l'allèle N (nocturne) est récessif par rapport à l'allèle D (diurne). En effet, il faut deux allèles N pour que la moyenne des cercaires nocturne soit plus élevée pour les QTLs des chromosomes 3 et 4.



**Figure 54 : Effet plot pour le QTL du chromosome 3 pour le croisement B interspécifique entre *Schistosoma mansoni* et *Schistosoma rodhaini* obtenu avec l'algorithme EM de la fonction scanone.** En abscisse, les génotypes homozygotes nocturnes (NN), hétérozygotes (ND) et homozygotes diurnes (DD) et en ordonnée la moyenne du pourcentage de cercaires nocturne. Les lettres « a », « b » et « c » désignent les  $p$ -values significativement différentes. Ainsi, s'il y a « a, b et c », les trois génotypes sont significativement différents. S'il y a uniquement « a » et « b », alors un seul génotype est significativement différent des deux autres. Pour le chromosome 3, seul le génotype homozygote nocturne (NN) est significativement différent des deux autres génotypes ( $W = 362$ ,  $p = 0,0007$  pour ND et  $W = 484$ ,  $p = 0,0007$  pour DD). Les hétérozygotes DN regroupent les génotypes DN et ND.



**Figure 55 : Effet plot pour le QTL du chromosome 4 pour le croisement B interspécifique entre *Schistosoma mansoni* et *Schistosoma rodhaini* obtenu avec l'algorithme EM de la fonction scanone.** En abscisse, les génotypes homozygotes nocturnes (NN), hétérozygotes (ND) et homozygotes diurnes (DD) et en ordonnée la moyenne du pourcentage de cercaires nocturne. Les lettres « a », « b » et « c » désignent les  $p$ -values significativement différentes. Ainsi, s'il y a « a, b et c », les trois génotypes sont significativement différents. S'il y a uniquement « a » et « b », alors un seul génotype est significativement différent des deux autres. Pour le chromosome 3, seul le génotype homozygote nocturne (NN) est significativement différent des deux autres génotypes ( $W = 446$ ,  $p = 0,0005$  pour ND et  $W = 331$ ,  $p = 0,003$  pour DD). Les hétérozygotes DN regroupent les génotypes DN et ND.

#### II.1.3.4. Gènes candidats

Pour le croisement A, il y a 497 gènes dans le QTL du chromosome 3. Pour le croisement B, il y a 736 gènes pour le QTL du chromosome 3 et 210 gènes pour le QTL du chromosome 4 (liste complète des gènes en Annexe 6). Les 497 gènes du croisement A sont inclus dans les 736 gènes du croisement B. Pour discerner ces différents gènes candidats, une analyse d'enrichissement GO et de Kegg pathway ont été réalisées. L'enrichissement GO a été fait avec Blast2GO version 5.2 en utilisant le Fisher's Exact Test ( $p < 0,05$ ). L'analyse d'enrichissement Kegg pathway a été faite sur le serveur KAAS en utilisant le programme Blast avec la méthode SBH (Single-directional Best Hit).

##### II.1.3.4.1. Les gènes de l'horloge (Basic Helix-Loop-Helix)

Parmi les 946 gènes QTLs, aucun n'est un gène canonique de l'horloge. Tout comme avec le croisement intraspécifique, on retrouve 5 gènes qui présentent un domaine BHLH (Basic Helix-Loop-Helix) de type Myc (Tableau\_33). Si on s'intéresse aux annotations HHsearch, on retrouve la protéine Mad qui lorsqu'elle forme l'hétérodimère Mad-Max est impliquée dans le recrutement d'histone désacétylase (Laherty, Yang et al., 1997). La protéine Mad est en compétition avec la protéine Myc pour se lier à la protéine Max.

Tableau 33 : Liste des gènes QTLs avec un motif BHLH (Basic Helix Loop Helix) pour les QTLs des croisements interspécifiques

Intérêt	Chr.	Identifiant (ID) du gène	Annotation GFF	Annotation HHsearch
Motif BHLH	3AB	Smp_021340	Myc-type, basic helix-loop-helix (bHLH) domain	Mad protein {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
		Smp_123570		Aryl hydrocarbon receptor nuclear translocator
		Smp_342580		Transcription factor E2-alpha, DNA-binding protein
		Smp_000620		DNA-binding protein inhibitor ID-1
		Smp_333620		Myc proto-oncogene protein {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}

Chr.: chromosome; A : croisement A; B: croisement B

#### II.1.3.4.2. Les rhodopsines et autres gènes impliqués dans la phototransduction

Tout comme dans les QTLs du croisement intraspécifique, on retrouve une rhodopsine (Smp\_117340) et 26 autres rhodopsine-like (Tableau\_34).

Tableau 34 : Liste des gènes QTLs rhodopsine et rhodopsine-like pour le QTL interspécifique

Intérêt	Chr.	Identifiant (ID) du gène	Annotation GFF	Annotation HHsearch
Rhodopsine	3B	Smp_117340	G protein-coupled receptor, rhodopsin-like	Rhodopsin
Rhodopsine-like	3AB	Smp_150180		Chimera of Proteinase-activated receptor 1
		Smp_159860		Cannabinoid receptor 1, Flavodoxin
	3B	Smp_132410		Lysozyme, Proteinase-activated receptor 2, Soluble cytochrome b562, Proteinase-activated
		Smp_083880		Neurotensin receptor type 1, Endolysin, NEUROTENSIN RECEPTOR
		Smp_177720		C-X-C chemokine receptor type 4
		Smp_083940		7TM_GPCR_Srw
		Smp_326640		7TM_GPCR_Srw
		Smp_241490		7TM_GPCR_Srw
		Smp_091950		7TM_GPCR_Srw
		Smp_326670		Platelet-activating factor receptor, Flavodoxin
		Smp_134960		Platelet-activating factor receptor, Flavodoxin
		Smp_023710		Platelet-activating factor receptor, Flavodoxin



		Smp_134820		Chimera of Proteinase-activated receptor 1	
	4B	Smp_343810	GPCR, rhodopsin-like, 7TM	Adenosine receptor A2a,Soluble cytochrome b562,Adenosine	
		Smp_043260	G protein-coupled receptor, rhodopsin-like	7tm_1	
		Smp_043270			
		Smp_336140			
		Smp_348040			
		Smp_043290			
		Smp_043300			
		Smp_348050			
		Smp_332070			human Chemokine Receptor 7
		Smp_043320			7tm_1
		Smp_043340			Muscarinic acetylcholine receptor M2,Vasopressin V2
		Smp_145540	Lysophosphatidic acid receptor 6a,Endolysin,receptor (E.C.3.2.1.17)		
		Smp_212750			

Chr.: chromosome; A : croisement A; B: croisement B

Il y a aussi dans ce croisement interspécifique la présence de certains acteurs de la phototransduction. En plus de la rhodopsine, il y a, dans les gènes QTLs, la protéine G (transducine), la phosphodiesterase, la sérine/thréonine kinase (rhodopsine kinase), des gènes contenant le domaine EF-hand qui est présent dans la recoverine, la protéine phosphatase 2A (PP2A) et la GTPase activating protein (Tableau\_34 et Tableau\_35). Contrairement au croisement intraspécifique, l'arrestine et la guanylyl cyclase ne sont pas présentes dans les gènes QTLs.

Tableau\_35 : Liste des gènes QTLs potentiellement liés à la phototransduction

Intérêt	Chr.	Identifiant (ID) du gène	Annotation GFF	Annotation HHsearch
Protéine G (Transducine)	3B	Smp_075130	G-protein, beta subunit	beta1-subunit of the signal-transducing G protein heterotrimer {Cow ( <i>Bos taurus</i> ) [TaxId: 9913]}
		Smp_018400	WD40 repeat	beta1-subunit of the signal-transducing G protein heterotrimer {Cow ( <i>Bos taurus</i> ) [TaxId: 9913]}
Phosphodiesterase	3AB	Smp_084890	Type I phosphodiesterase /nucleotide pyrophosphatase/p hosphate transferase	PHOSPHONOACETATE HYDROLASE (E.C.3.11.1.2)

	3B	Smp_018360	PLC-like phosphodiesterase, TIM beta/alpha-barrel domain superfamily	Crystal Structure of the Y247S/Y251S
	4B	Smp_135500	3'5'-cyclic nucleotide phosphodiesterase, catalytic domain	cGMP-specific 3',5'-cyclic phosphodiesterase (E.C.3.1.4.35)
		Smp_053550	U6 snRNA phosphodiesterase Usb1	U6 snRNA phosphodiesterase (E.C.3.1.4.-)
GPCR kinase, sérine/thréonine kinase (rhodospine kinase)	3AB	Smp_123610	Protein kinase domain	Phosphatidylinositol 3-kinase VPS34 (E.C.2.7.1.137), Serine/threonine-protein
		Smp_123640	GPCR kinase	G protein-coupled receptor kinase 4
	3B	Smp_144390	Ser/Thr protein kinase, TGFB receptor	B-Raf kinase {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
	4B	Smp_042920	Serine-threonine protein kinase 19	Stk19
		Smp_308680	Protein kinase domain	Serine/threonine-protein kinase Nek2 {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
domaine EF-hand (Recoverine)	3AB	Smp_106110	EF-hand domain	Kchip1, Kv4 potassium channel-interacting protein {Rat ( <i>Rattus norvegicus</i> ) [TaxId: 10116]}
		Smp_181460	EF-hand domain pair	DUF4205
		Smp_106060	EF-hand domain	ADP,ATP carrier protein {Cow ( <i>Bos taurus</i> ), heart isoform t1 [TaxId: 9913]}
	3B	Smp_000180	EF-hand domain	Frequenin (neuronal calcium sensor 1) {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
		Smp_134950		Myosin Essential Chain {Bay scallop ( <i>Aequipecten irradians</i> ) [TaxId: 31199]}
		Smp_018940		Myosin Essential Chain {Bay scallop ( <i>Aequipecten irradians</i> ) [TaxId: 31199]}
	4B	Smp_313310	Sarcoplasmic calcium-binding protein {Amphioxus ( <i>Branchiostoma lanceolatum</i> ) [TaxId: 7740]}	
	Protéine phosphatase 2A (PP2A)	3AB	Smp_308680	Protein phosphatase 2A regulatory subunit PR55
Smp_308690			NA	
Smp_308700			Serine/threonine-protein phosphatase 2A 65 kDa	

GTPase activating protein		Smp_138500	Arf GTPase activating protein	Ras-related protein Rab-35, Iporin	
		Smp_000670		ARFGAP with coiled-coil, ANK repeat	
		Smp_160610	Rho GTPase-activating protein domain	p50 RhoGAP domain {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}	
		Smp_138430	DEP domain	Regulator of G-protein signaling 7	
GTPase		Smp_246610	Small GTPase	Rad {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}	
		Smp_084940		Ras-related protein M-Ras (XRas) {Mouse ( <i>Mus musculus</i> ) [TaxId: 10090]}	
		Smp_132330		Rab30 {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}	
	3B	Smp_179910		GTP-binding protein RheB {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}	
		Smp_134940		GTP binding domain	YqeH GTPase
		Smp_342660		Small GTPase	Rad {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
		4B			Smp_313300

Chr.: chromosome; A : croisement A; B: croisement B

#### II.1.3.4.3. Les gènes impliqués dans l'acétylation et la méthylation des histones

Un autre point commun avec le QTL intraspécifique est la présence de gènes liés aux histones (Tableau\_36). Tout comme précédemment, on retrouve des gènes qui codent pour des histones et aussi des histones acétyltransférase (« writer ») mais pas d'histones désacétylases (« eraser »). Une spécificité du QTL interspécifique est qu'il y a aussi l'équivalent pour la méthylation des histones avec la présence du gène histone méthyltransférase (« writer ») et du gène histone déméthylase (« eraser »).

**Tableau 36 : Liste des gènes QTLs en lien avec les histones, l'acétylation des histones et la méthylation/déméthylation des histones**

Intérêt	Chr.	Identifiant (ID) du gène	Annotation GFF	Annotation HHsearch
Histone	3AB	Smp_337740	Histone H2A	Histone H2A {Chicken ( <i>Gallus gallus</i> ), erythrocytes [TaxId: 9031]}
	3B	Smp_000150	Zinc finger, PHD-type	Histone H3.2, Histone H4, Histone
	3B	Smp_027850	Chromatin SPT2	Histone H3.2, Histone H4, Protein
	3B	Smp_026880	Histone H3/CENP-A	Leishmania histone H3, Histone H4
Histone acétylase	3AB	Smp_105910	Zinc finger, TAZ-type	Histone acetyltransferase p300 (E.C.2.3.1.48), p53
	3B	Smp_083930	Chromatin modification-related protein Eaf6	Histone acetyltransferase ESA1 (E.C.2.3.1.48), Chromatin

Histone méthyl-transférase	3AB	Smp_210660	SET domain	Histone-lysine N-methyltransferase 2E (E.C.2.1.1.43)
	4B	Smp_150850	Methyl-CpG DNA binding	Histone-lysine N-methyltransferase SETDB1 (E.C.2.1.1.43), Histone
	4B	Smp_160700	PWWP domain	Histone-lysine N-methyltransferase SETD2 (E.C.2.1.1.43,2.1.1.-), H3.3S31phK36M(29-42)
Histone déméthylase	3AB	Smp_311980	Zinc finger C2H2-type	LYSINE-SPECIFIC HISTONE DEMETHYLASE 1A (E.C.1.-.-.-)
	3AB	Smp_000230	Zinc finger C2H2-type	LYSINE-SPECIFIC HISTONE DEMETHYLASE 1A (E.C.1.-.-.-)

Chr.: chromosome; A : croisement A; B: croisement B

#### II.1.3.4.4. Enrichissement GO : Processus biologiques

Les analyses d’enrichissement GO des gènes QTLs n’ont pas donné de résultats significatifs en utilisant la limite usuelle FDR < 0,05 (False Discovery Rate). Les résultats qui suivent ont été obtenus en utilisant la limite de  $p < 0,05$ .

On retrouve, parmi les processus biologiques mis en évidence par les analyses d’enrichissement GO, la réponse à la chaleur et la régulation du cycle cellulaire. L’annotation GO « réponse à la chaleur » correspond au gène **Smp\_096010**, une Heat shock protein DnaJ (HSP60) qui, en interagissant avec HSP70, empêche l’agrégation des protéines partiellement dénaturées causée par un stress environnemental (Ohtsuka & Hata, 2000). L’annotation GO « régulation du cycle cellulaire » contient le gène **Smp\_181450** qui code pour CCDC66 (Coiled-Coil Domain Containing 66). La déplétion de ce gène entraîne une rétinopathie pigmentaire (dégénérescence rétinienne) (Gerding, Shreiber et al., 2011).



Figure 56 : Treemap revigo des résultats de l’analyse d’enrichissement GO des processus biologiques obtenus pour les gènes QTL du croisement interspécifique.

Chaque couleur représente un grand ensemble de processus biologique. Il y a de gauche à droite : La régulation de la division cellulaire, le traitement de l'extrémité 3' du petit ARN nucléaire U2, transport du bicarbonate, processus métabolique des composés azotés, assemblage des particules régulatrices du protéasome, réponse à la chaleur, réparation des protéines, sulfatation, repliement des protéines « de novo », adhésion cellulaire, repliement des protéines, contraction des muscles, processus métabolique, réponse à des stimulus, communication cellulaire, mannosylation, signalisation, processus métabolique superoxyde, processus métabolique des espèces réactives de l'oxygène, glissement des microtubules et déphosphorylation. La treemap a été obtenue avec revigo (Supek, Bošnjak et al., 2011).

#### II.1.3.4.5. Kegg pathway

Les analyses Kegg pathway montrent que les gènes QTLs des croisements interspécifiques sont potentiellement impliqués dans 282 voies moléculaires différentes. Parmi toutes ces voies, sept retiennent notre attention, le rythme circadien, le rythme circadien chez la drosophile, l'entraînement circadien, la phototransduction, la phototransduction chez la drosophile et la transduction olfactive, la transduction gustative (Tableau\_37).

**Tableau 37: Liste des gènes QTLs correspondant aux Kegg pathways sélectionnés.** Les gènes marqués d'un astérisque (\*) sont des gènes correspondant à plusieurs Kegg pathways sélectionnés

Kegg pathway	KO annotation	Chr.	Identifiant (ID) du gène	Annotation GFF	Annotation HHsearch
Rythme circadien	K07199	4B	Smp_329580	5'-AMP-activated protein kinase subunit beta-1 (Fragment)	5'-AMP-activated protein kinase subunit beta-2
Rythme circadien drosophile	K23643	3B	Smp_023700	Anticodon-binding domain	AD
Entraînement du rythme circadien	K02183		Smp_018940	EF-hand domain	Myosin Essential Chain {Bay scallop ( <i>Aequipecten irradians</i> ) [TaxId: 31199]}
	K04539		Smp_075130	G-protein, beta subunit	beta1-subunit of the signal-transducing G protein heterotrimer {Cow ( <i>Bos taurus</i> ) [TaxId: 9913]}
Phototransduction	K02183		Smp_018940*	EF-hand domain	Myosin Essential Chain {Bay scallop ( <i>Aequipecten irradians</i> ) [TaxId: 31199]}
Phototransduction chez la drosophile		4B	Smp_135500	3'5'-cyclic nucleotide phosphodiesterase, catalytic domain	cGMP-specific 3',5'-cyclic phosphodiesterase (E.C.3.1.4.35)
Transduction olfactive	K18283				
Transduction gustative	K04153	3B	Smp_134820	G protein-coupled receptor, rhodopsin-like	Chimera of Proteinase-activated receptor 1

	K04131	4B	Smp_145540	G protein-coupled receptor, rhodopsin-like	Muscarinic acetylcholine receptor M2, Vasopressin V2
--	--------	----	------------	--	--

Chr.: chromosome; A : croisement A; B: croisement B.

On retrouve un seul gène dans le pathway rythme circadien qui est **Smp\_329580**, une sous-unité beta de la kinase 5'AMP-activatrice de protéine (AMPK). Il y a aussi un gène dans le pathway du rythme circadien chez la drosophile qui est **Smp\_023700**, un gène qui code pour une protéine qui contient un domaine AD (Anticodon-binding domain). Dans le pathway du rythme circadien chez la drosophile, ce gène correspond au LSM12 qui forme un complexe avec deux autres protéines pour stimuler la traduction du gène de l'horloge per (Lee, Yoo et al., 2017).

Le pathway de l'entraînement circadien compte deux gènes. Le premier est **Smp\_075130**, une sous-unité beta de la protéine G et le second est **Smp\_018940**, une putative calmoduline, un détecteur de calcium ubiquitaire qui a des implications dans la transduction du signal et le rythme circadien (Cavieres-Lepe & Ewer, 2021). Le gène Smp\_018940 est aussi présent dans le pathway de la phototransduction, de la phototransduction chez la drosophile et de la transduction olfactive.

Le pathway de la transduction olfactive contient aussi le gène **Smp\_135500** qui est une phosphodiesterase. Enfin, le pathway de la transduction gustative contient deux gènes. Le gène **Smp\_134820**, une rhodopsine-like qui correspond dans le pathway à un récepteur sérotoninergique ; c'est-à-dire un récepteur de la sérotonine, un neurotransmetteur impliqué, entre autres, dans le rythme circadien. Le second gène est le **Smp\_145540**, une autre rhodopsine-like qui correspond à un récepteur muscarinique acétylcholine. L'acétylcholine est considérée comme le neurotransmetteur de « l'éveil » puisque son taux est élevé pendant les périodes d'activité physique et d'éveil (Hut & Van der Zee, 2011).

### II.1.3.5. Conclusion

Le QTL mapping issu des croisements interspécifiques entre *S. mansoni* et *S. rodhaini* a mis en évidence 2 QTLs sur les chromosomes 3 (croisements A et B) et 4 (croisement B uniquement). Ces QTLs sur des chromosomes différents de ceux du QTL intraspécifique indiquent que les différences de rythme d'émission des cercaires reposent sur des régions génomiques différentes entre les espèces. Il n'est pas à exclure, cependant, que les gènes présents dans les QTLs intra et interspécifique jouent un rôle dans le rythme d'émission des cercaires, peu importe l'espèce de schistosome. Cela est d'autant plus vrai que l'on retrouve des gènes reposant sur des mécanismes similaires entre les deux expériences QTLs (Figure\_57). En effet, tout comme le croisement intraspécifique, il n'y a pas de gène canonique de l'horloge conformément avec Rawlinson, Reid et al. (2021) mais il y a des gènes avec domaine BHLH de type Myc. La sélection subjective basée sur les annotations GFF et HHsearch a aussi mis en évidence des gènes impliqués dans la phototransduction et plus largement dans la transduction des signaux environnementaux, avec la rhodopsine comme probable photorécepteur.

De plus, tout comme le croisement intraspécifique, la présence de gènes avec un motif BHLH qui recrute les histones acétylases et désacétylases ainsi que la présence de gène histone acétylase mais aussi d'histone méthyltransférase et déméthylase sont des résultats à mettre en regard des analyses épigénétiques qui font l'objet du prochain chapitre afin de voir l'implication de l'acétylation et de la méthylation des histones dans le rythme d'émission des cercaires.

Les QTLs intra- et interspécifiques partagent la réponse à des stimuli, la signalisation et la communication cellulaire comme processus biologiques mis en évidence par les analyses d'enrichissement GO (Figure\_56).

Enfin, les analyses de Kegg pathway ont confirmé l'implication fonctionnelle de certains gènes dans les voies moléculaires touchant à la régulation du rythme circadien et à la phototransduction, et plus largement dans la transduction des signaux environnementaux. Il est intéressant de noter que bien que les gènes soient impliqués dans les mêmes pathways, ils ne correspondent pas aux mêmes annotations fonctionnelles. Ces résultats indiquent que le rythme d'émission des cercaires nocturnes repose sur un mécanisme moléculaire similaire centré autour de rhodospine et la transduction de signaux environnementaux mais que les acteurs de régulation diffèrent entre les espèces de schistosomes.

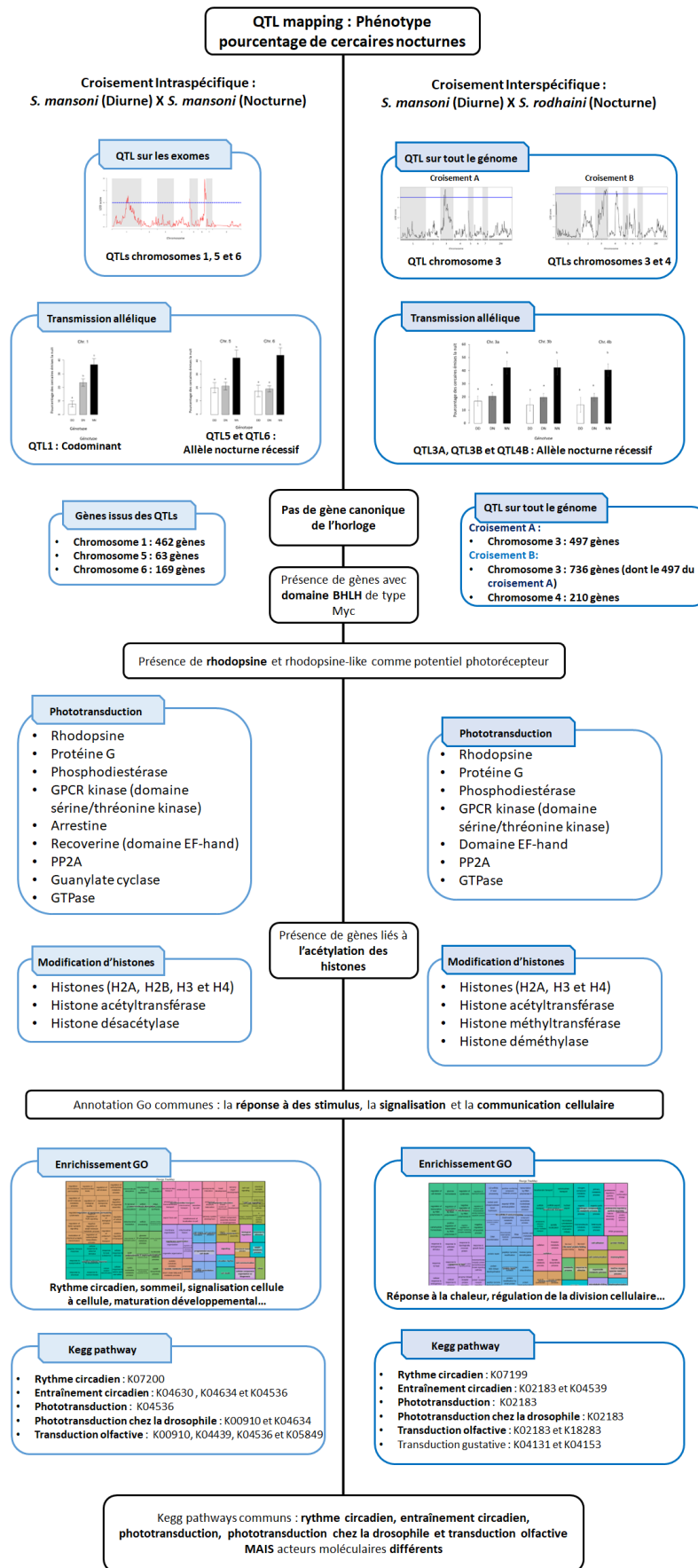


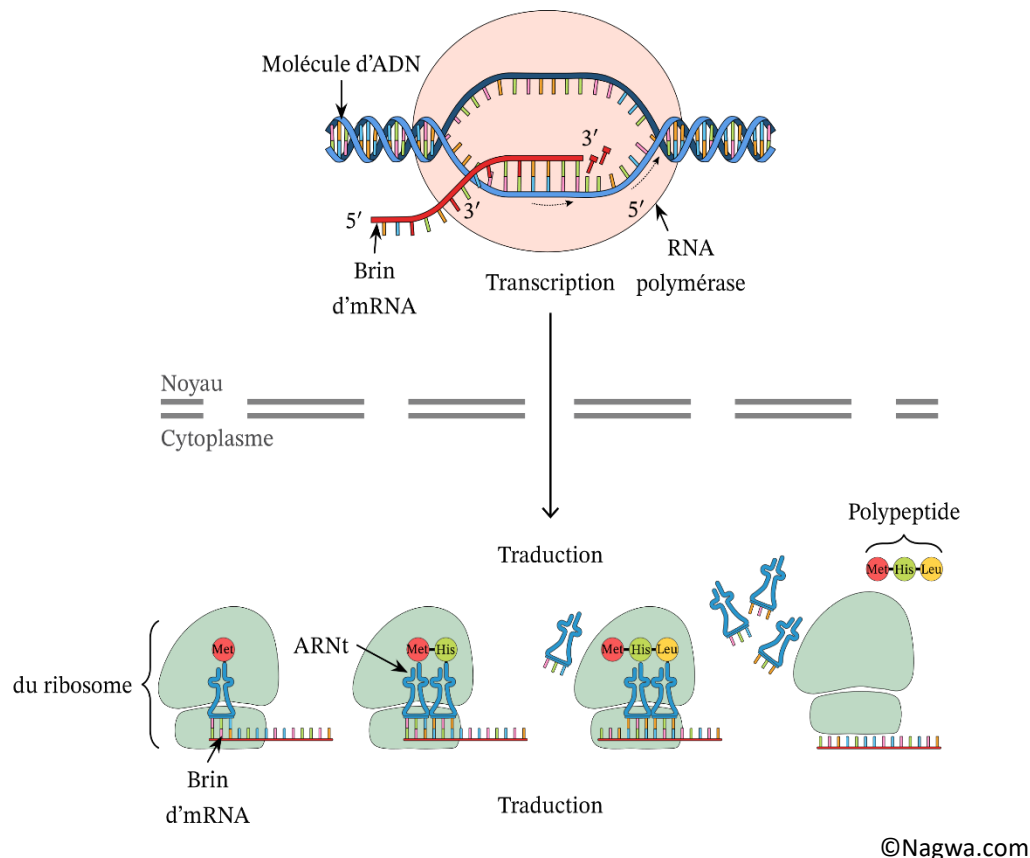
Figure 57 : Bilan des QTL mapping intra et interspécifiques.





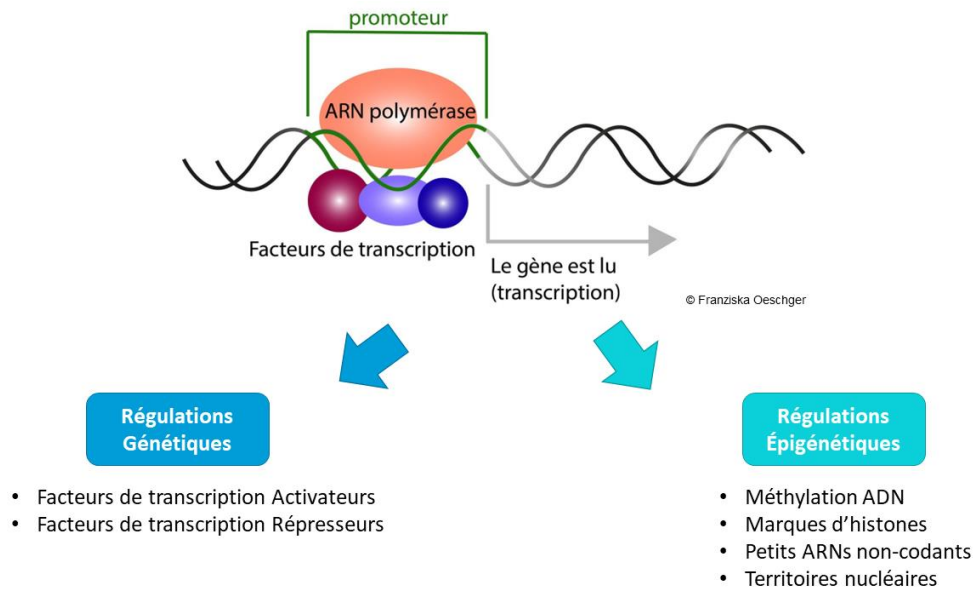
## II.2. Recherche des marqueurs rythmiques

La présence d'un gène dans le génome ne signifie pas forcément qu'il est impliqué dans un phénotype. Pour connaître son implication il est donc nécessaire d'étudier son expression. L'expression d'un gène passe par une étape de transcription de l'ADN double brin en ARN messager (ARNm) simple brin, puis d'une étape de traduction de cet ARNm en protéine qui donnera un phénotype (Figure\_58).



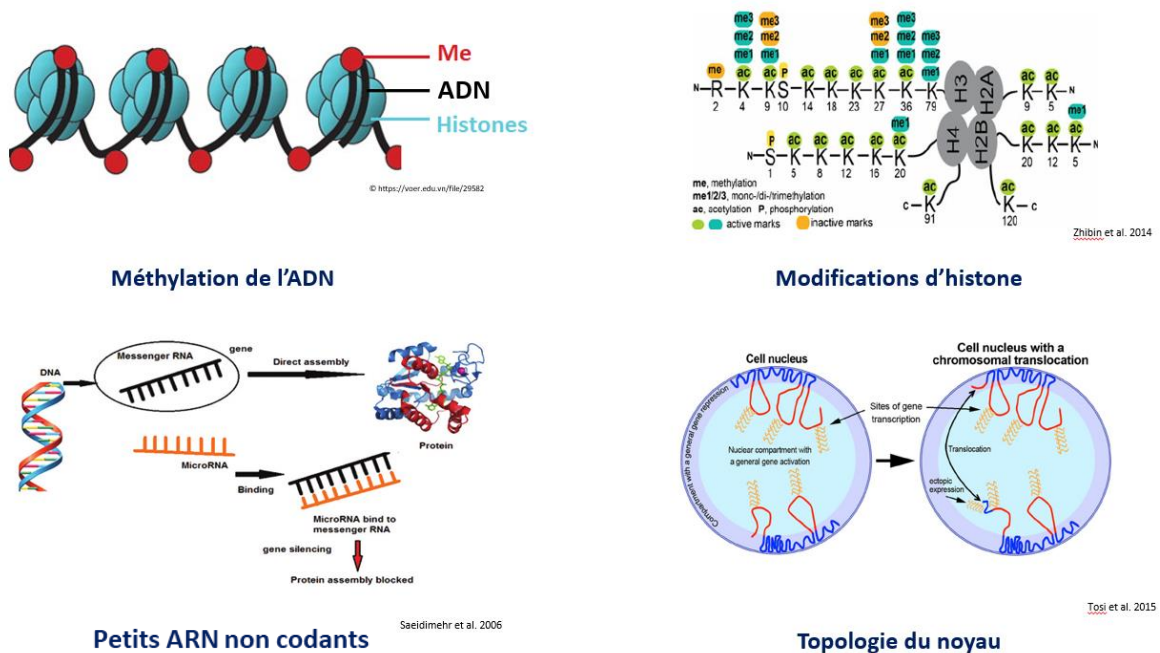
Figure\_58 : Schéma de l'expression d'un gène de la transcription jusqu'à la traduction.

La transcriptomique est l'étude des ARN messagers (ARNm) produits lors de la transcription. Ce phénomène peut être régulé par des acteurs génétiques (facteurs de transcription) et des acteurs épigénétiques (Figure\_59).



**Figure 59 : Schéma de la transcription d'un gène et des acteurs de la régulation.** De façon simplifiée lors de la transcription, l'ARN polymérase est recrutée par un promoteur en amont du gène. Une fois fixée, la polymérase synthétise un brin d'ARN en prenant comme modèle le brin d'ADN. Le recrutement de cette ARN polymérase peut être facilité ou empêché par des actions de régulations venant d'acteurs génétique et/ou épigénétique.

L'épigénétique est l'étude des caractères phénotypiques héréditaires qui n'impliquent pas de modification de la séquence d'ADN. Ces traits sont réversibles et héréditaires (mitotiquement et/ou méiotiquement). Il existe quatre principaux types de modification épigénétique : la méthylation de l'ADN, les modifications d'histone, la topologie du noyau et les petits ARN non-codants (Figure\_60) (Joly & Grunau, 2018).



**Figure 60 : Les quatre principaux types de modification épigénétique.**

A) La méthylation de l'ADN est l'ajout d'un groupement méthyl (Me en rouge) qui s'effectue uniquement sur les adénosines et les cytosines. De façon générale, la méthylation de l'ADN est associée à la répression des zones marquées. B) La modification d'histone s'opère sur les acides aminés des queues d'histones qui composent le nucléosome. En fonction de la marque, elle peut induire une activation ou une répression de la transcription en changeant l'état de compaction de la chromatine. C) Les petits ARN non codants sont inférieurs à 30 nucléotides et sont impliqués dans la répression de l'expression des gènes. D) La topologie du noyau consiste en des mouvements chromatiniens qui rapprochent ou éloignent physiquement des sites de transcription.

Chacune de ces modifications est associée à des phénomènes de régulation de l'horloge circadienne (Singh, Jha & Thakur, 2019). Par exemple la méthylation de l'ADN joue un rôle dans l'expression du gène clock *Frq* chez *Neurospora crassa* (Belden, Lewis et al., 2011). Chez la souris, l'acétylation de l'histone H3 dans les promoteurs de *Per1*, *Per2* et *Cry1* (gènes clock) suit un rythme circadien (Etchegaray, Lee et al., 2003). Des territoires chromosomiques circadiens ont pu être mis en évidence dans les fibroblastes embryonnaires des souris (Aguilar-Arnal, Hakim et al., 2013). Enfin, chez les mammifères, le miR-219 a un rythme circadien et son knock-down entraîne un rallongement de la période circadienne, de plus il est une cible du complexe CLOCK-BMAL1 (Cheng, Papp et al., 2007).

Sur les quatre types de modification épigénétique, deux sont au centre des recherches sur *S. mansoni* ; la modification d'histone et les petits ARN non codants (Azzi, Cosseau & Grunau, 2009 ; Simões, Lee et al., 2011 ; Picard, Boissier et al., 2016 ; Cosseau, Wolkenhauer et al., 2017). En effet, il existe très peu de méthylation de l'ADN chez *S. mansoni* (Fantappié, Gimba & Rumjanek, 2001) et aucune étude sur la topologie nucléaire n'a été faite pour cet organisme.

L'objectif de ce chapitre va être, dans un premier temps, d'identifier les marqueurs moléculaires rythmiques au travers de deux approches complémentaires, l'approche épigénétique et l'approche transcriptomique, car ayant été réalisées sur les mêmes échantillons biologiques ; dans un second temps, d'analyser les différents gènes candidats qui auront émergé de cette seconde partie de thèse consacrée aux déterminants moléculaires du rythme de l'émission des cercaires.

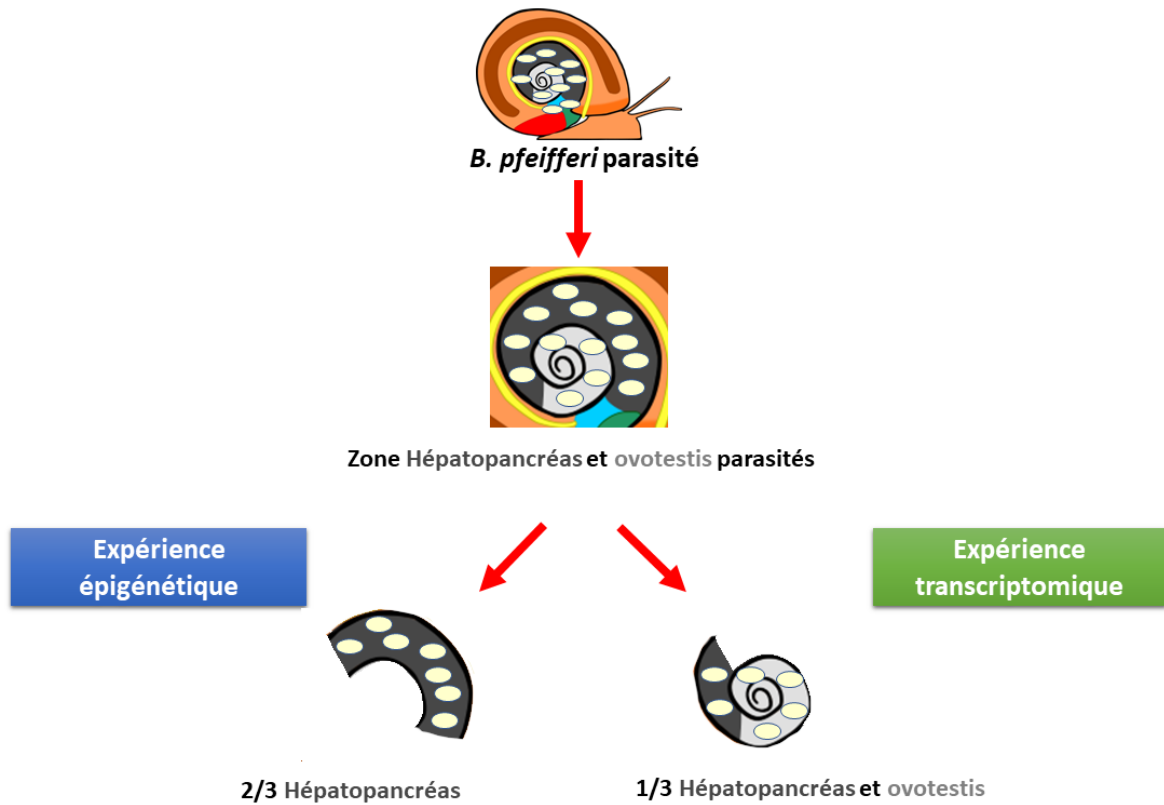
Le premier point développera l'expérience d'échantillonnage commune aux approches d'épigénétique et transcriptomique. Le second point abordera l'épigénétique au travers de la technique de ChIPmentation et le troisième point la transcriptomique avec la technique de RNA-seq. Enfin le dernier point sera consacré à une synthèse et analyse des divers gènes candidats.

## II.2.1. Echantillonnage commun pour les expériences d'épigénétique et de transcriptomique

### II.2.1.1. Méthode d'échantillonnage pour les expériences d'épigénétique et de transcriptomique

L'échantillonnage est commun aux expériences de ChIPmentation et de RNA-seq. Ces deux expériences partagent l'objectif d'identification d'un potentiel rythme (respectivement acétylation/méthylation d'histone et expression de gènes). L'échantillonnage comprend l'infestation des mollusques, le sexage des parasites, la chronobiologie des individus et la dissection des mollusques pour récupérer les hépatopancreas et ovotestis infestés. Chaque dissection se fait en retirant la



coquille et le pied du mollusque. Deux-tiers de l'hépatopancréas sont utilisés pour le ChIPmentation et le tiers restant ainsi que l'ovotestis sont utilisés pour le RNA-seq (Figure\_61) ; tous les échantillons sont plongés, dès la fin de la dissection, dans de l'azote liquide puis stockés au congélateur -80°C. Les deux chronotypes de *S. mansoni*, diurne et nocturne, sont analysés.



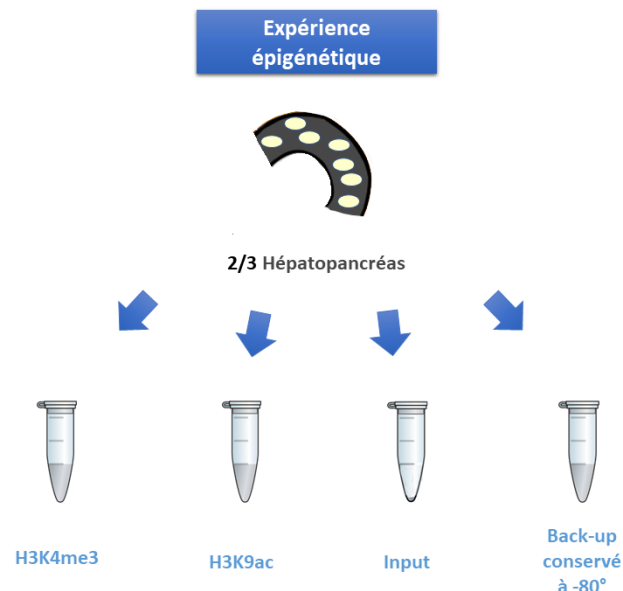
**Figure 61 : Répartition de l'échantillon entre les expériences de ChIPmentation et de RNA-seq.** Pour chaque *B. pfeifferi*, 2/3 de l'hépatopancréas sert pour le ChIPmentation et le 1/3 restant ainsi que l'ovotestis servent pour le RNA-seq. L'hépatopancréas est en gris foncé, l'ovotestis en gris clair et les parasites en ovales beiges.

Pour chaque chronotype, deux séries d'échantillonnage ont eu lieu. La première consiste à disséquer toutes les quatre heures pendant 48h trois *B. pfeifferi* infestés par des parasites mâles et trois *B. pfeifferi* infestés par des parasites femelles qui sont restés sous une photopériode classique (L-O). La deuxième série suit chronologiquement la première et consiste à disséquer toutes les quatre heures pendant 24h trois *B. pfeifferi* infestés par des parasites mâles et trois *B. pfeifferi* infestés par des parasites femelles qui ont été soumis, en plus, pendant 10 jours, à une photopériode inversée (O-L) (Tableau\_38).

**Tableau 38 : Description des deux séries d'échantillonnage pour les expériences d'épigénétique et de transcriptomique.** Les deux séries sont réalisées pour les deux chronotypes de *S. mansoni*. Les mâles et les femelles correspondent au sexe du parasite qui infeste les *B. pfeifferi*.

Série d'échantillonnage	1 <sup>ère</sup> série	2 <sup>ème</sup> série
Photopériode	L-O	O-L pendant 10 jours
Durée échantillonnage	48h 	24h 
Ecart entre les points	4h	
Nombre d'individus par point	3 mâles et 3 femelles	

Pour le ChIPmentation, l'extraction de la chromatine permet de pouvoir réaliser trois réactions d'immunoprécipitations différentes avec un input par morceau d'hépatopancreas (Figure\_62). Dans le cadre de nos expériences, deux réactions sont utilisées (H3K4me3 et H3K9ac) et la 3<sup>ème</sup> dose sert de back-up soit pour refaire une IP soit pour éventuellement faire le ChIPmentation sur une nouvelle marque.



**Figure 62 : Description des deux séries d'échantillonnage pour les expériences d'épigénétique et de transcriptomique.** Les deux séries sont réalisées pour les deux chronotypes de *S. mansoni*. Les mâles et les femelles correspondent au sexe du parasite qui infeste les *B. pfeifferi*.

Cet échantillonnage commun entre l'expérience d'épigénétique et de transcriptomique permet pour chaque individu/génome de parasite d'avoir deux profils : le profil épigénétique et l'expression des gènes associés.

## II.2.1.2. Chronobiologie

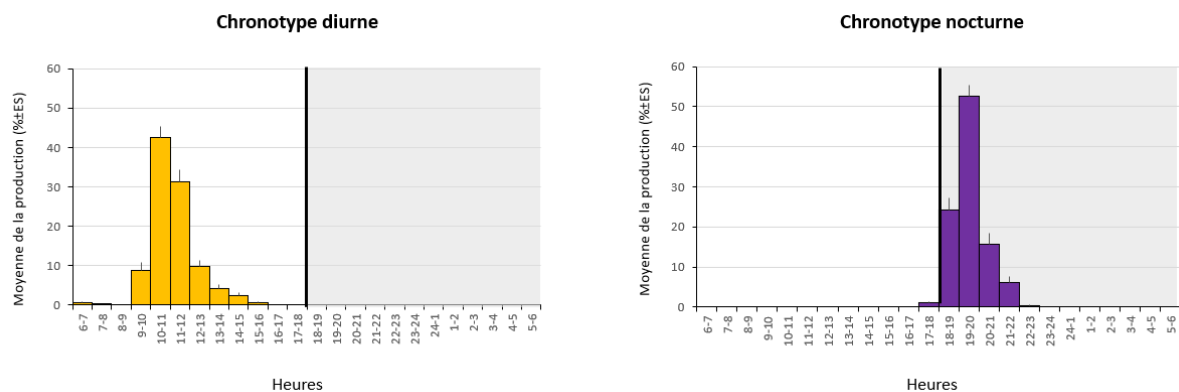
### II.2.1.2.1. Protocole

Sur les 480 *B. pfeifferi* infestés avec 1 miracidium de *S. mansoni* diurne, 99 se sont révélés positifs à l'infestation et ont été échantillonnés. Sur ces 99 *B. pfeifferi*, 57 ont été infestés par des mâles et 42 par des femelles. Sur les 504 *B. pfeifferi* infestés avec 1 miracidium de *S. mansoni* nocturne, 123 se sont révélés positifs à l'infestation et ont été échantillonnés. Sur ces 123 *B. pfeifferi*, 62 ont été infestés par des mâles et 61 par des femelles. La chronobiologie a été effectuée pendant deux jours consécutifs sur 12 profils d'infestations (6 mâles et 6 femelles) pour chaque chronotype, soit au total 24 profils diurnes et 24 profils nocturnes.

#### II.2.1.2.2. Chronotype diurne

Les résultats moyens, accompagnés des variations individuelles, sont présentés dans la Figure\_63 et le Tableau\_39. Sur les 24 profils diurnes réalisés, 16 ont pu être utilisés car leur émission totale sur les 24 heures était égale ou supérieure à 20 cercaires.

Le chronotype diurne a une émission unimodale dont l'heure du pic moyen est à 10-11h. La plage horaire moyenne d'émission est d'environ 4h avec une intensité de 92%. La plage d'émission a une asymétrie modérée sur la droite et une distribution normale. Il n'y a pas d'émission de cercaires nocturne.



**Figure 63 : Chronobiologie des chronotypes diurne (orange) et nocturne (violet) de *S. mansoni* pour les échantillons d'épigénétique et transcriptomique.** Chaque histogramme a, en ordonnées, la moyenne du pourcentage des cercaires émises  $\pm$  écart standard. L'axe des abscisses représente les 24h de la journée, débutant à 6h, début du jour, et finissant à 6h, fin de la nuit. L'arrière-plan gris correspond à l'obscurité. La ligne noire marque le début de l'obscurité à 18h.

**Tableau 39 : Paramètres des profils moyens d'émission, accompagnés des variations individuelles, pour les parasites diurne et nocturne de *S. mansoni* pour les échantillons de l'expérience épigénétique et transcriptomique**

Paramètres	Moyenne $\pm$ ES <i>Variations individuelles</i>		Test de Wilcoxon-Mann-Whitney
	Diurne (N=16)	Nocturne (N=20)	
Photopériode	L-O	L-O	
Nombre de pics	1 1	1 1	
Heure du pic	10-11h 10-11h à 11-12h	19-20h 18-19h à 20-21h	$\neq$
Plage d'émission (h)	3,6 $\pm$ 0,2 3 à 5	3,4 $\pm$ 0,1 3 à 5	W = 0, p = 0,53
Intensité de la plage (%)	92,3 $\pm$ 1,9 69,8 à 100	97,6 $\pm$ 0,5 94,3 à 100	W = 74, p = 0,006
Asymétrie	+0,7 $\pm$ 0,2 -1,4 à +1,8	+0,5 $\pm$ 0,3 -1,7 à +1,7	W = 166, p = 0,86
Aplatissement	0 $\pm$ 0,8 (N=8) -3,0 à +3,3	-0,7 $\pm$ 1,3 (N=7) -5,1 à +2,9	W = 30, p = 0,87
Cercaires la nuit (%)	0 0	99,0 $\pm$ 0,3 95,2 à 100	W = 177,5, p = 5,75e-08

#### II.2.1.2.3. Chronotype nocturne

Les résultats moyens, accompagnés des variations individuelles, sont présentés dans la Figure\_63 et le Tableau\_39. Sur les 24 profils nocturnes réalisés, 20 profils ont rempli cette condition d'émission d'au moins 20 cercaires.

Le chronotype nocturne a une émission unimodale dont l'heure du pic moyen est à 19-20h. La plage horaire moyenne d'émission est d'un peu plus de 3h avec une intensité de 97%. La plage d'émission moyenne est symétrique avec une distribution normale. L'émission nocturne est majoritaire avec en moyenne 99% des cercaires émises la nuit.

#### II.2.1.2.4. Comparaison des chronotypes

Les comparaisons sont présentées dans le Tableau\_39. Les différences et similitudes observées entre les deux chronotypes dans cette expérience d'échantillonnage sont les mêmes que celles établies dans la première partie (présentation des chronotypes). Les profils moyens des deux chronotypes se différencient par l'heure du pic (10-11h pour les diurnes et 19-20h pour les nocturnes) et le pourcentage de cercaires émises la nuit (0% pour les diurnes et 99% pour les nocturnes ;  $p < 0,01$ ). Il est à noter que l'intensité de l'émission de la plage est significativement plus forte chez le chronotype nocturne ( $p < 0,01$ ). Les deux chronotypes ont en commun la présence d'un seul pic. De plus, les plages d'émission, leurs coefficients d'asymétrie et d'aplatissement ne sont pas significativement différents entre les diurnes et les nocturnes ( $p > 0,01$ ).



## II.2.2. ChIPmentation

Dans le cadre de cette thèse nous allons nous concentrer uniquement sur la modification d'histone pour décortiquer le rythme d'émission des cercaires. Les modifications des histones impliquées dans l'état de la chromatine sont associées à l'activation et à la répression de la transcription. Les histones sont des protéines responsables de la compaction de l'ADN et donc, de l'accès ou non de la machinerie de transcription à l'ADN. La modification d'histones est une modification post-traductionnelle, qui a lieu sur les acides aminés constituant la queue des histones H2A, H2B, H3 et H4. Il existe quatre types de modification d'histone qui sont la méthylation, l'acétylation, la phosphorylation et l'ubiquitination (Figure\_64). On retrouve, pour chacune de ces modifications, des exemples d'implication dans la régulation du rythme circadien.

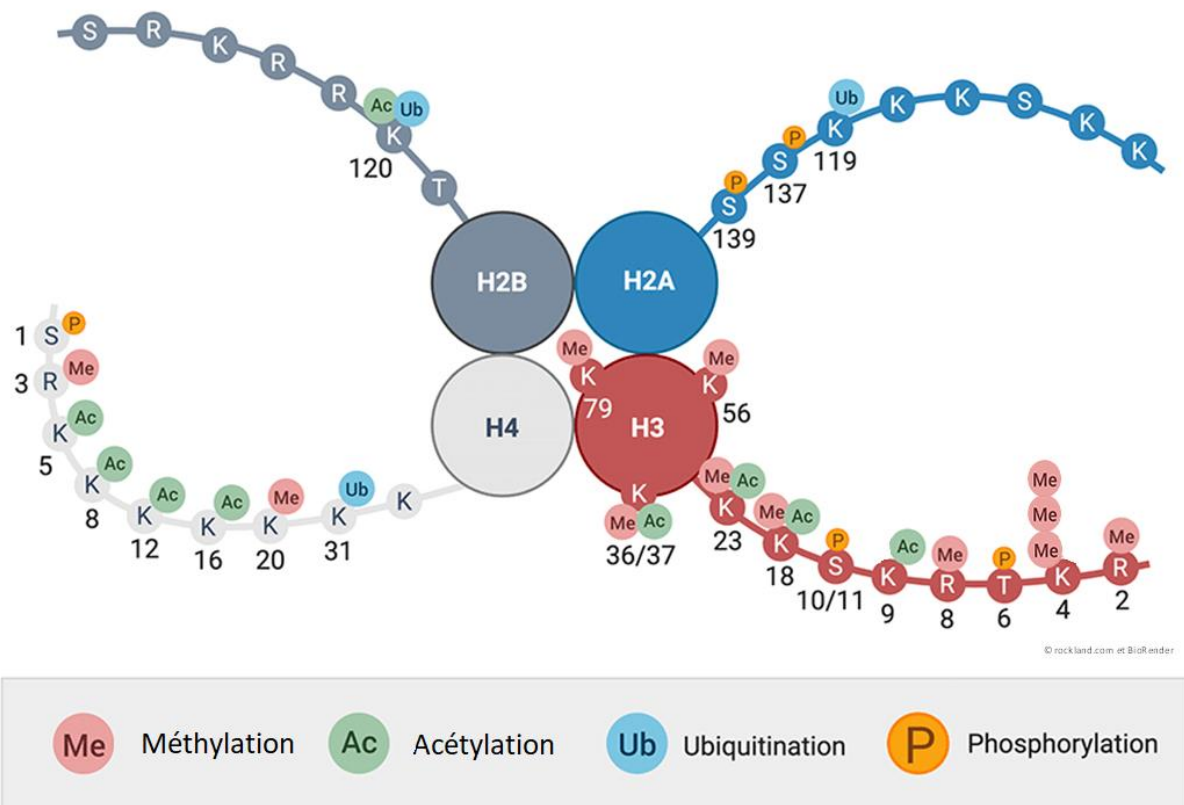


Figure 64 : Les quatre types de modification d'histone avec des exemples de marque.

Pour cette thèse, les marques H3K4me3 (triméthylation de l'histone H3 sur la lysine n°4) et H3K9ac (acétylation de l'histone H3 sur la lysine n°9) ont été testées. La marque H3K4me3 est présente chez *S. mansoni* (Roquis, Taudt et al., 2018). Chez d'autres espèces, cette marque est rythmique dans beaucoup de promoteurs de gènes de l'horloge (clock) ou de gènes contrôlés par les gènes de l'horloge (clock) (Koike, Yoo et al., 2012 ; Aguilar-Arnal, Hakim et al., 2013) et représente ainsi une piste intéressante. La marque H3K9ac est aussi présente chez *S. mansoni* et de manière plus intéressante dans le promoteur du gène *SmRho* (*smp\_104210*) au stade cercaire (Cosseau, Azzi et al., 2009). Ce gène code pour une rhodopsine (récepteur couplé à la protéine G), un photorécepteur impliqué notamment dans la vision. De façon plus générale, la position H3K9 est connue pour être un site préférentiel d'acétylation pour le gène *CLOCK* qui a une activité d'histone acétyl transférase (HAT) (Doi,

Hirayama & Sassone-Corsi, 2006). Chez les mammifères, les marques H3K4me3 et H3K9ac sont rythmiques dans les promoteurs de gènes clock (Aguilar-Arnal, Hakim et al., 2013 ; Takahashi, 2015).

L'objectif principal de ce sous chapitre va être de déterminer s'il existe un rythme d'acétylation ou de méthylation des histones dans les deux chronotypes de *S. mansoni* qui serait impliqué dans le rythme d'émission des cercaires.

Pour étudier les marques d'histone, la technique de ChIPmentation a été utilisée. Le premier point de ce chapitre présentera la technique et toutes les optimisations qui ont dû être réalisées pour l'adapter à *S. mansoni*. Les deuxième et troisième points seront consacrés respectivement aux résultats de la marque H3K4me3 et de la marque H3K9ac.

### II.2.2.1. Méthode et optimisations

Le ChIP-seq (Chromatin Immunoprecipitation-sequencing) pour immunoprécipitation de la chromatine est la technique qui permet d'étudier les protéines qui interagissent avec l'ADN et donc les modifications d'histone. Cette technique permet d'identifier les sites de liaisons des protéines associées à l'ADN, dans un premier temps, par la capture du complexe ADN-protéine via un anticorps spécifique (immunoprécipitation) puis, dans un second temps, par le séquençage de l'ADN issu du complexe. Dans le contexte d'étude d'un rythme d'acétylation et méthylation d'histone, un grand nombre d'échantillons doivent être générés pour établir une cinétique avec assez de points temporels pour chaque marque. Ce souci de quantité de matériel biologique est aussi étayé par la nature du modèle, ici *S. mansoni*, dont l'obtention des individus nécessite le sacrifice d'animaux. S'ajoute encore à cela le fait que la technique de ChIP-seq est très chronophage (Annexe 7). Pour pallier ces problèmes, le choix de la technique s'est porté sur le ChIPmentation, une version semi-automatisée du ChIP-seq qui nécessite de faibles quantités de chromatine (Schmidl, Rendeiro et al., 2015 ; Roels, Kuchmiy et al., 2020). Une partie du temps de cette thèse a été consacrée à l'adaptation du ChIPmentation sur le modèle *S. mansoni* et s'est concrétisé sous la forme d'un article de méthodologie, et en cours de review dans Wellcome Open Research (Lasica, de Carvalho Augusto et al., 2022). L'objectif de ce point est de présenter les différentes optimisations qu'a nécessité l'adaptation du ChIPmentation au modèle *S. mansoni*.

#### II.2.2.1.1. Méthode

La technique de ChIPmentation se divise en six étapes, l'échantillonnage biologique, la préparation de la chromatine, l'immunoprécipitation (ChIPmentation), la fabrication des bibliothèques, le séquençage et les analyses bioinformatiques des données. Dans le cadre de nos expériences, l'échantillonnage biologique consiste à récupérer des hépatopancreas de *B. pfeifferi* infestés par *S. mansoni*. La préparation de la chromatine consiste, dans un premier temps, en la lyse mécanique des échantillons, suivie de la fixation des protéines avec du formaldéhyde (cross-link). La chromatine est récupérée après une fragmentation enzymatique (Mnase) ou physique (sonication). L'étape d'immunoprécipitation consiste à isoler la chromatine contenant la marque histone d'intérêt en la ciblant avec un anticorps spécifique. Cette étape est réalisée par la machine IP-star de Diagenode® où se déroule aussi l'ajout des adaptateurs grâce à l'action de la transposase Tn5. Pour fabriquer les bibliothèques, il faut dans un premier temps éliminer les protéines (histones) pour ne garder que l'ADN par le décroisement du complexe. La détermination du nombre de cycles d'amplification pour chaque échantillon est faite à l'aide d'une qPCR. Une fois les paramètres définis, l'ADN est amplifié avec les

amorces qui serviront pour le séquençage. Les bibliothèques ainsi réalisées sont ensuite séquencées. Les analyses bioinformatiques consistent dans un premier temps à aligner les séquences contre les génomes, puis à effectuer une détection des pics (accumulation significative des séquences), enfin une analyse rythmique et fonctionnelle des gènes associés aux pics est ensuite réalisée.

### II.2.2.1.2. Optimisations

Le ChIPmentation qui a initialement été développé sur culture cellulaire humaine (Schmidl, Rendeiro et al., 2015 ; Roels, Kuchmiy et al., 2020) a nécessité un grand nombre d'optimisations pour s'adapter à notre modèle d'étude. Le détail de ces optimisations va faire l'objet de ce sous chapitre (Figure\_65).

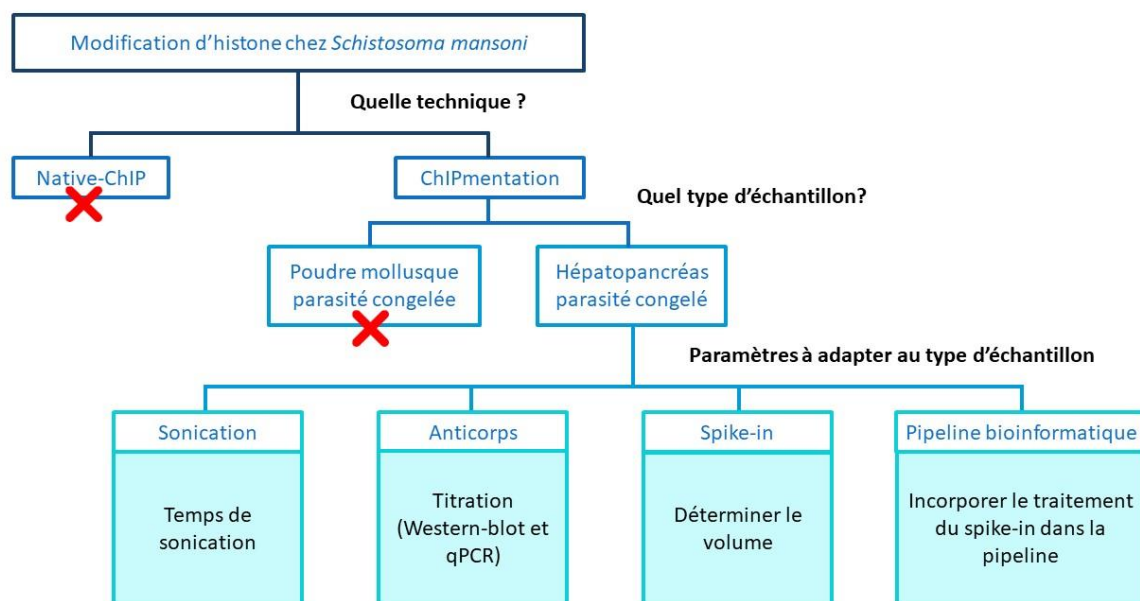


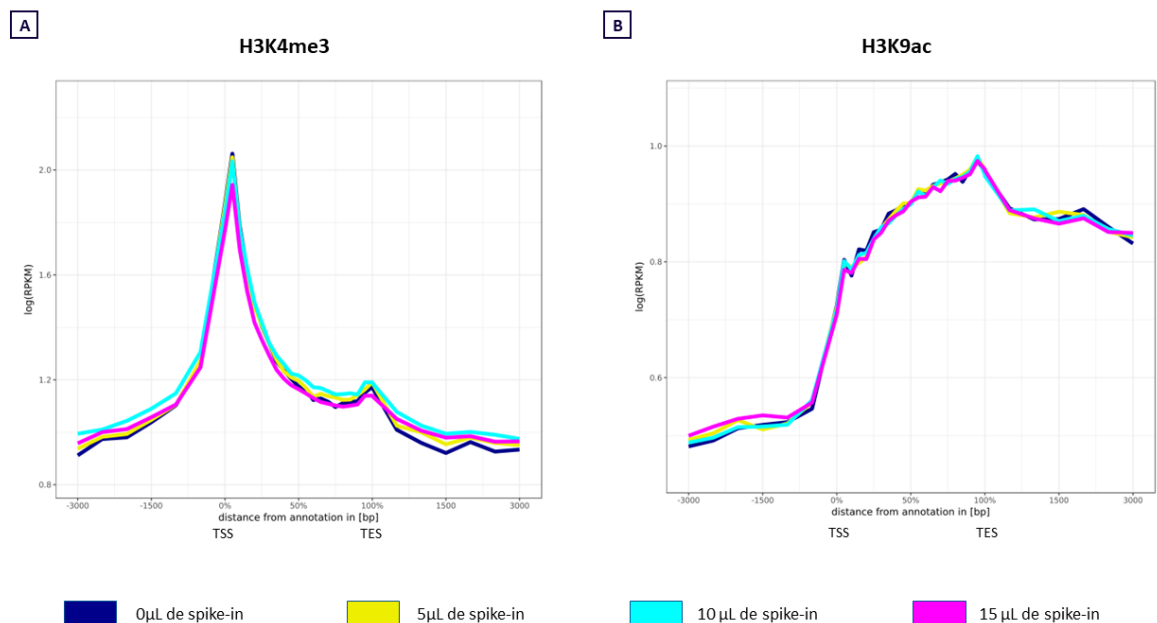
Figure 65 : Schéma récapitulatif des différentes optimisations pour l'étude des modifications d'histone chez *S. mansoni*.

L'optimisation du ChIPmentation a commencé par le choix du type d'échantillon. L'utilisation de tissu frais s'est révélé être la meilleure option car nos tests ont révélé un meilleur pourcentage d'alignement contre le génome de *S. mansoni*. Une fois choisi, il faut aussi déterminer le temps de sonication idéal et le volume d'anticorps à utiliser par échantillon. Toutefois, l'utilisation de tissu frais a fait remonter la problématique de variabilité biologique des échantillons. Chaque échantillon est un mollusque unique, ce qui implique la variation de la densité parasitaire et la taille du tissu. Si les échantillons ne sont pas normalisés, on ne pourra pas déterminer si la différence entre deux échantillons est due à une différence biologique ou technique.

Pour pallier ceci, il a été envisagé dans un premier temps de compter les noyaux d'un aliquot de l'échantillon avant l'étape de fixation (cross-link). Les diverses tentatives se sont révélées infructueuses car les noyaux étaient trop lysés suite à la fragmentation mécanique et donc impossibles à compter.

La solution s'est alors portée sur l'adaptation de la méthode de spike-in décrite par Bonhore, Bounova et al. (2014) où, pour quantifier des échantillons de souris, ils ajoutent un spike-in de culture cellulaire HeLa (humain). Un spike-in est l'ajout d'une quantité connue, dans notre cas de chromatine de drosophile, à tous les échantillons d'une expérience pour pouvoir les normaliser lors des analyses bioinformatiques. Dans le cas de Bonhore, Bounova et al. (2014), il y a immunoprécipitation des échantillons (chromatine souris + spike-in humain) avec deux anticorps, un pour l'expérience de CHIP et l'autre dirigé contre des régions connues pour être immunoprécipitées dans le spike-in (humain). Le taux d'immunoprécipitation du second anticorps avec le spike-in sert à calculer le facteur de normalisation qui sera appliqué aux échantillons. Dans notre cas expérimental, l'ajout d'un spike-in implique la coexistence de trois espèces différentes (mollusque, parasite et spike-in) au sein d'un même échantillon. Pour ne pas rajouter de la complexité, il a été décidé de ne garder qu'un seul anticorps par échantillon et d'utiliser de la chromatine d'embryon de drosophile diluée au  $10^6$ .

Différents volumes de chromatine de drosophile ont été testés. Le test des volumes de spike-in a consisté à faire des CHIPmentations en ajoutant 0, 5, 10 ou 15  $\mu$ L de spike-in après la sonication des échantillons pour les marques H3K4me3 et H3K9ac. Peu importe le volume utilisé, les profils méta gènes H3K4me3 et H3K9ac, ne sont pas affectés (Figure\_66).



**Figure 66 : Profils méta-gènes pour les marques H3K4me3 et H3K9ac sur le génome de *S. mansoni*, avec 0  $\mu$ L, 5  $\mu$ L, 10  $\mu$ L ou 15  $\mu$ L de spike-in drosophile. A) Marque H3K4me3, B) Marque H3K9ac.**

Pour déterminer le meilleur volume, on sélectionne visuellement tous les pics sur le génome de *Drosophila melanogaster* véritablement enrichis puis on calcule la profondeur de ces pics dans chaque condition. Les pics véritablement enrichis sont les pics qui s'alignent uniquement sur la version 6 du génome de *Drosophila melanogaster* (pic absent de la condition 0  $\mu$ L de spike-in) avec une profondeur supérieure à 10 reads. La détection des pics a été faite avec MACS2 sur galaxy version 2.1.1.20160309.5 (Zhang, Liu et al., 2008 ; Feng, Liu et al., 2012), la sélection visuelle avec IGV (Integrative genome viewer) version 2.8.2 et le comptage des reads par pic (profondeur) a été déterminé avec bedtool multicov sur galaxy version 2.29.0 (Quinlan & Hall, 2010).

Pour les deux marques, la meilleure condition est 15µL avec, respectivement, une profondeur moyenne de 69 reads pour 196 pics H3K4me3 drosophile et 42 reads pour 31 pics H3K9ac drosophile (Tableau\_40 et Tableau\_41). Même si le nombre de pics sélectionnés est inférieur aux 500 pics utilisés dans l'article de Bonhoure, Bounova et al. (2014), ces conditions restent satisfaisantes grâce à leur profondeur moyenne qui est bien supérieure aux 13 reads de l'article.

**Tableau 40 : Profondeur moyenne des pics uniques à *Drosophila melanogaster* en fonction du volume de spike-in pour la marque H3K4me3**

H3K4me3	0µL	5µL	10µL	15µL
Le nombre de pics uniques à <i>Drosophila</i> avec un véritable enrichissement sélectionné	0	196		
Profondeur moyenne des pics communs véritablement enrichis (nombre de reads par pics)	0	3,3 ± 1,3 0 à 237	15,9 ± 6,5 0 à 1251	69,1 ± 21,6 11 à 4090

**Tableau 41 : Profondeur moyenne des pics uniques à *Drosophila melanogaster* en fonction du volume de spike-in pour la marque H3K9ac**

H3K9ac	0µL	5µL	10µL	15µL
Le nombre de pics uniques à <i>Drosophila</i> avec un véritable enrichissement sélectionné	0	31		
Profondeur moyenne des pics communs véritablement enrichis (nombre de reads par pics)	0	9,6 ± 2,3 0 à 63	21,2 ± 4,4 0 à 120	42,3 ± 9,1 11 à 220

Une fois le volume de spike-in choisi, il a fallu inclure l'étape de normalisation dans le pipeline bioinformatique. Dans un premier temps, le calcul du facteur de spike-in et la normalisation étaient des étapes hors pipeline en suivant la formule issue de Bonhoure, Bounova et al. (2014). Suite à l'impossibilité de faire fonctionner Chromstar avec nos échantillons normalisés, il a été décidé de changer de pipeline bioinformatique et d'utiliser SnakePipe (Figure\_67) (Bhardwaj, Heyne et al., 2019). Ce pipeline offre l'avantage de prendre en compte les inputs et le spike-in. La normalisation est faite en appliquant une méthode similaire à celle implémentée dans DEseq2.

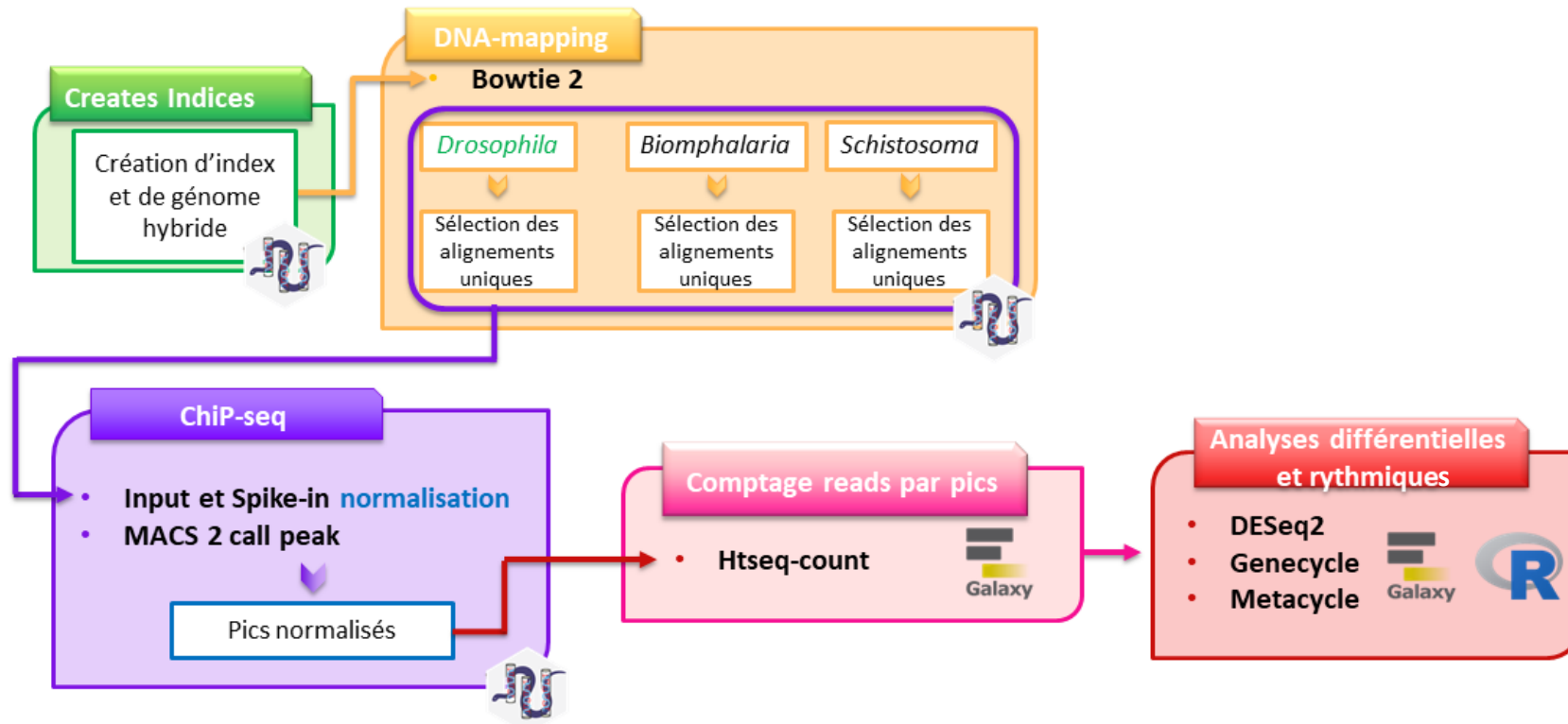


Figure 67 : Pipeline bioinformatique utilisé pour l'analyse finale des échantillons de ChIPmentation avec Snakepipe.

## II.2.2.1.3. Article ChIPmentation

Le protocole expérimental utilisé pour les expériences de ChIPmentation est détaillé dans le matériel et méthode de l'article (Annexe 7).

## II.2.2.2. H3K4me3

La moyenne et les variations individuelles du nombre total et du pourcentage d'alignement des reads par échantillon, selon chaque chronotype, sont présentées dans le Tableau\_42. Pour les analyses, seuls les échantillons avec un minimum de 1 500 000 reads alignés uniquement sur le génome de *S. mansoni* ont été gardés. Si initialement l'expérience devait compter au minimum 2 échantillons mâles et 2 échantillons femelles par point temporel, il a été décidé de réduire le nombre à 3 échantillons (sexes confondus) par point temporel. Le temps de la thèse, amputé par la pandémie, ne permettait pas de prendre le temps de refaire les ChIPmentation pour tous les échantillons en dessous de la limite des reads. Les expériences de chronobiologie qui montrent que le sexe du parasite n'a pas d'importance dans le rythme d'émission des cercaires justifient scientifiquement ce choix. Malheureusement, pour le chronotype diurne, deux échantillons sont en dessous de cette limite des 1 500 000 reads malgré la réalisation de ChIPmentation sur de nouveaux individus. Ces deux échantillons représentent tous les deux le point temporel 16h pour le 1<sup>er</sup> jour d'échantillonnage.

**Tableau 42 : Nombre et pourcentage d'alignement des reads moyens, accompagnés des variations individuelles, pour les parasites diurne et nocturne de *S. mansoni* pour les échantillons de l'expérience épigénétique H3K4me3**

H3K4me	Moyenne ± ES Variations individuelles		Test de Wilcoxon- Mann-Whitney
	Diurne (N=31)	Nocturne (N=33)	
Nombre total de reads	8 303 741 ± 411 309 4 610 796 à 12 106 301	6 683 639 ± 238 292 4 500 323 à 10 657 012	W = 2072, p = 0,0
Alignement sur <i>S. mansoni</i> (%)	46,90 ± 2,16 27,08 à 69,41	52,92 ± 2,12 29,69 à 75,73	W = 1059, p = 0,001
Alignement unique sur <i>S. mansoni</i> (%)	30,57 ± 1,67 13,24 à 44,53	37,72 ± 1,90 19,82 à 66,03	W = 1012, p = 0,0004
Alignement unique sur <i>S. mansoni</i> (nombre de reads)	2 448 792 ± 166 405 1 535 469 à 5 391 156	2 423 928 ± 110 007 1 612 973 à 3 998 228	W = 1154, p = 0,007
Alignement sur <i>B. glabrata</i> (%)	20,93 ± 1,77 6,79 à 41,84	13,31 ± 1,19 4,09 à 30,14	W = 2243, p = 0,0007
Alignement unique sur <i>B. glabrata</i> (%)	9,85 ± 0,87 2,76 à 19,38	6,02 ± 0,62 0,77 à 14,70	W = 2183, p = 0,002
Alignement unique sur <i>B.</i> <i>glabrata</i> (nombre de reads)	845 169 ± 98 113 194 746 à 2 246 369	396 528 ± 42 797 47 727 à 1 135 777	W = 2265, p = 0,0004
Alignement sur <i>D. melanogaster</i> , <i>spike-in</i> (%)	18,17 ± 2,00 4,57 à 36,59	23,83 ± 2,09 7,42 à 56,38	W = 1651, p = 0,94
Alignement unique sur <i>D. melanogaster</i> ,	11,58 ± 1,32 2,82 à 24,25	15,14 ± 1,35 4,47 à 35,95	W = 1654, p = 0,93

<i>spike-in</i> (%)			
Alignement unique sur <i>D. melanogaster</i> , <i>spike-in</i> (nombre de reads)	943 504 ± 116 076 198 766 à 2 182 448	1 060 953 ± 121 976 279 340 à 3 401 341	W = 1708, $p = 0,7$

*S. mansoni* est en bleu, *B. glabrata* en orange et *D. melanogaster* (*spike-in*) en vert.

Pour les chronotypes, les échantillons s'alignent majoritairement sur le génome de *S. mansoni*. Toutefois chez le chronotype nocturne, significativement plus de reads s'alignent sur le génome du *spike-in* que sur celui de *B. glabrata* ( $p < 0,01$ ). Cette situation n'est pas dérangeante puisque la part de l'hôte dans l'échantillon est trop faible pour être exploitée dans des analyses bioinformatiques. Cette faible part de l'hôte peut s'expliquer par le fait que l'hépatopancreas est rempli de sporocystes et que le ratio cellule hôte/parasite penche en faveur de *S. mansoni*. Il peut aussi être ajouté à cela que l'alignement se fait sur le génome d'une autre espèce que celle présente dans l'échantillon car le génome de *B. pfeifferi* n'existe pas. Côté *spike-in*, il n'y a aucune différence significative entre les deux chronotypes ( $p > 0,01$ ), ce qui permet de faire des comparaisons biologiques. Dans les échantillons nocturnes, l'immunoprécipitation du parasite est significativement plus importante. Ceci peut s'expliquer soit par le fait que le chronotype nocturne a une meilleure affinité avec l'anticorps, soit que les échantillons du chronotype nocturne contiennent plus de parasite que ceux du chronotype diurne. Si on s'intéresse aux inputs (chromatine non immunoprécipitée des échantillons), il n'y a pas de différence significative en termes de pourcentage d'alignement du parasite et du *spike-in* entre les deux chronotypes ( $p > 0,01$ ). Le pourcentage d'alignement sur le génome de *B. glabrata* n'est pas significativement différent avec un seuil fixé à 0,01 mais la  $p$ -value reste faible ( $w=1516$ ,  $p = 0,02$ ). Si on la considère comme significative, on pourrait alors penser que les mollusques utilisés pour le chronotype diurne étaient plus grands que pour le chronotype nocturne. Toutefois, comme le pourcentage de parasite n'est pas significativement différent entre les deux chronotypes, il serait plus plausible de penser que c'est tout simplement l'anticorps qui a une meilleure affinité avec le chronotype nocturne.

#### II.2.2.2.1. Rythme des pics

Pour détecter le rythme de méthylation des histones, le logiciel GeneCycle (Ahdesmäki, Fokianos & Strimmer, 2012) et le logiciel meta2d (Wu, Anafi et al., 2016) ont été utilisés. Les logiciels reposent sur l'utilisation du test exact de Fisher. Pour pallier au risque de faux positifs dus au grand nombre de tests à réaliser, chaque logiciel applique une correction de la  $p$ -value basée sur la méthode de Benjamini-Hochberg, appelée BH.Q ou FDR (False Discovery rate) avec un seuil fixé à 5% de risque (BH.Q/FDR < 0,05). Un pic est considéré comme rythmique s'il est détecté dans les deux logiciels avec au moins le seuil  $p$ -value < 0,05. Cette précaution est utilisée pour éviter une variabilité qui serait propre au logiciel (Takahashi, Kumar et al., 2015 ; Rijo-Ferreira, Acosta-Rodriguez et al., 2020).

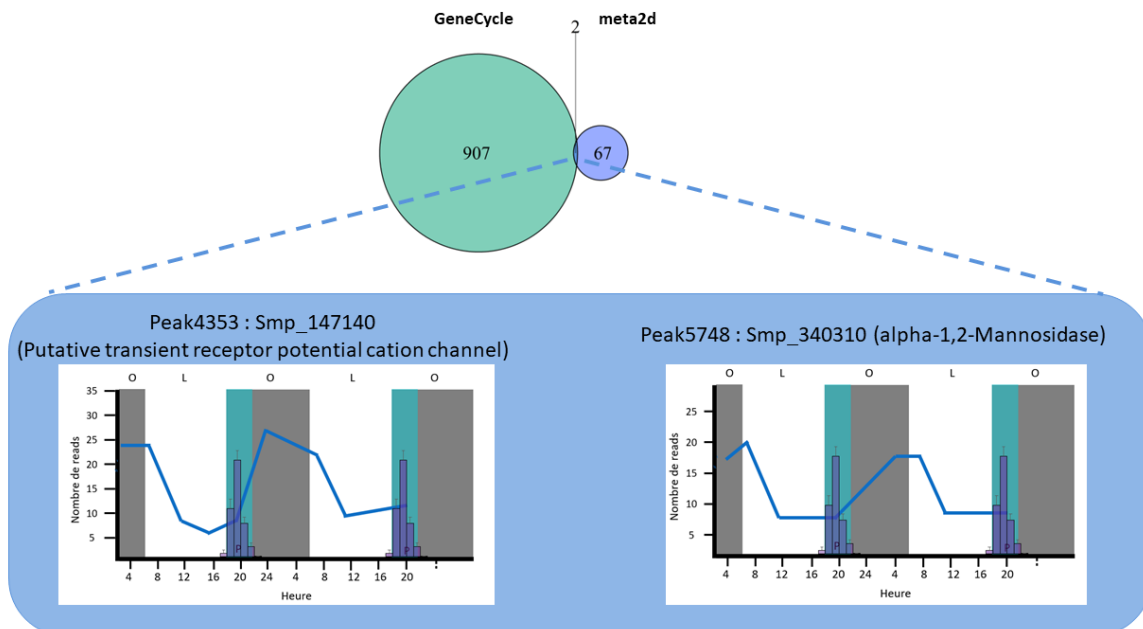
Les deux chronotypes confondus totalisent 6360 pics H3K4me3 détectés. Pour les analyses rythmiques, seuls les pics avec une profondeur moyenne supérieure à 10 reads sont utilisés. Le nombre de pics rythmiques par chronotype et logiciels sont présentés dans le Tableau\_43.



**Tableau 43 : Nombre de pics H3K4me3 détectés comme rythmiques et nombre de gènes associé en fonction du logiciel et du chronotype du parasite ( $p < 0,05$ ). Aucun gène n'est rythmique avec le FDR < 0,05.**

	H3K4me3	Diurne (6163 pics filtrés)	Nocturne (6090 pics filtrés)	Commun aux deux chronotypes
GeneCycle ( $p$ -value)	Pics	67	909	31
	Gènes associés	21	815	10
meta2d ( $p$ -value)	Pics	0	69	
	Gènes associés	0	63	

Aucun pic n'est détecté comme rythmique avec la FDR/BH.Q, peu importe le logiciel et le chronotype. Pour le chronotype diurne aucun pic H3K4me3 n'est considéré comme rythmique car aucun pic n'a été détecté avec le logiciel meta2d (Tableau\_43). Pour le chronotype nocturne, 909 pics ont été détectés comme rythmiques avec GeneCycle et 69 avec le logiciel meta2d ( $p < 0,05$ ). Sur tous ces pics, seuls deux sont communs aux deux logiciels (Figure\_68). Le premier pic est le pic n°4353 qui est présent sur le gène Smp\_147140. Ce gène code pour un canal TRP (Transient Receptor Potential) cationique. Chez la drosophile les canaux TRP sont impliqués dans les étapes intermédiaires de la phototransduction (Montell, Jones et al., 1985). Le pic des marques H3K4me3 est nocturne et à minuit, le point temporel juste après celui du pic d'émission des cercaire (20h). Le second gène est smp\_340310, une alpha-1,2-mannosidase impliqué dans la synthèse des glycoprotéines. Le pic des marques H3K4me3 est entre 4h et 8h matin.



**Figure 68 : Profils des deux marques H3K4me3 rythmiques chez le chronotype nocturne.** Chaque histogramme a, en abscisse, les heures (points temporels) et en ordonnées le nombre de reads pour le pic sélectionné. En fond, se trouve le profil d'émission des cercaires. Les rectangles gris symbolisent l'obscurité (O), les blancs la lumière (L) et les rectangles bleus la période du pic d'émission des cercaires.

#### II.2.2.2.2. Gènes associés aux marques H3K4me3

Si les analyses rythmiques n'ont pas été très probantes sur un rythme de méthylation des histones, cela ne signifie pas pour autant qu'il n'y a pas de lien avec le rythme d'émission des cercaires. Il a été montré qu'il y a une variation du taux de méthylation des histones en fonction du stade de vie de *S. mansoni* (Roquis, Taudt et al., 2018). Dans notre contexte d'étude, nous pouvons penser qu'il peut y avoir un set de marque d'histone (ici H3K4me3) qui pourrait être spécifique aux stades sporocystes/cercaires intrasporocystes et qui soit impliqué dans le rythme d'émission des cercaires. Voilà pourquoi il est intéressant de regarder de plus près les gènes associés aux marques d'histone H3K4me3. Sur les 6360 pics H3K4me3 détectés, 5537 coïncident avec des gènes.

Tout comme pour les gènes candidats QTL, une sélection subjective basée sur la recherche de gènes de l'horloge, de la phototransduction, de l'acétylation et la méthylation des histones a été faite. L'enrichissement GO a été fait avec Blast2GO version 5.2 en utilisant le Fisher's Exact Test ( $p < 0,05$ ). L'analyse d'enrichissement Kegg pathway a été faite sur le serveur KAAS en utilisant le programme Blast avec la méthode BBH (Bi-directional Best Hit).

#### II.2.2.2.3. Les gènes de l'horloge (Basic Helix-Loop-Helix)

Il n'y a aucun gène canonique de l'horloge dans les gènes marqués H3K4me3. Toutefois, nous retrouvons trois gènes du QTL interspécifique *mansoni X rodhaini* (Tableau\_33 partie génétique), Smp\_021340, Smp\_123570 et Smp\_333620 qui contiennent des motifs BHLH.

#### II.2.2.2.4. Les rhodopsines et autres gènes impliqués dans la phototransduction

On retrouve la rhodopsine (Smp\_117340) du QTL interspécifique. On retrouve aussi 6 rhodopsine-like ; parmi elles, deux sont issues du QTL *mansoni X mansoni* (Smp\_132730 et Smp\_127310) et deux autres du QTL interspécifique (Smp\_083880 et Smp\_326640). Les deux autres sont des putatives rhodopsine-like orphan (Tableau\_44), c'est-à-dire des rhodopsine-like dont le ligand n'a pas encore été identifié.

Tableau\_44 : Liste des gènes associés aux pics H3K4me3, rhodopsine et rhodopsine-like

Intérêt	Chr.	Identifiant (ID) du gène	Annotation GFF
Rhodopsine-like	1	Smp_170610	Putative rhodopsin-like orphan GPCR
		Smp_056080	Putative rhodopsin-like orphan GPCR

Parmi les différents acteurs de la phototransduction, nous retrouvons, dans les gènes marqués H3K4me3 : la rhodopsine, la protéine G, la phosphodiesterase, la GPCR kinase (sérine/thréonine kinase), l'arrestine, la recoverine, la protéine phosphatase 2A (PP2A), la GTPase activating protine et la guanylyl cyclase (Tableau\_45).

Tableau\_45 : Liste des gènes associés aux pics H3K4me3 potentiellement liés à la phototransduction

Intérêt	Chr.	Identifiant (ID) du gène	Annotation GFF	
Protéine G (Transducine)	1	Smp_098760	Transducin beta-like	
	4	Smp_150080	Transducin-like enhancer protein 1	
	QTL intraspé.	Smp_010230 et Smp_016630 (Tableau_27)		
	QTL interspé.	Smp_075130 et Smp_018400 (Tableau_35)		
Phosphodiesterase	2	Smp_146120	Phosphodiesterase	
	5	Smp_153340	Putative ectonucleotide pyrophosphatase/phosphodiesterase	
	6	Smp_141980	Phosphodiesterase	
	ZW	Smp_179590		
		Smp_151190	Retinal rod rhodopsin-sensitive cgmp 3',5'-cyclic phosphodiesterase, putative	
	QTL intraspé.	Smp_105040 (Tableau_27)		
QTL interspé.	Smp_018360 (Tableau_35)			
GPCR kinase, sérine/thréonine kinase (rhodospine kinase)	1	Smp_176620	Serine/threonine kinase	
		Smp_098840		
		Smp_068990		
		Smp_124130		
		Smp_097940		
		Smp_141230		
		Smp_073340		
		Smp_210820		Protein-serine/threonine kinase
		Smp_160760		Serine/threonine kinase
		Smp_090890		Serine/threonine protein kinase
		Smp_090980		Serine/threonine kinase
	Smp_158560			
	Smp_103950			
	Smp_040190			
	3	Smp_018290	Serine/threonine kinase	
		Smp_132260		
		Smp_181490		
		Smp_000720		
		Smp_134910		
		Smp_177120		
	4	Smp_049760	Receptor protein serine/threonine kinase	
		Smp_150860	Serine/threonine kinase	
		Smp_053560		
		Smp_126460		
		Smp_041770		
		5	Smp_334370	Receptor protein serine/threonine kinase
	Smp_171370		Serine/threonine kinase	
6	Smp_347070			
6	Smp_123290	Serine/threonine kinase		
	Smp_080730			

		Smp_169980		
		Smp_169950		
	7			Smp_155720
				Smp_180400
				Smp_127920
				Smp_057360
				Smp_125370
	ZW			Smp_152330
				Smp_097820
				Smp_156990
				Smp_068960
				Smp_210970
				Smp_094760
				Smp_047190
				Smp_163380
				Smp_058620
				Smp_094190
				Smp_176580
				Smp_096640
		Smp_194610		
	Smp_151140			
	Smp_151100			
	Smp_077180			
	QTL intraspé.	Smp_068060, Smp_009800 et Smp_005190 (Tableau_27)		
	QTL interspé.	Smp_123640, Smp_144390 et Smp_042920 (Tableau_35)		
Arrestine	1	Smp_152290	Putative beta-arrestin 1	
	2	Smp_122900		
		QTL intraspé.	Smp_126080 (Tableau_27)	
domaine EF-hand (Recoverine)	1	Smp_171010	Putative ef-hand domain (C-terminal) containing protein	
		Smp_159160		
		QTL intraspé.	Smp_005230, Smp_136640, Smp_083740 et Smp_137410 (Tableau_27)	
	QTL interspé.	Smp_181460, Smp_106060, Smp_134950, Smp_018940 (Tableau_35)		
Protéine phosphatase 2A (PP2A)	1	Smp_176600	Serine/threonine protein phosphatase 2A regulatory subunit	
		Smp_035430		
		Smp_166290		Serine/threonine-protein phosphatase 2A activator
	3	Smp_173810	Serine/threonine-protein phosphatase 2A 55 kDa regulatory subunit B	
	4	Smp_055770		
	5	Smp_153410	Putative serine/threonine protein phosphatase 2a regulatory subunit A	
	ZW	Smp_172280	Serine/threonine protein phosphatase 2a regulatory subunit A, putative	
		QTL intraspé.	Smp_165490 (Tableau_27)	
		QTL interspé.	Smp_308680 (Tableau_35)	
GTPase activating protein	1	Smp_069400	Putative regulator of G protein signaling	

		Smp_072660	Ras gtpase activating protein, putative
	3	Smp_079800	Putative gtpase activating protein
	3	Smp_127520	Putative regulator of G protein signaling
	4	Smp_155780	Gtpase activating protein-related
	QTL intraspé.	Smp_007210, Smp_180160, <b>Smp_028030</b> , <b>Smp_094890</b> , <b>Smp_319130</b> et Smp_165530 (Tableau_27)	
	QTL interspé.	Smp_138430, Smp_138500, Smp_000670 (Tableau_35)	
Guanylyl cyclase	1	Smp_142620	Guanylate cyclase
	ZW	Smp_019790	
		Smp_153500	
GTPAse	1	Smp_132620	Transmembrane gtpase mfn (Mitofusin 1) (Fzo homolog), putative
	QTL intraspé.	<b>Smp_025740</b> (Tableau_27)	
	QTL interspé.	Smp_084940 (Tableau_35)	

Chr. : chromosome ; intraspé.: intraspécifique; interspé. : interspécifique.

#### II.2.2.2.5. Les gènes impliqués dans l'acétylation et la méthylation des histones

Parmi les gènes associés à la marque H3K4me3, nous retrouvons aussi des gènes qui codent pour les histones H2A, H2B, H3 et H4, ainsi que les writer et eraser pour l'acétylation et la méthylation des histones (Tableau\_46).

**Tableau 46: Liste des gènes associés avec la marque H3K4me3 en lien avec les histones, l'acétylation/désacétylation des histones et la méthylation/déméthylation des histones**

Intérêt	Chr.	Identifiant (ID) du gène	Annotation GFF
Histone	1	Smp_053390	Histone H4
		Smp_086860	Histone H2A
	7	Smp_031720	
	ZW	Smp_082240	Histone H3
	QTL intraspé.	Smp_036220, Smp_210630, Smp_246990 et Smp_124840 (Tableau_27)	
QTL interspé.	Smp_000150 et Smp_027850 (Tableau_35)		
Histone acétylase	1	Smp_178700	Putative histone acetyltransferase type B catalytic subunit
	2	Smp_194520	Histone acetyltransferase
	4	Smp_053140	
	ZW	Smp_198670	
	QTL intraspé.	Smp_131320 (Tableau_27)	
QTL interspé.	Smp_105910 et Smp_083930 (Tableau_35)		
Histone désacétylase	1	Smp_069380	Putative histone deacetylase 4, 5
		Smp_091990	Putative histone deacetylase 1, 2,3
	2	Smp_138770	Histone deacetylase, putative
	5	Smp_093280	Histone deacetylase
QTL intraspé.	Smp_005210 (Tableau_27)		
Histone	1	Smp_140390	Histone-lysine N-methyltransferase

méthyltransférase	3	Smp_138030	
	4	Smp_070170	
		Smp_027300	
	5	Smp_043580	Putative histone-lysine n-methyltransferase, setb1
	6	Smp_165000	Histone-lysine N-methyltransferase, H3 lysine-79 specific
	ZW	Smp_210650	Histone-lysine N-methyltransferase
QTL interspé.	Smp_150850 (Tableau_35)		
Histone déméthylase	5	Smp_150560	Putative lysine-specific histone demethylase 1
	QTL interspé.	Smp_000230 (Tableau_35)	

Chr.: chromosome; intraspé.: intraspécifique; interspé. : interspécifique.

#### II.2.2.2.6. Enrichissement GO : Processus biologique

Parmi les gènes QTL mis en évidence avec l'analyse d'enrichissement GO, 4 sont associés avec la marque H3K4me3. Le gène Smp\_345020 (annotation GO sommeil) pour le QTL intraspécifique et les gènes Smp\_096010 (annotation GO réponse à la chaleur), Smp\_310920 et Smp\_181450 (annotation GO contraction musculaire).

Pour les 5537 gènes associés à la marque H3K4me3, les analyses d'enrichissement GO n'ont pas donné de résultats significatifs en utilisant la limite usuelle FDR < 0,05 (False Discovery Rate). Les résultats qui suivent ont été obtenus en utilisant la limite de  $p < 0,05$ .

Parmi toutes les annotations GO, les processus métabolique et biosynthétique du pigment intriguent. Le pigment impliqué n'est malheureusement pas photorécepteur puisqu'il s'agit de l'hème, précurseur de l'hémoglobine (Figure\_69). Ces deux annotations regroupent 10 gènes qui sont impliqués dans la synthèse de l'hème (Tableau\_47). Si l'hème est connu pour être impliqué dans divers processus biologiques, comme le transport d'oxygène, il a été montré qu'il pourrait aussi intervenir dans un nouveau mécanisme de régulation du rythme circadien (Freeman, Kwon et al., 2019). La fixation de l'hème sur la protéine clock perturberait la liaison entre la protéine clock et les E-box (enhancer box) qui permettent l'expression des éléments qui exercent un contrôle négatif de la boucle de rétrocontrôle du rythme circadien (ex : PER, CRY).

Partie II : Rythme-Epigénétique

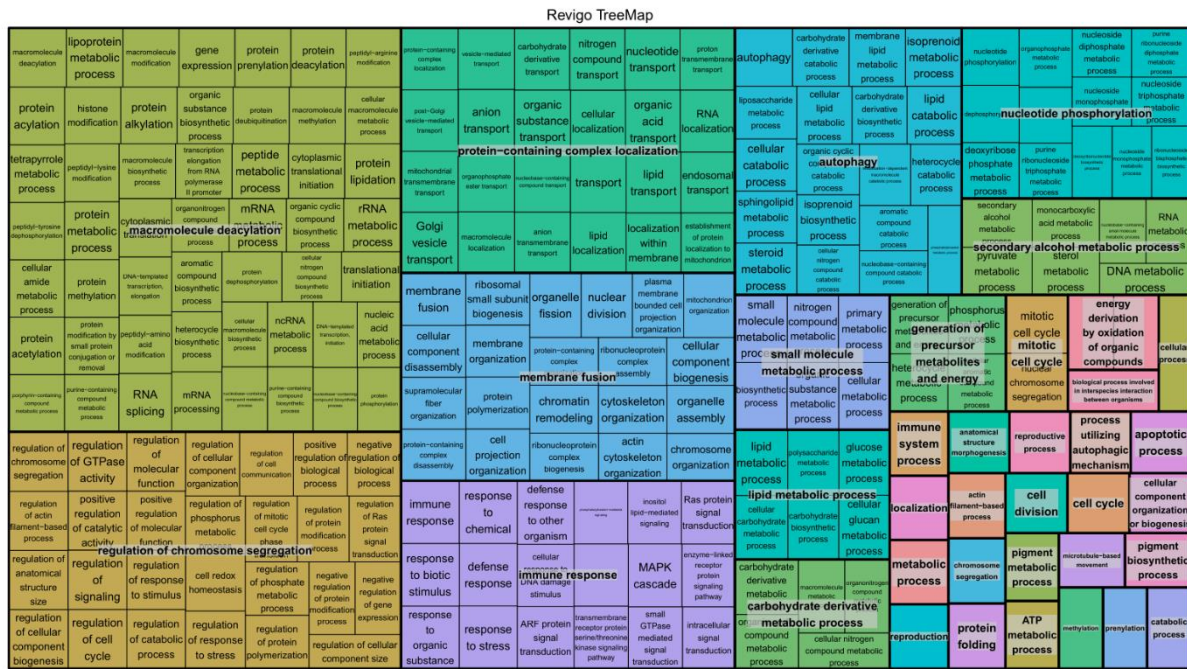


Figure 69 : Treemap revigo des résultats de l'analyse d'enrichissement GO des processus biologiques obtenus pour les gènes associés à la marque H3K4me3.

Chaque couleur représente un grand ensemble de processus biologiques. Il y a, de gauche à droite : désacétylation des macromolécules, localisation des complexes contenant des protéines, autophagie, phosphorylation des nucléotides, processus métabolique secondaire de l'alcool, fusion membranaire, processus métabolique des petites molécules, génération des métabolites précurseurs et énergie, cycle cellulaire mitotique, dérivation de l'énergie par l'oxydation des composés organiques, processus biologique impliqué dans les interactions inter-espèces entre les organismes, processus cellulaire, régulation de la ségrégation des chromosomes, réponse immunitaire, processus métabolique des lipides, processus du système immunitaire, morphogénèse des structures anatomiques, processus de reproduction, processus utilisant les mécanismes autophagiques, processus d'apoptose, localisation, processus basé sur les filament d'actines, division cellulaire, cycle cellulaire, organisation des composés cellulaires ou biogénèse, processus métaboliques des dérivés glucidiques, processus métabolique, ségrégation des chromosomes, processus métabolique des pigments, mouvement basés sur les microtubules, processus de biosynthétique des pigments, reproduction, repliement des protéines, processus métabolique de l'ATP, méthylation, prénylation et processus catabolique. La treemap a été obtenue avec revigo (Supek, Bošnjak et al., 2011).

Tableau 47 : Liste des gènes associés avec la marque H3K4me3 en lien avec l'annotation GO processus métabolique des pigments et processus de synthèse des pigments

GO	Chr.	Identifiant (ID) du gène	Annotation GFF	
Processus métabolique des pigments et processus de synthèse des pigments	1	Smp_140790	Ferrocélatase	
	2	Smp_045260	5-aminolevulinate synthase	
	3	Smp_079840	Uroporphyrinogen-III synthase	
	4	Smp_137190	Frataxin homolog, mitochondrial, putative	
	5	Smp_340920	previous_stable_id=Smp_057930	
	6	Smp_162280	Coproporphyrinogen oxidase	
	ZW		Smp_143740	Uroporphyrinogen decarboxylase
			Smp_333090	NA
			Smp_045810	Delta-aminolevulinic acid déshydratase
			Smp_095130	Putative porphobilinogen déaminase

Chr.: chromosome.

## II.2.2.2.7. Kegg pathway

Les analyses Kegg pathway montrent que les gènes associés à la marque H3K4me3 sont potentiellement impliqués dans 364 voies moléculaires différentes. Parmi toutes ces voies, neuf retiennent notre attention, le rythme circadien, le rythme circadien chez la drosophile, le rythme circadien chez la plante, l'entraînement circadien, la phototransduction, la phototransduction chez la drosophile, la transduction olfactive et la transduction gustative (Tableau\_48).

**Tableau\_48 : Liste des gènes associés avec la marque H3K4me3 en lien avec les Kegg pathway sélectionnés.** Les gènes et KO annotation qui sont marqué d'une étoile apparaissent dans plusieurs Kegg pathways.

Kegg pathway	KO annotation	Chr.	Identifiant (ID) du gène	Annotation GFF
Rythme circadien	K07200	1	Smp_341700	AMPKBI domain-containing protein
		3	Smp_052910	Skp1-related
			Smp_099930	Putative f-box and wd40 domain protein
			Smp_001030	5'-amp-activated protein kinase gamma-2 non-catalytic subunit transcript variant 2
		4	Smp_334010	NA
		ZW	Smp_170160	5-AMP-activated protein kinase, beta subunit, putative
			Smp_214100	Non-specific serine/threonine protein kinase
	QTL intraspé.	Smp_315790 (Tableau_27)		
	K07199	1	Smp_341700	AMPKBI domain-containing protein
		ZW	Smp_170160	5-AMP-activated protein kinase, beta subunit, putative
	K03094	3	Smp_052910	Skp1-related
	K03362		Smp_099930	Putative f-box and wd40 domain protein
K07198	ZW	Smp_214100	Non-specific serine/threonine protein kinase	
Rythme circadien chez la drosophile	K23643	QTL interspé.	Smp_023700 (Tableau_35)	
Rythme circadien chez la plante	K03097	ZW	Smp_099030	Protein kinase
	K03115	6	Smp_025010	Casein kinase II subunit beta
Entraînement du rythme circadien	K05208	ZW	Smp_245320	ANF_receptor domain-containing protein
	K02183*		Smp_026560*	Calmodulin 2
		QTL interspé.	Smp_018940* (Tableau_35)	
	K04515*	ZW	Smp_011660*	Protein kinase
	K04630	1	Smp_020080	Putative gtp-binding protein (I) alpha-1 subunit, gnai1
2		Smp_016250	Putative gtp-binding protein (I) alpha-2 subunit, gnai2	



	K04634*	7	Smp_005790*	Putative gtp-binding protein (Q) alpha-11 subunit, gna11
	K04536*	2	Smp_063640*	Putative g-protein, beta subunit
		QTL intraspé.	Smp_010230 *(Tableau_27)	
	K04539	QTL interspé.	Smp_075130 (Tableau_35)	
	K04371	4	Smp_142050	Mitogen-activated protein kinase
	K04958*		Smp_336900*	Inositol 1,4,5-trisphosphate receptor
	K04962	1	Smp_163570	Ryanodine receptor related
	K08049	4	Smp_102340	Adenylate cyclase type IX, putative
	K04345*	ZW	Smp_152330*	Serine/threonine kinase
			Smp_267270*	NA
	K04851	1	Smp_020270	Voltage-dependent L-type calcium channel subunit alpha
		3	Smp_159990	
	K04632	4	Smp_172640	Putative gtp-binding protein alpha subunit
	K05858*	2	Smp_245900*	Phosphoinositide phospholipase C
	K02677*	ZW	Smp_241420*	AGC-kinase C-terminal domain-containing protein
Phototransduction	K04536*	2	Smp_063640*	Putative g-protein, beta subunit
		QTL intraspé.	Smp_010230 *(Tableau_27)	
	K12322	2	Smp_125450	Guanylate cyclase
	K02183*	ZW	Smp_026560	Calmodulin 2
QTL interspé.		Smp_018940* (Tableau_35)		
Phototransduction chez la drosophile	K04634*	7	Smp_005790*	Putative gtp-binding protein (Q) alpha-11 subunit, gna11
	K05858*	2	Smp_245900*	Phosphoinositide phospholipase C
	K02677*	ZW	Smp_241420*	AGC-kinase C-terminal domain-containing protein
			Smp_026560	Calmodulin 2
	K02183*	QTL interspé.	Smp_018940* (Tableau_35)	
	K04958*	4	Smp_336900*	Inositol 1,4,5-trisphosphate receptor
	K00910*	QTL intraspé.	Smp_210820* *(Tableau_27)	
	K04515*	ZW	Smp_011660*	Protein kinase
K13806	Smp_128680		Lipase	
K13807	2	Smp_140770	Serine/threonine-protein phosphatase with EF-hands	
Transduction olfactive	K04439	1	Smp_152290	Putative beta-arrestin 1
		2	Smp_122900	
		QTL intraspé.	Smp_210820 (Tableau_27)	
	K05849		Smp_152830 (Tableau_27)	
K04536*	2	Smp_063640*	Putative g-protein, beta	

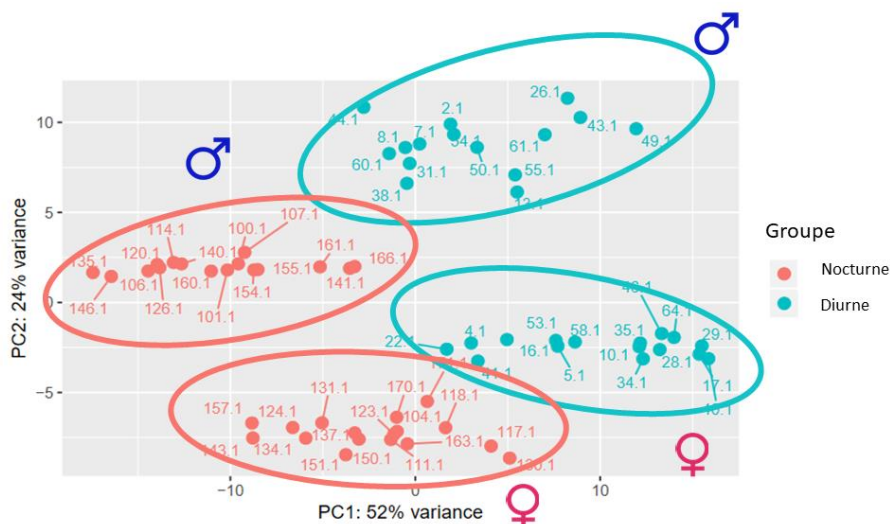
				subunit
		QTL intraspé.	Smp_010230* (Tableau_27)	
	K00910*	QTL intraspé.	Smp_210820* *(Tableau_27)	K00910*
	K04345*	ZW	Smp_152330*	Serine/threonine kinase
			Smp_267270*	NA
	K02183*	QTL interspé.	Smp_026560*	
	K04515*	ZW	Smp_018940* (Tableau_35)	
			Smp_011660*	Protein kinase
Transduction gustative	K04153	4	Smp_149770	Putative g-protein coupled receptor
	K04345*	ZW	Smp_152330*	Serine/threonine kinase
			Smp_267270*	NA
K05858*	2	Smp_245900*	Phosphoinositide phospholipase C	

Chr. : chromosome ; intraspé.: intraspécifique; interspé. : interspécifique.

### II.2.2.2.8. Analyses DEseq2

#### II.2.2.2.8.1. Comparaison des chronotypes

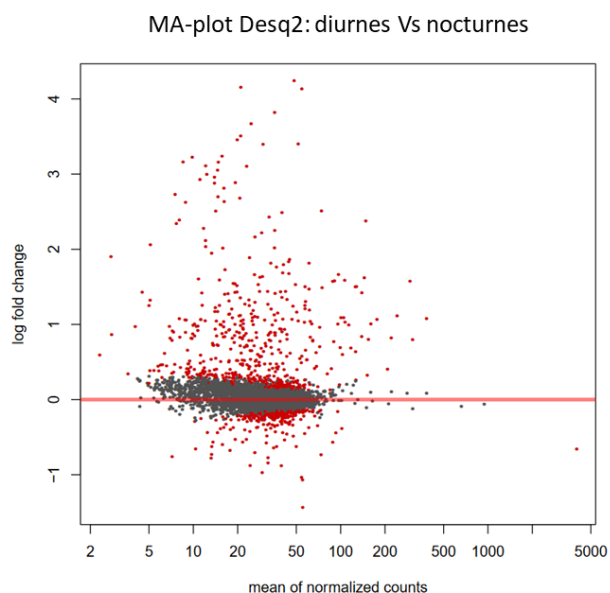
Pour déterminer s'il y a une différence de méthylation des histones entre les deux chronotypes de *S. mansoni*, des analyses différentielles ont été réalisées avec DEseq2 (Galaxy Version 2.11.40.2) (Love, Huber & Anders, 2014). Par convention, les pics différemment exprimés sont ceux qui ont une  $p$ -ajustée  $< 0,01$  et un  $\log_2(\text{foldchange})$  ( $\log_2\text{FC}$ ) supérieur à  $+1$  (le pic est surreprésenté chez le chronotype diurne) ou inférieur à  $-1$  (le pic est surreprésenté chez le chronotype nocturne). Lors des analyses DEseq2 entre les deux chronotypes, les échantillons se regroupent par chronotype mais aussi par sexe (Figure\_70). Comme le rythme d'émission des cercaires ne dépend pas phénotypiquement du sexe du parasite, un pic surreprésenté est un pic qui répond aux conditions citées précédemment lors d'une analyse DEseq2 entre les chronotypes, sexes confondus et sexes séparés (mâles diurnes Vs mâles nocturnes et femelles diurnes Vs femelles nocturnes).



**Figure 70 : Analyse en composante principale (ACP) issue de l'analyse DEseq2 entre les chronotypes diurne et nocturne de *S. mansoni* (sexe confondu).** La PC1, en abscisse, est le chronotype qui explique 52% de la variance des échantillons. La PC2, en ordonnée, est le sexe qui explique 24% de la variance des échantillons. Les échantillons diurnes sont représentés en bleu et les nocturnes en rose. Les échantillons mâles représentent les deux groupes du haut, tandis que les femelles sont les deux groupes du bas.

II.2.2.2.8.2. Gènes différentiellement marqués par H3K4me3 issus du DEseq2 entre les deux chronotypes

Les comparaisons entre les chronotypes diurne et nocturne révèlent qu'il y a plus de marques H3K4me3 chez le chronotype diurne (Tableau\_49 et Figure\_71).



**Figure 71: MA-plots issus des analyses DEseq2 diurnes Vs nocturnes pour H3K4me3.** Les abscisses représentent le comptage des moyennes normalisées et les ordonnées représentent le log(foldchange). Chaque point représente un pic. Tous les points ne sont pas différents entre les chronotypes. Les points rouges sont différentiellement présents entre les deux chronotypes. Dans ces MA-plots, les Log foldchange positifs sont surreprésentés chez les diurnes et les log (foldchange ) négatifs sont surreprésentés chez les nocturnes.

Sur les 105 surreprésentés chez le chronotype diurne, il a 31 gènes associés aux pics (Annexe 8).

**Tableau 49 : Tableau récapitulatif des analyses DEseq2 entre les chronotypes diurne et nocturne de *S. mansoni* pour la marque H3K4me3**

H3K4me3	Nombre de pics surreprésentés chez les diurnes	Nombre de pics surreprésentés chez les nocturnes
Desq2 Diurne (D) Vs Nocturne (N)	154	3
Desq2 Mâles D Vs Mâles N	155	4
Desq2 Femelles D Vs Femelles N	118	2
Pics communs aux trois Desq2	105	2

Parmi ces 31 gènes, aucun ne correspond à ceux sélectionnés dans les points II.2.2.2.2. à II.2.2.2.7. Toutefois, trois de ces gènes sont communs avec les QTLs intra et interspécifique (Tableau\_49).

Sur les deux pics H3K4me3 surreprésentés dans le chronotype nocturne, les deux sont associés à des gènes (Annexe 8). Aucun ne correspond aux gènes cités précédemment ; toutefois, un des gènes est commun au QTL intraspécifique (Tableau\_50).

**Tableau\_50 : Liste des gènes associés aux marques H3K4me3 différenciellement présentes entre les deux chronotypes et communs aux gènes QTL**

H3K4me3	Chr.	Identifiant (ID) du gène	Annotation GFF	Annotation HHsearch
Surreprésenté chez les diurnes	1	Smp_319050	Transmembrane protein 164	TMEM164
	3	Smp_085780	Ribosomal protein L10P	28S ribosomal protein S2, mitochondrial
		Smp_000370	Selenoprotein SelK/SelG	SelK_SelG
Surreprésenté chez les nocturnes	1	Smp_127370	Glycosyl transferase, family 14	beta-1,6-N-acetylglucosaminyltransferase (E.C.2.4.1.102)

Chr. : chromosome.

L'utilisation d'un spike-in avec la marque H3K4me3 ne perturbe pas l'immunoprécipitation de la chromatine du parasite puisque la majorité de l'alignement de l'échantillon se fait sur le génome de *S. mansoni*. La détection d'un rythme de la marque H3K4me3 s'est révélée infructueuse pour le chronotype diurne, et seulement deux pics ont été détectés pour le chronotype nocturne. Ces résultats peuvent s'expliquer de plusieurs manières. Soit, il n'y a tout simplement pas de rythme avec la marque H3K4me3, soit le rythme provient d'oscillations trop faibles pour qu'il soit détecté par les logiciels. Les logiciels de détection de rythme ont été initialement créés pour analyser des données de RNA-seq. Ces logiciels sont d'autant plus efficaces que le signal est fort (Laloum & Robinson-Rechavi, 2020). Une des solutions pour pallier à ce problème serait d'être moins sévère dans les critères de sélection des pics, au risque d'avoir des faux positifs. Une autre solution serait de refaire l'expérience avec plus de répliquas pour augmenter le signal. Enfin une autre solution est de voir le problème sous un autre angle et de coupler ces résultats avec le rythme d'expression des gènes qui codent pour les « writers » et les erasers » (ici histone méthyltransférase et histone déméthylase).

A défaut d'étudier un rythme H3K4me3, l'étude des gènes associés à la marque H3K4me3 s'est révélé intéressante puisque l'on retrouve certains gènes issus des QTLs intra- et interspécifiques. Tout comme dans la partie QTL mapping, aucun gène canonique de l'horloge n'a été détecté, conformément aux résultats de Rawlinson, Reid et al. (2021). La sélection subjective basée sur l'annotation du génome et les résultats HHsearch ont par contre mis en évidence d'autres gènes impliqués dans la phototransduction et plus largement dans la transduction des signaux environnementaux que ceux déjà mis en avant par les QTLs (Figure\_72). Cette sélection a aussi mis en évidence des gènes avec un motif BHLH de type Myc et aussi des gènes impliqués dans la modification des histones (acétylation et méthylation).

L'analyse différentielle des pics H3K4me3 entre les chronotypes a montré qu'il y avait plus de méthylation dans le chronotype diurne. Il est intéressant de voir que parmi les gènes associés aux

## Partie II : Rythme-Epigénétique

marques H3K4me3 surreprésentés dans l'un ou l'autre des chronotypes on retrouve des gènes issus des QTLs intra et interspécifiques.

Enfin, l'analyse différentielle entre les chronotypes a révélé qu'il y avait plus de marque H3K4me3 chez le diurne. Ces résultats sont intéressants dans une perspective de différenciation des parasites.

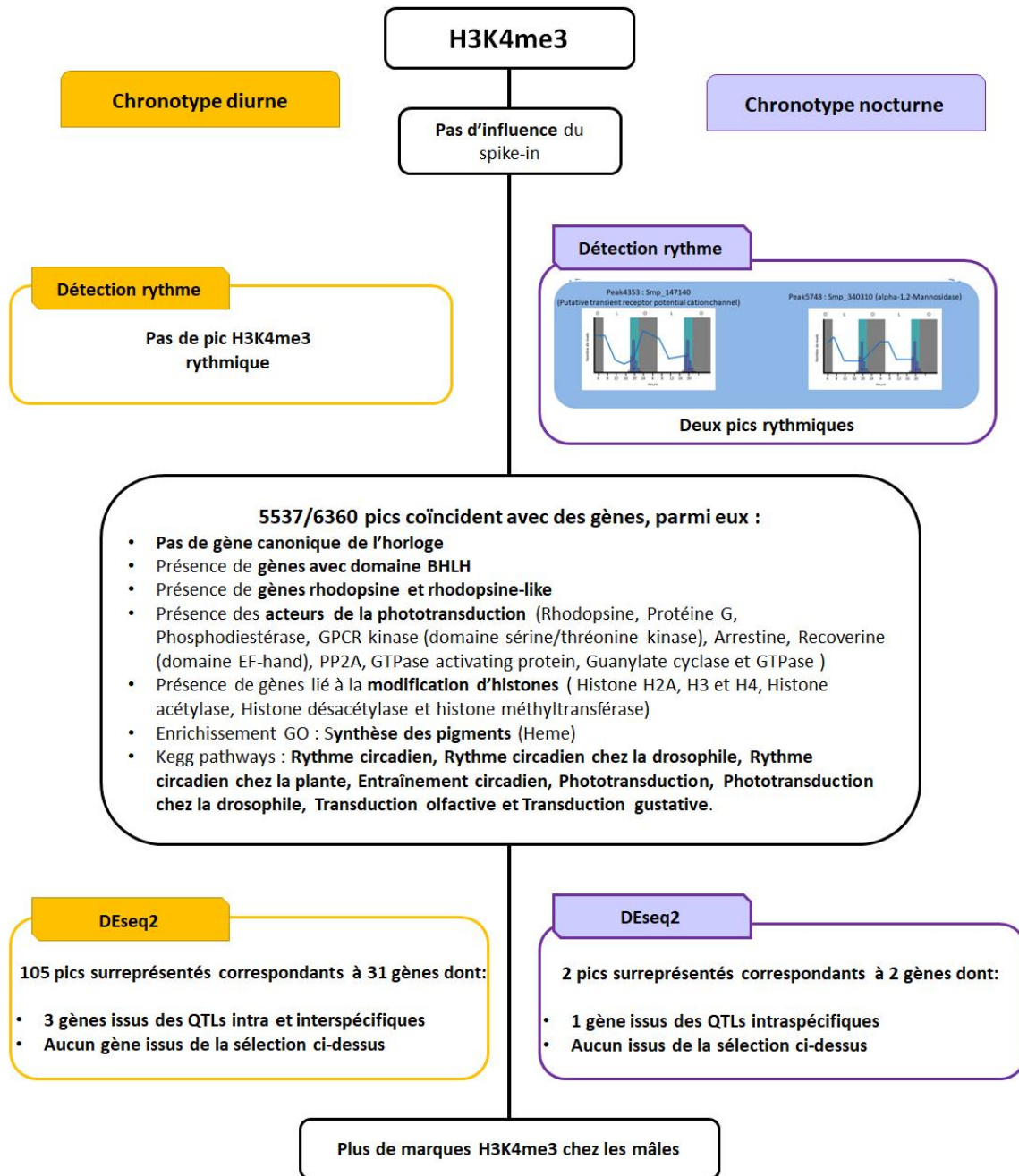


Figure 72 : Bilan de la marque H3K4me3 pour les chronotypes diurne et nocturne de *S. mansoni*.

## II.2.2.3. H3K9ac

La moyenne et les variations individuelles du nombre total et du pourcentage d'alignement des reads par échantillon selon chaque chronotype sont présentées dans le Tableau\_51. Pour les analyses, seuls les échantillons avec un minimum de 1 500 000 reads alignés uniquement sur le génome de *S. mansoni* ont été gardés (cf explication au point H3K4me3). Pour le chronotype diurne, deux échantillons sont en dessous de cette limite des 1 500 000 reads malgré la réalisation de ChIPmentation sur de nouveaux individus. Ces deux échantillons représentent tous les deux le point temporel 16h pour le 1<sup>er</sup> jour d'échantillonnage.

**Tableau 51 : Nombres et pourcentages d'alignement des reads moyens, accompagnés des variations individuelles, pour les parasites diurne et nocturne de *S. mansoni* pour les échantillons de l'expérience épigénétique H3K9ac.**

H3K9ac	Moyenne ± ES Variations individuelles		Test de Wilcoxon- Mann-Whitney
	Diurne (N=31)	Nocturne (N=33)	
Nombre total de reads	6 404 906 ± 218 958 4 150 315 à 10 542 122	5 951 099 ± 213 251 3 834 352 à 8 665 900	W =1413, p = 0,02
Alignement sur <i>S. mansoni</i> (%)	60,05 ± 2,96 35,40 à 87,78	69,60 ± 2,33 37,91 à 93,44	W = 652, p = 0,0007
Alignement unique sur <i>S. mansoni</i> (%)	40,86 ± 2,28 19,45 à 64,39	49,98 ± 1,89 26,29 à 71,11	W = 604, p = 0,0002
Alignement unique sur <i>S. mansoni</i> (nombre de reads)	2 664 144 ± 146 886 1 548 151 à 4 139 708	2 941 751 ± 133 257 1 551 133 à 4 375 208	W = 773, p = 0,01
Alignement sur <i>B. glabrata</i> (%)	17,31 ± 1,84 3,73 à 33,66	12,48 ± 1,30 1,91 à 34,40	W = 1475, p = 0,005
Alignement unique sur <i>B. glabrata</i> (%)	8,87 ± 0,93 1,59 à 17,18	6,23 ± 0,68 0,69 à 17,75	W = 1461, p = 0,006
Alignement unique sur <i>B. glabrata</i> (nombre de reads)	553 368 ± 63 308 105 697 à 1 363 188	380 489 ± 50 858 35 216 à 1 536 665	W = 1494, p = 0,003
Alignement sur <i>D. melanogaster</i> , <i>spike-in</i> (%)	7,91 ± 1,46 0,67 à 32,36	6,91 ± 0,95 0,76 à 19,49	W = 1308, p = 0,12
Alignement unique sur <i>D. melanogaster</i> , <i>spike-in</i> (%)	5,56 ± 1,09 0,35 à 23,61	4,86 ± 0,69 0,40 à 13,97	W = 1283, p = 0,17
Alignement unique sur <i>D. melanogaster</i> , <i>spike-in</i> (nombre de reads)	374 478 ± 89 342 17 981 à 2 487 191	294 899 ± 42 222 22 340 à 709 396	W = 1325, p = 0,09

*S. mansoni* est en bleu, *B. glabrata* en orange et *D. melanogaster* (*spike-in*) en vert.

Pour les chronotypes, les échantillons s'alignent majoritairement sur le génome de *S. mansoni*, puis vient celui de *B. glabrata* et enfin celui de *D. melanogaster*. Le nombre de reads qui s'aligne sur le génome de l'hôte est trop faible pour être exploité dans des analyses bioinformatiques. Cette faible part de l'hôte peut s'expliquer par le fait que l'hépatopancréas est rempli de sporocystes et que le ratio cellule hôte/parasite penche en faveur de *S. mansoni*. Il peut aussi être ajouté à cela que l'alignement se fait sur le génome d'une autre espèce que celle présente dans l'échantillon car celui de *B. pfeifferi* n'existe pas.

Côté spike-in, il n'y a aucune différence significative entre les deux chronotypes ( $p > 0,01$ ), ce qui permet de faire des comparaisons biologiques. Dans les échantillons nocturnes, l'immunoprécipitation du parasite est significativement plus importante ( $p < 0,01$ ). On peut tirer les mêmes conclusions qu'avec la marque H3K4me3 car les inputs sont les mêmes. Dans les deux expériences, la chromatine des échantillons nocturnes a une meilleure affinité avec l'anticorps que celle des échantillons diurnes.

### II.2.2.3.1. Rythme des pics

La détection de l'acétylation des histones se fait de la même façon qu'avec la marque H3K4me3 (cf II.2.2.2.1. Rythme des pics).

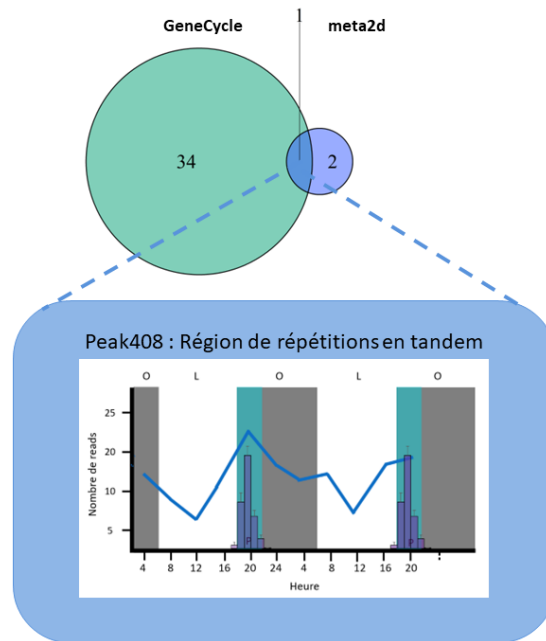
Les deux chronotypes confondus totalisent 566 pics H3K9ac détectés. Pour les analyses rythmiques, seuls les pics avec une profondeur moyenne supérieure à 10 reads sont utilisés. Le nombre de pics rythmiques par chronotype et logiciels sont présentés dans le Tableau\_52.

**Tableau\_52 : Nombre de pics H3K9ac détectés comme rythmiques et nombre de gènes associés en fonction du logiciel et du chronotype du parasite ( $p < 0,05$ )**

	H3K9ac	Diurne (453 pics filtrés)	Nocturne (483 pics filtrés)	Communs aux deux chronotypes
GeneCycle ( $p$ -value)	Pics	22	35 (dont 11 FDR < 0,05)	14
	Gènes associés	4	3 (dont 1 FDR < 0,05)	2
meta2d ( $p$ -value)	Pics	0	3	
	Gènes associés	0	0	

Aucun pic n'est détecté comme rythmique avec la FDR/BH.Q, sauf avec le logiciel GeneCycle pour le chronotype nocturne. Pour le chronotype diurne, aucun pic H3K9ac n'est considéré comme rythmique car aucun pic n'a été détecté avec le logiciel meta2d (Tableau\_51). Pour le chronotype nocturne, 35 pics ont été détectés comme rythmiques avec GeneCycle et 3 avec le logiciel meta2d ( $p < 0,05$ ). Sur tous ces pics, un seul est commun aux deux logiciels (Figure\_73), le n°408. Ce pic n'est pas associé à un gène mais à une région de répétition en tandem (ici région non codante qui contient des motifs nucléotidiques répétés). Le pic des marques H3K9ac est au point temporel 20h soit en même temps que le pic d'émission des cercaires.





**Figure 73 : Profil de la marque H3K9ac rythmique chez le chronotype nocturne.** Chaque histogramme a, en abscisse, les heures (points temporels) et, en ordonnées, le nombre de reads pour le pic sélectionné. En fond se trouve le profil d'émission des cercaires. Les rectangles gris symbolisent l'obscurité (O), les blancs la lumière (L) et les rectangles bleus la période du pic d'émission des cercaires.

#### II.2.2.3.2. Gènes associés aux marques H3K9ac

Sur les 566 pics H3K9ac détectés, 93 coïncident avec des gènes. Tout comme pour les gènes candidats QTL, une sélection subjective basée sur la recherche de gènes de l'horloge, de la phototransduction et de l'acétylation et la méthylation des histones a été faite. L'enrichissement GO a été fait avec Blast2GO version 5.2 en utilisant le Fisher's Exact Test ( $p < 0,05$ ). L'analyse d'enrichissement Kegg pathway a été faite sur le serveur KAAS en utilisant le programme Blast avec la méthode SBH (Single-directional Best Hit).

#### II.2.2.3.3. Les gènes de l'horloge (Basic Helix-Loop-Helix)

Il n'y a aucun gène canonique de l'horloge, ni de protéine au motif BHLH dans les gènes marqués H3K9ac.

#### II.2.2.3.4. Les rhodopsines et autres gènes impliqués dans la phototransduction

Il n'y a aucune rhodopsine dans les gènes associés aux marques H3K9ac mais on retrouve un gène rhodopsine-like **Smp\_134100** (chromosome 3) qui code pour un récepteur couplé aux protéines G. Parmi les différents acteurs de la phototransduction, nous retrouvons, dans les gènes marqués H3K9ac, uniquement une sous-unité  $\gamma$  (gamma) de la protéine G (**Smp\_205760**, chromosome 3).

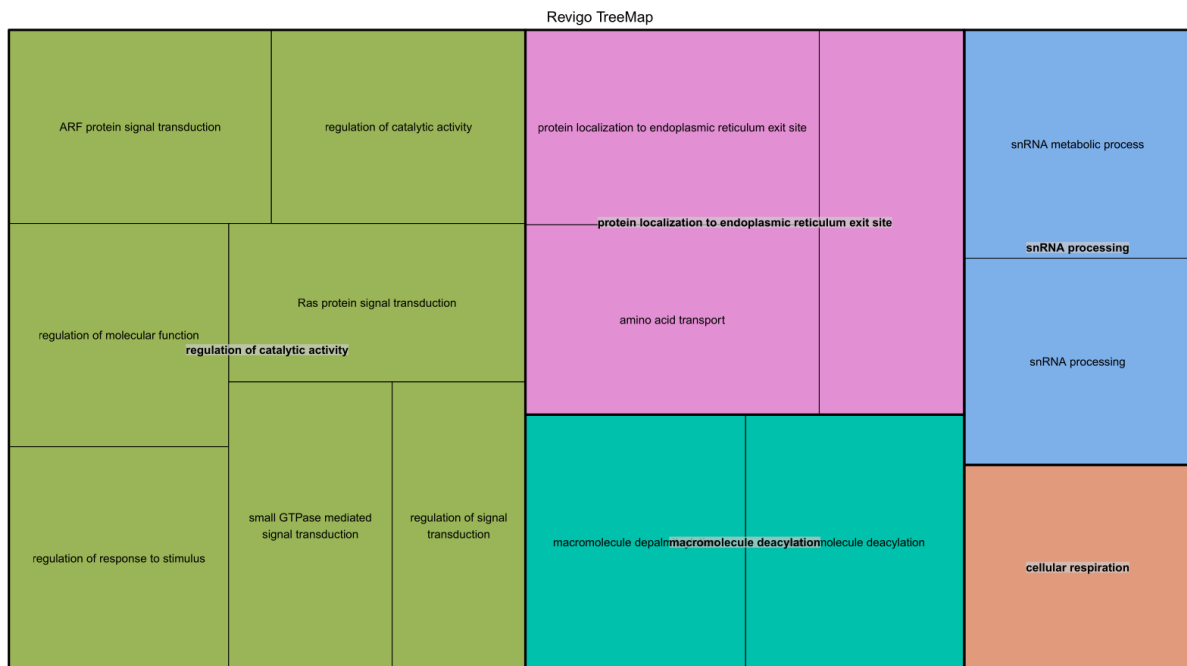
### II.2.2.3.5. Les gènes impliqués dans l'acétylation et la méthylation des histones

Parmi les gènes associés à la marque H3K9ac, aucun n'est en rapport avec les histones et la modification d'histone.

### II.2.2.3.6. Enrichissement GO : Processus biologique

Pour les gènes associés à la marque H3K9ac, les analyses d'enrichissement GO n'ont pas donné de résultats significatifs en utilisant la limite usuelle FDR < 0,05 (False Discovery Rate). Les résultats qui suivent ont été obtenus en utilisant la limite de  $p < 0,05$ .

Parmi toutes les annotations GO, le grand sous-ensemble régulation de l'activité catalytique est intéressant puisqu'il contient les annotations GO régulation de la réponse à un stimulus, régulation de la transduction du signal médié par des petites GTPases, transduction du signal médié par des petites GTPases et transduction du signal des protéines ARF (ADP-ribosylation factor) (Figure\_74). Ces quatre annotations regroupent deux gènes qui sont **Smp\_147020** dans le chromosome 6 (putative arf6 guanine nucleotide exchange factor) qui fait partie des gènes candidats du croisement intraspécifique et **Smp\_124660** dans le chromosome 2 (guanyl-nucleotide exchange factor). Les guanines exchange factor qui interagissent avec les protéines Arf sont impliqués dans le trafic vésiculaire. Il a été montré chez la souris que l'interaction entre la guanine exchange factor (EFA6A) et Arf6 pourrait avoir un rôle dans un sous-compartiment des processus photorécepteurs pérисynaptique (Katsumata, Honma et al., 2008).



**Figure 74 : Treemap revigo des résultats de l'analyse d'enrichissement GO des processus biologiques obtenus pour les gènes associés à la marque H3K9ac.** Chaque couleur représente un grand ensemble de processus biologiques. Il y a de gauche à droite : la régulation de l'activité catalytique, la localisation protéique au site de sortie du réticulum endoplasmique, le processus des nRNA (small nuclear RNA), la désacylation des macromolécules et la respiration cellulaire. La treemap a été obtenue avec revigo (Supek, Bošnjak et al., 2011).

Une autre annotation intéressante est la respiration cellulaire qui contient deux gènes différents. Le gène **Smp\_900070** qui code pour un cytochrome B et le gène **Smp\_900090** qui code pour la sous-unité 4 de la NADH déshydrogénase. Si de prime abord ces deux gènes sont impliqués dans la respiration cellulaire dans la mitochondrie, il a été montré chez le champignon *Schizosaccharomyces pombe* que le cytochrome B était un composant essentiel dans la photosynchronisation de l'horloge circadienne (Kippert, Ninnemann & Engelmann, 1991).

### II.2.2.3.7. Kegg pathway

Les analyses Kegg pathway montrent que les gènes associés à la marque H3K9ac sont potentiellement impliqués dans 23 voies moléculaires différentes. Parmi toutes ces voies, aucune ne concerne le rythme circadien, l'entraînement circadien ou encore la transduction des signaux lumineux, olfactifs et gustatifs.

### II.2.2.3.8. Analyses DEseq2

#### II.2.2.2.8.1. Comparaison des chronotypes

Pour déterminer s'il y a une différence d'acétylation des histones entre les deux chronotypes de *S. mansoni*, des analyses différentielles ont été réalisées avec DEseq2 (Galaxy Version 2.11.40.2) (Love, Huber & Anders, 2014) selon les mêmes conditions que la marque H3K4me3 (II.2.2.2.8. Analyses DEseq2). Lors des analyses DEseq2 entre les deux chronotypes, les échantillons se regroupent par chronotype mais aussi par sexe (Figure\_75). Comme le rythme d'émission des cercaires ne dépend pas phénotypiquement du sexe du parasite, un pic surreprésenté est un pic qui répond aux conditions citées précédemment lors d'une analyse DEseq2 entre les chronotypes, sexes confondus et sexes séparés (mâles diurnes Vs mâles nocturnes et femelles diurnes Vs femelles nocturnes).

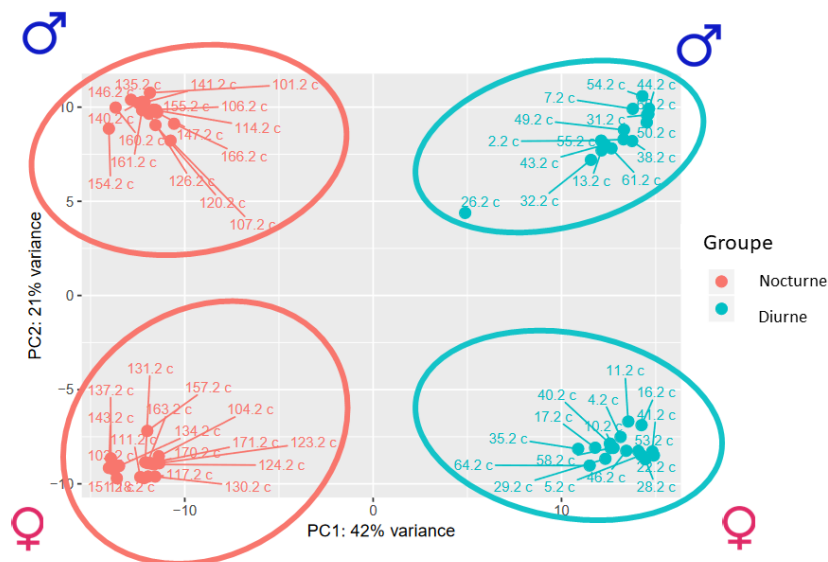
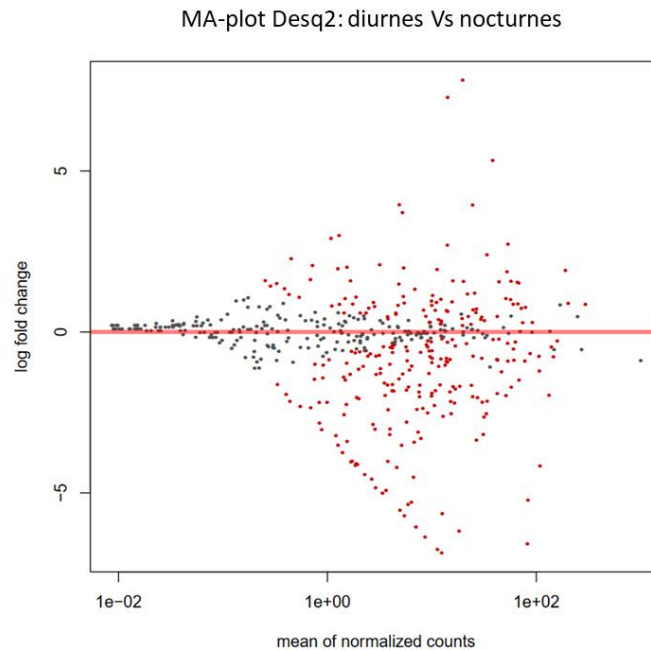


Figure 75 : Analyse en composante principale (ACP) issue de l'analyse DEseq2 entre les chronotypes diurne et nocturne de *S. mansoni* (sexes confondus) pour la marque H3K9ac. La PC1, en abscisse est le chronotype qui explique 52% de la variance des échantillons. La PC2, en ordonnée est le sexe qui explique 21% de la variance

des échantillons. Les échantillons diurnes sont représentés en bleu et les nocturnes en rose. Les échantillons mâles représentent les deux groupes du haut tandis que les femelles sont les deux groupes du bas.

### II.2.2.3.8.2. Gènes différentiellement marqués par H3K9ac entre les deux chronotypes

Les comparaisons entre les chronotypes diurne et nocturne révèlent qu'il y a plus de marque H3K9ac chez le chronotype nocturne (Tableau\_53 et Figure\_76).



**Figure 76 : MA-plots issus des analyses DEseq2 diurnes Vs nocturnes pour H3K9ac.** Les abscisses représentent le comptage des moyennes normalisées et les ordonnées représentent le log(foldchange). Chaque point représente un pic. Tous les points ne sont pas différents entre les chronotypes. Les points rouges sont différentiellement présents entre les deux chronotypes. Dans ces MA-plots, les Log foldchange positifs sont surreprésentés chez les diurnes et les log (foldchange ) négatifs sont surreprésentés chez les nocturnes.

Sur les 92 surreprésentés chez le chronotype nocturne, il y a 17 gènes associés aux pics (Annexe 9). Chez le chronotype diurne, sur les 32 pics surreprésentés, il y a 2 gènes associés (Annexe 9).

**Tableau\_53 : Tableau récapitulatif des analyses DEseq2 entre les chronotypes diurne et nocturne de *S. mansoni* pour la marque H3K9ac**

H3K9ac	Nombre de pics surreprésentés chez les diurnes	Nombre de pics surreprésentés chez les nocturnes
Desq2 Diurne (D) Vs Nocturne (N)	39	112
Desq2 Mâles D Vs Mâles N	36	99
Desq2 Femelles D Vs Femelles N	46	101
Pics communs aux trois Desq2	32	92

Parmi les gènes associés aux pics H3K9ac surreprésentés chez les nocturnes, nous retrouvons trois gènes sélectionnés dans les points précédents concernant H3K9ac (Tableau\_53). De plus, nous retrouvons aussi des gènes candidats des QTL intraspécifique et interspécifique (Tableau\_53). Il est aussi intéressant de noter que certains de ces gènes sont co-marqués par H3K9ac et H3K4me3 (Tableau\_54).

**Tableau 54 : Liste des gènes associés aux marques H3K9ac différenciellement présentes entre les deux chronotypes et communs aux gènes QTL et/ou sélectionnés dans les points précédents**

H3K9ac	Chr.	Identifiant (ID) du gène	Annotation GFF	Annotation HHsearch	Origine de la sélection
Sur-représenté chez les nocturnes	3	Smp_205760	Guanine nucleotide-binding protein subunit gamma		Potentiellement lié à la phototransduction (protéine G)
	7_MITO	Smp_900070	Cytochrome b (mitochondrion)		Annotation GO, respiration cellulaire
		Smp_900090	NADH dehydrogenase subunit 4 (mitochondrion)		
	1	Smp_127370	Glycosyl transferase, family 14	beta-1,6-N-acetylglucosaminyltransferase (E.C.2.4.1.102)	QTL intraspécifique
	3	Smp_308470	SEA domain	SEA	QTL interspécifique
Smp_084540		Forkhead-associated (FHA) domain	RNA-binding protein 5, Survival motor		
Sur-représenté chez les diurnes	6	Smp_147020	Sec7 domain	guanine nucleotide exchange protein	QTL intraspécifique

Chr. : chromosome.

Parmi les deux gènes associés aux pics H3K9ac surreprésentés chez les diurnes, nous retrouvons un gène sélectionné dans les points précédents qui fait partie des gènes candidats du QTL intraspécifique (Tableau\_54). Il est aussi intéressant de noter que ces deux gènes sont co-marqués par H3K9ac et H3K4me3 (Tableau\_55).

**Tableau 55 : Liste des gènes associés aux marques H3K9ac différenciellement présentes entre les deux chronotypes qui sont co-marqués avec H3K4me3**

H3K9ac et H3K4me3	Chr.	Identifiant (ID) du gène	Annotation GFF	Annotation HHsearch	
Sur-représenté chez les nocturnes H3K4me3 et H3K9ac	1 (QTL intraspé.)	Smp_127370	Glycosyl transferase, family 14	beta-1,6-N-acetylglucosaminyltransferase (E.C.2.4.1.102)	
Sur-représenté chez les nocturnes H3K9ac	1	Smp_158570	NA		
	2	Smp_011220	Putative glucosamine-fructose-6-phosphate aminotransferase		
	3	Smp_205760	Guanine nucleotide-binding protein subunit gamma		
	5	Smp_327560	Thioredoxin domain-containing protein		
	6	Smp_341570	NA		
	7	Smp_010380	Isochorismatase domain-containing protein		
	7_MITO		Smp_900070	Cytochrome b (mitochondrion)	
			Smp_900080	NA	
			Smp_900090	NADH dehydrogenase subunit 4 (mitochondrion)	
3 (QTL interspé.)	Smp_084540	Forkhead-associated (FHA) domain	RNA-binding protein 5, Survival motor		
Sur-représenté chez les diurnes H3K9ac	6	Smp_336310	NA		
	6 (QTL intraspé.)	Smp_147020	Sec7 domain	guanine nucleotide exchange protein	

Chr.: chromosome; intraspé.: intraspécifique; interspé. : interspécifique.

Tout comme la marque H3K4me3, l'utilisation d'un spike-in avec la marque H3K9ac ne perturbe pas l'immunoprécipitation de la chromatine du parasite puisque la majorité de l'alignement de l'échantillon se fait sur le génome de *S. mansoni*. De la même façon qu'H3K4me3, la détection d'un rythme de la marque H3K9ac s'est révélée infructueuse pour le chronotype diurne, et seulement un pic a été détecté pour le chronotype nocturne. Le pic rythmique atteint son sommet à 20h soit en même temps que le pic d'émission des cercaires du chronotype nocturne. Ce pic H3K9ac nocturne est associé à une région non codante de répétition en tandem du génome qui pourrait représenter une piste supplémentaire pour l'étude du rythme d'émission des cercaires. Cette très faible présence de pic H3K9ac rythmique s'explique de la même façon que pour la marque H3K4me3.

L'étude des gènes associés à la marque H3K9ac ne s'est pas révélée aussi prolifique que la marque H3K4me3 (Figure\_77). On notera tout de même la mise en évidence d'un autre gène rhodopsine-like.

Toutefois, il n'y a pas de marque H3K9ac associée au gène *SmRho* (smp\_104210) comme rapporté par Cosseau, Azzi et al., 2009. Cette différence avec leurs travaux peut s'expliquer par le fait que leurs expériences ont été menées sur des cercaires libres d'une autre souche de *S. mansoni*. La différence peut donc être interprétée soit comme une spécificité due à la souche du parasite (Oman ou Porto Rico) soit à une spécificité du stade (cercaire libre, contre sporocyste et cercaires intrasporocystes).

L'analyse différentielle des pics H3K9ac entre les chronotypes a montré qu'il y avait plus d'acétylation dans le chronotype nocturne. Il est intéressant de voir que, parmi les gènes associés aux marques H3K9ac surreprésentées dans l'un ou l'autre des chronotypes, on retrouve non seulement des gènes issus des QTLs intra et interspécifiques mais aussi des gènes qui sont co-marqués avec H3K4me3. Les marques H3K4me3 et H3K9ac sont toutes les deux des marques associées à l'activation de la transcription, ce qui peut mettre en avant l'importance du rôle de ces gènes co-marqués pour le parasite.

Enfin, l'analyse différentielle entre les chronotypes a révélé qu'il y avait plus de marque H3K9ac chez les nocturnes. Couplé avec les résultats de la marque H3K4me3, il est intéressant de voir que l'on peut différencier des individus parasites par leur chronotype. Ces résultats ouvrent la possibilité d'établir des profils de marques épigénétiques pour différencier les différents chronotypes de schistosomes.

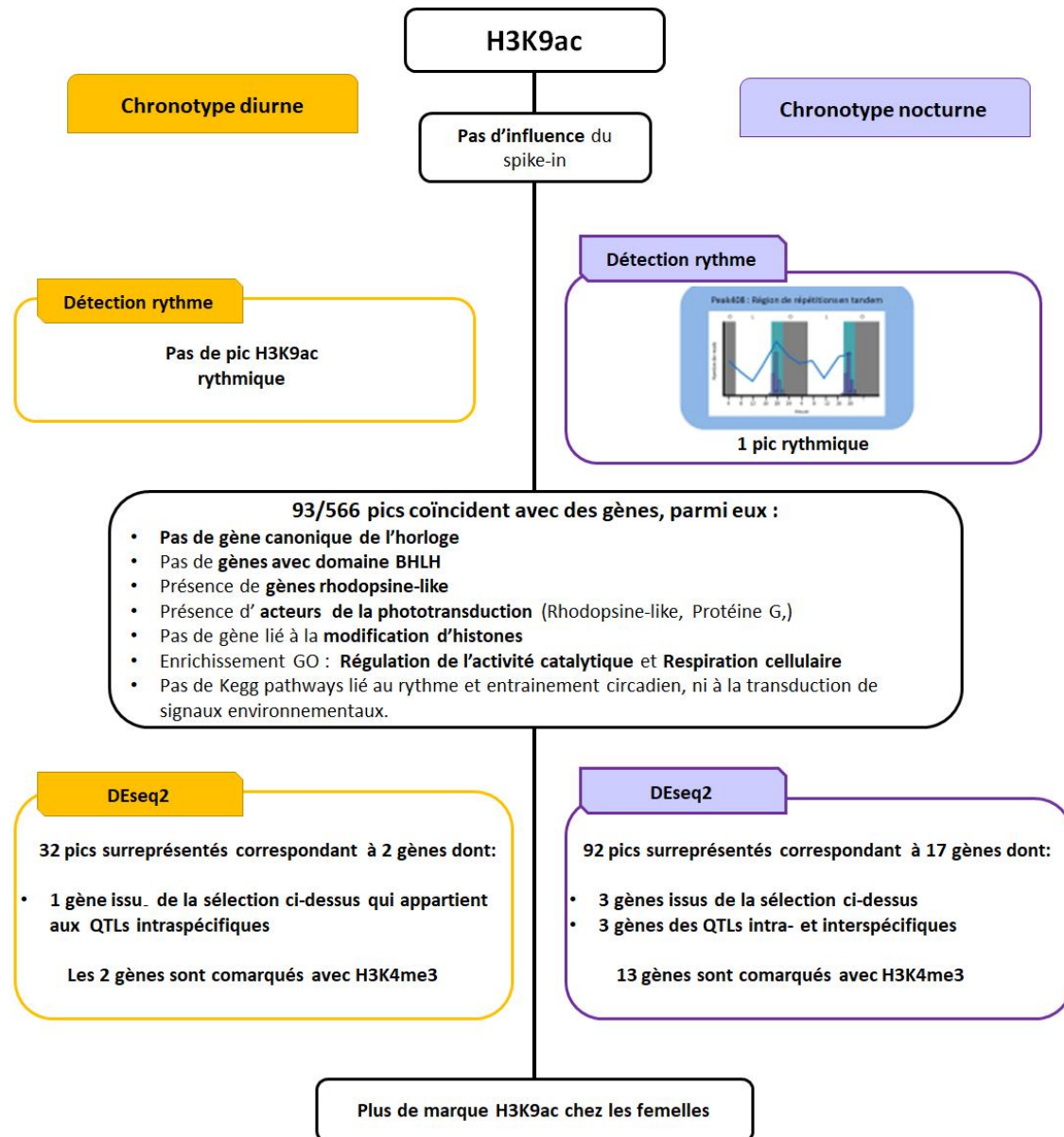


Figure 77 : Bilan de la marque H3K9ac pour les chronotypes diurne et nocturne de *S. mansoni*.



### II.2.3. RNA-seq

Dans le cadre de l'étude d'un rythme, la transcriptomique permet de mettre en évidence les gènes qui ont une expression rythmique, et donc d'identifier ceux potentiellement impliqués dans le phénotype. L'étude de cette expression se fait au travers de la technique de RNA-sequencing (RNA-seq) qui quantifie les ARNm présents dans un échantillon, couplée à un logiciel de détection des rythmes dans une cinétique d'échantillons. Chez *S. mansoni*, une seule étude de transcriptomique rythmique a été menée et c'était sur le stade adulte (Rawlinson, Reid et al., 2021). Cette étude révèle ainsi qu'il n'y a aucun gène principal de l'horloge présent dans le génome de *S. mansoni*.

L'objectif de ce chapitre est donc de déterminer quels sont les gènes rythmiques au stade sporocyste/cercaires intra-sporocystes et de croiser ces résultats avec les gènes candidats identifiés dans les approches de génétique et d'épigénétique. Pour cela, un RNA-seq a été effectué sur des morceaux d'hépatopancréas et d'ovotestis parasités par chacun des chronotypes de *S. mansoni*.

Le premier point de ce sous chapitre sera consacré à une présentation de la méthode utilisée, le second point présentera les gènes rythmiques et le dernier point fera le lien entre les trois approches moléculaires en étudiant les différents gènes candidats sélectionnés.

#### II.2.3.1. Méthode

La méthode de RNA-seq se divise en plusieurs étapes : l'échantillonnage biologique, la préparation moléculaire des échantillons et les analyses bioinformatiques des données. Ces expériences sont issues d'une collaboration avec l'équipe de Tim Anderson du Texas Biomedical Research Institute. L'échantillonnage biologique a été réalisé à Perpignan et est commun avec l'expérience d'épigénétique. La préparation moléculaire des échantillons a été réalisée au Texas par Robbie Diaz et Amanda Strickland de l'équipe de Tim Anderson. L'analyse bioinformatique a été réalisée entre les deux sites. L'échantillonnage biologique consiste à récupérer des hépatopancréas et ovotestis de *B. pfeifferi* infestés par *S. mansoni*. La préparation moléculaire des échantillons consiste à extraire l'ARN des échantillons, à isoler les ARNm grâce à leur queue polyA et à faire une transcription inverse (convertir ARN en ADN complémentaire). L'ADNc est ensuite séquencé. L'analyse bioinformatique consiste dans un premier temps à faire un alignement contre la version 7 du génome de *S. mansoni*, avec STAR (Dobin, Davis et al., 2013) puis de faire un comptage des transcrits pour chaque gène avec RSEM (Li & Dewey, 2011). La méthode détaillée de la préparation moléculaire ainsi que le détail du script utilisé pour l'alignement et le comptage des transcrits sont similaires à ceux utilisés dans Clec'h, Chevalier, Mattos et al. (2021). La détection des gènes rythmiques s'est faite avec 3 logiciels différents pour éviter les biais (Takahashi, Kumar et al., 2015) : GeneCycle (Ahdesmäki, Fokianos & Strimmer, 2012), empirical-JTK (Hutchinson, Maienschein-Cline et al., 2015) et meta2d (Wu, Anafi et al., 2016).

#### II.2.3.2. Gènes rythmiques

Deux RNA-seq ont été menés dans le cadre de cette thèse. L'expérience principale a été conduite sur des parasites ayant été laissés sous photopériode équilibrée classique (L-O). La seconde expérience utilise, quant à elle, des parasites qui ont subi 10 jours de photopériode inversée (O-L). Cette seconde expérience sert de complément à la première dans la sélection de gènes d'intérêt et ne sera donc pas développée.

Pour détecter le rythme d'expression des gènes de *S. mansoni*, le logiciel GeneCycle (Ahdesmäki, Fokianos & Strimmer, 2012), le logiciel empirical-JTK (Hutchinson, Maienschein-Cline et al., 2015) et le logiciel meta2d (Wu, Anafi et al., 2016) ont été utilisés. Les logiciels reposent sur l'utilisation du test exact de Fisher. Pour pallier au risque de faux positifs dus au grand nombre de tests à réaliser, chaque logiciel applique une correction de la *p*-value basée sur la méthode de Benjamini-Hochberg, appelée BH.Q ou FDR (False Discovery rate) avec un seuil fixé à 5% de risque (BH.Q/FDR < 0,05). Pour le cas du logiciel empirical-JTK, la correction de la *p*-value que nous avons utilisée est la méthode Bonferroni

Partie II : Rythme-Transcriptomique

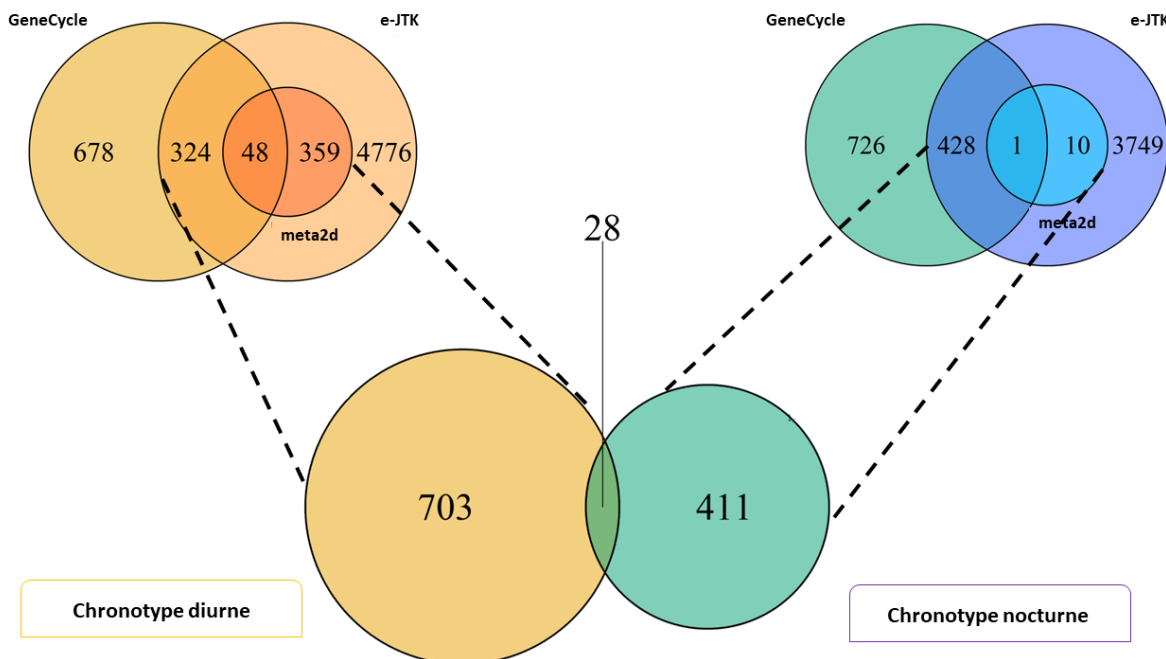
(appelé BF), avec un seuil fixé à 5% de risque ( $BF < 0,05$ ). La méthode de Bonferroni est plus restrictive que la méthode Benjamini-Hochberg. Un pic est considéré comme rythmique s'il est détecté dans au moins deux des trois logiciels avec au moins le seuil  $p$ -value  $< 0,05$ . Cette précaution est utilisée pour éviter une variabilité qui serait propre au logiciel (Takahashi, Kumar et al., 2015 ; Rijo-Ferreira, Acosta-Rodriguez et al., 2020).

Pour les analyses rythmiques, seuls les gènes avec une expression moyenne supérieure à 10 reads ont été utilisés. Le nombre de gènes rythmiques par chronotype et logiciel est présenté dans le Tableau\_56.

**Tableau 56 : Nombre de gènes détectés comme rythmiques en fonction du logiciel et du chronotype du parasite ( $p < 0,05$ ) pour les échantillons en photopériode équilibrée**

Condition photopériode équilibrée	Seuil	Diurne (9832 gènes exprimés)	Nocturne (9901 gènes exprimés)
GeneCycle (Nombre de gènes)	$p$ -val	1050	1155
	FDR	465	189
Empirical-JTK (Nombre de gènes)	$p$ -val	5507	4188
	BF	1369	297
Meta2d (Nombre de gènes)	$p$ -val	407	11
	BH.Q	32	0
Commun	$p$ -val	703	411

Environ 7% des gènes sont rythmiques dans le chronotype diurne contre 4% pour le chronotype nocturne. Sur les 1114 gènes détectés comme rythmiques, 28 sont communs au chronotype diurne et nocturne (Figure\_78), ce qui implique que la majorité des gènes rythmiques sont propres à chaque chronotype.



**Figure 78 : Diagramme des gènes rythmiques en fonction des logiciels et des chronotypes.**

## II.2.3.2.1. Les gènes communs aux deux chronotypes

Sur les 28 gènes communs, il y a un gène issu du QTL interspécifique (**Smp\_313310**) potentiellement impliqué dans la transduction et un gène marqué par H3K4me3 (**Smp\_045260**) (Tableau\_57). On retrouve aussi une putative otopetrin qui est un canal à proton impliqué dans la transduction du goût acide (Ramsey & DeSimone, 2018) dont le pic d'expression diurne est à 12h (le jour, pendant le pic) et le pic d'expression nocturne à 4h (la nuit). Les analyses d'enrichissement GO et de Kegg pathway n'ont pas donné de résultat intéressant à développer ici.

Tableau\_57 : Liste des gènes communs aux chronotypes diurne et nocturne avec le pic d'expression des gènes

Chr.	Gene ID	GFF annotation	HHsearch annotation	Pic expression Diurne (h)	Pic expression Nocturne (h)
1	Smp_121990	Glycerol-3-phosphate dehydrogenase		8	4
	Smp_149710	NA		16	24
	Smp_160710	Putative upstream stimulatory factor		8	16
	Smp_169360	Putative kinesin		12	4
	Smp_202260	NA		12	16
	Smp_314800	NA		12	4
2	Smp_045260	5-aminolevulinate synthase		8	4
	Smp_161210	NA		8	4
	Smp_310750	NA		8	4
	Smp_316380	NA		8	4
	Smp_328260	NA		20	20
4	Smp_050280	60S ribosomal protein L31		20	24
	Smp_172590	Carboxypeptidase		8	4
	Smp_197860	Putative gelsolin		12	4
	Smp_312580	Pribosyltran domain-containing protein		8	20
	Smp_313310	EF-hand domain	Sarcoplasmic calcium-binding protein {Amphioxus (Branchiostoma lanceolatum) [TaxId: 7740]}	16	12
5	Smp_125660	Heterogeneous nuclear ribonucleoprotein-related		12	4
6	Smp_183710	Putative actin		4	24
	Smp_203130	NA		8	4
7	Smp_125080	Mername-AA217 hypothetical peptidase (M14 family)		12	4
	Smp_130280	Cell polarity protein		12	4
	Smp_152940	Putative otopetrin		12	4
	Smp_337880	NA		8	20
ZW	Smp_091790	NA		16	12

## Partie II : Rythme-Transcriptomique

	Smp_128440	Putative run and fyve domain containing protein		<b>12</b>	<b>4</b>
	Smp_245290	NA		<b>12</b>	<b>4</b>
	Smp_306990	NA		8	4
	Smp_344820	NA		12	4

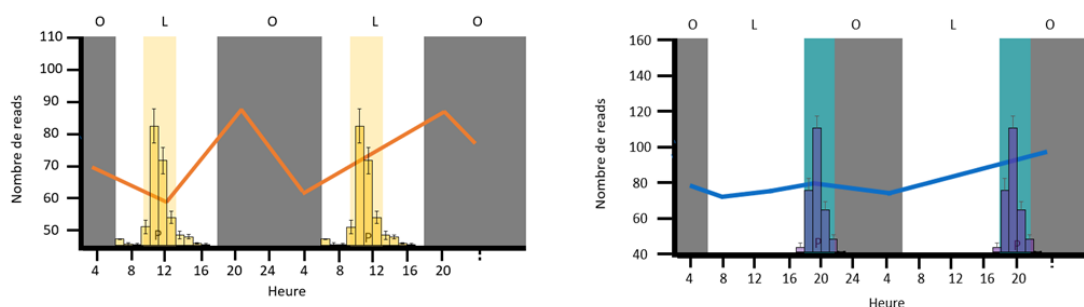
Chr. : Chromosome.

Chez le chronotype diurne, la majorité des pics d'expression des gènes communs a lieu à 8h et 12h (Tableau\_58), soit 4h avant le pic ou pendant le pic d'émission des cercaires. Chez le chronotype nocturne, l'écrasante majorité des pics d'expressions a lieu à 4h (Tableau\_57), soit 16h avant le pic d'émission des cercaires.

**Tableau 58 : Répartition des pics d'expression des gènes rythmiques dans les deux chronotypes**

Heures	Nombre des pics d'expression des gènes rythmiques Diurnes	Nombre des pics d'expression des gènes rythmiques Nocturnes
4	1	18
8	11	0
12	11	2
16	3	2
20	2	3
24	0	3
Total des pics d'expression diurne (8h, 12h, 16h)	25	4
Total des pics d'expression nocturne (4h, 20h, 0h)	3	24

Si on s'intéresse à l'heure des pics d'expression entre les chronotypes, on observe différents types de comportement. Il y a tout d'abord les gènes qui ont le même profil d'expression entre les deux chronotypes (**Smp\_328260**) (Figure\_79). Ce résultat peut indiquer que le gène est un acteur commun aux deux chronotypes dans le mécanisme du rythme d'émission des cercaires ou que ce gène occupe une fonction commune mais dans un processus biologique indépendant du rythme d'émission des cercaires. Un comportement inverse est observé où l'expression est diamétralement opposée (**Smp\_312580**).

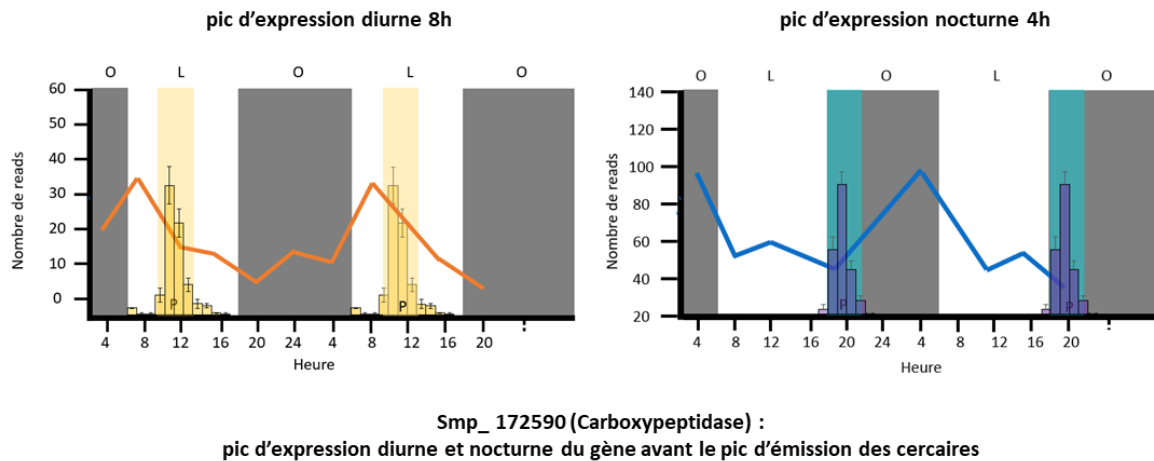


**Smp\_328260 (NA) : pic d'expression diurne et nocturne à 20h**

**Figure 79 : Exemple de profils d'expression de gène commun aux deux chronotypes (pic d'expression à la même heure).**

Chaque histogramme a, en abscisse, les heures (points temporels) et en ordonnées le nombre de reads pour le gène sélectionné. En fond, se trouve le profil d'émission des cercaires. Les rectangles gris symbolisent l'obscurité (O), les blancs la lumière (L), les rectangles jaunes la période du pic d'émission des cercaires chronotype diurne et les rectangles bleus la période du pic d'émission des cercaires chronotype nocturne.

Un autre comportement implique des pics d'expression avant le pic d'émission des cercaires (Figure\_80), soit 4h avant pour le diurne et 16h avant pour le nocturne (**Smp\_121990, Smp\_045260, Smp\_161210, Smp\_310750, Smp\_316380 et Smp\_172590**). Un autre comportement similaire, représente un pic d'expression diurne en même temps que le pic d'émission des cercaires diurne, tandis que le pic d'expression nocturne reste à 16h avant le pic d'émission des cercaires (Smp\_169360, Smp\_314800 et **Smp\_197860**). Ces deux cas de figure peuvent illustrer que les mêmes acteurs sont impliqués pour le rythme d'émission des cercaires dans les deux chronotypes mais qu'une mutation ou une modulation différente de l'expression des gènes concernés impliquent une différence d'heure d'émission des cercaires.



**Figure 80 : Exemple de profils d'expression de gène commun aux deux chronotypes (pic d'expression avant le pic d'émission des cercaires).** Chaque histogramme a, en abscisses, les heures (points temporels) et en ordonnées le nombre de reads pour le gène sélectionné. En fond, se trouve le profil d'émission des cercaires. Les rectangles gris symbolisent l'obscurité (O), les blancs la lumière (L), les rectangles jaunes la période du pic d'émission des cercaires chronotype diurne et les rectangles bleus la période du pic d'émission des cercaires chronotype nocturne.

### II.2.3.2.2. Les gènes diurnes rythmiques

Tout comme pour les gènes candidats QTL et marque histone, une sélection subjective basée sur la recherche de gènes de l'horloge, de la phototransduction et de l'acétylation et la méthylation des histones a été faite. L'enrichissement GO a été fait avec Blast2GO version 5.2 en utilisant le Fisher's Exact Test ( $p < 0,05$ ). L'analyse d'enrichissement Kegg pathway a été faite sur le serveur KAAS en utilisant le programme Blast avec la méthode SBH (Single-directional Best Hit). La liste complète des gènes diurnes rythmiques est présentée dans l'annexe 10.

#### II.2.3.2.2.1. Les gènes de l'horloge (Basic Helix-Loop-Helix)

Il n'y a aucun gène canonique de l'horloge ni de gène avec un domaine BHLH dans les gènes rythmiques diurnes.

#### II.2.3.2.2.2. Les rhodopsines et autres gènes impliqués dans la phototransduction

Il y a trois gènes rhodopsine-like (Tableau\_59), mais aucune rhodopsine. Chacun de ces trois gènes a un pic d'expression diurne différent, qui se déroule soit avant (8h), soit pendant (12h) soit après le pic d'émission des cercaires (16h).

Tableau 59 : Liste des gènes diurnes rythmiques rhodopsine-like

Chr.	Identifiant (ID) du gène	Annotation GFF	Heure du pic d'expression
1	Smp_072450	Putative rhodopsin-like orphan GPCR	16
ZW	Smp_203500	G_PROTEIN_RECEP_F1_2 domain-containing protein	8
	Smp_332930	G_PROTEIN_RECEP_F2_4 domain-containing protein	12

Parmi les différents acteurs de la phototransduction, nous retrouvons, dans les gènes diurnes rythmiques : la rhodopsine-like, la phosphodiesterase, la GPCR kinase (sérine/thréonine kinase), la recoverine (domaine EF-hand), la protéine phosphatase 2A (PP2A), la GTPase activating protéine et la GTPase (Tableau 60). Dans la majorité des cas, les pics d'expression sont diurnes et se déroulent au moment du pic d'émission des cercaires.

Tableau 60 : Liste des gènes diurnes rythmiques liés à la phototransduction

Intérêt	Chr.	Identifiant (ID) du gène	Annotation GFF	Annotation HHsearch	Commun	H pic exp
Phosphodiesterase	5	Smp_334600	Phosphodiesterase			12
	ZW	Smp_134140				12
GPCR kinase (sérine/thréonine kinase)	1	Smp_009600	Serine/threonine-protein kinase PLK			4
	ZW	Smp_194610	Serine/threonine kinase		H3K4me3	8
recoverine (domaine EF-hand),	6	Smp_341900	EF-hand domain-containing protein			12
	3	Smp_106110	EF-hand domain	Kchip1, Kv4 potassium channel-interacting protein {Rat ( <i>Rattus norvegicus</i> ) [TaxId: 10116]}	Kchip1, Kv4 potassium channel-interacting protein {Rat ( <i>Rattus norvegicus</i> ) [TaxId: 10116]}	12
		Smp_181460	EF-hand domain pair	DUF4205	DUF4205	12
PP2A	ZW	Smp_172280	Serine/threonine protein phosphatase 2a regulatory subunit A%2C putative			8
	5	Smp_165490	Calcineurin-like phosphoesterase domain, ApaH type	Protein phosphatase 2A catalytic	Protein phosphatase 2A catalytic subunit	4

				subunit alpha isoform, PP2A-alpha {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}	alpha isoform, PP2A-alpha {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}	
GTPase activating protein	ZW	Smp_152370	Putative arf gtpase-activating protein			24
GTPase	4	Smp_313300	Small GTPase	Rab21 {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}	Rab21 {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}	8

#### II.2.3.2.2.3. Les gènes impliqués dans l'acétylation des histones

Parmi les gènes rythmiques diurnes, nous retrouvons le gène **Smp\_053140** qui est une histone acétyltransférase. Ce gène est marqué avec H3K4me3 et a son pic d'expression à 4h (la nuit).

#### II.2.3.2.2.4. Enrichissement GO : Processus biologique

Les analyses d'enrichissement GO n'ont pas donné de résultats significatifs en utilisant la limite usuelle FDR < 0,05 (False Discovery Rate). Les résultats qui suivent ont été obtenus en utilisant la limite de  $p < 0,05$ .

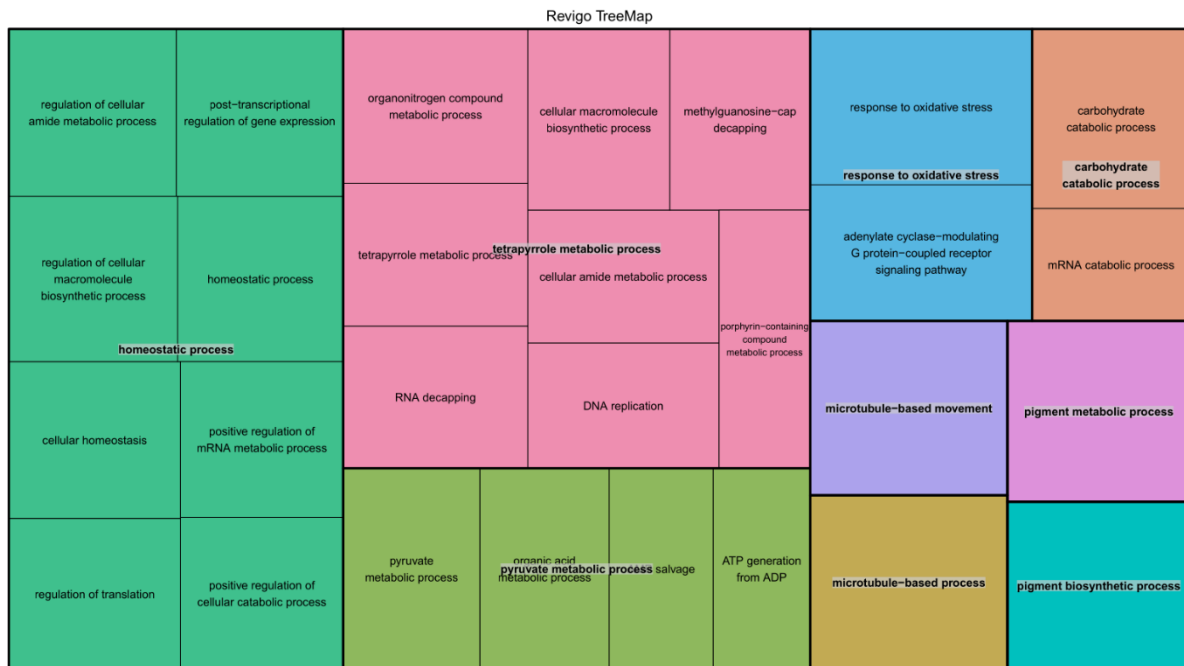
Les annotations processus métabolique/biosynthétique regroupent 3 gènes marqués avec H3K4me3 impliqués dans le processus de l'hème (Tableau\_61) (cf II.2.2.2.6 Enrichissement GO : Processus biologique).

**Tableau\_61 : Liste des gènes diurnes rythmiques issus des annotations GO sélectionnées**

GO	Chr.	Identifiant (ID) du gène	Annotation GFF	Commun	Heure du pic d'expression
Réponse à un stress oxydatif	1	Smp_342030	NA		12
	ZW	Smp_058690	Glutathione peroxidase		12
		Smp_058700			12
Processus homéostatique	2	Smp_047680	Ferritin		16
		Smp_047650			16
		Smp_046740	Dihydrolipoyl dehydrogenase		8
	4	Smp_043030	Hexokinase		8
	ZW	Smp_203970	Calcium load-activated calcium channel		20
Processus métabolique et biosynthétique du pigment	2	Smp_045260	5-aminolevulinate synthase	H3K4me3	8
	5	Smp_340920	NA		8
	ZW	Smp_095130	Putative porphobilinogen deaminase		20

## Partie II : Rythme-Transcriptomique

Parmi toutes les annotations GO (Figure\_81), la réponse au stress oxydatif et les processus homéostatiques sont intéressants. L'annotation réponse au stress oxydatif regroupe 3 gènes, dont deux Glutathione peroxidase. Il a été montré chez la souris que la glutathione peroxidase 4 (GPx4) était une enzyme antioxydante importante dans la maturation et la survie des cellules photoréceptrices (Ueta, Inoue et al., 2012).



**Figure\_81 : Treemap des résultats de l'analyse d'enrichissement GO des processus biologiques obtenus pour les gènes diurnes rythmiques.** Chaque couleur représente un grand ensemble de processus biologique. Il y a, de gauche à droite : Processus homéostatique ; Processus métabolique des tetrapyrroles ; Réponse à un stress oxydatif ; Processus catabolique des glucides ; Processus métabolique des pyruvates ; mouvement basé sur les microtubules ; Processus métabolique des pigments ; processus basé sur les microtubules et Processus biosynthétique des pigments.

L'annotation processus homéostatique contient deux gènes qui codent pour des ferritines. Ces protéines, qui permettent le stockage du fer, sont régulées par des cycles jour/nuit sous le contrôle de l'horloge circadienne chez *Ostreococcus tauri* ( Botebol, Lesuisse et al., 2015).

Dans la majorité des cas, le pic d'expression de ces gènes est diurne, soit avant le pic d'émission (8h) soit pendant le pic (12h).

### II.2.3.2.2.5. Kegg pathway

Les analyses Kegg pathway montrent que les gènes diurnes rythmiques sont potentiellement impliqués dans 308 voies moléculaires différentes. Parmi toutes ces voies, sept retiennent notre attention, le rythme circadien, l'entraînement circadien, la phototransduction, la phototransduction chez la drosophile, la transduction olfactive et la transduction gustative (Tableau\_62). La majorité de ces gènes ont des pics d'expressions diurnes, principalement avant le pic d'émission (8h).



Tableau 62 : Liste des gènes diurnes rythmiques issus des Kegg pathways sélectionnées

Kegg pathway	KO annotation	Chr.	Identifiant (ID) du gène	Annotation GFF	Commun	H pic exp
Rythme circadien	K08959	ZW	Smp_201220	Non-specific serine/threonine protein kinase		12
	K07199	1	Smp_341700	AMPKBI domain-containing protein	H3K4me3	12
Rythme circadien chez la drosophile	K09057	ZW	Smp_026400	Thyrotroph embryonic factor related		12
Entraînement circadien	K02183		Smp_033010*	Similar to 16 kDa calcium-binding protein		8
	K04345		Smp_307260	Protein kinase domain-containing protein		8
	K04630	1	Smp_020080	Putative gtp-binding protein (I) alpha-1 subunit, gna1	H3K4me3	12
2		Smp_016250	Putative gtp-binding protein (I) alpha-2 subunit, gna2	24		
Phototransduction	K02183	ZW	Smp_033010*	Similar to 16 kDa calcium-binding protein		8
Phototransduction chez la drosophile	K04967	1	Smp_344020	NA		12
	K02183	ZW	Smp_033010*	Similar to 16 kDa calcium-binding protein		8
Transduction olfactive	K04345		Smp_307260*	Protein kinase domain-containing protein		8
	K02183		Smp_033010*	Similar to 16 kDa calcium-binding protein		8
Transduction gustative	K04345		Smp_307260*	Protein kinase domain-containing protein		8

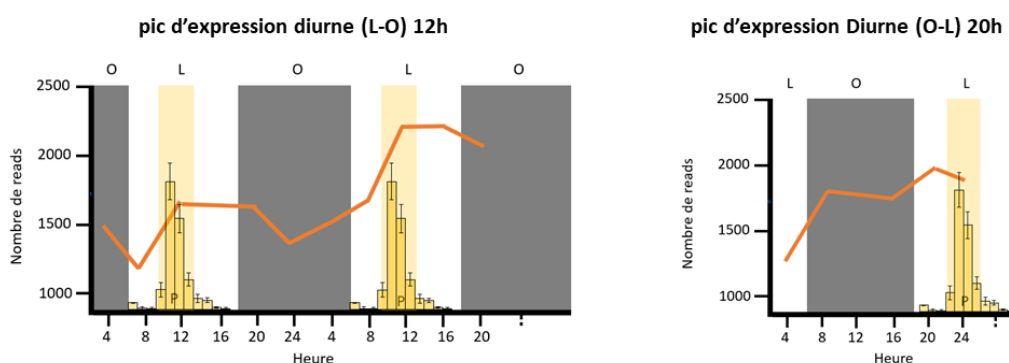
#### II.2.3.2.2.6. Les gènes rythmiques communs en photopériode classique et inversée

Parmi les 731 gènes diurnes rythmiques pour les échantillons en photopériode classique (L-O), 35 sont communs avec les gènes diurnes rythmiques pour les échantillons en photopériode inversée (O-L) (Tableau 63). Si la sélection subjective et l'annotation GO ne donnent pas matière à discuter, les analyses Kegg pathway révèlent que ces gènes sont potentiellement impliqués dans 67 pathways différents, dont le rythme circadien. Le pathway rythme circadien contient le gène **Smp\_201220**. Ce gène a des pics d'expression opposés entre les échantillons en photopériode classique (12h) et en condition photopériode inversée (20h) (Figure 82). Ces résultats font de ce gène un candidat très sérieux pour expliquer le rythme d'émission des cercaires.

Tableau\_63 : Liste des gènes diurnes rythmiques communs aux échantillons photopériode classique (L-O) et photopériode inversée (O-L). Les gènes aux pics d'expression diamétralement opposés sont marqués en rouge.

Chr.	Gene ID	GFF annotation	HHsearch annotation	Pic expression Diurne L-O (h)	Pic expression Diurne O-L (h)
SM_V7_1	Smp_073800	RecQ-mediated genome instability protein 1		12	8
SM_V7_1	Smp_246020	NA		16	20
SM_V7_3	Smp_167630	Solute carrier family%2C putative		16	16
SM_V7_3	Smp_326660	NA		8	12
<b>SM_V7_5</b>	<b>Smp_006440</b>	<b>Heterogeneous nuclear ribonucleoprotein I (Hnrnp I)%2C putative</b>		<b>8</b>	<b>20</b>
SM_V7_5	Smp_061570	Mediator of RNA polymerase II transcription subunit 19		12	12
<b>SM_V7_5</b>	<b>Smp_315750</b>	<b>Ubiquitin-conjugating enzyme E2</b>	<b>Ubiquitin-conjugating enzyme E2</b>	<b>12</b>	<b>24</b>
SM_V7_6	Smp_074330	PEST proteolytic signal-containing nuclear protein		12	12
SM_V7_7	Smp_152940	Putative otopetrin		12	16
SM_V7_ZW	Smp_006340	ER membrane protein complex subunit 4		20	0
<b>SM_V7_ZW</b>	<b>Smp_019820</b>	<b>NA</b>		<b>12</b>	<b>24</b>
SM_V7_ZW	Smp_022500	NA		20	16
SM_V7_ZW	Smp_037700	Nuclear transport factor 2-like protein%3B Nuclear transport factor%2C putative		20	16
SM_V7_ZW	Smp_046980	Pre-mRNA-splicing factor 38		20	20
<b>SM_V7_ZW</b>	<b>Smp_060210</b>	<b>Putative lipopolysaccharide-induced transcription factor regulating tumor necrosis factor alpha</b>		<b>8</b>	<b>16</b>
<b>SM_V7_ZW</b>	<b>Smp_091170</b>	<b>Ribosomal protein</b>		<b>12</b>	<b>24</b>
SM_V7_ZW	Smp_091840	Bestrophin homolog		20	16
<b>SM_V7_ZW</b>	<b>Smp_102240</b>	<b>Upf3 regulator of nonsense transcripts-like protein</b>		<b>12</b>	<b>24</b>

SM_V7_ZW	Smp_120380	Testis specific leucine rich repeat protein%2Cputative		12	8
<b>SM_V7_ZW</b>	<b>Smp_121130</b>	<b>Mitochondrial carrier protein%2C putative</b>		<b>16</b>	<b>24</b>
<b>SM_V7_ZW</b>	<b>Smp_128650</b>	<b>Nuclear movement protein related</b>		<b>12</b>	<b>20</b>
SM_V7_ZW	Smp_143730	Putative carboxypeptidase regulatory region-containing		24	4
SM_V7_ZW	Smp_165470	Receptor protein-tyrosine kinase		16	12
SM_V7_ZW	Smp_166500	Phosphatidylcholine-sterol acyltransferase (Lecithin-cholesterol acyltransferase)/ Phospholipase A		12	12
SM_V7_ZW	Smp_177790	Iron-sulfur cluster scaffold protein nfu-related		12	16
<b>SM_V7_ZW</b>	<b>Smp_201220</b>	<b>Non-specific serine/threonine protein kinase</b>		<b>12</b>	<b>20</b>
SM_V7_ZW	Smp_204380	NA		16	12
SM_V7_ZW	Smp_245400	NA		12	16
<b>SM_V7_ZW</b>	<b>Smp_315980</b>	<b>NA</b>		<b>12</b>	<b>20</b>
SM_V7_ZW	Smp_322240	NA		12	16
SM_V7_ZW	Smp_323230	NA		8	12
SM_V7_ZW	Smp_343530	NA		8	8
SM_V7_ZW	Smp_343890	NA		20	20
<b>SM_V7_ZW</b>	<b>Smp_346810</b>	<b>NA</b>		<b>12</b>	<b>24</b>
SM_V7_ZWH005	Smp_330950	NA		8	4



Smp\_201220 (Non-specific serine/threonine protein kinase) : pic d'expression Diurne (L-O) et Diurne photopériode inversée (O-L) opposés

Figure 82 : Exemple de profils d'expression de gène commun aux chronotypes Diurnes L-O et O-L (pic d'expression inversé). Chaque histogramme a, en abscisses, les heures (points temporels) et en ordonnées le nombre de reads pour le gène sélectionné. En fond, se trouve le profil d'émission des cercaires. Les rectangles gris symbolisent l'obscurité (O), les blancs la lumière (L), les rectangles jaunes la période du pic d'émission des cercaires chronotype diurne.

## Partie II : Rythme-Transcriptomique

Chez le diurne (L-O), la majorité des pics d'expression des gènes communs a lieu à 12h (Tableau\_64), soit pendant le pic d'émission des cercaires. Chez le diurne (O-L), la majorité des pics d'expressions à lieu à 16h (Tableau\_63), soit pendant la nuit, 8h avant le pic d'émission des cercaires (O-L).

**Tableau 64 : Répartition des pics d'expression des gènes rythmiques aux deux photopériodes du chronotype diurne**

Heures	Nombre des pics d'expression des gènes rythmiques Diurnes (L-O)	Nombre des pics d'expression des gènes rythmiques Diurnes (O-L)
4	0	2
8	6	3
12	17	7
16	5	9
20	6	7
24	1	7
Total des pics d'expression diurne (L-O : 8h, 12h, 16h ; O-L : 4h, 20h et 24h)	28	16
Total des pics d'expression nocturne (L-O : 4h, 20h, 24h ; O-L : 8h, 12h et 16h)	7	19

Si on s'intéresse à l'heure des pics d'expression entre les conditions de photopériode, on observe différents types de comportement. Il y a tout d'abord les gènes qui ont le même type d'expression entre les deux photopériodes. Ce résultat suggère que l'expression de ces gènes est indépendante de la photopériode. Le comportement inverse est aussi observé, c'est-à-dire des expressions qui se sont inversé avec la modification de la photopériode. Ces gènes représentent des candidats sérieux car leur expression est dépendante des variations de photopériode.

### II.2.3.2.3. Les gènes nocturnes

Comme précédemment, une sélection subjective basée sur la recherche de gènes de l'horloge, de la phototransduction et de l'acétylation et la méthylation des histones a été faite. L'enrichissement GO a été fait avec Blast2GO version 5.2 en utilisant le Fisher's Exact Test ( $p < 0,05$ ). L'analyse d'enrichissement Kegg pathway a été faite sur le serveur KAAS en utilisant le programme Blast avec la méthode SBH (Single-directional Best Hit). La liste complète des gènes nocturnes rythmiques est présentée dans l'annexe 11.

#### II.2.3.2.3.1. Les gènes de l'horloge (Basic Helix-Loop-Helix)

Il n'y a aucun gène canonique de l'horloge. On retrouve toutefois deux gènes avec un domaine BHLH issus des QTL interspécifiques (Tableau\_65) qui ont leur pic d'expression à 4h (la nuit).

Tableau 65 : Liste des gènes nocturnes rythmiques BHLH

Intérêt	Chr.	Identifiant (ID) du gène	Annotation GFF	Annotation HHsearch	Commun	Heure du pic d'expression
Domaine BHLH	3	Smp_021340	Myc-type, basic helix-loop-helix (bHLH) domain	Mad protein {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}	QTL interspécifique et H3K4me3	4
		Smp_000620		DNA-binding protein inhibitor ID-1	QTL interspécifique	4

## II.2.3.2.3.2. Les rhodopsines et autres gènes impliqués dans la phototransduction

Il y a quatre gènes rhodopsine-like tous identifiés lors des QTLs intra et interspécifique (Tableau 66). Trois d'entre eux ont leur pic d'expression à 4h (la nuit) et le quatrième a son pic à 12h (jour).

Tableau 66 : Liste des gènes nocturnes rythmiques rhodopsine-like

Intérêt	Chr.	Identifiant (ID) du gène	Annotation GFF	Annotation HHsearch	Commun	Hr pic exp
Rhodopsine-like	1	Smp_180140	G protein-coupled receptor, rhodopsin-like	P2Y purinoceptor 1, Rubredoxin	QTL intraspéc.	12
		Smp_140250		Platelet-activating factor receptor, Flavodoxin		4
	3	Smp_132410		Lysozyme, Proteinase-activated receptor 2, Soluble cytochrome b562, Proteinase-activated	QTL interspéc.	4
	4	Smp_145540		Muscarinic acetylcholine receptor M2, Vasopressin V2	QTL interspéc. et Kegg pathway transduction gustative	4

Partie II : Rythme-Transcriptomique

Parmi les différents acteurs de la phototransduction, nous retrouvons, dans les gènes nocturnes rythmiques : les rhodopsine-like, la GPCR kinase (sérine/thréonine kinase), l'arrestine, la recoverine (domaine EF-hand), la GTPase activating protein et la guanylate cyclase (Tableau\_67). Dans la majorité des gènes, le pic d'expression est à 4h.

Tableau 67 : Liste des gènes nocturnes rythmiques liés à la phototransduction

Intérêt	Chr.	Identifiant (ID) du gène	Annotation GFF	Annotation HHsearch	Commun	H Pic exp
GPCR kinase (sérine/thréonine kinase)	1	Smp_210820	Protein-serine/threonine kinase			4
	2	Smp_310770	Serine/threonine kinase		H3K4me3	8
	1	Smp_097940				4
	3	Smp_134910				4
	7	Smp_057360				4
	ZW	Smp_094190				4
	ZW	Smp_156990				4
Arrestine	3	Smp_176980		Putative beta-arrestin 1		
Recoverine (domaine EF-hand)	1	Smp_171010	Putative ef-hand domain (C-terminal) containing protein		H3K4me3	16
	4	Smp_313310	EF-hand domain	Sarcoplasmic calcium-binding protein {Amphioxus ( <i>Branchiostoma lanceolatum</i> ) [TaxId: 7740]}	QTL interspé.	12
	7	Smp_318260	EF-hand domain-containing protein			4
GTPase activating protein	1	Smp_166310	Putative ran gtpase-activating protein			20
		Smp_081190	Putative rho3 GTPase%3B Rho3 GTPase			4
	2	Smp_154850	Putative rap gtpase-activating protein			4
	ZW	Smp_164680	Synaptic ras gtpase activating protein%2C syngap%2Cputative			4
	1	Smp_319130	FCH domain	SLIT-ROBO Rho GTPase-activating protein 2	QTL intraspé. et H3K4me3	4

		Smp_094890	Rho GTPase-activating protein domain	Beta-chimaerin, C-terminal domain {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}	QTL intraspé. et H3K4me3	4
Guanylyl cyclase	7	Smp_140040	Guanylate cyclase domain-containing protein			4

#### II.2.3.2.3.3. Les gènes impliqués dans l'acétylation des histones

Contrairement au chronotype diurne, il n'y a pas de gènes rythmiques impliqués dans l'acétylation ou la méthylation des histones pour le chronotype nocturne. Toutefois, il est intéressant de noter la présence de deux gènes rythmiques de type myc dont un potentiel gène Mad qui recrute des histones désacétylases (Smp\_021340 et Smp\_000620).

#### II.2.3.2.3.4. Enrichissement GO : Processus biologique

Les analyses d'enrichissement GO des gènes détectés comme rythmique dans le chronotype nocturne n'ont pas donné de résultats significatifs en utilisant la limite usuelle FDR < 0,05 (False Discovery Rate). Les résultats qui suivent ont été obtenus en utilisant la limite de  $p < 0,05$ .

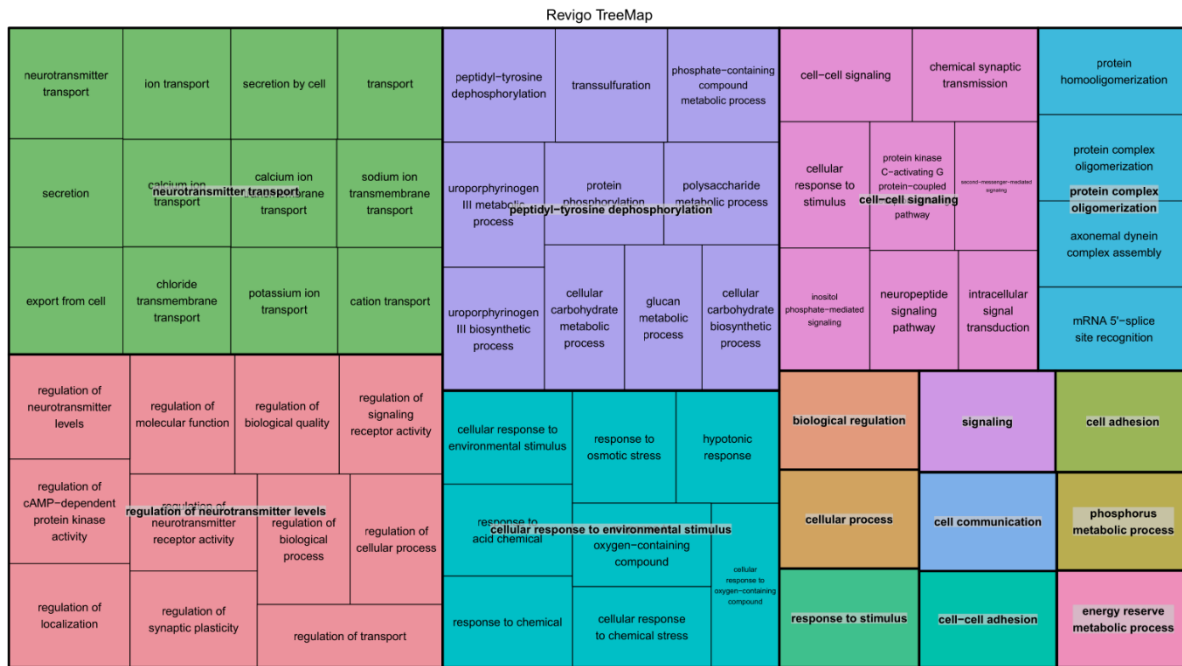
Parmi les gènes nocturnes rythmiques, on retrouve deux gènes marqués avec H3K4me3 impliqués dans le métabolisme de l'hème (Tableau\_68).

**Tableau 68 : Liste des gènes nocturnes rythmiques issus des annotations GO sélectionnées**

GO	Chr.	Identifiant (ID) du gène	Annotation GFF	Commun	Heure du pic d'expression
Réponse à un stimulus environnemental	ZW	Smp_244250	NA		4
Régulation des niveaux de neurotransmetteur	2	Smp_316380	NA		4
	4	Smp_126990	Putative synapsin		4
	ZW	Smp_175900	Synaptotagmin		4
Processus métabolique et biosynthétique du pigment	2	Smp_045260	5-aminolevulinate synthase	H3K4me3	4
	SM_V7_3	Smp_079840	Uroporphyrinogen -III synthase		4

Parmi toutes les annotations GO (Figure\_83), la réponse cellulaire à un stimulus environnemental et la régulation des niveaux de neurotransmetteurs sont intéressantes. L'annotation « réponse cellulaire à un stimulus environnemental » correspond à un gène non identifié. L'annotation « régulation des niveaux de neurotransmetteur » regroupe 3 gènes, avec, parmi eux, une synaptotagmine. Il a été montré que la synaptotagmine régule la transmission synaptique dépendant des ions Ca<sup>2+</sup> (Grassmeyer, Cahill et al., 2019). Tous les gènes de ce point ont leur pic d'expression à 4h (la nuit).

## Partie II : Rythme-Transcriptomique



**Figure 83 : Treemap des résultats de l'analyse d'enrichissement GO des processus biologiques obtenus pour les gènes nocturnes rythmiques.** Chaque couleur représente un grand ensemble de processus biologiques. Il y a, de gauche à droite : Transport des neurotransmetteurs ; déphosphorylation peptidyl-tyrosine ; signalisation cellule à cellule ; Oligomérisation du complexe protéique ; Régulation du niveau des neurotransmetteurs ; Réponse cellulaire à un stimulus environnemental ; Régulation biologique ; Signalisation ; Adhésion cellulaire ; Processus cellulaire ; Communication cellulaire ; processus métabolique du phosphore ; Réponse aux stimulus ; Adhésion cellule à cellule ; Processus métabolique des réserves d'énergie.

### II.2.3.2.3.5. Kegg pathway

Les analyses Kegg pathway montrent que les gènes nocturnes rythmiques sont potentiellement impliqués dans 220 voies moléculaires différentes. Parmi toutes ces voies, quatre retiennent notre attention, l'entraînement circadien, la phototransduction chez la drosophile, la transduction olfactive et la transduction gustative (Tableau\_69). On retrouve des gènes déjà identifiés avec les QTL intra et interspécifique. La majorité des pics d'expression a lieu à 4h, la nuit, et le reste a lieu à 12h.

**Tableau 69 : Liste des gènes diurnes rythmiques issus des Kegg pathways sélectionnés.** Les gènes marqués avec une étoile (\*) sont présents dans plusieurs pathways

Kegg pathway	KO annotation	Chr.	Identifiant (ID) du gène	Annotation GFF	Annotation HHsearch	Commun	H
Entraînement circadien	K05208	ZW	Smp_245310	NA			1 2
	K04962	1	Smp_163570	Ryanodine receptor related		H3K4me3	4
	K04851	3	Smp_159990	Voltage-dependent L-type calcium channel subunit alpha			4



	K04958	4	Smp_336900 *	Inositol 1,4,5-trisphosphate receptor			4
	K02677	ZW	Smp_241420 *	AGC-kinase C-terminal domain-containing protein			4
Phototransduction chez la drosophile	K04958	4	Smp_336900 *	Inositol 1,4,5-trisphosphate receptor			4
	K02677	ZW	Smp_241420 *	AGC-kinase C-terminal domain-containing protein			4
	K00910	1	Smp_210820 *	GPCR kinase	Beta-adrenergic receptor kinase 1 (E.C.2.7.11.15)	QTL intraspécifique et H3K4me3	4
Transduction olfactive	K04439	3	Smp_176980	Putative beta-arrestin 1			1 2
	K05849	1	Smp_152830	Na-Ca exchanger/integrin-beta4	Na/Ca exchange protein		4
	K00910	1	Smp_210820 *	GPCR kinase	Beta-adrenergic receptor kinase 1 (E.C.2.7.11.15)	QTL intraspécifique et H3K4me3	4
Transduction gustative	K04131	4	Smp_145540	G protein-coupled receptor, rhodopsin-like	Muscarinic acetylcholine receptor M2, Vasopressin V2	QTL interspécifique	4

#### II.2.3.2.3.6. Les gènes rythmiques communs en photopériode classique et inversée

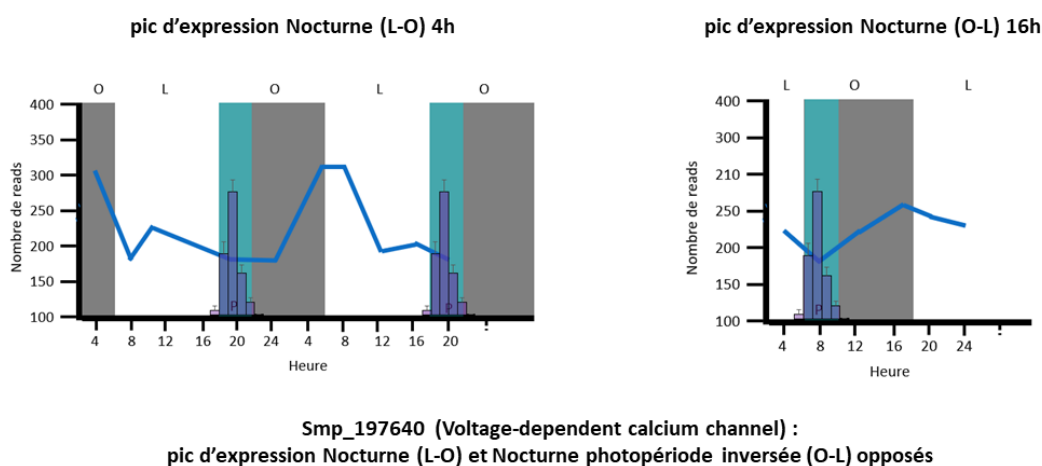
Parmi les 439 gènes nocturnes rythmiques pour les échantillons en photopériode classique (L-O), 7 sont communs avec les gènes nocturnes rythmiques pour les échantillons en photopériode inversée (O-L) (Tableau\_70). Parmi ces 7 gènes, on retrouve deux gènes issus des QTLs intraspécifiques qui codent pour des rhodopsine-like. Dans le cas du gène **Smp\_140250**, le pic d'expression en photopériode classique est à 4h (nuit) tandis qu'il est à 20h (jour) en condition photopériode inversée. On retrouve aussi un gène sérine/thréonine kinase (Smp\_310770) qui ne change pas d'heure de pic d'expression entre les deux conditions. Cela peut signifier que l'expression de ce gène n'est pas influencée par la photopériode. Ce cas de figure arrive pour deux autres gènes (Tableau\_70).

Le gène **Smp\_197640**, qui code pour un canal calcique voltage dépendant, a son rythme d'expression totalement inversé entre les deux conditions de photopériode (Tableau\_70 et Figure\_84). Cela implique que les mécanismes Ca<sup>2+</sup> dépendants sont importants dans le rythme d'émission des cercaires.

Les analyses Kegg pathways révèlent que ces gènes sont potentiellement impliqués dans 11 pathways différents. Aucun pathway rythme/entraînement circadien ou transduction de signaux environnementaux ne ressort de ces analyses. Toutefois le gène **Smp\_180140** (K04144) apparaît dans les pathways concernant la voie de signalisation du calcium, la voie de signalisation de l'AMPC, les synapses dopaminergiques et même le pathway de l'addiction à la cocaïne.

**Tableau 70 : Liste des gènes nocturnes rythmiques communs aux échantillons photopériode classique (L-O) et photopériode inversée (O-L).** Les gènes aux pics d'expression diamétralement opposés sont marqués en rouge.

Chr	Gene ID	GFF annotation	HHsearch annotation	Pic expression Nocturne L-O (h)	Pic expression Nocturne O-L (h)
SM_V7_1	Smp_140250	G protein-coupled receptor, rhodopsin-like	Platelet-activating factor receptor, Flavodoxin	4	20
SM_V7_3	Smp_162660	Glutamine-dependent NAD(+) synthetase		12	20
SM_V7_1	Smp_163430	C2H2-type domain-containing protein		24	8
SM_V7_1	Smp_180140	G protein-coupled receptor, rhodopsin-like	P2Y purinoceptor 1, Rubredoxin	12	16
<b>SM_V7_3</b>	<b>Smp_197640</b>	<b>Voltage-dependent calcium channel</b>		<b>4</b>	<b>16</b>
SM_V7_ZW	Smp_212240	NA		16	16
SM_V7_2	Smp_310770	Serine/threonine kinase		8	8



**Figure 84 : Exemple de profils d'expression de gène commun aux chronotypes Nocturnes L-O et O-L (pic d'expression inversé).** Chaque histogramme a, en abscisses, les heures (points temporels) et en ordonnées le nombre de reads pour le gène sélectionné. En fond, se trouve le profil d'émission des cercaires. Les rectangles gris symbolisent l'obscurité (O), les blancs la lumière (L), les rectangles bleus la période du pic d'émission des cercaires chronotype nocturne.

Chez le nocturne (L-O), la majorité des pics d'expression des gènes communs a lieu à 4h (nuit) et 12h (jour) (Tableau\_71), soit 16h et 8h avant le pic d'émission des cercaires. Chez le nocturne (O-L), la majorité des pics d'expression a lieu à 16h (Tableau\_70) la nuit, soit 16h avant le pic d'émission des cercaires (O-L).

**Tableau 71 : Répartition des pics d'expression des gènes rythmiques aux deux photopériodes du chronotype nocturne**

Heures	Nombre des pics d'expression des gènes rythmiques Nocturnes (L-O)	Nombre des pics d'expression des gènes rythmiques Nocturnes (O-L)
4	2	2
8	1	3
12	2	7
16	1	9
20	0	7
0	1	7
Total des pics d'expression diurne (L-O : 8h, 12h, 16h ; O-L : 4h, 20h et 24h)	4	2
Total des pics d'expression nocturne (L-O : 4h, 20h, 24h ; O-L : 8h, 12h et 16h)	3	5

Tout comme le chronotype diurne, on observe différents types de comportement dans l'expression des gènes entre les deux photopériodes. Il y a tout d'abord les gènes qui ont le même type d'expression entre les deux photopériodes. Ce résultat suggère que l'expression de ces gènes est indépendante de la photopériode. Le comportement inverse est aussi observé, c'est-à-dire des expressions qui se sont inversé avec la modification de la photopériode. Ces gènes représentent des candidats sérieux car leur expression est dépendante des variations de photopériode.

## II.2.3.3. Comparaison des chronotypes

La majorité des gènes rythmiques sont propres à chaque chronotype. Il est intéressant de noter que chez le chronotype diurne, la majorité des pics d'expression sont diurnes avec le maximum à 8h soit quatre heures avant le pic d'émission des cercaires (Tableau\_72). A l'inverse, chez le chronotype nocturne, la majorité des pics d'expression sont nocturnes, avec une concentration de pic à 4h (la nuit), soit 16h avant l'heure du pic (Tableau\_72).

**Tableau 72 : Répartition des pics d'expression des gènes rythmiques des deux chronotypes en photopériode classique (L-O)**

Heures	Nombre des pics d'expression des gènes rythmiques Diurne (N=703)	Nombre des pics d'expression des gènes rythmiques Nocturnes (N=411)
4	66	316
8	246	7
12	186	37
16	64	10
20	95	5
0	46	36
Total des pics d'expression diurne (8h, 12h et 16h)	496	54
Total des pics d'expression nocturne (4h, 20h et 24h )	207	357

Il est intéressant de noter que même si les gènes rythmiques présents dans les deux chronotypes sont différents, on retrouve des fonctions communes avec notamment la présence de rhodopsine-like et d'acteur de la phototransduction (Figure\_85).

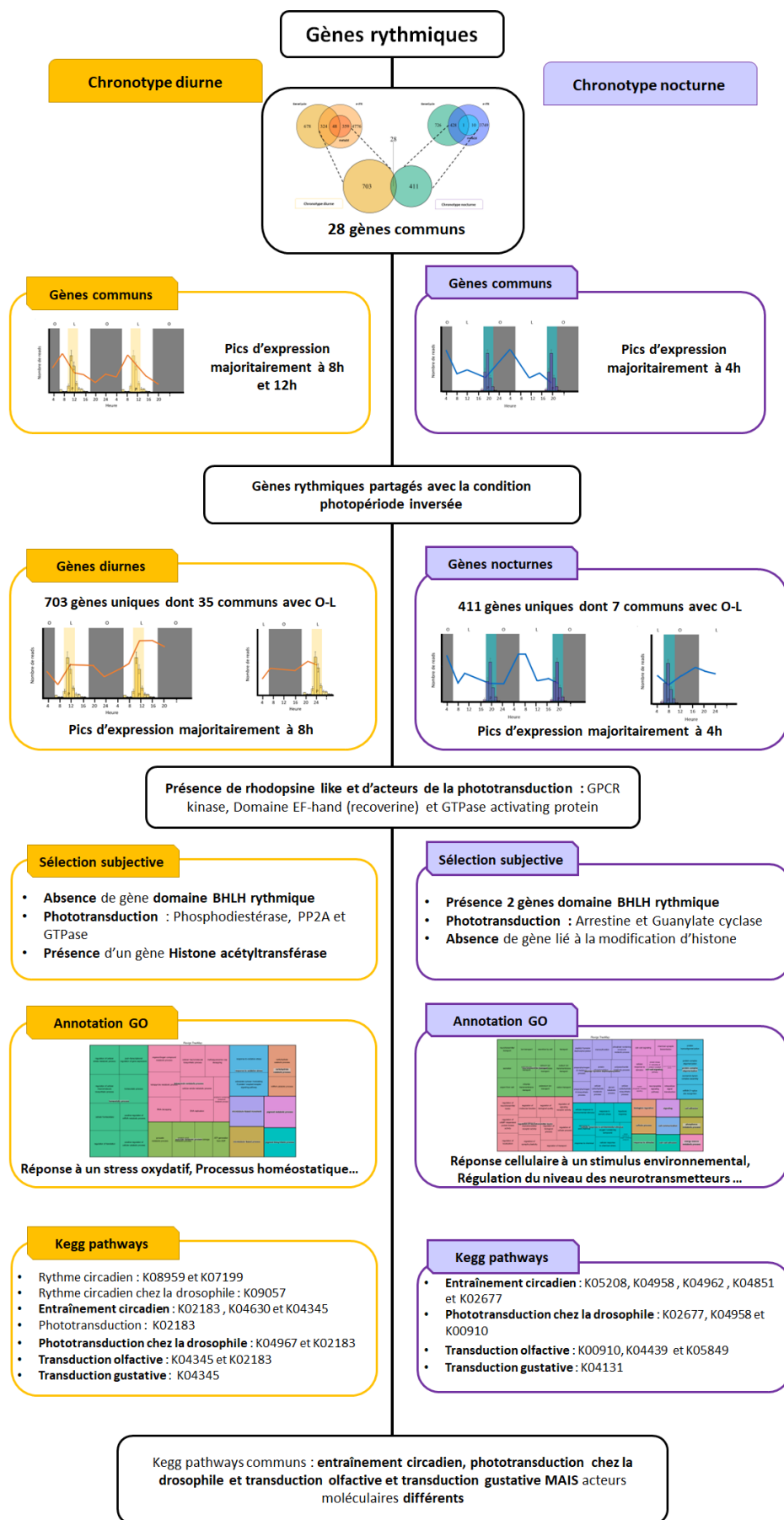


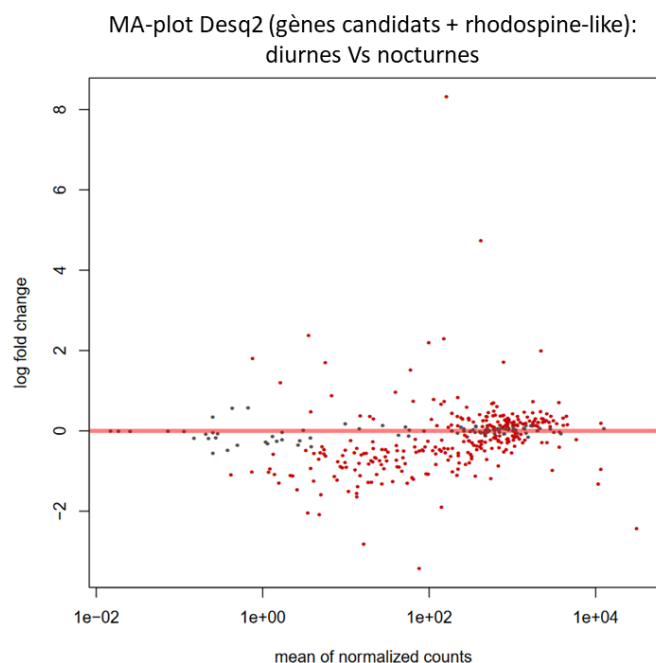
Figure 85 : Bilan des gènes rythmiques.

## II.2.3.4. Intégration des résultats de génétique et épigénétique

Les approches de génétique, épigénétique et transcriptomique ont permis de mettre en lumière 353 gènes candidats. Contrairement à l'approche transcriptomique, les approches génétique et épigénétique ne donnent pas accès à l'expression des gènes sélectionnés. Pour affiner cette liste de 353 gènes, une vérification du comptage a été faite. Ainsi tous les gènes qui ont une moyenne inférieure à 10 reads pour tous les échantillons des deux chronotypes sont considérés comme non exprimés. Ce filtrage des gènes candidats retire 3 gènes à la liste. Le premier gène est Smp\_124470, une adénylate/guanylate cyclase qui est issue du QTL intraspécifique. Les deux autres gènes sont Smp\_083940 et Smp\_134960, deux rhodopsine-like, issus du QTL interspécifique. Ce résultat intéressant pourrait signifier que si aucun des deux chronotypes de *S. mansoni* n'exprime significativement ces gènes, alors ces deux rhodopsine-like sont spécifiques de *S. rodhaini*. Ces résultats pourraient aussi aller dans le sens de l'hypothèse que le chronotype des schistosomes dépend des rhodopsines/rhodopsine-like exprimées dans le parasite.

Pour vérifier cette hypothèse de spécificité de rhodopsine et plus largement des gènes candidats sélectionnés, une analyse différentielle DEseq2 (Galaxy Version 2.11.40.2) (Love, Huber & Anders, 2014) a été menée entre tous les échantillons des deux chronotypes (sans distinction de temps). Par convention, les gènes différentiellement exprimés sont ceux qui ont une  $p$ -ajustée  $< 0,01$  et un  $\log_2(\text{foldchange})$  ( $\log_2\text{FC}$ ) supérieur à +1 (le gène surexprimé chez le chronotype diurne) ou inférieur à -1 (le gène surexprimé chez le chronotype nocturne). Comme le rythme d'émission des cercaires ne dépend pas phénotypiquement du sexe du parasite, un gène surreprésenté est un gène qui répond aux conditions citées précédemment lors d'une analyse DEseq2 entre les chronotypes, sexes confondus et sexes séparés (mâles diurnes Vs mâles nocturnes et femelles diurnes Vs femelles nocturnes). En plus des 353 gènes sélectionnés par nos analyses, nous avons ajouté 28 gènes rapportés comme rhodopsine-like (opsine-like et photolyase) dans la littérature ou dans le génome de *S. mansoni* qui n'ont pas été sélectionnés au cours de ces travaux (Cosseau, Azzi et al., 2009 ; Hoffman, Davis et al., 2001 ; Liang, Zhao et al., 2016 ; Phan, Liang et al., 2022) (Annexe\_12).

La comparaison des chronotypes révèle que les gènes candidats sont plus exprimés dans le chronotype nocturne (Figure\_86 et Tableau\_73). Ce résultat peut s'expliquer en partie par le fait que les analyses QTL ont été faites sur le phénotype pourcentage de cercaires nocturne.



**Figure 86 :** MA-plots issus des analyses DEseq2 diurnes Vs nocturnes pour les gènes candidats et les 28 « rhodopsine-like ». Les abscisses représentent le comptage des moyennes normalisées et les ordonnées représentent le  $\log(\text{foldchange})$ . Chaque point représente un pic. Tous les points ne sont pas différents entre les chronotypes. Les points rouges sont différentiellement présents entre les deux chronotypes. Dans ces MA-plots, les  $\log(\text{foldchange})$  positifs sont surreprésentés chez les diurnes et les  $\log(\text{foldchange})$  négatifs sont surreprésentés chez les nocturnes.

Parmi les 24 gènes surexprimés chez les nocturnes, nous retrouvons 10 rhodopsine-like contre une seule pour les 9 gènes surexprimés chez les diurnes (Annexe 13). L'expression plus forte de rhodopsine implique une sensibilité plus forte à la lumière pour le chronotype nocturne. Il est intéressant de voir que, sur les 10 rhodopsines surexprimées chez le nocturne, seules deux ont été détectées comme rythmiques. Parmi ces deux, une est rythmique uniquement chez les nocturnes (**Smp\_145540**) et l'autre uniquement chez les diurnes (**Smp\_203500**). Le gène **Smp\_203500** n'est pas le seul à être surexprimé chez le nocturne et à avoir été détecté comme rythmique uniquement chez les diurnes (Annexe 13).

**Tableau 73 : Tableau récapitulatif des analyses DEseq2 entre les chronotypes diurne et nocturne de *S. mansoni* pour les gènes candidats sélectionnés**

H3K4me3	Nombre de gènes surreprésentés chez les diurnes	Nombre de gènes surreprésentés chez les nocturnes
Deseq2 Diurne D Vs Nocturne N	10	38
Deseq2 Mâles D Vs Mâles N	10	42
Deseq2 Femelles D Vs Femelles N	10	30
Pics communs aux trois Deseq2	9	24

Chez le chronotype diurne, il est intéressant de voir qu'il y a deux acteurs de la phototransduction que l'on retrouve aussi dans les gènes surexprimés chez les nocturnes (recoverine (EF-hand domain) et PP2A). Ces deux protéines sont directement impliquées dans la régulation de l'état de la rhodopsine pour enclencher ou non un nouveau cycle de phototransduction. En fonction de la concentration en Ca<sup>2+</sup>, la recoverine inhibe la phosphorylation de la rhodopsine activée par la lumière en retenant captive la rhodopsine-kinase. De son côté, la PP2A, déphosphoryle la rhodopsine ce qui lui permet d'être à nouveau disponible pour interagir de nouveau avec la lumière et enclencher une nouvelle cascade.

L'ensemble de ces résultats suggèrent que le rythme d'émission des cercaires des chronotypes diurne et nocturne de *S. mansoni* reposent sur un même principe de phototransduction basé sur la rhodopsine et que le nombre de rhodopsine/rhodopsine-like exprimée ainsi que les régulateurs de la phototransduction impliqués aient un rôle majeur dans la différence de phénotype.

---

## Discussion et conclusion générales

L'objectif principal de cette thèse est d'approfondir la connaissance de la transmission de *Schistosoma mansoni* du mollusque aux mammifères à travers le trait de vie du rythme d'émission des cercaires. L'étude de ce phénomène s'étend du phénotype, au travers d'approches de chronobiologie et d'histologie, jusqu'aux gènes impliqués, au travers d'approches de biologie moléculaire (génétique, épigénétique et transcriptomique). Les modèles biologiques principaux de ces recherches sont deux chronotypes opposés d'émission des cercaires (diurne et nocturne) issus de la même espèce de schistosomes, *S. mansoni*, tous deux originaires de la même région du Dhofar à Oman. La situation sanitaire de cette région peut être comparée à la Corse en termes de faible endémie et de prise en charge rapide de la bilharziose.

La première partie de cette thèse a été consacrée à la description des phénotypes des émissions journalières des deux chronotypes diurne et nocturne de *S. mansoni*. Cette étape a été nécessaire pour établir une base solide de phénotype à utiliser dans les études de biologie moléculaire. Elle s'axe sur deux grands objectifs qui sont de déterminer quelles sont les caractéristiques du phénotype du rythme d'émission des cercaires chez *S. mansoni* et si ce rythme d'émission des cercaires est circadien ? Dans la littérature qui traite de ces problématiques, aucun consensus n'a été établi en termes de protocole et de description des résultats. Ainsi chaque article relève de sa propre méthode et analyse empirique rendant ainsi la possibilité de faire des comparaisons fines entre les expériences difficile. C'est donc dans cette optique qu'il a été primordial de faire un protocole avec un comptage de cercaires toutes les heures sur une durée minimale de 48h. Et également d'établir des paramètres de description précis et quantitatifs pour donner une dimension « statistique » aux observations et ainsi permettre aux non-initiés de l'émission des cercaires de comprendre les analyses.

Pour le premier point de cette partie, sept paramètres ont ainsi été définis, à savoir, le nombre de pic, l'heure du pic, la plage horaire d'émission, l'intensité de la plage horaire d'émission, le coefficient d'asymétrie (skewness), le coefficient d'aplatissement (Kurtosis) et le pourcentage de cercaires nocturne. Un huitième paramètre, qui est le nombre total de cercaires émises sur 24h, n'a pas été pris en compte dans la notion de « rythme d'émission » des cercaires. Il est toutefois un paramètre capital dans la notion de « production cercarienne quantitative » notamment en termes de transmission et d'infectivité du parasite. Les diverses expériences de chronobiologie menées dans cette thèse en condition de photopériode équilibrée L-O (12-12) ont montré que l'espèce de l'hôte et le sexe du parasite n'avaient pas d'influence sur le rythme d'émission des cercaires. Elles ont aussi démontré que les profils journaliers d'émission des cercaires des chronotypes diurne et nocturne de *S. mansoni* se distinguent par trois paramètres qui sont l'heure du pic, l'intensité de la plage d'émission et le pourcentage de cercaires nocturne, illustrant ainsi l'unicité des deux chronotypes. Ces distinctions viennent s'ajouter à celles mises en évidence par Mouahid, Idris *et al.*, 2012 et Mouahid, Mintsa Nguema *et al.*, 2019.

Le deuxième point soulevé dans cette première partie de thèse est le questionnement sur le caractère circadien du rythme d'émission des cercaires. Pour répondre à cette interrogation, le rythme doit répondre à trois caractéristiques, l'entraînabilité du rythme, le maintien du rythme en condition environnementale constante et la compensation du rythme au changement de température.

L'entraînabilité du rythme par le zeitgeber revient à se questionner sur la capacité du rythme à s'adapter à une inversion de photopériode. Les travaux de cette thèse montrent que, peu importe les chronotypes, le rythme d'émission des cercaires s'adapte à l'inversion de photopériode. Dans le cas du chronotype diurne, cette adaptation est immédiate. Ces résultats sont en accord avec les expériences menées sur *S. mansoni* diurne originaire de Porto Rico (Luttermoser, 1955 ; Asch, 1972 ; and Glaudel & Edges, 1973) même si les expériences ne peuvent pas être finement comparées. Dans le cas du chronotype nocturne, l'adaptation passe par une étape intermédiaire d'émission bimodale avant de s'adapter totalement à l'inversion. Aucune expérience dans la littérature n'a été menée sur des schistosomes avec un chronotype nocturne pour avoir un point de comparaison.



Le maintien du rythme en condition environnementale constante est, dans notre cas, la capacité à maintenir le rythme d'émission des cercaires en lumière ou obscurité continue (L-L (12-12) ou O-O (12-12)). Cette question a été traitée pour les deux chronotypes, en condition de lumière continue et obscurité continue dans des expériences indépendantes. Pour éviter une redondance face à la similarité des résultats entre les deux conditions (L-L (12-12) ou O-O (12-12)), seule la lumière continue pour le chronotype diurne et l'obscurité continue pour le chronotype nocturne ont été présentés dans cette thèse. Que ce soit pour le chronotype diurne ou nocturne, le maintien des organismes en condition de photopériode continue implique une perte de rythme et une baisse significative du nombre de cercaires émises voire un arrêt total de l'émission des cercaires dans la majorité des cas. Le rythme d'émission des cercaires est aussitôt rétabli dès qu'il y a le retour de la photopériode équilibrée L-O (12-12). Ces résultats sont en accord avec les expériences menées sur *S. mansoni* diurne Porto Rico par Valle, Pellegrino & Alvarenga, 1973 et Williams, Wessels & Gilbertson, 1984. L'observation de l'arrêt d'émission cercarienne en condition continue suivi d'un retour à la normale avec le retour à la photopériode équilibrée a amené à s'interroger sur le devenir des cercaires à l'intérieur du mollusque pendant cette période. L'observation de coupes histologiques a permis de découvrir qu'en condition de photopériode continue il y a une accumulation des cercaires matures à l'intérieur des sporocystes, couplée à un arrêt de la cercariogenèse. Ces expériences d'histologie mettent ainsi en lumière un potentiel rôle du sporocyste dans le rythme d'émission des cercaires qui est jusque-là très peu étudié dans la littérature.

Enfin la dernière caractéristique d'un rythme circadien est la compensation du rythme à la température. C'est-à-dire, du point de vue expérimental, de voir si le rythme reste inchangé sous différentes températures. Dans le cas des travaux de cette thèse, limitée par le temps mais aussi par la nécessité de la génération d'assez de matériel biologique, ce dernier caractère n'a pas été testé. En effet, dans l'environnement naturel de *S. mansoni* Oman (zone tropicale), la variation de la température de l'eau n'est pas significative au cours d'une journée. Dans la littérature, le sujet de la compensation du rythme d'émission des cercaires à la température n'est quasiment par traité, à l'exception de travaux G. H. Frank, 1966. Malheureusement, la méthode utilisée ne permet pas de tirer de conclusion définitive (période d'expérimentation trop courte, points temporels de comptage trop éloignés). Ainsi, mises bout à bout, nos expériences ne mettent pas en évidence un rythme de type circadien pour l'émission des cercaires des deux chronotypes de *S. mansoni* (Oman).

Les résultats de la première partie de cette thèse ont permis de poser les bases du phénotype du rythme d'émission des cercaires mais aussi d'orienter les expériences de la partie de biologie moléculaire. De ce fait, cette première partie a démontré l'unicité de chacun des chronotypes diurne et nocturne de *S. mansoni* (Oman), ce qui a rendu possible l'utilisation de la technique de QTL mapping. La non-évidence de la nature circadienne du rythme et l'identification de l'alternance jour/nuit comme zeitgeber suggère que, lors des analyses moléculaires, il n'y aura pas de gène de l'horloge et que les mécanismes de phototransduction seront importants. Les expériences de chronobiologie en photopériode équilibrée L-O (12-12) ont permis d'exclure le sexe du parasite et l'espèce de mollusque comme paramètres de variabilité du rythme d'émission des cercaires de nos expériences ainsi que de minimiser le rôle de l'hôte dans le phénomène. Enfin, la mise en lumière du rôle potentiel des sporocystes par l'approche d'histologie nous a permis d'inclure ce stade parasitaire dans les expérimentations de biologie moléculaire.

Si aucune évidence de rythme circadien n'a été montrée, les divers résultats de la première partie ont identifié l'alternance jour/nuit comme le zeitgeber du rythme d'émission des cercaires. Les expériences d'histologie ont mis en lumière le rôle des sporocystes et les expériences de chronobiologie ont permis d'exclure le sexe du parasite, l'espèce de l'hôte mollusque et la lumière rouge comme paramètres de variation.

La deuxième partie de cette thèse a été consacrée à l'identification des déterminants moléculaires impliqués dans le rythme d'émission des cercaires. Cette problématique a été abordée avec trois approches de biologie moléculaire différentes et complémentaires (génétique, épigénétique et transcriptomique).

L'approche de génétique qui fait office de premier chapitre avec la technique de QTL mapping s'est concentrée sur l'identification des loci impliqués dans le rythme d'émission des cercaires chez *S. mansoni* Oman avec l'expérience intraspécifique issue des croisements entre les deux chronotypes. Les résultats ont montré que le rythme d'émission des cercaires était multigénique, avec des QTL détectés sur les chromosomes 1, 5 et 6. Parmi tous ces gènes, aucun gène canonique de l'horloge n'a été trouvé, ce qui est en accord avec les travaux de Rawlinson, Reid et al., 2021. Il y a toutefois la présence de facteurs de transcription contenant le domaine bhlh présent dans certains gènes de l'horloge. L'analyse de ces gènes candidats a révélé l'implication de certains dans la phototransduction avec notamment la présence de rhodopsine et rhodopsine-like. L'implication de la rhodopsine dans ce phénomène n'est pas surprenante puisque des rhodopsines et certains GPCR (récepteurs couplés aux protéines G) sont impliqués dans d'autres stades du parasite comme le miracidium (Hoffmann, Davis et al., 2001 ; El-Shehabi, Taman et al., 2012 ; Liang, Zao et al., 2016; Phan, Liang et al., 2022). La présence de gènes candidats impliqués dans la modification d'histone (ici acétylation) représente une découverte encourageante pour l'approche épigénétique portant sur les modifications d'histones dans le rythme d'émission des cercaires.

Une expérience de QTL mapping interspécifique issue des croisements entre *S. mansoni* diurne et *S. rodhaini* nocturne a aussi été menée dans le but, cette fois ci, de déterminer si les loci impliqués dans le rythme d'émission sont les mêmes entre espèces de schistosomes. Les résultats présentés sont des résultats QTL préliminaires réalisés sur les croisements réciproques séparés. Contrairement à l'expérience intraspécifique réalisée sur les exomes (exome capture), l'expérience interspécifique a été réalisée sur le génome entier (WGS, whole genome sequencing). Ce choix de technique se justifie par le fait que les croisements font intervenir deux génomes d'espèces différentes (*S. mansoni* et *S. rodhaini*). Avoir ainsi accès à l'ensemble des génomes parents et recombinés permet de détecter les régions incompatibles ou de faible recombinaison et d'avoir une meilleure détection des variants. Les résultats préliminaires ont montré qu'il y avait un QTL dans le chromosome 3 pour les deux croisements réciproques et un QTL dans le chromosome 4 uniquement pour le croisement B. Ce QTL dans le chromosome 4 n'est pas certain car il n'est pas présent dans le croisement A (très loin du seuil de significativité). Cela peut s'expliquer par le fait que ce QTL chromosome 4 est un artefact ou un phénomène propre au croisement B que l'on peut écarter de notre phénotype d'intérêt, dans le sens où il n'y a pas de différence significative dans la chronobiologie des individus des croisements réciproques A et B. L'analyse des gènes candidats a révélé des résultats similaires avec l'expérience intraspécifique, même si les QTLs ne sont pas sur les mêmes chromosomes. Ainsi, aucun gène canonique de l'horloge n'est présent mais il y a des facteurs de transcription avec un domaine bhlh, des gènes impliqués dans la phototransduction dont des rhodopsines et rhodopsines-like et des gènes liés à la modification d'histone (acétylation et méthylation). Avec ces résultats préliminaires du QTL interspécifique, nous pouvons dire que le rythme d'émission des cercaires (au travers du phénotype du nombre de cercaires nocturnes), repose sur des régions génomiques différentes, mais sur des mécanismes similaires.

L'intégration des résultats des approches de chronobiologie et d'histologie avec ceux de l'approche de génétique fait apparaître des liens. Ainsi la non évidence de rythme circadien en chronobiologie est confirmée par l'absence de gène canonique de l'horloge par l'approche moléculaire. L'identification du principal zeitgeber (alternance jour/nuit) est validée par la présence de gènes candidats impliqués dans la phototransduction. Côté photorécepteur, l'approche de chronobiologie avait écarté la lumière rouge (et donc par extension les photorécepteurs de type phytochrome). Ceci est confirmé avec d'une part l'absence de phytochrome et phytochrome-like dans le génome de *S. mansoni* et d'autre part avec l'identification de la rhodopsine comme potentiel photorécepteur qui absorbe la lumière verte-bleue. Il est aussi intéressant de noter que la non différence des phénotypes dans les croisements réciproques des expériences de génétique confirme qu'il n'y a pas d'effet du sexe dans le rythme d'émission des cercaires, comme observé avec la chronobiologie. Indépendamment de ces liens, les résultats de génétique ont conforté l'intérêt de l'étude de la modification d'histone en mettant en évidence les « writer » et « eraser » de l'acétylation et méthylation des histones.

Le second chapitre de la deuxième partie s'est centré sur l'approche d'épigénétique avec pour but de déterminer s'il y avait un rythme d'acétylation et de méthylation pour les marques H3K4me3 et H3K9ac pour les deux chronotypes de *S. mansoni* Oman. Ces deux marques présentent un intérêt particulier car elles ont été identifiées comme étant

rythmiques dans les promoteurs des gènes de l'horloge (Koike, Yoo et al., 2012 ; Aguilar-Arnal, Hakim et al., 2013 Takahashi, 2015). De plus la marque H3K9ac chez *S. mansoni* est présente sur un gène codant pour une rhodopsine (smp\_104210). Cette approche d'épigénétique a nécessité l'optimisation de la technique de CHIPmentation au modèle biologique de *S. mansoni*, valorisée sous la forme d'un article (Lasica, de Carvalho Augusto et al., 2022).

Pour le chronotype diurne, aucune évidence de rythme n'a été détectée pour les deux marques. En ce qui concerne le chronotype nocturne, seul deux pics pour H3K4me3 et un seul pic pour H3K9ac ont été détectés comme rythmiques. Ces résultats peuvent s'expliquer de plusieurs manières. La première et la plus simple est qu'il n'y a tout simplement pas de rythme dans ces deux marques histones. La deuxième raison est centrée sur les logiciels de détection qui pourraient sous-estimer les pics rythmiques. En effet, les logiciels de détection des rythmes sont des outils qui ont été créés en premier lieu pour les analyses transcriptomiques qui ont des signaux beaucoup plus forts que les expériences d'épigénétique. Ceci est d'autant plus vrai qu'il a été montré que les logiciels rythmiques détectent plus facilement le rythme avec un signal fort (Laloum & Robinson-Rechavi, 2020). La quantité de matériel biologique étant souvent un facteur limitant pour les modèles atypiques, diminuant la possibilité de fortifier le signal avec une augmentation du nombre d'échantillons, il est intéressant de s'interroger sur la sévérité des seuils de détection ou alors de développer des logiciels capables de détecter le rythme d'un signal faible.

Si les évidences de rythme ne sont pas très probantes pour cette approche d'épigénétique, l'étude des gènes associés aux marques histones s'est révélée intéressante, en retrouvant des gènes QTLs, mais aussi des gènes impliqués dans la phototransduction et des rhodopsine-like, ce qui conforte l'intérêt de ces voies avec les résultats des autres approches. L'analyse différentielle des pics entre les chronotypes diurne et nocturne a montré qu'il y avait une différence significative pour les deux marques. On retrouve de ce fait plus de marque H3K4me3 dans le chronotype diurne et plus de marque H3K9ac dans le chronotype nocturne, ce qui vient ajouter un nouveau paramètre de différenciation entre les deux chronotypes en plus de ceux mis en avant dans l'approche de chronobiologie.

La dernière approche de la partie de biologie moléculaire est la transcriptomique dont le but est de déterminer s'il y a des gènes avec une expression rythmique. Les résultats ont montré que la grande majorité des gènes rythmiques était spécifique à chaque chronotype. Il est intéressant de voir que, pour certains gènes communs, le pic d'expression est différent entre les deux chronotypes. Cette différence globale de pic d'expression se retrouve sur l'ensemble des gènes rythmiques des deux chronotypes. Chez le chronotype diurne, la majorité des pics d'expression est à 8h, soit 4h avant le pic d'émission tandis que chez le chronotype nocturne le pic d'expression est à 4h, soit 16h avant le pic d'émission. Cette différence peut s'expliquer par le fait que le signal environnemental perçu par les deux chronotypes n'est pas de la même puissance. Les analyses différentielles des gènes candidats entre les deux chronotypes ont montré qu'il y avait plus de rhodopsine exprimée chez le chronotype nocturne. On peut donc en déduire que le chronotype nocturne est plus sensible à la lumière et plus largement aux signaux environnementaux. Les pics d'expression nocturnes à 4h pourraient signifier la perception d'un signal environnemental faible, puis l'apparition de la lumière entraînerait l'activation de plusieurs rhodopsines qui déclencheraient un signal fort dans la cellule et une inhibition des mécanismes qui permettraient la sortie des cercaires. Dès l'apparition de la nuit, le fort signal d'inhibition enclenché par les rhodopsines est levé, démarrant ainsi l'émission des cercaires. La présence de plusieurs rhodopsines qui agiraient de concert pour faire un fort signal d'inhibition levé dès la nuit expliquerait le phénotype très abrupt d'émission de cercaires nocturnes qui débute immédiatement à la nuit. Pour le chronotype diurne, à l'inverse, qui a un pic d'expression majoritaire à 8h et exprime moins de rhodopsine, nous pouvons penser qu'il faut un signal environnemental fort pour enclencher l'émission des cercaires. En d'autres termes, s'il y a moins de rhodopsine/rhodopsine-like exprimée dans les cellules, il faut, pour que le niveau de signal cellulaire soit assez fort, attendre une plus forte intensité lumineuse qui arriverait à 8h et enclencherait ainsi l'émission des cercaires.

L'analyse des gènes rythmiques pour les deux chronotypes a révélé des gènes rhodopsine, rhodopsine-like, des gènes impliqués dans la phototransduction, l'acétylation des histones (chronotype diurne uniquement), des facteurs de transcription (chronotype nocturne uniquement) dont certains sont communs avec les expériences de QTL et de

ChIPmentation. L'ensemble des résultats de biologie moléculaire fait émerger des voies intéressantes mais surtout communes aux trois approches, avec l'absence de gène canonique de l'horloge, la mise en avant de gènes impliqués dans la phototransduction avec la rhodopsine comme photorécepteur, et même la modification d'histone. Intégrés avec la première partie de la thèse, les différents résultats indiquent que le rythme d'émission des cercaires repose au minimum sur un mécanisme de phototransduction, où la rhodopsine joue un rôle central. Ce mécanisme semble être certainement commun entre les différentes espèces de schistosomes mais reposerait sur l'action de régions génomiques différentes. Par exemple, le nombre de gènes rhodopsine/rhodopsine-like exprimés pourrait être un facteur qui expliquerait la différence de chronotype. En effet, le chronotype diurne de *S. mansoni* Oman exprime moins de rhodopsine et a besoin d'un signal environnemental/lumineux fort pour enclencher une transduction assez forte pour enclencher le rythme d'émission des cercaires (Figure\_87). A l'inverse, le chronotype nocturne, qui a plus de rhodopsine/rhodopsine-like exprimés, perçoit les signaux environnementaux faibles (ex : les tout premiers rayons du soleil) pour enclencher une transduction qui bloquera l'émergence des cercaires jusqu'à la levée de cette inhibition par le retour de la nuit (Figure\_87). Ce scénario est renforcé par la forme des profils d'émission des cercaires qui s'enclenche dès le début de la nuit comme un interrupteur.

# Discussion et conclusion générales

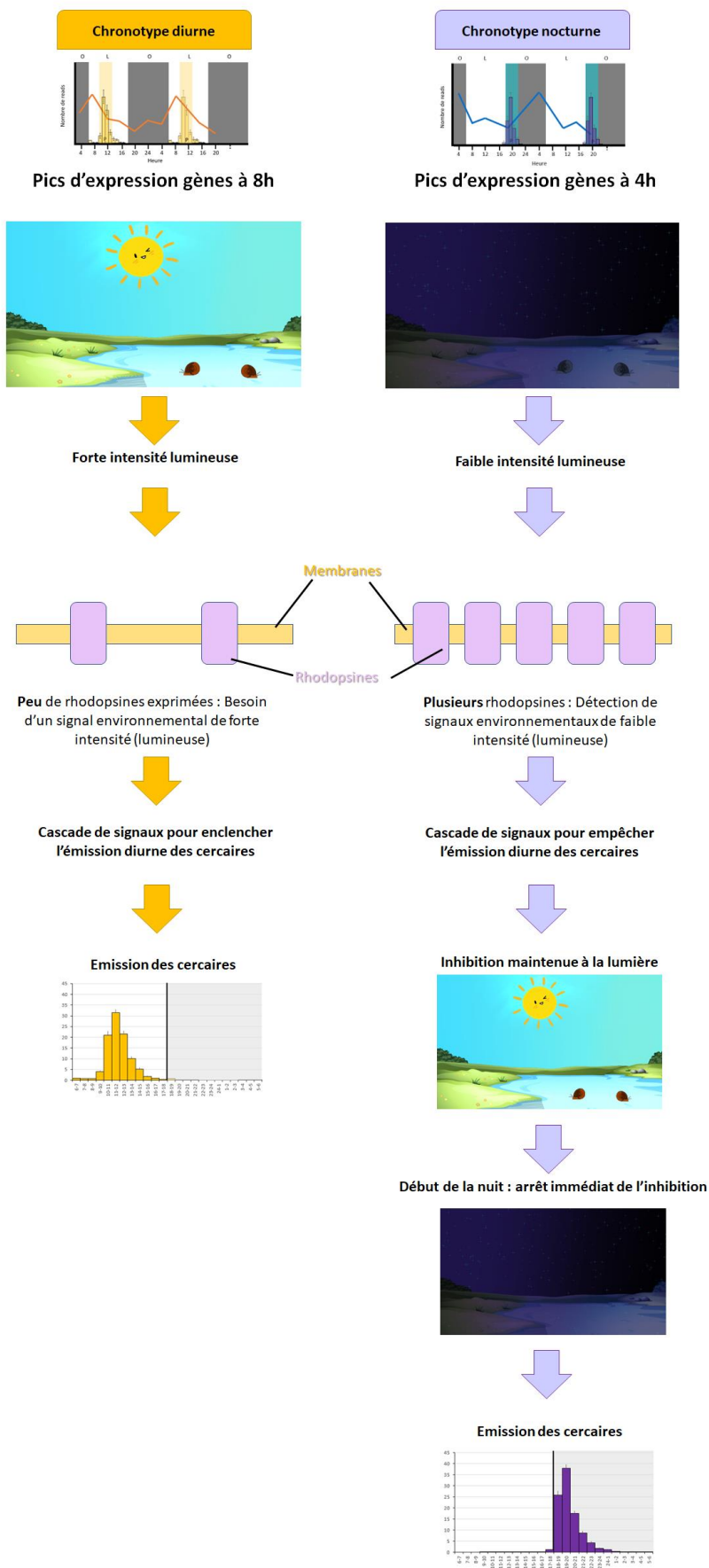


Figure 87 : Proposition d'un modèle chronobiologique et moléculaire du rythme de l'émission des cercaires chez les deux chronotypes de *S. mansoni*.

Pour résumer, les approches phénotypiques et moléculaires convergent vers une même interprétation du rythme d'émission des cercaires chez *S. mansoni*, illustrant ainsi l'intérêt et la complémentarité d'une vision intégrative du phénomène. Ces approches ont mis en lumière des caractéristiques propres à chaque chronotype, comme l'heure du pic, l'intensité de la plage d'émission et le pourcentage de cercaires nocturne avec la chronobiologie, ou encore l'acétylation et la méthylation des histones avec l'épigénétique et aussi le pic du rythme de l'expression de gènes rythmiques avec la transcriptomique. La différence d'acétylation et de méthylation des histones entre les deux chronotypes présente une piste intéressante dans l'identification des schistosomes. Si les techniques de détection avec les séquences COX et ITS sont robustes et éprouvées, la problématique des individus hybrides amène à complexifier l'interprétation des résultats. Les analyses différentielles de l'acétylation et de la méthylation des histones avec H3K9ac et H3K4me3 sont un résultat qui doit être creusé. Ainsi, il faudrait effectuer de nouvelles expériences de ChIPmentation, sur différentes espèces de schistosomes « pures/modèle » et aussi hybrides pour essayer d'identifier des profils de marques histones. Ce potentiel outil d'identification épigénétique permettrait ainsi une identification plus fine des schistosomes hybrides.

Les résultats des approches phénotypiques et moléculaires convergent vers la non évidence de rythme circadien. Des nuances peuvent être apportées à cette conclusion. En effet, la perte totale de rythme d'émission des cercaires observée avec les expériences de chronobiologie ne permet pas d'affirmer ou d'infirmer la nature du rythme. Les conditions drastiques de photopériode continue pourraient avoir perturbé d'autres fonctions biologiques comme par exemple la production cercarienne et faire disparaître le rythme d'émission. Nous pouvons ajouter à cela l'absence de gène canonique de l'horloge dans le génome de *S. mansoni* qui ne signifie pas qu'il n'existe pas d'autres gènes au rôle similaire mais à l'heure actuelle inconnus. Pour statuer de façon claire sur la nature circadienne ou non du rythme, de nouvelles expériences de lumière pulsée en chronobiologie doivent être faites. Ces expériences permettront ainsi de déterminer si le rythme d'émission des cercaires est bien exogène ou s'il est endogène en suivant soit le modèle du sablier soit le modèle de la coïncidence externe.

La transversalité des approches a mis en lumière l'alternance jour/nuit comme zeitgeber, soutenue par la sélection subjective des gènes impliqués dans le phénomène de phototransduction. Toutefois, les acteurs des transductions d'autres signaux environnementaux ne sont pas à exclure, bien au contraire. Par exemple, chez le némathelminthe, *C. elegans*, la phototransduction passe par la protéine LITE-1 appartenant à une famille de récepteurs gustatifs chez les invertébrés (Liu, Ward et al., 2010). On peut donc naturellement penser que les rhodopsines ne sont pas forcément les seuls photorécepteurs chez *S. mansoni* mais aussi que la perception de la lumière peut passer par d'autres signaux, par exemple un changement dans le pH gastrique du mollusque médié par un rythme circadien de la digestion. Si la sélection subjective présentée dans cette thèse s'est concentrée autour de la réception du signal lumineux et de la cascade associée, la prochaine étape dans les analyses des gènes est d'y inclure les gènes responsables des messages nerveux qui résultent de la réception de la lumière, les récepteurs de ces messages et les actions qui en découlent.

Nos expériences d'histologie ont mis en lumière l'intérêt de la forme sporocyste qui pourrait jouer un rôle dans le rythme d'émission des cercaires. En réponse à cet intérêt, les sporocystes ont été inclus dans les expériences de biologie moléculaire (épigénétique et transcriptomique). Si avoir plusieurs stades du parasite dans une seule analyse moléculaire peut être préjudiciable dans l'interprétation des résultats, l'état actuel des connaissances et la technique ne nous permettent pas d'exclure le stade sporocyste. En effet, il faut savoir que le phénomène de sortie des cercaires du sporocyste reste encore inconnu et qu'exclure le sporocyste des analyses est une potentielle perte d'information non négligeable dans la compréhension du phénomène. De plus, il est impossible de dissocier physiquement un sporocyste des cercaires qu'il contient, ne permettant pas une analyse séparée des stades. Ce trou dans les connaissances, généré par la difficulté du modèle biologique, est un sujet de recherche à approfondir. Pour cela de nouvelles expériences d'histologie pourraient être menées notamment en réalisant des coupes histologiques sur des mollusques prélevés toutes les heures sur minimum 24h afin d'avoir une chronologie de développement du parasite dans les tissus. Ces expériences d'histologie doivent être menées à la fois sur les schistosomes du groupe des œufs avec éperon terminal, comme *S. mansoni* qui possède des « birth pore » (Théron, 2015) et les schistosomes d'autres

## Discussion et conclusion générales

groupes comme *S. bovis*, *S. haematobium* ou *S. intercalatum* où aucune structure de sortie sur les sporocystes n'avait été observée.

Enfin, les diverses expériences de chronobiologie et d'histologie ont montré leur intérêt et leur importance dans cette thèse pour compléter et éclaircir les approches de biologie moléculaire. Toutefois l'ancienneté des articles portant sur la chronobiologie de l'émission des cercaires montrent un désintérêt et une méconnaissance du sujet. Pourtant aujourd'hui avec l'émergence de plus en plus fréquente d'hybrides, rendant petit à petit caduques les connaissances anciennes, il est primordial de s'intéresser de nouveau à cet aspect de chronobiologie de l'émission des cercaires. Ceci est notamment renforcé par le réchauffement climatique et l'émergence du parasite en Corse dépassant ainsi les limites géographiques des zones tropicales. Des expériences de chronobiologie supplémentaires doivent être menées notamment sur la capacité du rythme d'émergence des cercaires à évoluer ou non en fonction de la température. En d'autres termes, de voir si le parasite sous conditions hivernales émettra ces cercaires aux mêmes horaires qu'en conditions optimales de température ?

Les travaux de thèse soulèvent beaucoup de nouveaux questionnements et de champs de recherche à approfondir. Ils présentent ainsi les bases chronobiologiques et moléculaires du rythme d'émission des cercaires chez les deux chronotypes de *S. mansoni* et illustre la nécessité d'aborder les phénomènes biologiques avec une approche transversale pour comprendre les subtilités et les diverses nuances de la relation phénotype/génotype.

---

## Participation aux congrès, manifestations scientifiques et Distinctions

### En national

- 10ème CONGRÈS DES DOCTORANTS - Communication orale + poster + vidéo. 19 novembre 2021 3ème prix pour la présentation orale, 1er prix pour la présentation de Poster et de vidéo.
- Fête de la science 2018 (Promotion de faire une thèse auprès de classes de lycée, stand Upvdoc)
- Fête de la science 2019 et 2020 (stand du laboratoire IHPE)

### En international

Summer Internship 2019 - Texas Biomedical research institute (2 présentations orales en tant qu'invitée) juin à septembre **2019**

4 congrès internationaux:

- 85th Cold Spring Harbor Laboratory Symposium on Quantitative Biology: Biological Time Keeping – **Lasica** et al., Poster- 1 juin **2021**
- 18th Ecology and Evolution of Infectious Diseases (online) meeting - **Lasica** et al., Poster- 14 juin **2021** + 2ème prix du meilleur teaser video
- 15th International Conference on Molecular Epidemiology and Evolution Genetics of Infectious Diseases (MEEGID XV) – **Lasica** et al., Présentation orale- 2 novembre **2021**
- Molecular Helminthology: An integrated Approach, San Antonio, Tx, USA, Poster, Mouahid G., Chevalier F., Languard J., **Lasica** C. ... & Moné H. April 7-10th **2019**.

Teaser vidéo (anglais) : <https://vimeo.com/557965346>

Teaser vidéo (Français) : <https://mediaserver.univ-perp.fr/permalink/v1261ca3c6d19ckewcyb/iframe/>



---

## Références

- Aguilar-Arnal, L., Hakim, O., Patel, V. R., Baldi, P., Hager, G. L. & Sassone-Corsi, P. (2013). Cycles in spatial and temporal chromosomal organization driven by the circadian clock. *Nature Structural & Molecular Biology*, 20(10), 1206-1213. doi:10.1038/nsmb.2667
- Ahdesmäki, M., Fokianos, K. & Strimmer, K. (2012). GeneCycle: identification of periodically expressed genes. In: R package version.
- Anderson, T. J., LoVerde, P. T., Le Clec'h, W. & Chevalier, F. D. (2018). Genetic crosses and linkage mapping in schistosome parasites. *Trends in Parasitology*, 34(11), 982-996. doi:10.1016/j.pt.2018.08.001
- Andolfatto, P., Davison, D., Ereyilmaz, D., Hu, T. T., Mast, J., Sunayama-Morita, T. & Stern, D. L. (2011). Multiplexed shotgun genotyping for rapid and efficient genetic mapping. *Genome Research* 21(4), 610-617. doi:10.1101/gr.115402.110
- Asch, H. L. (1972). Rhythmic emergence of *Schistosoma mansoni* cercariae from *Biomphalaria glabrata*: control by illumination. *Experimental Parasitology*, 31(3), 350-355. doi:10.1016/0014-4894(72)90096-3
- Aschoff, J. (1960). *Exogenous and endogenous components in circadian rhythms*. Paper presented at the Cold Spring Harbor symposia on quantitative biology, Cold Spring Harbor Laboratory Press.
- Azzi, A., Cosseau, C. & Grunau, C. (2009). *Schistosoma mansoni*: developmental arrest of miracidia treated with histone deacetylase inhibitors. *Experimental parasitology*, 121, 288-291. doi:10.1016/j.exppara.2008.11.010
- Belden, W. J., Lewis, Z. A., Selker, E. U., Loros, J. J. & Dunlap, J. C. (2011). CHD1 remodels chromatin and influences transient DNA methylation at the clock gene frequency. *PLoS genetics*, 7(7), e1002166. doi:10.1371/journal.pgen.1002166
- Bell-Pedersen, D., Cassone, V. M., Earnest, D. J., Golden, S. S., Hardin, P. E., Thomas, T. L. & Zoran, M. J. (2005). Circadian rhythms from multiple oscillators: lessons from diverse organisms. *Nature Reviews Genetics*, 6(7), 544-556. doi:10.1038/nrg1633
- Berry, A., Moné, H., Iriart, X., Mouahid, G., Aboo, O., Boissier, J., . . . & Valentin, A. (2014). Schistosomiasis *haematobium*, Corsica, France. *Emerging infectious diseases*, 20(9), 1595. doi:10.3201/eid2009.140928
- Bhardwaj, V., Heyne, S., Sikora, K., Rabbani, L., Rauer, M., Kilpert, F., . . . & Manke, T. (2019). snakePipes: facilitating flexible, scalable and integrative epigenomic analysis. *Bioinformatics*, 35(22), 4757-4759. doi:10.1093/bioinformatics/btz436
- Boissier, J. & Moné, H. (2000). Experimental observations on the sex ratio of adult *Schistosoma mansoni*, with comments on the natural male bias. *Parasitology*, 121(4), 379-383. doi:10.1017/S0031182099006393
- Boissier, J., Morand S. & Moné H. (1999). A review of performance and pathogenicity of male and female *Schistosoma mansoni* during the life-cycle. *Parasitology*, 119 (5), 447-454. doi:10.1017/s0031182099005004
- Boissier, J., Mouahid, G. & Moné, H. (2019). *Schistosoma* spp. In: In: J.B. Rose and B. Jiménez-Cisneros, (eds) Global Water Pathogen Project. <http://www.waterpathogens.org> (<http://www.waterpathogens.org/>) (Robertson, L (eds) Part 4 Helminths) <http://www.waterpathogens.org/book/shistosoma> (<http://www.waterpathogens.org/book/shistosoma>) Michigan State University, E. Lansing, MI, UNESCO.
- Bonhoure, N., Bounova, G., Bernasconi, D., Praz, V., Lammers, F., Canella, D., . . . & Delorenzi, M. (2014). Quantifying ChIP-seq data: a spiking method providing an internal reference for sample-to-sample normalization. *Genome research*, 24(7), 1157-1168. doi:10.1101/gr.168260.113
- Botebol, H., Lesuisse, E., Šuták, R., Six, C., Lozano, J.-C., Schatt, P., . . . & Léger, T. (2015). Central role for ferritin in the day/night regulation of iron homeostasis in marine phytoplankton. *Proceedings of the National Academy of Sciences*, 112(47), 14652-14657. doi:10.1073/pnas.1506074112
- Boyle, J. P. & Yoshino, T. P. (2005). Serotonin-induced muscular activity in *Schistosoma mansoni* larval stages: importance of 5-HT transport and role in daughter sporocyst production. *The Journal of Parasitology*, 91(3), 542-550. doi:10.1645/GE-432R
- Broman, K. W. & Sen, S. (2009). *A Guide to QTL Mapping with R/qtl*, Springer, 46, 133 p.
- Cavieres-Lepe, J. & Ewer, J. (2021). Reciprocal relationship between calcium signaling and circadian clocks: implications for calcium homeostasis, clock function, and therapeutics. *Frontiers in Molecular Neuroscience*, 14, 666673. doi:10.3389/fnmol.2021.66667

- Chabre, M. (1998). Actualité moléculaire de la phototransduction des bâtonnets rétiens. MS. *Médecine Sciences*, 14(12), 1315-1321.
- Cheng, H.-Y. M., Papp, J. W., Varlamova, O., Dziema, H., Russell, B., Curfman, J. P., . . . & Impey, S. (2007). microRNA modulation of circadian-clock period and entrainment. *Neuron*, 54(5), 813-829. doi:10.1016/j.neuron.2007.05.017
- 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(1), 70-76. doi:10.1016/j.molbiopara.2009.02.015
- Cosseau, C., Wolkenhauer, O., Padalino, G., Geyer, K. K., Hoffmann, K. F. & Grunau, C. (2017). (Epi) genetic inheritance in *Schistosoma mansoni*: a systems approach. *Trends in Parasitology*, 33(4), 285-294. doi:10.1016/j.pt.2016.12.002 2
- Cupit, P. M. & Cunningham, C. (2015). What is the mechanism of action of praziquantel and how might resistance strike? *Future medicinal chemistry*, 7(6), 701-705. doi:10.4155/fmc.15.11
- Delaunay, F. & Laudet, V. (1998). Rythme circadien : des horloges dans les organes périphériques et dans des fibroblastes en culture. MS. *Médecine Sciences*, 14(10), 1114-1117.
- Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., . . . & Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, 29(1), 15-21. doi:10.1093/bioinformatics/bts635
- Doi, M., Hirayama, J. & Sassone-Corsi, P. (2006). Circadian regulator CLOCK is a histone acetyltransferase. *Cell*, 125(3), 497-508. doi:10.1016/j.cell.2006.03.033
- Dortous de Mairan, J.-J. (1729). Observation botanique. *Histoire de l'Académie royale des Sciences*, 35-36.
- Dunlap, J. C., Loros, J. J. & DeCoursey, P. J. (2004). *Chronobiology: biological timekeeping*: Sinauer Associates.
- El-Shehabi, F., Taman, A., Moali, L. S., El-Sakkary, N. & Ribeiro, P. (2012). A novel G protein-coupled receptor of *Schistosoma mansoni* (SmGPR-3) is activated by dopamine and is widely expressed in the nervous system. *PLoS Neglected Tropical Diseases*, 6(2), e1523. doi:10.1371/journal.pntd.0001523
- Erren, T. C. (2018). A chronology of chronobiology. *Chronobiology and Sleep*, (113), 32-35. Retrieved from <https://www.physoc.org/magazine-articles/a-chronology-of-chronobiology/>
- Etchegaray, J.-P., Lee, C., Wade, P. A. & Reppert, S. M. (2003). Rhythmic histone acetylation underlies transcription in the mammalian circadian clock. *Nature*, 421(6919), 177-182. doi:10.1038/nature01314
- Fantappiè, M. R., Gimba, E. R. P. & Rumjanek, F. D. (2001). Lack of DNA methylation in *Schistosoma mansoni*. *Experimental parasitology*, 98(3), 162-166. doi:10.1006/expr.2001.4630
- Frank, G. H. (1966). The effect of temperature on the rate of development and emergence of schistosome cercariae. *Zoologica Africana*, 2:2, 211-221. doi:10.1080/00445096.1966.11447344
- Feldman, J. F. & Hoyle, M. N. (1973). Isolation of circadian clock mutants of *Neurospora crassa*. *Genetics*, 75(4), 605-613. doi:10.1093/genetics/75.4.605
- Feng, J., Liu, T., Qin, B., Zhang, Y. & Liu, X. S. (2012). Identifying ChIP-seq enrichment using MACS. *Nature Protocols*, 7(9), 1728-1740. doi:10.1038/nprot.2012.101
- Freeman, S. L., Kwon, H., Portolano, N., Parkin, G., Venkatraman Girija, U., Basran, J., . . . & Moody, P. C. (2019). Heme binding to human CLOCK affects interactions with the E-box. *Proceedings of the National Academy of Sciences*, 116(40), 19911-19916.
- Gerding, W. M., Schreiber, S., Schulte-Middelmann, T., de Castro Marques, A., Atorf, J., Akkad, D. A., . . . & Gal, A. (2011). Ccdc66 null mutation causes retinal degeneration and dysfunction. *Human Molecular Genetics*, 20(18), 3620-3631. doi:10.1093/hmg/ddr282
- Glaudel, R. J. & Etges F. J. (1973). The effect of photoperiod inversion upon *Schistosoma mansoni* cercarial emergence from *Biomphalaria glabrata*. *International Journal for Parasitology*, 3(5), 619-622. doi:10.1016/0020-7519(73)90086-6
- Grassmeyer, J. J., Cahill, A. L., Hays, C. L., Barta, C., Quadros, R. M., Gurumurthy, C. B., & Thoreson, W. B. (2019). Ca<sup>2+</sup> sensor synaptotagmin-1 mediates exocytosis in mammalian photoreceptors. *Elife*, 8, e45946. Doi:10.7554/eLife.45946
- Green, C. B. & Besharse, J. C. (1996). Identification of a novel vertebrate circadian clock-regulated gene encoding the protein nocturnin. *Proceedings of the National Academy of Sciences*, 93(25), 14884-14888. doi:10.1073/pnas.93.25.14884

- Hasegawa, K., Saigusa, T. & Tamai, Y. (2005). *Caenorhabditis elegans* opens up new insights into circadian clock mechanisms. *Chronobiology international*, 22(1), 1-19. doi:10.1081/cbi-200038149
- Hawking, F. (1967). The 24-hour periodicity of microfilariae: biological mechanisms responsible for its production and control. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, 169(1014), 59-76. doi:10.1098/rspb.1967.0079
- Hawking, F., Worms, M. J. & Gammage, K. (1968). 24-and 48-hour cycles of malaria parasites in the blood; their purpose, production and control. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 62(6), 731-760. doi:10.1016/0035-9203(68)90001-1
- Hevia, M. A., Canessa, P., Müller-Esparza, H. & Larrondo, L. F. (2015). A circadian oscillator in the fungus *Botrytis cinerea* regulates virulence when infecting *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences*, 112(28), 8744-8749. doi:10.1073/pnas.1508432112
- Hoffmann, K. F., Davis, E. M., Fischer, E. R. & Wynn, T. A. (2001). The guanine protein coupled receptor rhodopsin is developmentally regulated in the free-living stages of *Schistosoma mansoni*. *Molecular and Biochemical Parasitology*, 112(1), 113-123.
- Holtfreter, M., Moné, H., Müller-Stöver, I., Mouahid, G. & Richter, J. (2014). *Schistosoma haematobium* infections acquired in Corsica, France, August 2013. *Eurosurveillance*, 19(22), 20821. doi:10.2807/1560-7917.es2014.19.22.20821
- Howe, K. L., Bolt, B. J., Cain, S., Chan, J., Chen, W. J., Davis, P., . . . & Grove, C. (2016). WormBase 2016: expanding to enable helminth genomic research. *Nucleic acids research*, 44(D1), D774-D780. doi:10.1093/nar/gkv1217
- Howe, K. L., Bolt, B. J., Shafie, M., Kersey, P. & Berriman, M. (2017). WormBase ParaSite– a comprehensive resource for helminth genomics. *Molecular and biochemical parasitology*, 215, 2-10. doi:10.1016/j.molbiopara.2016.11.005
- Hut, R. & Van der Zee, E. (2011). The cholinergic system, circadian rhythmicity, and time memory. *Behavioural Brain Research*, 221(2), 466-480. doi:10.1016/j.bbr.2010.11.039
- Hutchison, A. L., Maienschein-Cline, M., Chiang, A. H., Tabei, S. A., Gudjonson, H., Bahroos, N., . . . & Dinner, A. R. (2015). Improved statistical methods enable greater sensitivity in rhythm detection for genome-wide data. *PLoS Computational Biology*, 11(3), e1004094. doi:10.1371/journal.pcbi.1004094
- Joly, D. & Grunau, C. (2018). Prospective Épigénétique en Écologie et Environnement de l'Institut Ecologie et Environnement du CNRS. *Les Cahiers Prospectives (CNRS-INEE)* n° 8 Novembre, 80 pages.
- Kageyama, R., Ohtsuka, T. & Kobayashi, T. (2007). The Hes gene family: repressors and oscillators that orchestrate embryogenesis. doi:10.1242/dev.000786
- Katsumata, O., Honma, T., Sanda, M., Kamata, A., Takeda, S. i., Kondo, H. & Sakagami, H. (2008). Predominant localization of EFA6A, a guanine nucleotide exchange factor for ARF6, at the perisynaptic photoreceptor processes. *Brain research*, 1234, 44-49. doi:10.1016/j.brainres.2008.07.093
- Kazibwe, F., Makanga, B., Rubaire-Akiiki, C., Ouma, J., Kariuki, C., Kabatereine, N., Vennervald, B., Rollinson, D. & Stothard, J. (2010). Transmission studies of intestinal schistosomiasis in Lake Albert, Uganda and experimental compatibility of local *Biomphalaria* spp. *Parasitology international*, 59(1), 49-53. doi: 10.1016/j.parint.2009.10.004
- King, D. P. & Takahashi, J. S. (2000). Molecular genetics of circadian rhythms in mammals. *Annual Review of Neuroscience*, 23, 713-742. doi:10.1146/annurev.neuro.23.1.713
- King, D. P., Zhao, Y., Sangoram, A. M., Wilsbacher, L. D., Tanaka, M., Antoch, M. P., . . . & Lowrey, P. L. (1997). Positional cloning of the mouse circadian clock gene. *Cell*, 89(4), 641-653. doi:10.1016/s0092-8674(00)80245-7
- Kippert, F., Ninnemann, H. & Engelmann, W. (1991). Photosynchronization of the circadian clock of *Schizosaccharomyces pombe*: mitochondrial cytochrome b is an essential component. *Current genetics*, 19(2), 103-107.
- Klarsfeld, A. (2013). Aux aurores de la chronobiologie. *Bibnum, Sciences de la vie*, 12 p. Retrieved from <http://journals.openedition.org/bibnum/511>
- Koike, N., Yoo, S.-H., Huang, H.-C., Kumar, V., Lee, C., Kim, T.-K. & Takahashi, J. S. (2012). Transcriptional architecture and chromatin landscape of the core circadian clock in mammals. *Science*, 338(6105), 349-354. doi:10.1126/science.1226339
- Konopka, R. J. & Benzer, S. (1971). Clock mutants of *Drosophila melanogaster*. *Proceedings of the National Academy of Science U S A*, 68(9), 2112-2116. doi:10.1073/pnas.68.9.2112
- Kuhlman, S. J., Craig, L. M. & Duffy, J. F. (2018). Introduction to chronobiology. *Cold Spring Harbor Perspectives in Biology*, 10(9), a033613. doi:10.1101/cshperspect.a033613

- Laherty, C. D., Yang, W.-M., Sun, J.-M., Davie, J. R., Seto, E. & Eisenman, R. N. (1997). Histone deacetylases associated with the mSin3 corepressor mediate mad transcriptional repression. *Cell*, *89*(3), 349-356. doi:10.1016/s0092-8674(00)80215-19
- Laloum, D. & Robinson-Rechavi, M. (2020). Methods detecting rhythmic gene expression are biologically relevant only for strong signal. *PLoS Computational Biology*, *16*(3), e1007666. doi:10.1371/journal.pcbi.1007666
- Lasica, C., de Carvalho Augusto, R., Moné, H., Mouahid, G., Chaparro, C., Veillard, A.-C., . . . & Grunau, C. (2021). Automated ChIPmentation procedure on limited biological material of the human blood fluke *Schistosoma mansoni*. *bioRxiv*. doi:10.1101/2021.06.02.446743
- Lasica, C., de Carvalho Augusto, R., Moné, H., Mouahid, G., Chaparro, C., Veillard, A.-C., . . . & Grunau, C. (2022). Automated ChIPmentation procedure on limited biological material of the human blood fluke *Schistosoma mansoni* [version 1; peer review: 1 approved with reservations]. *Wellcome Open Research*, *7*,133. doi:10.12688/wellcomeopenres.17779.1.
- Le Clec'h, W., Chevalier, F. D., Mattos, A. C. A., Strickland, A., Diaz, R., McDew-White, M., . . . & Anderson, T. J. (2021). Genetic analysis of praziquantel response in schistosome parasites implicates a Transient Receptor Potential channel. *Science Translational Medicine*, *13*(625), eabj9114. doi:10.1126/scitranslmed.abj9114
- Le Clec'h, W., Chevalier, F. D., McDew-White, M., Menon, V., Arya, G.-A. & Anderson, T. J. (2021). Genetic architecture of transmission stage production and virulence in schistosome parasites. *Virulence*, *12*(1), 1508-1526. doi:10.1080/21505594.2021.1932183
- Lee, Y., & Kim, E.-K. (2013). AMP-activated protein kinase as a key molecular link between metabolism and clockwork. *Experimental & Molecular Medicine*, *45*(7), e33-e33. doi:10.1038/emm.2013.65
- Lee, J., Yoo, E., Lee, H., Park, K., Hur, J.-H. & Lim, C. (2017). LSM12 and ME31B/DDX6 define distinct modes of posttranscriptional regulation by ATAXIN-2 protein complex in *Drosophila* circadian pacemaker neurons. *Molecular Cell*, *66*(1), 129-140. e127. doi:10.1016/j.molcel.2017.03.004
- Li, B., & Dewey, C. N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*, *12*(1), 1-16. doi:10.1186/1471-2105-12-323
- Liang, D., Zhao, M., Wang, T., McManus, D. P. & Cummins, S. F. (2016). GPCR and IR genes in *Schistosoma mansoni* miracidia. *Parasites & Vectors*, *9*(1), 1-12. doi:10.1186/s13071-016-1837-2
- Liu, J., Ward, A., Gao, J., Dong, Y., Nishio N., Inada H.,... & Xu, X. Z. S. (2010). *C. elegans* phototransduction requires a G protein-dependent cGMP pathway and a taste receptor homolog. *Nature Neuroscience*, *13*(6), 715-722. doi:10.1038/nn.2540
- Love, M. I., Huber, W. & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology*, *15*(12), 1-21. doi:10.1186/s13059-014-0550-8
- Luttermoser, G. W. (1955). Studies on the chemotherapy of experimental schistosomiasis. III. Harvest of *Schistosoma mansoni* cercariae by forced nocturnal emergence from *Australorbis glabratus*. *Journal of Parasitology*, *41*(2), 201-208. doi:10.2307/3273793
- Maienschein-Cline, M., Lei, Z., Gardeux, V., Abbasi, T., Machado, R. F., Gordeuk, V., ... & Lussier, Y. (2014). ARTS: automated randomization of multiple traits for study design. *Bioinformatics*, *30*(11), 1637-1639.
- Marsaudon, É. (2006). La chronobiologie, une conception dynamique du fonctionnement corporel. *Les Tribunes de la santé*, (4), 39-44. doi:10.3917/seve.013.0039
- Matsumoto, E., Ishihara, A., Tamai, S., Nemoto, A., Iwase, K., Hiwasa, T., . . . & Takiguchi, M. (2010). Time of day and nutrients in feeding govern daily expression rhythms of the gene for sterol regulatory element-binding protein (SREBP)-1 in the mouse liver. *Journal of Biological Chemistry*, *285*(43), 33028-33036. doi:10.1074/jbc.M109.089391
- McClung, C. R. (2006). Plant circadian rhythms. *The Plant Cell*, *18*(4), 792-803. doi:10.1105/tpc.106.040980
- Merrow, M., Brunner, M. & Roenneberg, T. (1999). Assignment of circadian function for the *Neurospora* clock gene frequency. *Nature*, *399*(6736), 584-586. doi:10.1038/21190
- Moné, H., Holtfreter, M. C., Allienne, J.-F., Mintsa-Nguéma, R., Ibikounlé, M., Boissier, J., . . . & Mouahid, G. (2015). Introgressive hybridizations of *Schistosoma haematobium* by *Schistosoma bovis* at the origin of the first case report of schistosomiasis in Corsica (France, Europe). *Parasitology Research*, *114*(11), 4127-4133. doi:10.1007/s00436-015-4643-4
- Montell, C., Jones, K., Hafen, E. & Rubin, G. (1985). Rescue of the *Drosophila* phototransduction mutation trp by germline transformation. *Science*, *230*(4729), 1040-1043. doi:10.1126/science.3933112
- Morgan, E. & Last, V. (1982). The behaviour of *Bulinus africanus*: a circadian profile. *Animal Behaviour*, *30*, 557-567. doi:10.1016/S0003-3472(82)80069-9

- Mouahid, A. & Théron, A. (1986). *Schistosoma bovis*: patterns of cercarial emergence from snails of the genera *Bulinus* and *Planorbarius*. *Experimental Parasitology*, 62(3), 389-393. doi:10.1016/0014-4894(86)90047-0
- Mouahid, G., Idris, M. A., Verneau, O., Théron, A., Shaban, M. M. & Moné, H. (2012). A new chronotype of *Schistosoma mansoni*: adaptive significance. *Tropical Medicine & International Health*, 17(6), 727-732. doi:10.1111/j.1365-3156.2012.02988.x
- Mouahid, G., Mintsu Nguema, R., Al Mashikhi, K. M., Al Yfae, S. A., Idris, M. A. & Moné, H. (2019). Host-parasite life-histories of the diurnal vs. nocturnal chronotypes of *Schistosoma mansoni*: adaptive significance. *Tropical Medicine & International Health*, 24(6), 692-700. doi:10.1111/tmi.13227
- Moulager, M., Monnier, A., Jesson, B., Bouvet, R., Mosser, J., Schwartz, C.,... & Bouget, F.-Y. (2007). Light-dependent regulation of cell division in *Ostreococcus*: evidence for a major transcriptional input. *Plant physiology*, 144(3), 1360-1369. doi:10.1104/pp.107.096149
- Niemann, G. M. & Lewis, F. A. (1990). *Schistosoma mansoni*: influence of *Biomphalaria glabrata* size on susceptibility to infection and resultant cercarial production. *Experimental Parasitology*, 70(3), 286-292. doi: 10.1016/0014-4894(90)90110-x
- Nishimura, H., Nakahira, Y., Imai, K., Tsuruhara, A., Kondo, H., Hayashi, H., . . . & Kondo, T. (2002). Mutations in KaiA, a clock protein, extend the period of circadian rhythm in the cyanobacterium *Synechococcus elongatus* PCC 7942. *Microbiology*, 148(9), 2903-2909. doi:10.1099/00221287-148-9-2903
- Nojima, H. & Sato, A. (1982). *Schistosoma mansoni* and *Schistosoma haematobium*: emergence of schistosome cercariae from snails with darkness and illumination. *Experimental Parasitology*, 53(2), 189-198. doi:10.1016/0014-4894(82)90060-1
- Ohtsuka, K. & Hata, M. (2000). Molecular chaperone function of mammalian Hsp70 and Hsp40-a review. *International Journal of Hyperthermia*, 16(3), 231-245. doi:10.1080/026567300285259
- Patocka, N., Sharma, N., Rashid, M. & Ribeiro, P. (2014). Serotonin signaling in *Schistosoma mansoni*: A Serotonin-activated G protein-coupled receptor controls parasite movement. *PLoS Pathogens*, 10(1), e1003878. doi:10.1371/journal.ppat.1003878
- Phan, P., Liang, D., Zhao, M., Wyeth, R. C., Fogarty, C., Duke, M. G., . . . Cummins, S. F. (2022). Analysis of rhodopsin G protein-coupled receptor orthologs reveals semiochemical peptides for parasite (*Schistosoma mansoni*) and host (*Biomphalaria glabrata*) interplay. *Scientific Reports*, 12(1), 1-11. doi.org/10.1038/s41598-022-11996-x
- Picard, M. A., Boissier, J., Roquis, D., Grunau, C., Allienne, J.-F., Duval, D., . . . & Long, T. (2016). Sex-biased transcriptome of *Schistosoma mansoni*: host-parasite interaction, genetic determinants and epigenetic regulators are associated with sexual differentiation. *PLoS Neglected Tropical Diseases*, 10(9), e0004930. doi:10.1371/journal.pntd.0004930
- Protasio, A. V., Tsai, I. J., Babbage, A., Nichol, S., Hunt, M., Aslett, M. A., . . . & Clark, R. C. (2012). A systematically improved high quality genome and transcriptome of the human blood fluke *Schistosoma mansoni*. *PLoS neglected tropical diseases*, 6(1), e1455. doi:10.1371/journal.pntd.0001455
- Queiroz, O. (1979). Les horloges biologiques. Organisation du métabolisme chez les végétaux supérieurs et adaptation au climat. *Bulletin de la Société Botanique de France. Actualités Botaniques*, 126(1), 5-19. doi:10.1080/01811789.1979.10826368
- Quinlan, A. R. & Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26(6), 841-842. doi:10.1093/bioinformatics/btq033
- Ramsey, I. S. & DeSimone, J. A. (2018). Otopetrin-1: a sour-tasting proton channel. *Journal of General Physiology*, 150(3), 379-382. doi:10.1085/jgp.201812003
- Rawlinson, K. A., Reid, A. J., Lu, Z., Driguez, P., Wawer, A., Coghlan A., ... & McCarthy C. (2021). Daily rhythms in gene expression of the human parasite *Schistosoma mansoni*. *BMC Biology*, 19(1), 1-21. doi:10.1186/s12915-021-01189-9
- Reece, S. E., Prior, K. F. & Mideo, N. (2017). The life and times of parasites: rhythms in strategies for within-host survival and between-host transmission. *Journal of Biological Rhythms*, 32(6), 516-533. doi:10.1177/0748730417718904
- Reinberg, A. (2003). Chronobiologie médicale. *Chronothérapeutique. 2nd edition. Paris: Médecine-Sciences Flammarion*, 298 p.
- Rensing, L., Meyer-Grahe, U. & Ruoff, P. (2001). Biological timing and the clock metaphor: oscillatory and hourglass mechanisms. *Chronobiology international*, 18(3), 329-369. doi:10.1081/CBI-100103961
- Rijo-Ferreira, F., Takahashi, J. S. & Figueiredo, L. M. (2017). Circadian rhythms in parasites. *PLoS pathogens*, 13(10), e1006590. doi:10.1371/journal.ppat.1006590

- Rijo-Ferreira, F., Pinto-Neves, D., Barbosa-Morais, N. L., Takahashi, J. S. & Figueiredo, L. M. (2017). *Trypanosoma brucei* metabolism is under circadian control. *Nature microbiology*, 2(6), 1-10. doi:10.1038/nmicrobiol.2017.32
- Rijo-Ferreira, F., Acosta-Rodriguez, V. A., Abel, J. H., Kornblum, I., Bento, I., Kilaru, G., . . . & Takahashi, J. S. (2020). The malaria parasite has an intrinsic clock. *Science*, 368(6492), 746-753. doi:10.1126/science.aba2658
- Ritchie, L. S., Jachowski, L. A. & Ferguson, F. F. (1960). "Bilharzia"—A Military Hazard in Puerto Rico. *Military Medicine*, 125(4), 253-257. doi:10.1093/milmed/125.4.253
- Roels, J., Kuchmiy, A., De Decker, M., Strubbe, S., Lavaert, M., Liang, K. L., . . . & Van Vlierberghe, P. (2020). Distinct and temporary-restricted epigenetic mechanisms regulate human  $\alpha\beta$  and  $\gamma\delta$  T cell development. *Nature Immunology*, 21(10), 1280-1292. doi:10.1038/s41590-020-0747-9
- Roquis, D., Taudt, A., Geyer, K. K., Padalino, G., Hoffmann, K. F., Holroyd, N., . . . & Grunau, C. (2018). Histone methylation changes are required for life cycle progression in the human parasite *Schistosoma mansoni*. *PLoS Pathogens*, 14(5), e1007066. doi:10.1371/journal.ppat.1007066
- Saigusa, T., Ishizaki, S., Watabiki, S., Ishii, N., Tanakadate, A., Tamai, Y. & Hasegawa, K. (2002). Circadian behavioural rhythm in *Caenorhabditis elegans*. *Current Biology*, 12(2), R46-47. doi:10.1016/s0960-9822(01)00669-8
- Savassi, B. A., Mouahid, G., Lasica, C., Mahaman, S.-D. K., Garcia, A., Courtin, D., . . . & Moné, H. (2020). Cattle as natural host for *Schistosoma haematobium* (Bilharz, 1852) Weinland, 1858 x *Schistosoma bovis* Sonsino, 1876 interactions, with new cercarial emergence and genetic patterns. *Parasitology Research*, 119(7), 2189-2205. doi:10.1007/s00436-020-06709-0
- Savassi, B., Dobigny, G., Etougbéché, J. R., Avocegan, T. T., Quinsou, F. T., Gauthier, P., . . . & Mouahid, G. (2021). *Mastomys natalensis* (Smith, 1834) as a natural host for *Schistosoma haematobium* (Bilharz, 1852) Weinland, 1858 x *Schistosoma bovis* Sonsino, 1876 introgressive hybrids. *Parasitology Research*, 120(5), 1755-1770. doi:10.1007/s00436-021-07099-7
- Sawa, M., Nusinow, D. A., Kay S. A. & Imaizumi, T. (2007). FKF1 and GIGANTEA complex formation is required for day-length measurement in *Arabidopsis*. *Science*, 318(5848), 261-265. doi:10.1126/science.1146994
- Schmidl, C., Rendeiro, A. F., Sheffield, N. C. & Bock, C. (2015). ChIPmentation: fast, robust, low-input ChIP-seq for histones and transcription factors. *Nature Methods*, 12(10), 963-965. doi:10.1038/nmeth.3542
- Sharma, V. K. (2003). Adaptive significance of circadian clocks. *Chronobiology International*, 20(6), 901-919. doi:10.1081/cbi-120026099
- Simões, M. C., Lee, J., Djikeng, A., Cerqueira, G. C., Zerlotini, A., da Silva-Pereira, R. A., . . . & Oliveira, G. (2011). Identification of *Schistosoma mansoni* microRNAs. *BMC Genomics*, 12(1), 1-17. doi:10.1186/1471-2164-12-47
- Singh, K., Jha, N. K. & Thakur, A. (2019). Spatiotemporal chromatin dynamics—A telltale of circadian epigenetic gene regulation. *Life Sciences*, 221, 377-391. doi:10.1016/j.lfs.2019.02.006
- Steinegger, M., Meier, M., Mirdita, M., Vöhringer, H., Haunsberger, S. J. & Söding, J. (2019). HH-suite3 for fast remote homology detection and deep protein annotation. *BMC bioinformatics*, 20(1), 1-15. doi:10.1186/s12859-019-3019-7
- Supek, F., Bošnjak, M., Škunca, N. & Šmuc, T. (2011). REVIGO summarizes and visualizes long lists of gene ontology terms. *PloS One*, 6(7), e21800. doi:10.1371/journal.pone.0021800
- Takahashi, J. S. (2015). Molecular components of the circadian clock in mammals. *Diabetes, Obesity and Metabolism*, 17, 6-11. doi:10.1093/hmg/ddl207
- Takahashi, J. S., Kumar, V., Nakashe, P., Koike, N., Huang, H.-C., Green, C. B. & Kim, T.-K. (2015). ChIP-seq and RNA-seq methods to study circadian control of transcription in mammals. *In Methods in enzymology* (Vol. 551, pp. 285-321): Elsevier. doi:10.1016/bs.mie.2014.10.059
- Tan, S. & Ahana, A. (2007). Theodor Bilharz (1825-1862): discoverer of schistosomiasis. *Singapore medical journal*, 48(3), 184-185.
- Théron, A. (1980). Mise en évidence de races chronobiologiques de *Schistosoma mansoni*, agent de la bilharziose, à partir des cinétiques d'émission cercarienne. *Comptes-Rendus de l'Académie des Sciences, Paris*, 291, 279-282.
- Théron, A. (1982). *Le compartiment cercaire dans le cycle de Schistosoma mansoni Sambon, 1907: écologie de la transmission bilharzienne en Guadeloupe*: Thèse d'état.
- Théron, A. (1984). Early and late shedding patterns of *Schistosoma mansoni* cercariae: ecological significance in transmission to human and murine hosts. *The Journal of Parasitology*, 70(5), 652-655. doi:10.2307/3281744
- Théron, A. (1989). Hybrids between *Schistosoma mansoni* and *S. rodhaini*: characterization by cercarial emergence rhythms. *Parasitology*, 99(2), 225-228. doi:10.1017/s0031182000058674
- Théron, A. (2015). Chronobiology of trematode cercarial emergence: from data recovery to epidemiological, ecological and evolutionary implications. *Advances in Parasitology*, 88, 123-164. doi:10.1016/bs.apar.2015.02.003

- Théron, A. & Combes, C. (1983). Genetic analysis of the shedding pattern of *Schistosoma mansoni* cercariae by hybridization of species with early and late emission peaks. *Comptes-Rendus de l'Académie des Sciences III, Paris*, 297(12), 571-574.
- Théron, A. & Combes, C. (1988). Genetic analysis of cercarial emergence rhythms of *Schistosoma mansoni*. *Behavior Genetics*, 18(2), 201-209. doi:10.1007/bf01067842
- Théron, A., Mouahid, G. & Moné, H. (1997). *Schistosoma mansoni*: cercarial shedding patterns from a mixed infection of *Biomphalaria glabrata* with two (early and late) chronobiological variants. *Parasitology Research* 83(4),3 56-358.
- Thommen, Q., Pfeuty, B., Corellou, F., Bouget, F. Y. & Lefranc, M. (2012). Robust and flexible response of the *Ostreococcus tauri* circadian clock to light/dark cycles of varying photoperiod. *Federation of European Biochemical Societies Journal*, 279(18), 3432-3448. doi:10.1111/j.1742-4658.2012.08666.x
- Ueta, T., Inoue, T., Furukawa, T., Tamaki, Y., Nakagawa, Y., Imai, H. & Yanagi, Y. (2012). Glutathione peroxidase 4 is required for maturation of photoreceptor cells. *Journal of Biological Chemistry*, 287(10), 7675-7682. doi:10.1074/jbc.M111.335174
- Valle, C. M., Pellegrino, J. & Alvarenga, N. (1973). Rhythmic emergence of *Schistosoma mansoni* cercariae from *Biomphalaria glabrata*: influence of the temperature. *Revista do Instituto de Medicina Tropical de São Paulo*, 15(4), 195-201.
- Wang, S. R., Zhu, Y. J., Ge, Q. P., Yang, M. J., Huang, J. L., Huang, W. Q.,...& Lu, D. B. (2015). Effect of photoperiod change on chronobiology of cercarial emergence of *Schistosoma japonicum* derived from hilly and marshy regions of China. *Experimental Parasitology*, 159,227-232. doi:10.1016/j.exppara.2015.10.004
- Weasner, B., Salzer, C. & Kumar, J. P. (2007). Sine oculis, a member of the SIX family of transcription factors, directs eye formation. *Developmental biology*, 303(2), 756-771. doi:10.1016/j.ydbio.2006.10.040
- WHO (2021). Lutter contre les maladies tropicales négligées pour atteindre les objectifs de développement durable: feuille de route pour les maladies tropicales négligées 2021–2030. 40 p. Retrieved from <https://www.who.int/fr/publications-detail/9789240019027>
- WHO (2022). Fact sheet. Retrieved from <https://www.who.int/news-room/fact-sheets/detail/schistosomiasis>
- Williams, C. L. & Gilbertson, D. E. (1983). Altered feeding response as a cause for the altered heartbeat rate and locomotor activity of *Schistosoma mansoni*-infected *Biomphalaria glabrata*. *The Journal of Parasitology*, 671-676.
- Williams, C. L., Wessels, W. S. & Gilbertson, D. E. (1984). Comparison of the rhythmic emergence of *Schistosoma mansoni* cercariae from *Biomphalaria glabrata* in different lighting regimens. *The Journal of Parasitology*, 70(3), 450-452. doi:10.2307/3281585
- Wu, G., Anafi, R. C., Hughes, M. E., Kornacker, K. & Hogenesch, J. B. (2016). MetaCycle: an integrated R package to evaluate periodicity in large scale data. *Bioinformatics*, 32(21), 3351-3353. doi:10.1093/bioinformatics/btw405
- Xu, Y., Mori, T. & Johnson, C. H. (2003). Cyanobacterial circadian clockwork: roles of KaiA, KaiB and the kaiBC promoter in regulating KaiC. *Embo j*, 22(9), 2117-2126. doi:10.1093/emboj/cdg168
- Zhang, Y., Liu, T., Meyer, C., Eeckhoute, J., Johnson, D., Bernstein, B., . . . & Li, W. (2008). Model-based analysis of chip-seq (macs). *Genome Biology*, 9(9). R137. doi:10.1186/gb-2008-9-9-r137
- Ziskind, B. (2009). Urinary schistosomiasis in ancient Egypt. *Néphrologie & Thérapeutique*, 5(7), 658-661. doi:10.1016/j.nephro.2009.06.001





# Annexe 1 : Article sur la chronobiologie comme outil de mise en évidence d'un événement d'hybridation

Parasitology Research  
<https://doi.org/10.1007/s00436-020-06709-0>

HELMINTHOLOGY - ORIGINAL PAPER



## Cattle as natural host for *Schistosoma haematobium* (Bilharz, 1852) Weinland, 1858 x *Schistosoma bovis* Sonsino, 1876 interactions, with new cercarial emergence and genetic patterns

Boris A.E.S. Savassi<sup>1,2</sup> · Gabriel Mouahid<sup>1</sup> · Chrystelle Lasica<sup>1</sup> · Samoussou-Dine K. Mahaman<sup>2</sup> · André Garcia<sup>3</sup> · David Courtin<sup>4</sup> · Jean-François Allienne<sup>1</sup> · Moudachirou Ibikounlé<sup>2</sup> · Hélène Moné<sup>1</sup>

Received: 16 December 2019 / Accepted: 5 May 2020  
 © Springer-Verlag GmbH Germany, part of Springer Nature 2020

### Abstract

Schistosomiasis remains a parasitic infection which poses serious public health consequences around the world, particularly on the African continent where cases of introgression/hybridization between human and cattle schistosomiasis are being discovered on a more frequent basis in humans, specifically between *Schistosoma haematobium* and *S. bovis*. The aim of this paper is to analyze the occurrence of *S. bovis* in cattle and its relationship with *S. haematobium* in an area where cattle and humans share the same site in Benin (West Africa). We used the chronobiology of cercarial emergence as an ecological parameter and both molecular biology (COI mtDNA and ITS rDNA) of the larvae and morphology of the eggs as taxonomic parameters. The results showed a chronobiological polymorphism in the cercarial emergence rhythm. They showed for the first time the presence of *S. bovis* in Benin, the presence of introgressive hybridization between *S. bovis* and *S. haematobium* in domestic cattle, and the presence of atypical chronobiological patterns in schistosomes from cattle, with typical *S. haematobium* shedding pattern, double-peak patterns, and nocturnal patterns. Our results showed that the chronobiological life-history trait is useful for the detection of new hosts and also may reveal the possible presence of introgressive hybridization in schistosomes. Our results, for the first time, place cattle as reservoir host for *S. haematobium* and *S. bovis* x *S. haematobium*. The consequences of these results on the epidemiology of the disease, the transmission to humans, and the control of the disease are very important.

**Keywords** *Schistosoma bovis* · *Schistosoma haematobium* · Introgressive hybridization · Cercarial emergence pattern · Cattle · Benin

### Introduction

Schistosomiasis or Bilharziasis remains a parasitic infection which poses serious public health consequences around the

world, particularly on the African continent which harbors 85% of the total number of infected people worldwide (Chitsulo et al. 2000; Engels et al. 2002). In Africa, humans can be infected with four species of *Schistosoma*: *Schistosoma mansoni* Sambon, 1907, *S. haematobium*, *S. intercalatum* Fisher, 1934 and *S. guineensis* Pagès, Jourdan, Southgate & Tchuem Tchuente, 2003 where *S. haematobium* is the most prevalent species affecting around 112 million people (WHO 2019). Cattle schistosomiasis affects approximately 165 million domestic cattle worldwide and the disease is of veterinary and economic significances (De Bont and Vercruyse 1997). In Africa, species such as *S. bovis*, *S. curassoni* Brumpt, 1931, *S. matthei* Veglia & Le Roux 1929, and *S. leiperi* Le Roux, 1955 are responsible for inducing severe infections in animals (De Bont and Vercruyse 1998). Africa is also the continent where cases of introgression/hybridization between human and cattle schistosomes are being discovered on a more frequent basis, specifically between *S. haematobium* and

Section Editor: Hiroshi Sato

✉ Hélène Moné  
 mone@univ-perp.fr

<sup>1</sup> University of Perpignan Via Domitia, IHPE UMR 5244, CNRS, IFREMER, University of Montpellier, 58 avenue Paul Alduy, Bât. R, F-66860 Perpignan, France

<sup>2</sup> Faculté des Sciences et Techniques, Université d'Abomey-Calavi, 01BP526 Cotonou, Benin

<sup>3</sup> UMR 261 MERIT, Institut de Recherche pour le Développement (IRD), Cotonou, Bénin

<sup>4</sup> UMR 261 MERIT, Institut de Recherche pour le Développement (IRD), Université de Paris, Paris, France

*S. bovis*. All but one of these cases were found in humans in West Africa: in Niger (introgression of *S. haematobium* by genes of *S. bovis*, Brémond et al. 1993; Leger et al. 2016), in Senegal (bidirectional introgression, Huyse et al. 2009; Webster et al. 2013), in Mali (hybrids, Soentjens et al. 2016), in Côte d'Ivoire (Angora et al. 2019), and in Benin (introgressive hybridization, Moné et al. 2015). The outlying case was from Corsica (France) (introgressive hybridization, Moné et al. 2015). Hybridization between *S. haematobium* and *S. bovis* in a non-human host was found in one rodent, *Mastomys huberti* Wroughton, 1909 in Senegal, which was shown to harbor a hybrid female (Catalano et al. 2018).

In Benin, two species of human schistosomes are reported in the literature: *S. haematobium* and *S. mansoni* (Moné et al. 2010): *S. haematobium* is widely distributed, with a national prevalence of 17.6%, while *S. mansoni* has a narrow distribution, with a prevalence of 2.5% (Onzo-Aboki et al. 2019). Introgressive hybridization between *S. bovis* and *S. haematobium* was already reported for humans in Benin (Moné et al. 2015) but, as a species, *S. bovis* was never reported in this country. This is interesting given that all of Benin's neighboring countries including Togo, Burkina Faso, Niger, and Nigeria were known to harbor this bovine species (Moné et al. 1999). In many villages in Benin, humans and cattle live in close proximity. Since the snails from the genus *Bulinus* are known to act as intermediate hosts for both *S. haematobium* and *S. bovis*, a logical question to consider is whether cattle in these areas could be infected by *S. bovis* or *S. haematobium* and if introgressive hybridization exists between the two species.

The aim of this paper is to analyze the occurrence of *S. bovis* in cattle and its relationship with *S. haematobium* in an area where cattle and humans share the same site (Fig. 1). We used the chronobiology of cercarial emergence as an ecological parameter, as the cercarial emission is related to the definitive host behavior (Théron et al. 1997; Ibikounlé et al. 2012; Mouahid et al. 2012; Mintsá-Nguema et al. 2014), the emission pattern of *S. bovis* cercariae is different (early diurnal, Mouahid et al. 1991) from that of *S. haematobium* cercariae (midday to late diurnal, Mouahid et al. 1991), and since schistosome cercarial emission has been shown to have a genetic base (Théron and Combes 1988). We also used both molecular biology (cytochrome *c* oxidase (COI mtDNA) and internal transcribed spacer (ITS rDNA)) of the larvae and morphology of the eggs as taxonomic parameters.

## Materials and methods

### Geographical sampling position

The study was conducted on cows and schoolchildren (elementary school; latitude N: 6°34'39.84"; longitude E: 2°31'

9.03") from 2017 to 2018 in Kessounou Village, located in the department of Ouémé in southeastern Benin (Fig. 2). The district of Kessounou is a fluvio-lacustrine zone characterized by a four-season subequatorial climate (Abou et al. 2018): (i) a long rainy season from March to June, (ii) a short dry season in July and August, (iii) a short rainy season in September and October, and finally (iv) a long dry season that runs from November to February. The Ouémé River crosses Kessounou and periodically floods the entire area during its flood period (mid-July to early November). During the period where river water levels are low (December to June), almost all of the water disappears giving way to ponds in various places. The climatic characteristics observed at Kessounou favor the seasonal installation of freshwater snails, especially during the flood period. Adults engage in fishing, agriculture, and livestock breeding. Schoolchildren regularly fish and search for crabs and shrimps for sale, which places them in permanent contact with contaminated waters. Cattle breeding is done in collectivity. Cows are left to wander even during periods of flooding and they never leave Kessounou unless they are sold. According to the breeders, most of the cows were born in Kessounou and some have spent more than 20 years in the one location.

### Schistosoma egg and miracidium recoveries from cattle

The feces from 48 cows were collected individually (from 150 to 250 g), immediately after defecation. The infection status of each fecal sample was analyzed using the miracidial hatching test from 50 g of feces. The filtration technique consisted of placing the 50 g of feces in a container with half a liter of NaCl 9‰ saline solution. The solution was then homogenized using an electric hand mixer, passed through a series of decreasing mesh vacuum sieves (315 µm, 180 µm, 106 µm, and 45 µm), then washed with NaCl 9‰ saline solution using a pressure pump. The residues retained on the 45 µm pore size sieve were placed in a beaker containing drilling water, at room temperature under light, for miracidial hatching. Among the 48 cows, 35 were found positive (73%). We chose to continue the analyses with the four fecal samples that gave the most miracidia (BK19, BK24, BK26, and BK29). The rest of the fecal sample from each of the four cows was then filtered as above and separated into two subsamples: the first subsample was once again passed through the 45 µm pore size sieve, and the eggs were placed in a beaker containing drilling water for miracidial hatching. Several miracidia from each cow were collected individually onto FTA® Classic Cards (Whatman, GE Healthcare companies, Little Chalfont, United Kingdom) and then transferred to University of Perpignan (France) for molecular analyses; the rest of the miracidia from the four cows were mixed and used for snail infection; the second subsamples from the four cows were once again individually

**Fig. 1** Kessounou sampling area where cows access water in close proximity to houses

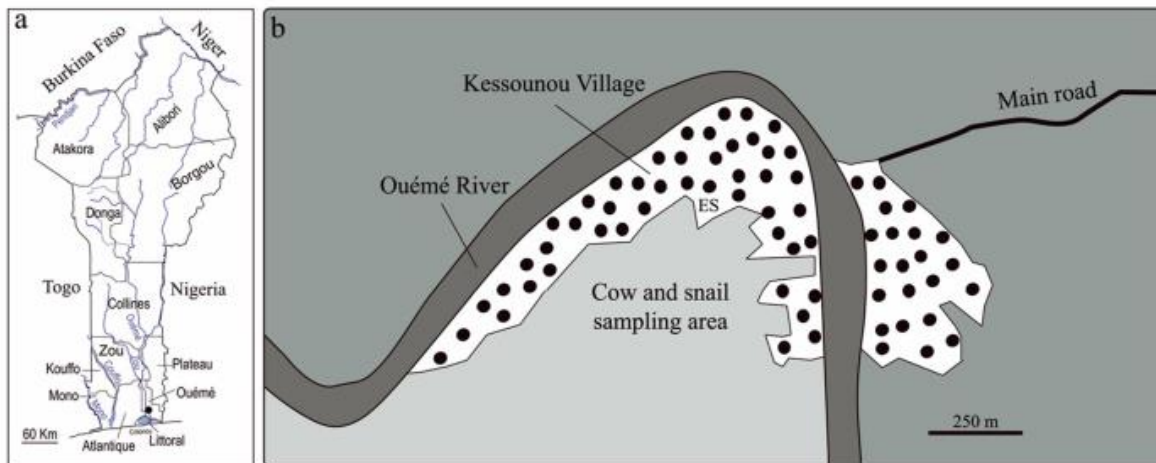


passed through the 45  $\mu\text{m}$  pore size sieve in 9‰ NaCl saline solution and the eggs from the four subsamples were mixed for an egg morphology study.

#### **Schistosoma egg and miracidium recoveries from schoolchildren**

Urine samples from 76 schoolchildren were collected individually. The infection status of each urine sample was analyzed using the syringe filtration technique (Boko et al. 2016; Onzo-Aboki et al. 2019) after staining each filter with Lugol solution. One of the 18 girls' samples (5.6%) and 4 of the 58 boys (6.9%) were found positive. The rest of the

five positive urine-egg samples (DK20, DK23, DK54, KE14, and KE15) were individually filtered according to the same procedure as above, and the filters were placed in saline solution. Two subsamples were made from each urine sample: the first subsample was passed through a 45  $\mu\text{m}$  pore size sieve and the eggs were placed in a beaker containing drilling water for miracidial hatching. Several miracidia were collected individually for each urine sample in Eppendorf tubes (only one miracidium per tube), containing 95° alcohol, then transferred to the University of Perpignan (France) for molecular analyses. The second subsamples from the five urine samples were mixed in saline solution for an egg morphology study.



**Fig. 2** Sampling area. **a** Map of Benin; Kessounou Village (black circle) in Ouémé Department. **b** Kessounou Village (white with black circles). ES: Elementary School

The stools from the four urine-egg-positive boys were collected individually. Each stool was passed through a series of sieves (315, 180, 106, and 45 pore size) and washed through with 9‰ NaCl saline solution with a pressure pump.

The eggs that were retained on the 45 µm pore size sieve were collected in a beaker containing saline solution. The presence of eggs was ascertained under a binocular microscope. Only one stool was found positive and two subsamples were made: the first subsample was passed through the 45 µm pore size sieve and the eggs were placed in a beaker containing drilling water for miracidial hatching. Several miracidia were collected individually in Eppendorf tubes (only one miracidium per tube) containing 95° alcohol, then transferred to the University of Perpignan (France) for molecular analyses. The second subsample was kept in 9‰ NaCl saline solution for an egg morphology study.

### Snail infection

Two species of freshwater snails, *Bulinus forskalii* (Ehrenberg, 1831) and *Bulinus globosus* (Morelet, 1866), were collected in Kessounou Village and used in this study. These species were identified using shell morphology (Brown 1994). All of the collected snails were tested for infection with schistosomes or other trematodes for a period of 55 days. None of the snails used for the infection were found to be naturally infected. We exposed 43 *Bulinus forskalii* (6.3 to 8.2 mm in height) and 62 *Bulinus globosus* (4.4 to 7.6 mm in height), individually, to 3 miracidia obtained from the eggs originating from a mixture of feces from the 4 cows (BK19, BK24, BK26, and BK29). The snails of each species were maintained separately, fed with dry and fresh salad in water at 28 °C with a balanced photoperiod (12 h light, 12 h dark) and a photophase from 6:00 am to 6:00 pm. The snails were then transferred to Perpignan (France) for cercarial emission and molecular analyses.

### Cercarial emergence pattern

The snails were maintained individually in glass vessels containing 150 ml of drilling water, at a constant temperature (25 °C), balanced photoperiod (light/dark: 12 h/12 h), and photophase (6:00 am to 6:00 pm). They were given fresh washed lettuce ad libitum. The cercarial emission rhythm (chronobiology) was studied for five consecutive days for 13 infected snails, 11 *Bulinus forskalii* (Kf1 to Kf11), and 2 *Bulinus globosus* (Kg1 and Kg2) to show the stability of the emission pattern, from the 6th day after the first day of the patent period, i.e., 32 to 38 days after snail exposure to the miracidia. Each hour, each snail, with its lettuce, was transferred into a new vessel. The water left in the vessel, containing the cercariae, was transferred to a container where a few drops of Lugol solution was added to kill the cercariae and

red-color them; this solution was then filtered through a Nitrel polyamide filter (25 µm pore size). The cercariae retained on the filter were counted under the binocular microscope.

Cercariae emitted by four infected *Bulinus forskalii* (Kf5, Kf6, Kf8, and Kf9) and the two *Bulinus globosus* (Kg1 and Kg2) were individually stored onto FTA® Classic Cards (Whatman, GE Healthcare companies, Little Chalfont, United Kingdom), as described above, for molecular studies.

## Molecular analyses

### DNA extraction

DNA from miracidia collected in alcohol was extracted according Beltran et al. (2008) and stored at -20 °C. DNA from either miracidia or cercariae individually stored on FTA® Classic Cards (Whatman, GE Healthcare companies, Little Chalfont, United Kingdom) was extracted as follows. A 3-mm disk was removed with a Craft Punch from the FTA® Classic Cards at the center of where the miracidium or cercaria was loaded. The disks were deposited in Eppendorf tubes (1.5 ml) and an initial washing step with 100 µl of Milli-Q water was performed. After a 10-min incubation at room temperature, water was removed and replaced with 80 µl of 5% Chelex® 100 Molecular Biology Grade Resin solution (Bio-Rad Laboratories, Hercules, California, USA). The sample was then heated to 65 °C for 30 min at a stirring speed of 800 rpm followed by a second heating at 99 °C for 8 min without stirring. DNA was finally collected by centrifugation at 14,000 rpm for 2 min. Fifty microliters of the supernatant (DNA) of each sample was taken and stored at -20 °C for further molecular analyses.

### Rapid diagnostic multiplex PCR

A rapid diagnostic multiplex PCR was used to target the COI mtDNA gene. We used the technique of Webster et al. (2010), optimized by Van den Broeck et al. (2011) and Angora et al. (2019). The primers were one universal reverse primer (Shmb.R: 5'-CAA GTA TCA TGA AAY ART ATR TCT AA-3') and three species-specific forward primers (for *S. haematobium* (120 bp) Sh.F: 5'-GGT CTC GTG TAT GAG ATC CTA TAG TTT G-3', for *S. bovis* (260 bp) Sb.F: 5'-GTT TAG GTA GTG TAG TTT GGG CTC AC-3' and for *S. mansoni* (215 bp) Sm.F: 5'-CTT TGA TTC GTT AAC TGG AGT G-3'). Each PCR amplification of partial COI mtDNA was carried out in a total volume of 10 µl, containing 2 µl of DNA template, 2 µl of the GoTaq® Flexy buffer (Promega; Madison, Wisconsin, USA), 0.6 µl of 25 mM MgCl<sub>2</sub>, 1 µl of 10× primer mix (4 µl of 100 µM universal reverse primer, 4 µl of each 100 µM forward primer, and 84 µl of milli-Q water), 0.2 µl solutions of dNTP at 10 mM each, 0.2 µl of the GoTaq® G2 Hot Start Taq Polymerase

(Promega), and 4 µl milli-Q water. The PCR conditions involved an initial phase of activation of GoTaq Polymerase at 95 °C for 3 min followed by 45 amplification cycles. Each cycle comprised DNA denaturation step at 95 °C for 10 s, followed by primer annealing at 52 °C for 30 s, and an elongation step at 72 °C for 10 s. The program ended with a final extension at 72 °C for 2 min. The PCR products were examined on 2.5% agarose gels at 135 V for 40 min using the 100 bp DNA size marker (Promega) for size estimation.

#### COI mtDNA PCR

Partial COI mtDNA amplification was performed by PCR using the forward primer, *Cox1\_schist* F: 5'-TCTT TRGATCATAAGCG-3', and the reverse primer, *Cox1\_schist*-R: 5'-TAATGCATMGGAAAAAACA-3' (Lockyer et al. 2003). PCR conditions used were those as described in Moné et al. (2015).

#### ITS rDNA PCR

PCR amplification of ITS rDNA (partial 18S, ITS1, 5.8S, and ITS2) was performed using the primers of Barber et al. (2000): ITS4F: 5'-TCCTCCGCTTATTGATATGC-3'; ITS5R: 5'-GGAAGTAAAAGTCGTAACAAGG-3'. The amplification was carried out in a total reaction volume of 25 µl, containing 2 µl of DNA; 1.5 µl of MgCl<sub>2</sub> (25 mM), 5 µl of the 5× buffer (Promega), 1 µl of each of the primers (10 µM), 0.5 µl of each dNTP (10 mM), 0.2 µl of Go Taq polymerase (Promega), and 13.8 µl of H<sub>2</sub>O using a thermocycler (Techne TC-Plus, Bibby Scientific, Staffordshire, UK). The reaction conditions included an activation step at 95 °C for 3 min, followed by 45 amplification cycles. Each cycle comprised a DNA denaturation step at 95 °C for 40 s, followed by primer annealing at 48 °C for 40 s and an elongation step at 72 °C for 1 min 10 s and a final extension at 72 °C for 5 min. One percent agarose gel electrophoresis was used to visualize the ITS rDNA-PCR products.

#### Sequencing

Partial COI mtDNA gene and ITS rDNA region were sequenced (Genoscreen; Lille, France) using the reverse primers; some regions were also sequenced using the forward primer in order to confirm the sequence. The sequences were manually edited using Sequencher 4.5 (Gene Codes Corporation, Ann Arbor, USA). For the analysis of nuclear gene sequences (ITS rDNA), the sequence polymorphism was verified and confirmed by visualization of the raw sequence chromatograms. Since this gene has biparental transmission and segregation sites (polymorphic sites) between *S. haematobium* and *S. bovis*, the sequence chromatograms were carefully checked to identify the presence of possible

heterozygosity. At each polymorphic site where two chromatogram peaks overlapped (indicating the genetic signature of both parents), the IUPAC ambiguity symbols were used to indicate the individual nucleotide polymorphisms. Thus, Y indicates the presence of the bases T and C, rather than an ambiguous reading between T and C. Similarly, R indicates the presence of the bases A and G and not an ambiguous reading between A and G. The different haplotypes and the different profiles are presented in Table 1 for the COI mtDNA gene and the ITS rDNA region, respectively.

#### Partial COI mtDNA phylogenetic analysis

DNA multiple sequence alignments were performed using Muscle program (Edgar 2004) on the MEGA 7.0 software (Kumar et al. 2016) and refined by Gblocks 0.91b (Castresana 2000; Dereeper et al. 2008, 2010). The probabilistic model of sequence evolution (Nei and Kumar 2000) and the gamma distribution (G) to approximate rate heterogeneity among haplotypes was performed using MEGA 7.0 software. The best model with the lowest BIC score (Bayesian Information Criterion) was HKY + G (Hasegawa-Kishino-Yano with the gamma distribution). A phylogenetic tree using the maximum likelihood method was constructed. The maximum likelihood method was performed using MEGA 7.0 software. Reliability for internal branch was assessed using bootstrapping procedure (1000 replicates). The topology of the tree was rooted by *S. intercalatum*. The comparison was also performed using the percentage of difference calculated from the pairwise distances data obtained by using MEGA 7.0 software.

#### Egg morphology

The eggs were mounted individually at random in a 9% NaCl solution beneath glass cover-slips on glass slides using a Pasteur pipette. They were photographed and measured (length, width, spine length, and length/width ratio) by microscopy. Only the eggs which contained a living miracidium were measured. In total, 113 eggs were measured: 47 from cow feces, 48 from schoolchildren urines, and 18 from the schoolchildren stool.

#### Statistical analyses

Means and standard errors were calculated. The Mann-Whitney test was used for measure comparisons and the Fisher's exact test was used for proportion comparisons using the BiostaTGV <https://biostatgv.sentiweb.fr>. The probability values (*p*) less than 5% were considered statistically significant.

**Table 1** Origin and sample label of the miracidia and cercariae used for the molecular analyses. Number of haplotypes/number of sequenced larval stages (either miracidia or cercariae) for the COI mtDNA gene and number of profiles/number of sequenced larval stages (either miracidia or cercariae) for the ITS rDNA region

Host_Biological material_Larval stage	Sample label	COI mtDNA	ITS rDNA
Cow_feces_miracidium	BK19	3/5	2/5
	BK24	2/5	3/5
	BK26	2/5	3/5
	BK29	3/5	2/5
<i>B. forskalii</i> exposed to miracidia from cows_Cercaria	Kf5	1/5	1/5
	Kf6	1/3	1/3
	Kf8	1/3	1/3
	Kf9	2/3	1/3
<i>B. globosus</i> exposed to miracidia from cows_Cercaria	Kg1	1/5	1/5
	Kg2	1/5	1/5
Child_urine_miracidium	DK20	1/3	–
	DK23	4/6	5/6
	DK54	2/2	1/2
	KE14	2/11	6/14
	KE15	4/11	5/10
Child_stool_miracidium	KE14	3/14	7/13

BK, Kf, and Kg: cow origin; DK and KE: schoolchild origin; “–”: not available

## Results

### Snail infection

Of the 43 *B. forskalii* that were exposed, 16 survived and 11 shed cercariae (69%) (Kf1 to Kf11); of the 62 *B. globosus* that were exposed, 52 survived and only 2 shed cercariae (2.9%) (Kg1 and Kg2). The percentage of infection was significantly higher for *B. forskalii* compared to *B. globosus* (Fisher's exact test;  $p < 0.0001$ ). Prepatent periods lasted 26 days (from the *B. forskalii* Kf1, Kf2, Kf3, Kf4, Kf5, Kf6, Kf7, and Kf8 and from the *B. globosus* Kg1), 30 days (from the *B. forskalii* Kf9 and Kf10), or 32 days (from the *B. forskalii* Kf11 and from the *B. globosus* Kg2).

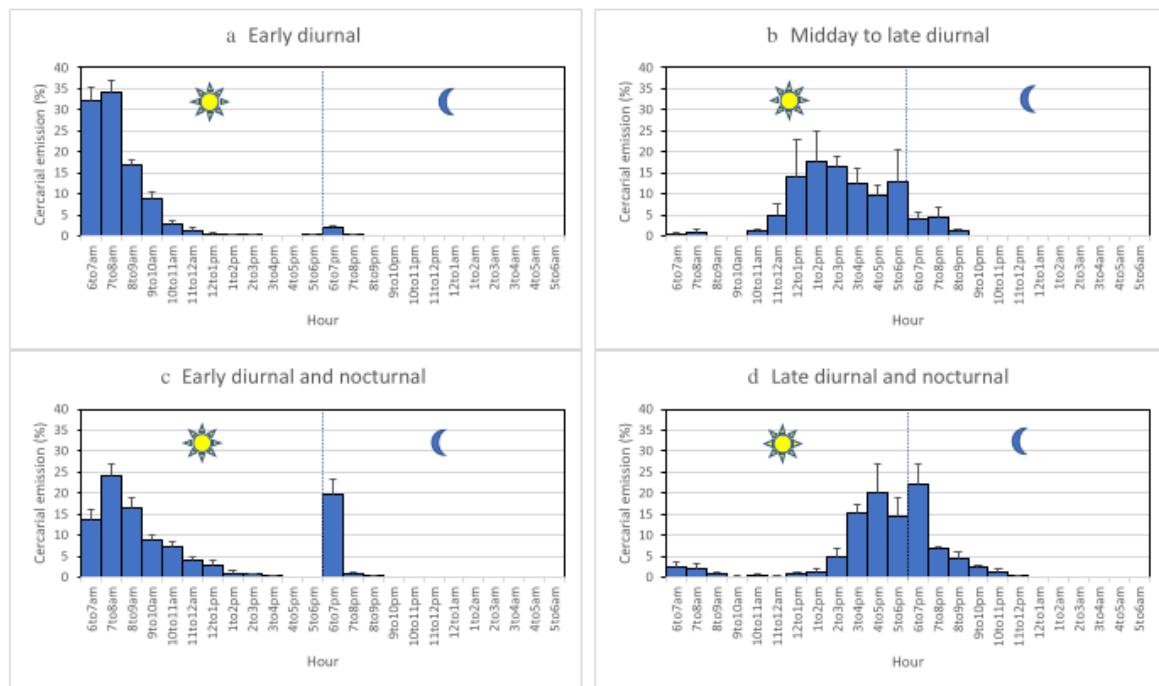
### Cercarial emission patterns

The cercarial emission patterns were analyzed for the 13 positive snails (11 *B. forskalii* and 2 *B. globosus*) during 5 consecutive days (65 occurrences in total) from the 6th day after the first day of the patent period, i.e. 32 to 38 days after snail exposure to the miracidia. The results showed a variability in the patterns; we identified four different patterns (Fig. 3), and each snail harbored only one pattern.

a- An early diurnal pattern was observed for 7 of the 13 snails (53.8%). Cercarial emission began at 6 am, immediately after the start of the light period; it increased very rapidly to reach an emission peak at either 7 am (for Kf1

and Kf11) or 8 am (for Kf2, Kf3, Kf6, Kf7, and Kf8), then decreased gradually to stop at 2 pm. Immediately after the beginning of the dark period, very low cercarial emissions were observed at 7 pm, and sometimes at 8 pm as well. The mean pattern showed that more than 97% of the cercariae were shed during the day and a peak occurred at 8 am (Fig. 3a);

- b- A midday to late diurnal pattern was found for 1 of the 13 snails (7.7%). More than 90% of the cercariae were shed during the day. Cercarial emission mainly occurred from 12 am to 6 pm with an emission peak at 2 pm (Kg2), followed by a gradual decrease during the daylight hours and then a rapid decrease at the beginning of the dark period (Fig. 3b);
- c- An early diurnal and nocturnal pattern was found for 4 of the 13 snails (30.8%). The percentages of cercariae shed during the day were highly variable and represented from more than 90% (as for the two previous chronotypes) to less than 60%. The diurnal cercarial emission occurred during the 4 to 5 first hours of light with peaks at 7 am (for Kf5), 8 am (for Kf4 and Kf9), or 9 am (for Kf10). The nocturnal cercarial emission occurred mainly during the first hour of darkness where another peak was observed for the 4 snails for which Kf5 represented more than 40% of the total cercarial daily emission. The mean pattern had more than 79% of the cercariae that were shed during the day and harbored two peaks, an early peak at 8 am and a nocturnal peak at 7 pm (Fig. 3c);
- d- A late diurnal and nocturnal pattern for 1 of the 13 snails (7.7%). Roughly 60% of the cercariae were shed during



**Fig. 3** Cercarial emission patterns from infected snails, each exposed to 3 miracidia of *Schistosoma* from cow feces. **a** Early diurnal pattern for *B. forskalii* (Kf1, Kf2, Kf3, Kf6, Kf7, Kf8 and Kf11). **b** Midday to late

diurnal pattern for *B. globosus* (Kg2). **c** Early diurnal and nocturnal pattern for *B. forskalii* (Kf4, Kf5, Kf9, Kf10). **d** Late diurnal and nocturnal pattern for *B. globosus* (Kg1)

the day. Cercarial emission occurred mainly from 4 pm to 7 pm with an emission peak at 7 pm (Kg1) (Fig. 3d).

and K\_Cow\_Hap6: accession number MT159594) or schoolchild (K\_Child\_Hap10: accession number MT159598, K\_Child\_Hap12: accession number MT159600 and K\_Child\_Hap13: accession number MT159601)) or was shared by different cows (K\_Cow\_Hap3: accession number MT159591, K\_Cow\_Hap4: accession number MT159592 and K\_Cow\_Hap5: accession number MT159593) or different schoolchildren (K\_Child\_Hap7: accession number MT159595, K\_Child\_Hap8: accession number MT159596, K\_Child\_Hap9: accession number MT159597 and K\_Child\_Hap11: accession number MT159599). Three different haplotypes were found from the snails among the six recovered in the cows; they were different between *B. forskalii* (K\_Cow\_Hap1: accession number MT159589 and K\_Cow\_Hap2: accession number MT159590) and *B. globosus* (K\_Cow\_Hap4: accession number MT159592). Two haplotypes (K\_Child\_Hap8: accession number MT159596 and K\_Child\_Hap9: accession number MT159597) were shared by miracidia that came from the same child in both the urine and the stool.

### COI mtDNA (894 bp)

Ninety-one sequences were analyzed. The numbers of haplotypes on the numbers of sequenced larval stages for each sample (4 cows, 6 snails, and 5 schoolchildren) are presented in Table 1. We found from 2 to 3 haplotypes per cow, from 1 to 2 haplotypes per snail, and from 1 to 4 haplotypes per child. In total, 13 haplotypes were found: 6 haplotypes from cows (K\_Cow\_Hap1 to K\_Cow\_Hap6) which contained 13 mutation sites, all different from the 7 haplotypes from schoolchildren (K\_Child\_Hap7 to K\_Child\_Hap13), containing 16 mutation sites (Table 2). Sequence data were deposited in the NCBI GenBank database under the accession numbers MT159589 to MT159601. The percentages of differences ranged between 0.11 and 1.36% between the cow and the human haplotypes. Site polymorphism was shared between cows and schoolchildren in seven different positions while six positions were polymorph only in cows and only eight in schoolchildren. One haplotype was harbored by only one individual (either cow (K\_Cow\_Hap1: accession number MT159589, K\_Cow\_Hap2: accession number MT159590

### ITS rDNA (946 bp)

Eighty-nine sequences were analyzed for the 18S, ITS1, 5.8S and ITS2 genes. The first 22 bp of our ITS rDNA sequences belonged to the 18S gene and no variability was found

**Table 2** The 13 haplotypes of COI mtDNA gene, harbored by cow (BK, Kf and Kg) and schoolchild (DK and KE) in Kessounou Village, Benin

Host	Haplotype	22 polymorphic positions (presented in vertical numbers)																						Sample label (Number of sequenced larval stages)																																		
		2	4	8	4	14	14	4	15	0	19	5	2	9	4	3	0	9	3	2	4	3	3		3	6	9	4	3	2	5	0	1	5	9	4	6	3	0	6	9	0	6	9	9	7	2	6	7	7	7	7	7	8	2	7	9	2
Cow	K_Cow_Hap1	A	A	G	A	C	G	A	G	T	G	C	G	A	T	T	G	T	T	C	G	T	BK19(2); Kf9(2)																																			
	K_Cow_Hap2	.	.	A	.	T	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	BK19(1); Kf5(5); Kf6(3); Kf8(3); Kf9(1)																																			
	K_Cow_Hap3	.	.	A	G	T	A	.	.	C	T	.	.	G	.	C	.	.	.	.	.	.	BK19(2); BK26(1); BK29(2)																																			
	K_Cow_Hap4	.	.	A	G	T	A	.	.	C	T	.	.	.	.	C	.	C	.	.	.	.	BK24(4); BK26(4); Kg1(5); Kg2(5)																																			
	K_Cow_Hap5	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	BK24(1); BK29(2)																																			
	K_Cow_Hap6	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	BK29(1)																																			
Child	K_Child_Hap7	G	A	G	T	A	.	.	.	C	T	.	.	.	C	.	.	.	C	.	.	DK20(3); DK23(1); KE15(1)																																				
	K_Child_Hap8	.	.	A	G	T	A	.	.	C	T	.	.	.	.	.	C	.	C	.	.	.	DK23(3); DK54(1); KE14 (urine(3) and stool(2))																																			
	K_Child_Hap9	.	.	A	G	T	A	.	.	C	T	T	.	.	.	C	.	C	.	.	.	.	DK23(1); KE14 (urine(9) and stool(6)); KE15(8)																																			
	K_Child_Hap10	.	.	A	G	T	A	.	.	C	T	.	T	.	.	.	C	.	C	.	.	.	DK23(1)																																			
	K_Child_Hap11	.	.	A	G	T	A	.	.	C	T	T	.	.	.	C	.	C	.	.	A	.	DK54(1); KE14 (stool(5))																																			
	K_Child_Hap12	.	.	A	.	T	A	G	.	.	.	.	.	.	.	.	.	A	.	.	.	C	KE15(1)																																			
	K_Child_Hap13	G	.	A	.	T	A	G	.	.	.	.	.	.	.	.	A	.	.	.	.	C	KE15(1)																																			



between any of our samples. We thus no longer gave this gene further consideration. Eleven profiles were found with the ITS1, 5.8S and ITS2 sequences, with 2 to 3 profiles per cow, only 1 profile per snail and 1 to 7 profiles per child (Table 1). Sequence data were deposited in the NCBI GenBank database under the accession numbers MT158872 to MT158882.

Table 3 shows that the 11 profiles may differ by their ITS1 gene (458 bp: from position 23 to position 480), with 5 genetic variants at three positions (TGT, TAT, TGY, TRT, and YAT), by their 5.8S gene (155 bp: from position 481 to position 635), with 2 genetic variants at 1 position (C and Y), and by their ITS2 gene (311 bp: from position 636 to position 946), with 3 genetic variants at 4 positions (ATAT, GCGC and RYRY). Among these 11 profiles, 1 profile corresponding to *S. bovis* ITS rDNA was found only in cows (profile 1: accession number MT158872), and 1 profile corresponding to *S. haematobium* ITS rDNA was found only in schoolchildren (profile 2: accession number MT158873). All of the other 9 profiles showed atypical sequences in ITS rDNA. Two were found only in cows (profiles 3: accession number MT158874 and 4: accession number MT158875) and showed double peaks in the ITS1 gene. Five were found only in schoolchildren (profiles 5: accession number MT158876, 6: accession number MT158877, 7: accession number MT158878, 8: accession number MT158879 and 9: accession number MT158880) and showed double peaks in 1 or 2 of the ITS1, 5.8S and ITS2 genes. Two were found both in cows and schoolchildren, where the first one showed a typical *S. haematobium* ITS1 marker together with a typical *S. bovis* ITS2 marker (profile 10: accession number MT158881) and the second showed a typical ITS1 *S. haematobium* marker

together with double peaks in both the 5.8S and ITS2 genes (profile 11: accession number MT158882).

### COI mtDNA phylogeny (881 bp)

The 13 haplotypes obtained from the COI mtDNA sequences (haplotype code/COI mtDNA accession No: K\_Cow\_Hap1/MT159589, K\_Cow\_Hap2/MT159590, K\_Cow\_Hap3/MT159591, K\_Cow\_Hap4/MT159592, K\_Cow\_Hap5/MT159593, K\_Cow\_Hap6/MT159594, K\_Child\_Hap7/MT159595, K\_Child\_Hap8/MT159596, K\_Child\_Hap9/MT159597, K\_Child\_Hap10/MT159598, K\_Child\_Hap11/MT159599, K\_Child\_Hap12/MT159600, K\_Child\_Hap13/MT159601) were compared to 23 published nucleotide sequences of *Schistosoma* (12 sequences from *S. haematobium* (haplotype code/COI mtDNA accession No: Sh\_Gambia/JQ397349; Sh\_Liberia/JQ397350; Sh\_Mali/AY157209; Sh\_Egypt/JQ397368; Sh\_Guinea Bissau/JQ397351; Sh\_Toho\_Benin/KT354661; Sh\_Cameroon/JQ397365; Sh\_Melen\_Gabon/KT354660; Sh\_Ekouk\_Gabon/KT354659; Sh\_Kenya/JQ397378; Sh\_South Africa/JQ397397; Sh\_Madagascar/JQ397399), 7 from *S. bovis* (haplotype code/COI mtDNA accession No: SbxSh\_Corsica1\_France/KT354656; SbxSh\_Corsica2\_France/KT354657; SbxSh\_Corsica3\_France/KT354658; SbxSh\_Sô-Tchanhoué\_Benin/KT354662; Sb\_Senegal/AJ519521; Sb\_Kenya/FJ897160; Sb\_Tanzania/AY157212), 2 from *S. guineensis* (haplotype code/COI mtDNA accession No: Sg\_Sao Tome and Principe/AJ519517 and Sg\_Cameroon/AJ519522), 1 from *S. curassoni* (haplotype code/COI mtDNA accession No: Sc\_Senegal/AJ519516), and 1 from *S. intercalatum* (haplotype code/COI mtDNA accession No:

**Table 3** The 11 profiles of the ITS rDNA region, and their positions, harbored by cow (BK, Kf, and Kg) and schoolchild (DK and KE) in Kessounou Village, Benin

Species	Profile	ITS1			5.8	ITS2				Host	Sample (Number of sequenced larval stages)
		41	73	257	556	725	780	830	900		
<i>S. bovis</i>	1	T	G	T	C	A	T	A	T	Cow	BK19(4); BK24(1); BK26(1); BK29(4); Kf5(5); Kf6(3); Kf8(3); Kf9(3)
<i>S. haematobium</i>	2	T	A	T	C	G	C	G	C	Schoolchild	KE14 (urine(1) and stool(1))
<i>S. bovis x S. haematobium</i>	3	T	G	Y	C	A	T	A	T	Cow	BK19(1); BK29(1)
	4	T	R	T	C	A	T	A	T		BK24(3); BK26(3)
	5	T	A	T	Y	G	C	G	C	Schoolchild	DK54(2); KE14 (urine(4) and stool(2)); KE15(2)
	6	T	A	T	C	R	Y	R	Y		DK23(1); KE14 (stool(1)); KE15(1)
	7	Y	A	T	C	G	C	G	C		DK23(2); KE14 (urine(1) and stool(1))
	8	Y	A	T	Y	G	C	G	C		DK23(1); KE14 (urine(4) and stool(4)); KE15(4)
	9	Y	A	T	C	R	Y	R	Y	Cow and Schoolchild	DK23(1); KE14 (stool(2)); KE15(2)
	10	T	A	T	C	A	T	A	T		BK24(1); BK26(1); Kg1(5); KE14 (urine(2))
	11	T	A	T	Y	R	Y	R	Y		Kg2(5); DK23(1); KE14 (urine(2) and stool(2)); KE15(1)

Y indicates the presence of the bases T and C, rather than an ambiguous reading between T and C. Similarly, R indicates the presence of the bases A and G and not an ambiguous reading between A and G

Si\_Democratic Republic of the Congo/AJ519515) obtained at the GenBank database. Maximum likelihood tree topology of COI mtDNA showed that all 13 haplotypes which came from cows and schoolchildren belonged to the *S. bovis* clade and not to the *S. haematobium* or to other *Schistosoma* species clades (Fig. 4). The percentages of differences ranged between 0.23 and 1.70% between our 13 haplotypes and the *S. bovis* haplotypes from Senegal, Kenya, and Tanzania. We found 0% difference between K\_Child\_Hap11 and the haplotype from Sô-Tchanhoué, Benin. We also found 0% difference between K\_Child\_Hap13 and the Corsica3 haplotype.

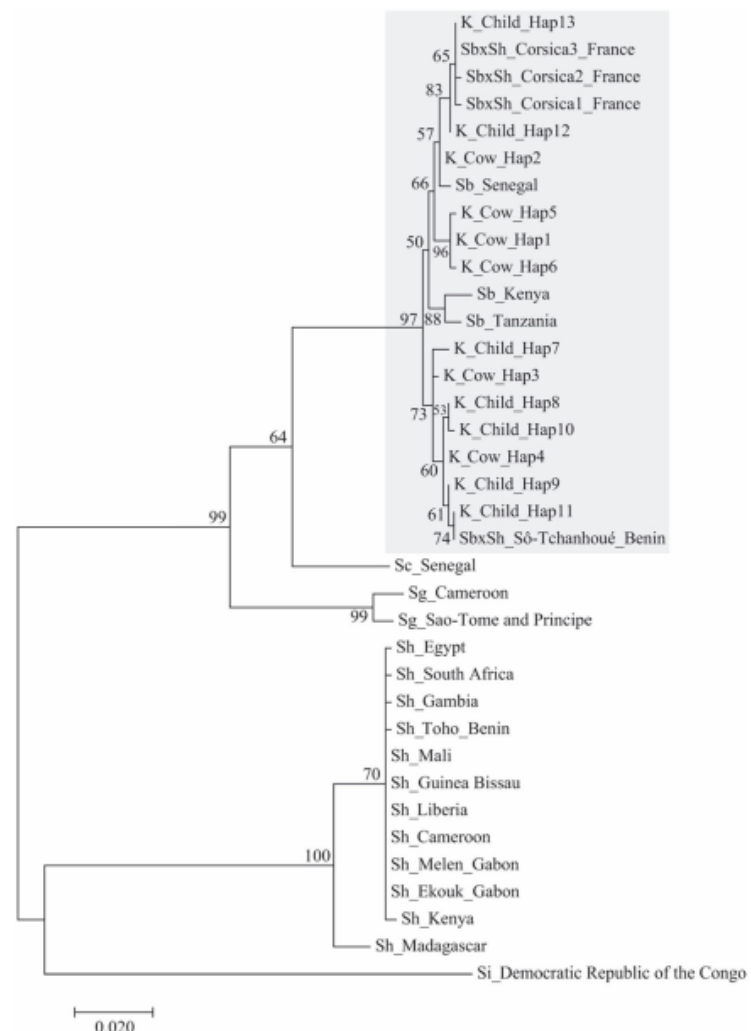
### Egg morphometry

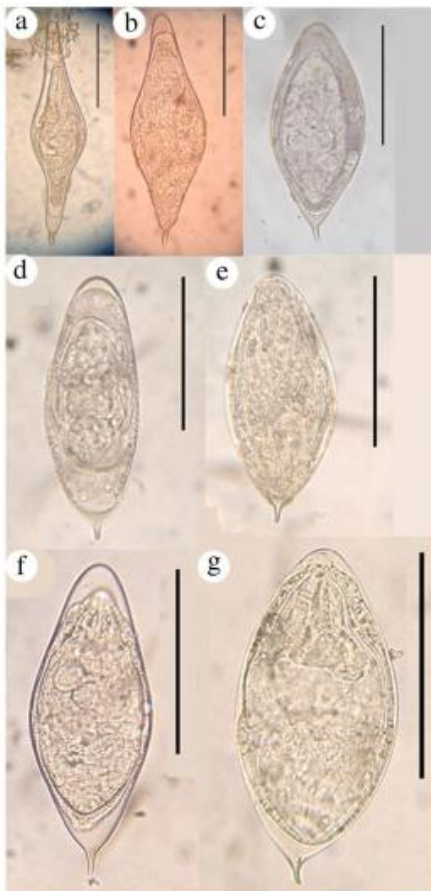
Two or three morphotypes were observed according to the host and the excreta (urine or feces/stool). In cow feces, three

morphotypes were found: *S. bovis*, *S. haematobium*, and their intermediate morphotype (Fig. 5a–c); in schoolchild stool (Fig. 5d, e) and urine (Fig. 5f, g), two morphotypes were found: *S. haematobium* and an intermediate morphotype.

The percentages of each morphotype are presented in Fig. 6 for each origin. The percentage of *S. bovis* morphotype was high (63.8%) in the cow feces, and null in schoolchild (stool and urine origins) ( $p < 0.05$ ). The percentages of intermediate morphotype were not different between the origins ( $p > 0.05$ ). The percentage of *S. haematobium* morphotype observed from cow feces origin was significantly lower than those from both stool and urine of schoolchild origins ( $p < 0.05$ ). For each origin, the percentages of *S. bovis*, intermediate and *S. haematobium* morphotypes were all significantly different from each other ( $p < 0.05$ ).

**Fig. 4** Maximum likelihood tree topology built with the 13 haplotypes COI mtDNA (881 bp) showing that *S. haematobium* from Kessounou (recovered either from cows or schoolchildren) belongs to the *S. bovis* clade. Haplotype code/COI mtDNA accession No: K\_Cow\_Hap1/MT159589, K\_Cow\_Hap2/MT159590, K\_Cow\_Hap3/MT159591, K\_Cow\_Hap4/MT159592, K\_Cow\_Hap5/MT159593, K\_Cow\_Hap6/MT159594, K\_Child\_Hap7/MT159595, K\_Child\_Hap8/MT159596, K\_Child\_Hap9/MT159597, K\_Child\_Hap10/MT159598, K\_Child\_Hap11/MT159599, K\_Child\_Hap12/MT159600, K\_Child\_Hap13/MT159601). The scale shows the number of nucleotide substitutions per site





**Fig. 5** Egg morphology. Cow feces, **a** *S. bovis* morphotype, **b** intermediate morphotype, **c** *S. haematobium* morphotype. Schoolchild stool, **d** intermediate morphotype, **e** *S. haematobium* morphotype. Schoolchild urine, **f** intermediate morphotype, **g** *S. haematobium* morphotype. Bar represents 100  $\mu\text{m}$

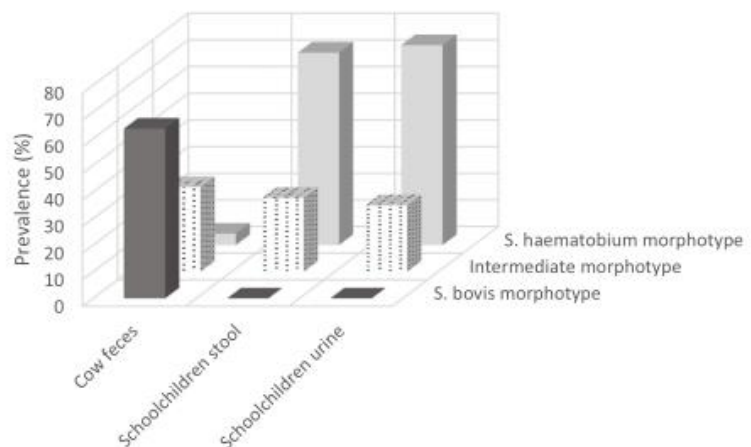
The morphometry of the eggs from cows and schoolchildren is presented in Table 4 and Fig. 7. For the *S. bovis* morphotype, no comparison could be made since this morphotype was not present in schoolchildren. For the intermediate morphotype, mean egg length and spine length were significantly higher for the cow feces origin ( $p < 0.05$ ) and no difference was observed between the urine and the stool of the schoolchildren ( $p > 0.05$ ). The mean length/width ratio was significantly smaller for the eggs obtained from schoolchild urine ( $p < 0.05$ ) and no differences were observed between the schoolchild and cow feces ( $p > 0.05$ ). For the *S. haematobium* morphotype, no major difference was observed between the origins ( $p > 0.05$ ). The comparisons between morphotypes indicated that mean egg length and egg length/width ratio from urine or from stool were significantly higher in *S. bovis* morphotype compared to the intermediate one and higher in the intermediate morphotype compared to the *S. haematobium* one ( $p < 0.05$ ). The plot of the egg width ( $\mu\text{m}$ ) and egg length ( $\mu\text{m}$ ) for schoolchild urine (green diamonds), schoolchild stool (blue squares), and cow feces (orange triangles) showed that the eggs from schoolchild urine and stool were grouped together but that those of cow feces constituted a separate group, with some individuals making the link between the two groups showing a higher variability in the eggs from cows compared to the eggs from schoolchildren.

**Discussion**

The results showed a chronobiological polymorphism in the cercarial emergence rhythm from snails exposed to miracidia collected from cow feces, with four different patterns.

The first pattern, early diurnal, exhibited a typical *S. bovis* cercarial emission; it was obtained in 7 *B. forskalii* among 11 (Kf1, Kf2, Kf3, Kf6, Kf7, Kf8, and Kf11). Such a pattern was found in *S. bovis* from Sardinia (Italy), Sudan and Spain

**Fig. 6** Percentages of each egg morphotype (*S. bovis*, black; intermediate, dots; *S. haematobium*, grey) according to origin (cow feces, schoolchild stool, schoolchild urine)



**Table 4** Egg morphometry (mean ( $\mu\text{m}$ )  $\pm$  standard error)

Morphotype	<i>S. bovis</i>			Intermediate			<i>S. haematobium</i>		
	CF	SS	SU	CF	SS	SU	CF	SS	SU
Length	247.83 <sup>b</sup>	na	na	223.17 <sup>a, b</sup>	153.00	148.54 <sup>b</sup>	140.00	138.85	132.01
SE	4.03	na	na	7.51	5.48	3.82	42.43	4.68	1.66
N	30.00	na	na	15.00	5.00	12.00	2.00	13.00	36.00
Min	205.00	na	na	170.00	140.00	135.00	110.00	120.00	117.50
Max	300.00	na	na	262.50	170.00	170.00	170.00	175.00	170.00
Width	69.58 <sup>b</sup> int	na	na	83.17 <sup>a</sup>	56.00 <sup>a, b</sup>	68.13 <sup>a</sup>	67.50	68.85	68.96
SE	1.56	na	na	2.19	3.26	2.04	17.68	3.25	1.62
N	30.00	na	na	15.00	5.00	12.00	2.00	13.00	36.00
Min	57.50	na	na	65.00	50.00	60.00	55.00	50.00	50.00
Max	90.00	na	na	100.00	65.00	80.00	80.00	95.00	85.00
Spine length	17.38	na	na	17.31 <sup>a</sup> SU	13.33	10.94	17.50	10.31 <sup>a</sup>	12.59
SE	0.68	na	na	1.04	4.45	0.49	3.54	0.61	0.48
N	20.00	na	na	13.00	3.00	8.00	2.00	8.00	29.00
Min	12.50	na	na	10.00	7.50	10.00	15.00	7.50	10.00
Max	22.50	na	na	22.50	20.00	12.50	20.00	12.50	20.00
Length/Width	3.59 <sup>b</sup>	na	na	2.68 <sup>b</sup>	2.75 <sup>b</sup>	2.19 <sup>a, b</sup>	2.06 <sup>b</sup>	2.04 <sup>b</sup>	1.94
SE	0.07	na	na	0.05	0.11	0.07	0.09	0.07	0.04
N	30.00	na	na	15.00	5.00	12.00	2.00	13.00	36.00
Min	2.93	na	na	2.31	2.50	1.93	2.00	1.71	1.63
Max	4.17	na	na	2.93	3.10	2.72	2.13	2.50	2.60

CF cow feces, SS schoolchild stool, SU schoolchild urine, SE standard error, N number of eggs

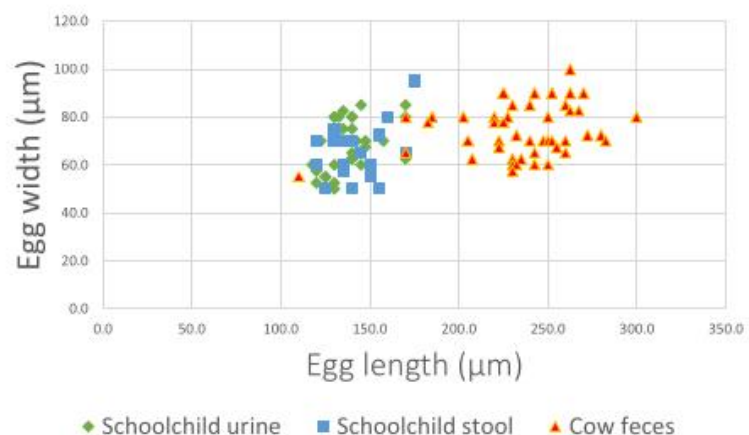
<sup>a</sup>Mann-Whitney test: significantly different from the other origins of the same morphotype

<sup>b</sup>Mann-Whitney test: significantly different from the other morphotypes of the same origin

(Mouahid et al. 1987, 1991), and Niger (Mouchet et al. 1992). The COI mtDNA molecular analyses, conducted on cercariae from 2 snails among the 7 (Kf6 and Kf8), showed that the DNA sequences (K\_Cow\_Hap2) belonged to the *S. bovis* clade. *Schistosoma bovis* has an extensive geographical distribution (Moné et al. 1999), and in their review, Moné et al. showed that this species was naturally found in all of the neighboring

countries of Benin: Togo, Burkina Faso, Niger, and Nigeria. However, while this is the first time that this species has been found naturally in Benin, this result is not surprising. Furthermore, if the presence of *S. bovis* was not surprising in cows, the presence of COI mtDNA *S. bovis* haplotypes in all of the schoolchildren samples confirms that introgressive hybridization between *S. bovis* and *S. haematobium* exists in humans in

**Fig. 7** Plot of the egg width ( $\mu\text{m}$ ) and egg length ( $\mu\text{m}$ ) for schoolchild urine (diamonds), schoolchild stool (squares) and cow feces (triangles)



Benin, as shown previously by Moné et al. (2015). The phylogenetic analysis showed that 2 of the 6 haplotypes recovered from cows (K\_Cow\_Hap3 and K\_Cow\_Hap4) and 5 of the 7 haplotypes recovered from schoolchildren (K\_Child\_Hap7, K\_Child\_Hap8, K\_Child\_Hap9, K\_Child\_Hap10, and K\_Child\_Hap11) grouped with the one from Sô-Tchanhoué from Benin, confirms previous results on introgressive hybridization between *S. bovis* and *S. haematobium* with respect to humans in this area (Moné et al. 2015). The four last haplotypes recovered from cows (K\_Cow\_Hap1, K\_Cow\_Hap2, K\_Cow\_Hap5, and K\_Cow\_Hap6) grouped with the *S. bovis* haplotypes. The last two haplotypes recovered in humans (K\_Child\_Hap12 and K\_Child\_Hap13) grouped with those from Corsica, with the haplotype number 13 being exactly the same as the Corsica 3 haplotype. This result provides insight into the origin of the Corsican hybrids which was previously attributed to Senegal (Boissier et al. 2016) on the basis of 407 positions. The phylogenetic tree in the present work, based on 881 positions of the COI mtDNA, showed that the origin of Corsican haplotypes could be attributed to Benin.

The second pattern, midday to late diurnal, was new for cercarial emission in cows while it was similar to what has been published on *S. haematobium* in humans from Algeria (Kéchemir and Théron 1997), Morocco (Mouahid et al. 1991), Niger (Pagès and Théron 1990; Mouchet et al. 1992), and Gabon (Mintsa-Nguema et al. 2014). This is the first time that a typical *S. haematobium* pattern has been observed in cows as *S. haematobium* has always been thought to have the highest specificity to humans. This atypical pattern from cows was not due to the fact that the cercariae were shed by *B. globosus* (Kg2) since it has been shown that the taxa of the snail from which the cercariae emerge does not determine the cercarial emergence pattern of *S. bovis* (Mouahid and Théron 1986). Furthermore, hybridization in schistosomes is known to be accompanied by heterosis (hybrid vigor), including the capability of hybrid schistosomes to open their snail intermediate host spectrum (Wright and Ross 1980; Webster and Southgate 2003; Huyse et al. 2009). *Bulinus forskalii* is the main snail host for *S. bovis* in Sub-Saharan Africa while *B. globosus* is the main snail host for *S. haematobium* in this area (Brown 1994). The fact that a *B. globosus* was shedding this atypical pattern could stem from the opening of the snail host spectrum by this *S. bovis* x *S. haematobium* new pattern. The molecular and phylogenetic results presented in this paper support this idea since this *S. bovis* COI mtDNA haplotype (K\_Cow\_Hap4) grouped with the one from Sô-Tchanhoué from Benin harboring a *S. haematobium* x *S. bovis* genetic pattern (Moné et al. 2015). The molecular analyses also showed an atypical ITS rDNA profile for this sample (profile 11), with a *S. haematobium* pattern for ITS1 and double peaks for both 5.8S and ITS2 (Table 3). These results show that in this area, introgressive hybridization exists between *S. bovis* and *S. haematobium* in cows as well. This is the first time that

such an introgressive hybridization has been identified in domestic bovines. Previous results showed introgression between *S. bovis* and *S. haematobium* in humans (Brémond et al. 1993; Huyse et al. 2009; Webster et al. 2013; Moné et al. 2015; Leger et al. 2016; Soentjens et al. 2016; Angora et al. 2019) and rodents (Catalano et al. 2018).

The third pattern corresponded to a new pattern for *S. bovis*, with both a typical early pattern for *S. bovis* accompanied by a nocturnal emergence peak at 7 pm. This pattern was obtained from 4 of the 11 *B. forskalii* snails (Kf4, Kf5, Kf9, and Kf10). This is the first time that a double peak in cercarial emergence is shown for *S. bovis*. Another animal schistosome, *S. margrebowiei* Le Roux, 1933 also showed two emergence peaks per day, with the first peak occurring 1 h after the onset of daylight, and the second peak 1 h after the onset of darkness (Raymond and Probert 1991). This is also the first time that a nocturnal cercarial emergence is shown for *S. bovis*. In another species of schistosome, *S. mansoni*, atypical chronotypes were found with nocturnal shedding. A three-peak cercarial emergence was found in southern Benin in both natural and experimental infections, with one peak in the early morning, a second around midday, and a third one at the beginning of the night (Ibikounlé et al. 2012). Another chronotype that displayed a strictly nocturnal emergence was shown in Oman and was related to the presence of reservoir hosts (rodents) in the transmission sites (Mouahid et al. 2012). However, this particular *S. mansoni* nocturnal chronotype was purely nocturnal and was not accompanied by diurnal cercarial emergence from the same individual snail. Several hypotheses could be proposed for the two peaks in diurnal-nocturnal cercarial emergence. First, the peaks may coincide with the behavior of the definitive host where cattle would go twice a day to drink and *S. bovis* would thus be adapted to this behavior. This explanation could also apply in the case of *S. margrebowiei*. Second, the peaks may be due to the presence of nocturnal reservoir hosts for *S. bovis*, like rodents, with nocturnal watering behaviors. This explanation would apply in the case of *S. mansoni* from Oman (Mouahid et al. 2012). Lastly, the peaks may have resulted in response to the presence of introgressive hybridization between *S. bovis* and a nocturnal population of *S. bovis* or another species of schistosome. It is difficult, however, to get a nocturnal pattern with two diurnal species of schistosomes (in our case, *S. bovis* and *S. haematobium*), even if they can hybridize. Pagès and Théron (1990) conducted crossing experiments between *S. bovis* from Spain and *S. haematobium* from Niger and the F1 hybrids harbored the same chronobiological pattern as the parental *S. bovis*, i.e., with an early peak. Unfortunately, these authors did not analyze the F2 generations. Furthermore, double infections were experimentally done with both *S. bovis* from Sardinia (Italy) and *S. haematobium* from Morocco and both species kept their own cercarial shedding rhythm in the doubly infected snails (Mouahid et al. 1991). This was also

the case for *S. mansoni* with both early and late diurnal patterns (Théron et al. 1997). The cercariae from 2 of the 4 snails (Kf5 and Kf9) used for cercarial emergence pattern were analyzed for molecular biology and showed *S. bovis* COI mtDNA haplotypes (K\_Cow\_Hap1 and K\_Cow\_Hap2). Both COI mtDNA haplotypes were found from different cercariae found in a single snail sample suggesting that, perhaps, 2 miracidia succeeded in developing in this one snail, leading to the double peak in cercarial emergence. However, as the other snail sample had only 1 COI mtDNA haplotype (K\_Cow\_Hap2) and yielded the same double peak pattern, this hypothesis is unsupported. The phylogenetic analysis showed that these COI mtDNA haplotypes (K\_Cow\_Hap1 and K\_Cow\_Hap2) grouped with the *S. bovis* haplotypes from other countries. The molecular analyses also showed typical ITS rDNA *S. bovis* profiles for these samples (profile 1).

The fourth chronotype, obtained from one *B. globosus* snail (Kg1), corresponded to a new chronotype for *S. bovis*, with both a typical *S. haematobium* late diurnal pattern accompanied by a nocturnal emergence peak at 7 pm. This is the first time that a midday to late diurnal peak together with a nocturnal peak was identified in *S. bovis*. This result shares similarities with the pattern with the previous one, with the presence of nocturnal cercarial shedding, but is different in that the early pattern in *S. bovis* was replaced by a late *S. haematobium* pattern. The molecular studies showed that this sample had the same COI mtDNA haplotype (K\_Cow\_Hap4) as the one linked to the second cercarial emergence pattern (*S. haematobium* midday to late diurnal pattern) and the phylogenetic analyses grouped this haplotype with the one from Sô-Tchanhoué, Benin. Furthermore, the ITS rDNA profile of all the cercariae in this sample (profile 10) showed a typical *S. haematobium* ITS1 marker together with a typical *S. bovis* ITS2 marker. Further studies on the behavior of the cattle and the presence of rodents as reservoir hosts are needed to explain this atypical pattern of cercarial emergence.

The results showed a polymorphism in egg morphology in both cows and schoolchildren, with 3 morphotypes. The first, the *S. bovis* morphotype, represented over 60% of the eggs in cow feces and no *S. bovis* morphotype was found in either the stool or urine of schoolchildren. Still, this result shows that 40% of the eggs recovered in the cows did not have a *S. bovis* morphotype and leads us to also suspect hybridization in cows between *S. bovis* and *S. haematobium*. With respect to egg morphometry, mean egg length of our *S. bovis* morphotype (247 µm) was similar to those recorded for *S. bovis* from cow feces (from 179 to 260 µm (Alves (1949); Pitchford (1965))). The second morphotype, *S. haematobium*, represented around 4% of the eggs in cow feces, compared to over 70% of the eggs in both stool and urine of the schoolchildren. This result showed that some eggs which harbored the *S. haematobium* morphotype were present in cows and that 30% of the eggs recovered in the schoolchildren lacked a *S. haematobium*

morphotype, leading us to also suspect hybridization between *S. bovis* and *S. haematobium* in both cows and humans. Mean egg lengths of our *S. haematobium* morphotypes (132 to 140 µm) were similar to those recorded for *S. haematobium* eggs (from 131 to 146 µm; see Pitchford (1965), Loker (1983), Richard-Lenoble et al. (1993), and Moné et al. (2012)). The third chronotype, an intermediate morphotype between *S. bovis* and *S. haematobium*, sustained the hybridization between *S. bovis* and *S. haematobium*, as we would have suspected from the cercarial emergence results. Even if the shape of the eggs produced is not necessarily a guide to the genetic constitution of the enclosed larvae, as stated by Wright and Ross (1980), these authors also suspected natural hybrids between *S. haematobium* and *S. mattheei* thanks to *S. mattheei*-like eggs in a human infection. Mean egg lengths for our intermediate morphotypes, between *S. bovis* and *S. haematobium* (148 to 223 µm), also suggest the presence of hybridization between these two species in both cows and humans. However, much smaller intermediate egg sizes were found in the hybrids we found between *S. bovis* and *S. haematobium* in Corsica, France (Moné et al. 2015). A higher variability in the morphometry of the eggs was observed in cow feces compared to schoolchild stool or urine and thus place the cow host as the host which provides the largest panel of eggs for schistosome transmission in the area.

In conclusion, our results showed, for the first time, (i) the presence of *S. bovis* in Benin, (ii) the presence of introgressive hybridization between *S. bovis* and *S. haematobium* in domestic cattle, and (iii) the presence of atypical chronobiological patterns in schistosomes from cattle, with typical *S. haematobium* shedding pattern, double-peak patterns and nocturnal patterns. They also confirmed that introgressive hybridization between *S. bovis* and *S. haematobium* exists in humans in Benin. Very little comparison is possible between the present findings and the literature, because until now, the researchers who worked on introgressive hybridization had never analyzed the cercarial emergence patterns from naturally-infected cows. Our results showed that the chronobiological life-history trait is useful for the detection of new hosts and also may reveal the possible presence of introgressive hybridization in schistosomes. They also showed that another life-history trait, the prepatent period, which is known to be short for *Schistosoma bovis* (21–22 days, Mouahid and Théron 1987) and long for *Schistosoma haematobium* (32 to 43 days, Ibikounlé et al. 2013) was intermediate for all our samples (26 to 32 days).

Our results, for the first time, place cattle as reservoir host for *S. haematobium* and *S. bovis* x *S. haematobium*. The proximity between humans and cattle in Kessounou, like in many villages in Africa, is favorable for the interactions between their schistosomes, *S. haematobium* for humans and *S. bovis* for cattle. The consequences of these results on the epidemiology of the disease, the transmission to humans, and the control of the disease

are very important. They show that the definitive host spectrum of *S. haematobium* may be opened to include cattle; further research should be done regarding the importance of rodents, especially in nocturnal schistosomiasis.

Our results are not congruent with two recently published papers which suggest that hybridization may occur quite infrequently between *S. bovis* and *S. haematobium*. Boon et al. (2019) found strong differentiation between *S. bovis* and *S. haematobium* in human populations of the Senegal River Basin, suggesting that there is minimal gene flow between them. However, in their paper, they found a miracidium harboring a double *S. bovis* and *S. haematobium* ITS rDNA profile from a neighborhood village which was genetically highly differentiated from the others. They also questioned about the previous findings by Huysse et al. (2009) and Webster et al. (2013) showing the existence of contemporary hybrid crosses between *S. bovis* and *S. haematobium* in humans in Senegal. Platt et al. (2019) studied the exomes of miracidia coming from humans in Niger and Zanzibar and their data did not reveal any evidence for contemporary hybridization. Instead, they showed ancient introgression of some *S. bovis* alleles into the genome of *S. haematobium* in Niger. We are aware that our data were genotyped with just two loci (COI mtDNA and ITS rDNA) and that they may provide little power to detect recent hybridization. However, unlike other recent results, our data showed that double *S. bovis* and *S. haematobium* ITS rDNA profiles were not rare in Kessounou village suggesting a recent hybridization, as the ITS rDNA marker can retain both parental copies for several generations before they are homogenized by concerted evolution (Sang et al. 1995). Our results show that, in some places in Benin, gene flow between *S. bovis* and *S. haematobium* is present.

**Acknowledgments** We would like to thank M. Marius CODJO, chief of Kessounou village and all the breeders and also M. Jean-Pierre TOKPO, Director of Kessounou school and all the schoolchildren for their participation to the study.

**Authors' contributions** Conceptualization: Hélène Moné, Gabriel Mouahid, Moudachirou Ibikounlé.

Material preparation, data collection and analysis: Boris A.E.S. Savassi, Gabriel Mouahid, Chrystelle Lasica, Samoussou-Dine K. Mahaman, Jean-François Allienne, Moudachirou Ibikounlé, Hélène Moné.

Funding acquisition: Gabriel Mouahid, André Garcia, David Courtin, Moudachirou Ibikounlé, Hélène Moné.

Investigation: Boris A.E.S. Savassi, Gabriel Mouahid, Chrystelle Lasica, Samoussou-Dine K. Mahaman, Jean-François Allienne, Moudachirou Ibikounlé, Hélène Moné.

Methodology: Hélène Moné, Gabriel Mouahid, Moudachirou Ibikounlé.

Project administration: Gabriel Mouahid, André Garcia, David Courtin, Moudachirou Ibikounlé, Hélène Moné.

Resources: Boris A.E.S. Savassi, Gabriel Mouahid, Chrystelle Lasica, Samoussou-Dine K. Mahaman, André Garcia, David Courtin, Jean-François Allienne, Moudachirou Ibikounlé, Hélène Moné.

Supervision: Gabriel Mouahid, Hélène Moné.

Validation: Gabriel Mouahid, Hélène Moné.

Visualization: Gabriel Mouahid, Hélène Moné.

Writing—original draft: Boris A.E.S. Savassi, Gabriel Mouahid, Hélène Moné.

Writing—review & editing: Boris A.E.S. Savassi, Gabriel Mouahid, Chrystelle Lasica, André Garcia, David Courtin, Jean-François Allienne, Moudachirou Ibikounlé, Hélène Moné.

All authors read and approved the final manuscript.

**Funding information** This research was funded by the French Ministry of Foreign Affairs, the “Laboratoire Mixte International” (Mixed International Laboratory) on the project CONS-HELM « Infections helminthiques: traitements et conséquences sur la santé et le développement au Sud », and the French National Agency for Research (ANR) [grant ANR-17-CE12-0005-01] CHRONOGET and CNRS. BS is a student fellow from the SCAC (Service de Coopération et d’Action Culturelle), French Embassy in Benin; CL is a student fellow from the Occitanie Region, France.

## Compliance with ethical standards

**Conflict of interest** The authors declare that they have no conflict of interest.

**Ethical approval** Ethical permission (No119, 17/07/2019/CER-ISBA) was obtained from the Ethic Committee of the « Institut des Sciences Biomédicales Appliquées, Faculté de Médecine, Université d’Abomey-Calavi, 01BP526, Cotonou, Benin.

**Informed consent** The Director of Kessounou school obtained informed consent from the parents of all the schoolchildren who participated to the study.

**Data set availability** Sequence data were deposited in the NCBI GenBank database under the accession numbers MT159589 to MT159601 for COI mtDNA and MT158872 to MT158882 for ITS rDNA.

## References

- Abou M, Yabi I, Yolou I, Ogouwale E (2018) Caractérisation des systèmes de production sur les sites d’aménagements hydro-agricoles dans le doublet Dangbo-Adjohoun au sud du Bénin. *Int J Biol Chem Sci* 12:462–478
- Alves W (1949) The eggs of *Schistosoma bovis*, *S. matthei* and *S. haematobium*. *J Helminthol* 23:127–134
- Angora EK, Allienne J-F, Rey O, Menan H, Touré AO, Coulibaly JT, Raso G, Yavo W, N’Goran EK, Utzinger J, Balmer O, Boissier J (2019) High prevalence of *Schistosoma haematobium* × *Schistosoma bovis* hybrids in schoolchildren in Côte d’Ivoire. *Parasitology* 147:287–294
- Barber K, Mkoji G, Loker E (2000) PCR-RFLP analysis of the ITS2 region to identify *Schistosoma haematobium* and *S. bovis* from Kenya. *Am J Trop Med Hyg* 62:434–440
- 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. *Mem Inst Oswaldo Cruz* 103:501–503
- Boissier J, Grech-Angelini S, Webster BL, Allienne J-F, Huysse T, Mas-Coma S, Toulza E, Barré-Cardi H, Rollinson D, Kincaid-Smith J, Oleaga A, Galinier R, Foata J, Rognon A, Berry A, Mouahid G, Henneron R, Moné H, Noel H, Mitta G (2016) Outbreak of

- urogenital schistosomiasis in Corsica (France): an epidemiological case study. *Lancet Infect Dis*. [https://doi.org/10.1016/S1473-3099\(16\)00175-4](https://doi.org/10.1016/S1473-3099(16)00175-4)
- Boko PM, Ibikounlé M, Onzo-Aboki A, Tougoue J-J, Sissinto Y, Batcho W, Kindé-Gazard D, Kabore A (2016) Schistosomiasis and soil transmitted helminths distribution in Benin : a baseline prevalence survey in 30 districts. *PLoS One* 11:e0162798
- Boon NAM, Mbow M, Paredis L, Moris P, Sy I, Maes T, Webster BL, Sacko M, Volckaert FAM, Polman K, Huysse T (2019) No barrier breakdown between human and cattle schistosome species in the Senegal River Basin in the face of hybridisation. *Int J Parasitol* 49:1039–1048
- Brémond P, Sellin B, Sellin E, Naméoua B, Labbo R, Théron A, Combes C (1993) Arguments for the modification of the genome (introgression) of the human parasite *Schistosoma haematobium* by genes from *S. bovis*, in Niger. *C R Acad Sci Série III, Sciences de la vie* 316:667–670
- Brown DS (1994) Freshwater snails of Africa and their medical importance, revised 2nd edition. Taylor & Francis Ltd, London, pp 608
- Castresana J (2000) Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol* 17:540–552
- Catalano S, Sene M, Diouf ND, Fall CB, Borlase A, Leger E, Ba K, Webster JP (2018) Rodents as natural hosts of zoonotic *Schistosoma* species and hybrids: an epidemiological and evolutionary perspective from West Africa. *J Infect Dis* 218:429–433
- Chitsulo L, Engels D, Montresor A, Savioli L (2000) The global status of schistosomiasis and its control. *Acta Trop* 77:41–51
- De Bont J, Vercrucyue J (1997) The epidemiology and control of cattle schistosomiasis. *Parasitol Today* 13:255–262
- De Bont J, Vercrucyue J (1998) Schistosomiasis in cattle. *Adv Parasitol* 41:285–364
- Dereeper A, Audic S, Claverie JM, Blanc G (2010) BLAST-EXPLORER helps you building datasets for phylogenetic analysis. *BMC Evol Biol* 10:8. <https://doi.org/10.1186/1471-2148-10-8>
- Dereeper A, Guignon V, Blanc G, Audic S, Buffet S, Chevenet F, Dufayard JF, Guindon S, Lefort V, Lescot M, Claverie JM, Gascuel O (2008) Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res* 36(Web Server issue):W465–W469. <https://doi.org/10.1093/nar/gkn180>
- Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32:1792–1797
- Engels D, Chitsulo L, Montresor A, Savioli L (2002) The global epidemiological situation of schistosomiasis and new approaches to control and research. *Acta Trop* 82:139–146
- Huysse T, Webster BL, Geldof S, Stothard R, Diaw OT, Polman K, Rollinson D (2009) Bidirectional introgressive hybridization between a cattle and human schistosome species. *PLoS Pathog* 5(9):e1000571
- Ibikounlé M, Moné H, Abou Y, Kinde-Gazard D, Sakiti NG, Mouahid G, Massougbdji A (2012) Premier cas de chronobiologie des émissions cercariennes de type infradien chez *Schistosoma mansoni* dans deux foyers du sud-Bénin. *Int J Biol Chem Sci* 6:1081–1089
- Ibikounlé M, Mouahid G, Mints-Nguema R, Sakiti N, Massougbdji A, Moné H (2013) Snail intermediate host/*Schistosoma haematobium* relationships from three transmission sites in Benin (West Africa). *Parasitol Res* 112:227–233
- Kéchemir N, Théron A (1997) Intraspecific variation in *Schistosoma haematobium* from Algeria. *J Helminthol* 71:29–33
- Kumar S, Stecher G, Tamura K (2016) MEGA7: molecular evolutionary genetics analysis version 7.0. *Mol Biol Evol* 33:1870–1874
- Leger E, Garba A, Hamidou AA, Webster BL, Pennance T, Rollinson D, Webster JP (2016) Introgressed animal schistosomes *Schistosoma curassoni* and *S. bovis* naturally infecting humans. *EID* 22:2212
- Lockyer A, Olson P, Østergaard P, Rollinson D, Johnston D, Attwood S, Southgate V, Horak P, Snyder S, Le T (2003) The phylogeny of the Schistosomatidae based on three genes with emphasis on the inter-relationships of *Schistosoma* Weinland, 1858. *Parasitology* 126:203–224
- Loker ES (1983) A comparative study of the life-histories of mammalian schistosomes. *Parasitology* 87:343–369
- Mints-Nguema R, Moné H, Ibikounlé M, Mengué-Ngou-Milama K, Kombila M, Mouahid G (2014) Cercarial emergence pattern of *Schistosoma haematobium* from Libreville, Gabon. *Parasite* 21:3. <https://doi.org/10.1051/parasite/2014004>
- Moné H, Holtfreter MC, Allienne J-F, Mints-Nguema R, Ibikounlé M, Boissier J, Berry A, Mitta G, Richter J, Mouahid G (2015) Introgressive hybridizations of *Schistosoma haematobium* by *Schistosoma bovis* at the origin of the first case report of schistosomiasis in Corsica (France, Europe). *Parasitol Res* 114:4127–4133
- Moné H, Ibikounlé M, Massougbdji A, Mouahid G (2010) Human schistosomiasis in the Economic Community of West African States; epidemiology and control. *Adv Parasitol* 71:33–91
- Moné H, Minguez S, Ibikounlé M, Allienne J-F, Massougbdji A, Mouahid G (2012) Natural interactions between *S. haematobium* and *S. guineensis* in the Republic of Benin. *Sci World J* 2012:793420
- Moné H, Mouahid G, Morand S (1999) The distribution of *Schistosoma bovis* Sonsino, 1876 in relation to intermediate host mollusc-parasite relationships. *Adv Parasitol* 44:99–138
- Mouahid A, Théron A (1986) *Schistosoma bovis*: patterns of cercarial emergence from snails of the genera *Bulinus* and *Planorbis*. *Exp Parasitol* 62:389–393
- Mouahid A, Théron A (1987) *Schistosoma bovis*: variability of cercarial production as related to the snail hosts: *Bulinus truncatus*, *B. wrighti* and *Planorbis metidjensis*. *Int J Parasitol* 17:1431–1434
- Mouahid G, Idris MA, Verneau O, Théron A, Shaban MM, Moné H (2012) A new chronotype of *Schistosoma mansoni*: adaptive significance. *Tropical Med Int Health* 17:727–732
- Mouahid A, Moné H, Arru E, Chassé J-L, Théron A, Combes C (1987) Analyse comparative du rythme d'émission des cercaires de trois souches de *Schistosoma bovis*. *Parassitologia* 29:79–85
- Mouahid A, Moné H, Chaïb A, Théron A (1991) Cercarial shedding patterns of *Schistosoma bovis* and *Schistosoma haematobium* from single and mixed infections of *Bulinus truncatus*. *J Helminthol* 65:8–14
- Mouchet F, Théron A, Brémond P, Sellin E, Sellin B (1992) Pattern of cercarial emergence of *Schistosoma curassoni* from Niger and comparison with three sympatric species of schistosomes. *J Parasitol* 78:61–63
- Nei M, Kumar S (2000) Molecular evolution and phylogenetics. Oxford University Press, New York
- Onzo-Aboki A, Ibikounlé M, Boko PM, Savassi BS, Doritchamou J, Siko EJ, Daré A, Batcho W, Massougbdji A, Tougoue JJ (2019) Human Schistosomiasis in Benin: countrywide evidence of *Schistosoma haematobium* predominance. *Acta Trop* 191:185–197
- Pagès J, Théron A (1990) Analysis and comparison of cercarial emergence rhythms of *Schistosoma haematobium*, *S. intercalatum*, *S. bovis*, and their hybrid progeny. *Int J Parasitol* 20:193–197
- Pitchford R (1965) Differences in the egg morphology and certain biological characteristics of some African and Middle Eastern schistosomes, genus *Schistosoma*, with terminal-spined eggs. *Bull World Health Organ* 32:105–120
- Platt RN, McDew-White M, Le Clec'h W, Chevalier F, Allan F, Emery AM, Garba A, Hamidou AA, Ame SM, Webster JP, Rollinson D, Webster BL, Anderson TJC (2019) Ancient hybridization and adaptive introgression of an invadysin gene in schistosome parasites. *Mol Biol Evol* 36:2127–2142
- Raymond K, Probert AJ (1991) The daily cercarial emission rhythm of *Schistosoma margrebowiei* with particular reference to dark period stimuli. *J Helminthol* 65:159–168

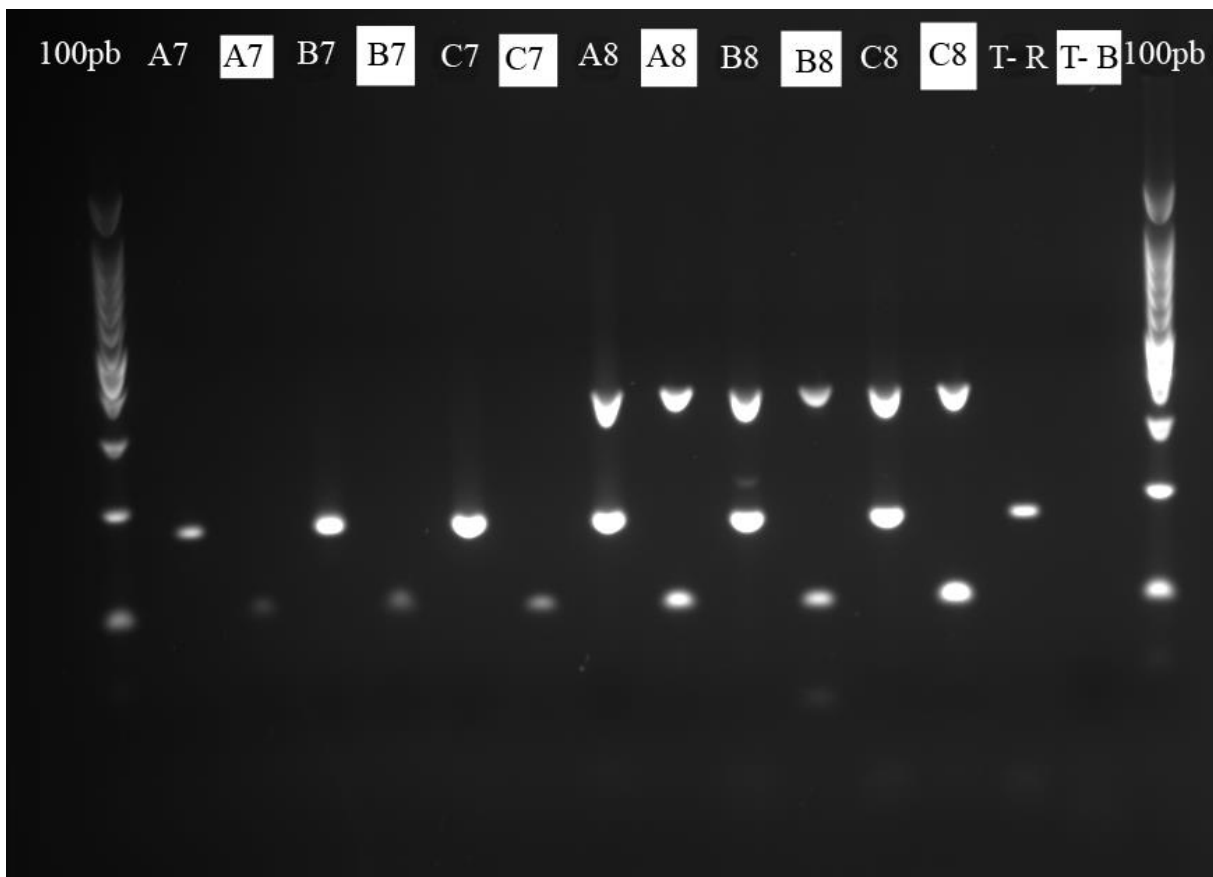


- Richard-Lenoble D, Kombila M, Duong TH, Gendrel D (1993) *Schistosoma intercalatum* schistosomiasis. A recent and forgotten schistosomiasis. *Rev Praticien* 43:432–439
- Sang T, Crawford DJ, Stuessy TF (1995) Documentation of reticulate evolution in peonies (*Paeonia*) using internal transcribed spacer sequences of nuclear ribosomal DNA: implications for biogeography and concerted evolution. *Proc Natl Acad Sci U S A* 92:6813–6817
- Soentjens P, Cnops L, Huyse T, Yansouni C, De Vos D, Bottieau E, Clerinx J, Van Esbroeck M (2016) Diagnosis and clinical management of *Schistosoma haematobium*–*Schistosoma bovis* hybrid infection in a cluster of travelers returning from Mali. *Clin Infect Dis* 63:1626–1629
- Théron A, Combes C (1988) Genetic analysis of cercarial emergence rhythms of *Schistosoma mansoni*. *Behavior Gen* 18:201–209
- Théron A, Mouahid G, Moné H (1997) *Schistosoma mansoni*: cercarial shedding patterns from a mixed infection of *Biomphalaria glabrata* with two (early and late) chronobiological variants. *Parasitol Res* 83:356–358
- Van den Broeck F, Geldof S, Polman K, Volckaert F, Huyse T (2011) Optimal sample storage and extraction protocols for reliable multilocus genotyping of the human parasite *Schistosoma mansoni*. *Infect Genet Evol* 11:1413–1418
- Webster BL, Southgate VR (2003) Compatibility of *Schistosoma haematobium*, *S. intercalatum* and their hybrids with *Bulinus truncatus* and *B. forskalii*. *Parasitology* 127:231–242
- Webster BL, Diaw OT, Seye MM, Webster JP, Rollinson D (2013) Introgressive hybridization of *Schistosoma haematobium* group species in Senegal: species barrier breakdown between ruminant and human schistosomes. *PLoS Negl Trop Dis* 7(4):e2110. <https://doi.org/10.1371/journal.pntd.0002110>
- Webster BL, Rollinson D, Stothard JR, Huyse T (2010) Rapid diagnostic multiplex PCR (RD-PCR) to discriminate *Schistosoma haematobium* and *S. bovis*. *J Helminthol* 84(107):114
- World Health Organization (2019) Schistosomiasis. Fact sheet 17 April 2019
- Wright C, Ross G (1980) Hybrids between *Schistosoma haematobium* and *S. matthei* and their identification by isoelectric focusing of enzymes. *Trans R Soc Trop Med Hyg* 74:326–332

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Annexe 2 : Test de la PCR multiplex pour le sexage avec les couples d'amorces Scaff\_02739 et Rhodopsine 3\_BIS

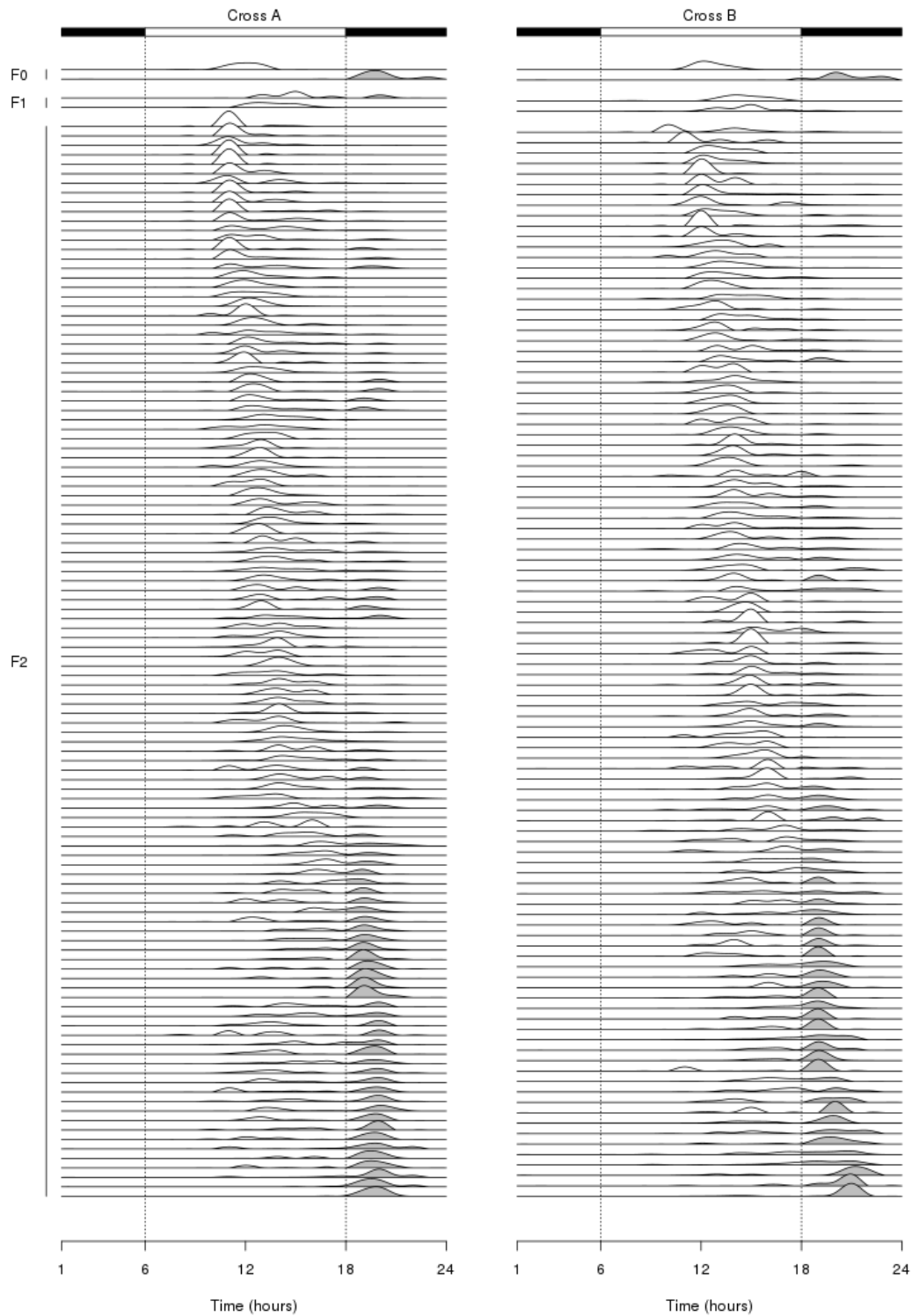
Lors du sexage pour le chronotype diurne, 6 plaques 96 puits ont été réalisées. Les deux dernières plaques ont montré que quatre puits vides avaient une bande à 200 pb correspondant à la Rhodopsine sur le gel d'électrophorèse, alors que les deux témoins négatifs et d'autres puits vides étaient sans bandes. Pour éliminer les diverses hypothèses de problème de contamination d'ADN entre les puits ou une mauvaise manipulation de la part de l'expérimentatrice, plusieurs PCR multiplex ont été refaites à partir d'une nouvelle dilution de l'ADN mère, des nouvelles solutions et mix, des nouvelles pipettes ainsi que par un autre expérimentateur. Les contrôles négatifs restant positifs, il a été décidé de remplacer le couple d'amorces Rhodopsine 3 par la Rhodopsine 3\_BIS (cf méthode 4.1). Une comparaison a été effectuée entre les couples d'amorces de Rhodopsine en utilisant l'ADN de trois mâles et trois femelles connus et des témoins négatifs (Figure\_A2).



**Figure A2 : Gel électrophorèse : comparaison des sexes réalisés avec les amorces Rhodopsine 3 et Rhodopsine 3\_BIS.** Les échantillons A7, B7 et C7 sont des mâles tandis que les échantillons A8, B8 et C8 sont des femelles. Le T- R est le témoin négatif du sexage réalisé avec les amorces Scaff\_02739/Rhodopsine 3. Le T- B est le témoin négatif du sexage réalisé avec les amorces Scaff\_02739/Rhodopsine 3\_BIS. Les échantillons dans un carré blanc sont des sexes réalisés avec les amorces Scaff\_02739/Rhodopsine 3\_BIS.

Pour le sexage réalisé avec les amorces Scaff\_02739/Rhodopsine 3, le diagnostic des sexes est correct. Toutefois le témoin négatif (T- R) présente une bande à 200 pb et est donc diagnostiqué comme un mâle. Bien que l'on ait un biais de mâle avec les amorces Scaff\_02739/Rhodopsine 3, le résultat des sexes du chronotype diurne n'est pas remis en cause. Pour le sexage réalisé avec les amorces Scaff\_02739/Rhodopsine 3\_BIS, le diagnostic des sexes est correct. De plus, le témoin négatif (T- B) est bien négatif. De ce fait, le couple Scaff\_02739/Rhodopsine 3\_BIS devient le nouveau couple d'amorces pour réaliser le sexage en routine.

### Annexe 3 : Répartition des profils chronobiologiques des générations F0, F1 et F2 des croisements intraspécifiques



**Figure A3 : Joyplots des profils chronobiologiques des générations F0, F1 et F2 des croisements intraspécifiques pour le QTL mapping.** Le croisement A est à gauche et le croisement B à droite. L'aire blanche sous la courbe correspond à une émission diurne des cercaires. L'aire grise sous la courbe correspond à une émission nocturne des cercaires.

## Annexe 4 : Liste des gènes candidats issus du QTL mapping des croisements intraspécifiques

Chr.	Gene ID	GFF annotation	HHsearch annotation
1	Smp_345740	Oligomycin resistance ATP-dependent permease Yor1	ABC_membrane_2
1	Smp_083740	EF-hand domain	Myosin Essential Chain {Bay scallop ( <i>Aequipecten irradians</i> ) [TaxId: 31199]}
1	Smp_319580	NA	Maltose transport protein MalK, N-terminal domain { <i>Archaeon Thermococcus litoralis</i> [TaxId: 2265]}
1	Smp_319570	NA	3-deoxy-D-arabino-heptulosonate-7-phosphate synthase (DAHP synthase, AroG) { <i>Escherichia coli</i> , phenylalanine-regulated isozyme [TaxId: 562]}
1	Smp_083730	NA	NA
1	Smp_083720	Mitochondrial substrate/solute carrier	ADP,ATP carrier protein {Cow ( <i>Bos taurus</i> ), heart isoform t1 [TaxId: 9913]}
1	Smp_083710	Helicase, C-terminal	PROBABLE ATP-DEPENDENT RNA HELICASE DDX58
1	Smp_167590	ASCH domain	Hypothetical protein TTHA0113 { <i>Thermus thermophilus</i> [TaxId: 274]}
1	Smp_083680	Ribosome biogenesis protein Nop53/GLTSCR2	Nop53
1	Smp_083650	Cys/Met metabolism, pyridoxal phosphate-dependent enzyme	Methionine gamma-lyase, MGL { <i>Pseudomonas putida</i> [TaxId: 303]}
1	Smp_337600	NA	NA
1	Smp_246990	Transcription factor CBF/NF-Y/archaeal histone domain	Histone H3.1, Histone H4, Histone
1	Smp_196840	Fibrillar collagen, C-terminal	Collagen alpha-1(I) chain, Collagen alpha-2(I)
1	Smp_103830	Zinc finger, HIT-type	Vacuolar protein sorting-associated protein 72
1	Smp_180140	G protein-coupled receptor, rhodopsin-like	P2Y purinoceptor 1, Rubredoxin
1	Smp_319540	NA	NA
1	Smp_158520	Diacylglycerol acyltransferase	DAGAT
1	Smp_344950	Diacylglycerol acyltransferase	DAGAT
1	Smp_210570	JAB1/MPN/MOV34 metalloenzyme domain	40S ribosomal protein SA, 40S
1	Smp_212850	LSM domain, eukaryotic/archaea-type	D1 core SNRNP protein {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
1	Smp_212840	PQ-loop repeat	Bidirectional sugar transporter SWEET2b (E.C.1-215)
1	Smp_158490	Zinc finger C2H2-type	Zinc finger protein 297b {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
1	Smp_158480	AMP-dependent synthetase/ligase	Acetyl-CoA synthetase {Baker's yeast ( <i>Saccharomyces cerevisiae</i> ) [TaxId: 4932]}

## Annexe 4

1	Smp_158470	NA	DUF3896
1	Smp_158460	SH3 domain	Hypothetical protein YFR024c {Baker's yeast ( <i>Saccharomyces cerevisiae</i> ) [TaxId: 4932]}
1	Smp_067140	Small GTPase	RhoA {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
1	Smp_158450	BTB/POZ domain	S-phase kinase-associated protein 1, F-box/LRR-repeat
1	Smp_159070	NA	Herpes_LMP1
1	Smp_159060	SRCR domain	Lysyl oxidase homolog 2 (E.C.1.4.3.13)
1	Smp_159050	Thrombospondin type-1 (TSP1) repeat	Semaphorin-6A
1	Smp_319520	NA	RIFIN
1	Smp_068120	Tetratricopeptide-like helical domain superfamily	FKBP51, C-terminal domain {Monkey ( <i>Saimiri boliviensis</i> ) [TaxId: 27679]}
1	Smp_068110	DNA double-strand break repair protein Mre11	Mre11
1	Smp_068100	Micro-fibrillar-associated protein 1, C-terminal	Pre-mRNA-processing-splicing factor 8, Thioredoxin-like protein
1	Smp_337420	Ribosomal protein S23, mitochondrial	MRP-S23
1	Smp_068070	Nucleoside phosphorylase superfamily	MTA/SAH nucleosidase (E.C.3.2.2.16,3.2.2.9)
1	Smp_068060	Protein kinase domain	Serine/threonine protein kinase TAO2 {Rat ( <i>Rattus norvegicus</i> ) [TaxId: 10116]}
1	Smp_159020	NA	Otopetrin
1	Smp_210820	GPCR kinase	Beta-adrenergic receptor kinase 1 (E.C.2.7.11.15)
1	Smp_210810	Arsenical pump ATPase, ArsA/GET3	Arsenite-translocating ATPase ArsA { <i>Escherichia coli</i> [TaxId: 562]}
1	Smp_319500	Small subunit of serine palmitoyltransferase-like	SPT_ssu-like
1	Smp_092770	Clathrin/coatomer adaptor, adaptin-like, N-terminal	ADP-RIBOSYLATION FACTOR 1, COATOMER SUBUNIT
1	Smp_092780	snRNA-activating protein complex, subunit 3	zf-SNAP50_C
1	Smp_060480	C2 domain	Copine
1	Smp_060490	Zinc finger, RanBP2-type	E3 ubiquitin-protein ligase RING2 (E.C.6.3.2.-)
1	Smp_154620	WD40 repeat	DUF1513
1	Smp_154640	High mobility group box domain	Transcription activator BRG1 (E.C.3.6.4.-), AT-rich
1	Smp_154650	FCP1 homology domain	Carboxy-terminal domain RNA polymerase II polypeptide A small phosphatase 1, NRAMP1 {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
1	Smp_319490	NA	Nucleoporin p54
1	Smp_332520	RIB43A	Tubulin beta, Tubulin alpha, PACRG
1	Smp_060560	Leucine-rich repeat	Ribonuclease inhibitor {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}

## Annexe 4

1	Smp_060570	Protein of unknown function DUF4735	DUF4735
1	Smp_154670	Glycosyl transferase, family 14	beta-1,6-N-acetylglucosaminyltransferase (E.C.2.4.1.102)
1	Smp_154680	Glycosyl transferase, family 14	beta-1,6-N-acetylglucosaminyltransferase (E.C.2.4.1.102)
1	Smp_346560	NA	NA
1	Smp_060620	Pyrimidine 5'-nucleotidase, eukaryotic	Cytosolic 5'-nucleotidase III {Mouse ( <i>Mus musculus</i> ) [TaxId: 10090]}
1	Smp_332510	Pyrimidine 5'-nucleotidase, eukaryotic	Cytosolic 5'-nucleotidase III {Mouse ( <i>Mus musculus</i> ) [TaxId: 10090]}
1	Smp_154690	Ribosomal protein L16	28S ribosomal protein S2, mitochondrial
1	Smp_319480	NA	NA
1	Smp_332500	DIX domain	Axin 1 protein
1	Smp_187360	Glycosyl transferase, family 14	beta-1,6-N-acetylglucosaminyltransferase (E.C.2.4.1.102)
1	Smp_142550	NA	Intimin adherence protein
1	Smp_142560	Homeobox domain	LIM domain-binding protein 1, LIM/homeobox
1	Smp_037780	NA	Lipase L1 { <i>Bacillus stearothermophilus</i> [TaxId: 1422]}
1	Smp_142600	High mobility group box domain	Lymphoid enhancer-binding factor, LEF1 {Mouse ( <i>Mus musculus</i> ) [TaxId: 10090]}
1	Smp_310870	Protein kinase domain	Cell division protein kinase 7, CDK7 {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
1	Smp_201000	NA	NA
1	Smp_319470	NA	NA
1	Smp_142620	Protein kinase domain	G protein-coupled receptor kinase 6
1	Smp_241440	NA	NA
1	Smp_343540	I/LWEQ domain	I_LWEQ
1	Smp_142640	YjeF N-terminal domain	Hypothetical protein YNL200c (YNUO_YEAST) {Baker's yeast ( <i>Saccharomyces cerevisiae</i> ) [TaxId: 4932]}
1	Smp_142650	PDZ domain	Neurabin-i {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
1	Smp_142670	Intraflagellar transport complex B protein 46	IFT46_B_C
1	Smp_037900	Beta-lactamase-related	D-Amino acid amidase DaaA { <i>Ochrobactrum anthropi</i> [TaxId: 529]}
1	Smp_037910	Succinate dehydrogenase [ubiquinone] cytochrome b small subunit, CybS	CybS
1	Smp_084620	Transcriptional coactivator p15 (PC4)	Transcriptional coactivator PC4 C-terminal domain {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
1	Smp_084610	UBX domain	UBX domain-containing protein 7 {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}

## Annexe 4

1	Smp_084600	Sugar phosphate transporter	Glycerol-3-phosphate transporter { <i>Escherichia coli</i> [TaxId: 562]}
1	Smp_168000	Major facilitator superfamily	Monocarboxylate transporter 2
1	Smp_319460	NA	NA
1	Smp_319450	NA	NA
1	Smp_319440	NA	NA
1	Smp_319430	NA	NA
1	Smp_319420	NA	NA
1	Smp_141630	Armadillo-like helical	Splicing factor 3B subunit 5
1	Smp_319410	Dynein light chain, type 1/2	Dynein light chain 1 (DLC1) {Fruit fly ( <i>Drosophila melanogaster</i> ) [TaxId: 7227]}
1	Smp_141640	NA	NA
1	Smp_201030	Dynein light chain, type 1/2	Dynein light chain 1 (DLC1) {Fruit fly ( <i>Drosophila melanogaster</i> ) [TaxId: 7227]}
1	Smp_036180	NA	NA
1	Smp_141650	Resistance to inhibitors of cholinesterase protein 3	RIC3
1	Smp_141660	Voltage-dependent calcium channel, L-type, beta subunit	Guanylate kinase-like domain of the L-type calcium channel {Rabbit ( <i>Oryctolagus cuniculus</i> ) [TaxId: 9986]}
1	Smp_036220	Histone H2B	Histone H2B {Chicken ( <i>Gallus gallus</i> ), erythrocytes [TaxId: 9031]}
1	Smp_343460	NA	Ankyrin-2, Spectrin beta chain
1	Smp_141680	FAS1 domain	Periostin
1	Smp_141690	NA	NA
1	Smp_036250	PRELI/MSF1 domain	LTP1, LTP1BP
1	Smp_036270	RNA recognition motif domain	Nucleolysin TIA-1 isoform p40
1	Smp_141710	NA	Rod_C
1	Smp_343470	Short-chain dehydrogenase/reductase SDR	17-beta-hydroxysteroid dehydrogenase type XI {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
1	Smp_301800	Cadherin-like	Protocadherin-10
1	Smp_092850	SprT-like	SprT-like domain-containing protein Spartan
1	Smp_104980	NA	PHD finger protein 7 (NYD-SP6) {Mouse ( <i>Mus musculus</i> ) [TaxId: 10090]}
1	Smp_180800	Learning associated protein 18-like	Nucleolar GTP-binding protein 2, 60S
1	Smp_180810	Pancreatic trypsin inhibitor Kunitz domain	CARBOXYPEPTIDASE A4 (E.C.3.4.17.-), CARBOXYPEPTIDASE INHIBITOR
1	Smp_105020	Immunoglobulin subtype 2	Titin
1	Smp_301790	NA	NA
1	Smp_105040	Endonuclease/exonuclease/phosphatase	2',5'-phosphodiesterase 12 (E.C.3.1.4.-,3.1.13.4)
1	Smp_315760	NA	Nucleoid-associated protein Rv3716c

## Annexe 4

1	Smp_347990	DNA/RNA non-specific endonuclease	Nuclease EXOG, mitochondrial/DNA (5'-D(*CP*GP*GP*GP*AP*TP*AP*TP*CP*CP*CP*G)-3') Complex
1	Smp_105090	Nuclear hormone receptor, ligand-binding domain	Hepatocyte nuclear factor 4-alpha, Nuclear
1	Smp_105100	Ribonuclease kappa	ATPase H+-transporting V1 subunit A
1	Smp_212230	Proteasome component (PCI) domain	COP9 signalosome complex subunit 1
1	Smp_332490	Alpha carbonic anhydrase domain	Carbonic anhydrase {Human ( <i>Homo sapiens</i> ), erythrocytes, isozyme I [TaxId: 9606]}
1	Smp_205000	NA	NA
1	Smp_243790	NA	Parathyroid hormone/parathyroid hormone-related peptide receptor
1	Smp_094890	Rho GTPase-activating protein domain	Beta-chimaerin, C-terminal domain {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
1	Smp_031920	Luc7-related	U1 small nuclear ribonucleoprotein 70
1	Smp_139470	NA	NA
1	Smp_301780	Protein kinase domain	Dual specificity mitogen-activated protein kinase kinase 1, Mek1 {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
1	Smp_031950	Programmed cell death protein 10	Programmed cell death protein 10
1	Smp_139480	Protein kinase domain	ephb2 receptor tyrosine kinase {Mouse ( <i>Mus musculus</i> ) [TaxId: 10090]}
1	Smp_031990	Amidase	Glutamyl-tRNA(Gln) amidotransferase subunit A { <i>Thermotoga maritima</i> [TaxId: 2336]}
1	Smp_139500	NA	MYEOV2
1	Smp_266940	NA	DUF3967
1	Smp_032000	Peptidase C14, p20 domain	Apaf-1 related killer DARK, Caspase
1	Smp_343320	G-patch domain	Pre-mRNA-splicing factor 8, Pre-mRNA-splicing factor
1	Smp_139520	SH3 domain	Hypothetical protein FLJ21935
1	Smp_032060	RNA recognition motif domain	Pre-mRNA branch site protein p14 {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
1	Smp_139530	p53 tumour suppressor family	p53 tumor suppressor, DNA-binding domain {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
1	Smp_139540	NA	NA
1	Smp_316850	G protein-coupled receptor, rhodopsin-like	C-C chemokine receptor type 2
1	Smp_212830	NA	DX
1	Smp_032120	NA	Ribonucleases P/MRP protein subunit POP6
1	Smp_139580	NA	Ribonucleases P/MRP protein subunit POP6
1	Smp_139590	NA	NPL4
1	Smp_032150	Snf7 family	Charged multivesicular body protein 3
1	Smp_139600	Zinc finger C2H2-type	Zinc finger protein 568/DNA Complex



## Annexe 4

1	Smp_309610	Protein Lines, N-terminal	LINES_N
1	Smp_032180	Small GTP-binding protein domain	ADP-ribosylation factor {Human ( <i>Homo sapiens</i> ), ARF6 [TaxId: 9606]}
1	Smp_319390	Protein Lines, N-terminal	LINES_N
1	Smp_139620	PDZ domain	Glutamate receptor interacting protein 2, GRIP2 (KIAA1719) {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
1	Smp_319380	NA	NA
1	Smp_213160	Protein kinase domain	WaaY
1	Smp_213170	NA	CT398
1	Smp_032230	Tetratricopeptide-like helical domain superfamily	Mitochondria fission protein Fis1 {Baker's yeast ( <i>Saccharomyces cerevisiae</i> ) [TaxId: 4932]}
1	Smp_319370	Tetratricopeptide-like helical domain superfamily	Mitochondria fission protein Fis1 {Baker's yeast ( <i>Saccharomyces cerevisiae</i> ) [TaxId: 4932]}
1	Smp_032240	MOZART2 family	MOZART2
1	Smp_032250	Beta-lactamase-related	Carboxylesterase (E.C.3.1.1.1)
1	Smp_032260	Ribosomal protein L15e	uL2, uS2, uL15, eS26, uL3
1	Smp_009650	Peptidase M16, zinc-binding site	Cytochrome b-c1 complex subunit 1
1	Smp_204580	G protein-coupled receptor, rhodopsin-like	Neurotensin receptor type 1, Neurotensin
1	Smp_009630	Homeobox domain	SIX1_SD
1	Smp_127240	Homeobox domain	Maltose-binding periplasmic protein, Homeobox protein
1	Smp_315890	JmjC domain	PHD FINGER PROTEIN 8
1	Smp_009600	Protein kinase domain	Aurora-related kinase 1 (aurora-2) {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
1	Smp_127220	NA	Corticotropin-releasing factor receptor 1
1	Smp_004990	Chaperonin TCP-1, conserved site	T-COMPLEX PROTEIN 1 SUBUNIT ALPHA
1	Smp_005000	PH-like domain superfamily	AP-2 complex subunit alpha-2, AP-2
1	Smp_004980	Ribosomal protein L20	28S ribosomal protein S2, mitochondrial
1	Smp_213790	Papain-like cysteine peptidase superfamily	Cysteine protease ATG4A {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
1	Smp_124820	Importin-beta, N-terminal domain	Chromosome region maintenance 1 (CRM1)
1	Smp_004940	Ribosomal protein L29/L35	Ribosomal protein L8, putative, Uncharacterized
1	Smp_340120	Malectin domain	Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit STT3B (E.C.2.4.99.18)
1	Smp_005070	DnaJ domain	Large T antigen, the N-terminal J domain {Simian virus 40, Sv40 [TaxId: 10633]}
1	Smp_005080	SAM-dependent methyltransferase RsmB/NOP2-type	60S ribosomal protein L13-A, 60S
1	Smp_124840	Zinc finger, UBR-type	Histone H3.2, Histone H4, Histone
1	Smp_319330	SAM-dependent methyltransferase RsmB/NOP2-type	60S ribosomal protein L13-A, 60S
1	Smp_124850	Protein kinase domain	Bruton's Tyrosine Kinase (E.C.2.7.10.2)

## Annexe 4

1	Smp_337760	Selenoprotein F/M	GLI-Krueppel family member HKR3 {Mouse ( <i>Mus musculus</i> ) [TaxId: 10090]}
1	Smp_124860	Methyltransferase Ppm1/Ppm2/Tcmp	Leucine carboxy methyltransferase Ppm1 {Baker's yeast ( <i>Saccharomyces cerevisiae</i> ) [TaxId: 4932]}
1	Smp_005140	Thrombospondin type-1 (TSP1) repeat	Vitelline membrane outer protein-I (VMO-I) {Hen ( <i>Gallus gallus</i> ) [TaxId: 9031]}
1	Smp_005160	Zinc finger, CCCH-type	Muscleblind-like protein 1/RNA Complex
1	Smp_005180	Transcription factor, GTP-binding domain	uL2, uL3, uL4, 60S ribosomal
1	Smp_124870	NA	Activity-regulated cytoskeleton associated protein 1
1	Smp_319320	Acylphosphatase-like domain	Acylphosphatase 2 (Cg18505) {Fruit fly ( <i>Drosophila melanogaster</i> ) [TaxId: 7227]}
1	Smp_005190	Protein kinase domain	STE20-like serine/threonine-protein kinase, SLK {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
1	Smp_005210	Histone deacetylase family	METASTASIS-ASSOCIATED PROTEIN MTA1, HISTONE DEACETYLASE
1	Smp_093440	Ribosomal protein S5/S7	Ribosomal protein S7 {Archaeon <i>Pyrococcus horikoshii</i> [TaxId: 53953]}
1	Smp_173360	NA	DUF2613
1	Smp_124890	NA	DUF2613
1	Smp_005230	EF-hand domain	Calcium-dependent protein kinase sk5 CLD {Soybean ( <i>Glycine max</i> ) [TaxId: 3847]}
1	Smp_342080	Stonin homology	Second domain of Mu2 adaptin subunit (ap50) of ap2 adaptor {Rat ( <i>Rattus norvegicus</i> ) [TaxId: 10116]}
1	Smp_124920	Endoplasmic reticulum oxidoreductin 1	ERO1-like protein alpha (E.C.1.8.4.-)
1	Smp_028170	NA	zf_C2H2_10
1	Smp_246940	LIS1 homology motif	LisH_2
1	Smp_028140	FCH domain	Growth arrest-specific protein 7
1	Smp_028130	Protein FAM78	Uncharacterized protein PA3332 { <i>Pseudomonas aeruginosa</i> [TaxId: 287]}
1	Smp_028120	Dynamin, GTPase domain	Dynamin-like 120 kDa protein, mitochondrial,OPA1
1	Smp_137410	EF-hand domain pair	M-CALPAIN LARGE AND SMALL SUBUNITS
1	Smp_028100	Ubiquitin-conjugating enzyme E2	Ubiquitin conjugating enzyme, UBC {Nematode ( <i>Caenorhabditis elegans</i> ), E2-19 kDa [TaxId: 6239]}
1	Smp_137400	NA	Brain-specific angiogenesis inhibitor 1-associated protein
1	Smp_137390	Histone RNA stem-loop-binding protein SLBP1/SLBP2	SLBP/RNA Complex

## Annexe 4

1	Smp_028080	Aminotransferase, class I/classII	5-aminolevulinate synthase { <i>Rhodobacter capsulatus</i> [TaxId: 1061]}
1	Smp_028070	Bicaudal-D protein	BicD
1	Smp_028060	Surfeit locus 6	SURF6
1	Smp_028050	NA	NA
1	Smp_028030	Rho GTPase-activating protein domain	SLIT-ROBO Rho GTPase-activating protein 2
1	Smp_137380	Forkhead-associated (FHA) domain	Pre-mRNA-processing-splicing factor 8, 116 kDa
1	Smp_317370	NA	NA
1	Smp_118040	G protein-coupled receptor, rhodopsin-like	Chimera of Proteinase-activated receptor 1
1	Smp_137360	NA	Collectrin
1	Smp_027990	Homeobox domain	Homeobox protein hox-a9 {Mouse ( <i>Mus musculus</i> ) [TaxId: 10090]}
1	Smp_027970	Helicase, C-terminal	RNA-binding motif protein, X-linked 2
1	Smp_137330	RNA recognition motif domain	L-serine dehydratase {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
1	Smp_137320	G protein-coupled receptor, rhodopsin-like	Srg
1	Smp_027940	G protein-coupled receptor, rhodopsin-like	ADENOSINE RECEPTOR A2A
1	Smp_137310	G protein-coupled receptor, rhodopsin-like	7TM_GPCR_Srw
1	Smp_137300	G protein-coupled receptor, rhodopsin-like	7TM_GPCR_Srw
1	Smp_319310	NA	7TM_GPCR_Srab
1	Smp_178420	GPCR, rhodopsin-like, 7TM	ADENOSINE RECEPTOR A2A
1	Smp_168080	RmlC-like cupin domain superfamily	Hypothetical protein MJ0764 { <i>Archaeon Methanococcus jannaschii</i> [TaxId: 2190]}
1	Smp_332480	TNFR/NGFR cysteine-rich region	Tumor necrosis factor ligand superfamily
1	Smp_180150	Rho GTPase-activating protein domain	Beta-chimaerin, C-terminal domain {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
1	Smp_180160	Rho GTPase-activating protein domain	Graf {Chicken ( <i>Gallus gallus</i> ) [TaxId: 9031]}
1	Smp_180170	Ubiquitin-conjugating enzyme E2	Ubiquitin conjugating enzyme, UBC {Human( <i>Homo sapiens</i> ), E2 S [TaxId: 9606]}
1	Smp_175680	Tudor domain	Survival motor neuron protein 1, smn {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
1	Smp_213980	K Homology domain	neuro-oncological ventral antigen 1/RNA Complex
1	Smp_213990	NA	Membrane-associated protein slr1513
1	Smp_124470	Adenylyl cyclase class-3/4/guanylyl cyclase	Soluble guanylyl cyclase alpha-1 subunit
1	Smp_097460	Adenylyl cyclase class-3/4/guanylyl cyclase	Soluble guanylyl cyclase alpha-1 subunit

## Annexe 4

1	Smp_175710	Helix-turn-helix motif	Homeobox protein hox-b13 {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
1	Smp_097490	P25-alpha	Tubulin polymerization-promoting protein family member
1	Smp_097500	F-box domain	S-phase kinase-associated protein 1/F-box only
1	Smp_319300	NA	NA
1	Smp_319290	NA	Corona_S2
1	Smp_332470	B-box-type zinc finger	E3 UBIQUITIN/ISG15 LIGASE TRIM25 (E.C.6.3.2.19)
1	Smp_132830	Helix-loop-helix DNA-binding domain superfamily	Pho4 B/HLH domain {Baker's yeast ( <i>Saccharomyces cerevisiae</i> ) [TaxId: 4932]}
1	Smp_319280	NA	NA
1	Smp_019490	Retrieval of early ER protein Rer1	Rer1
1	Smp_132810	Myc-type, basic helix-loop-helix (bHLH) domain	Myc proto-oncogene protein {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
1	Smp_132800	Myc-type, basic helix-loop-helix (bHLH) domain	Aryl hydrocarbon receptor nuclear translocator-like
1	Smp_019460	Myc-type, basic helix-loop-helix (bHLH) domain	SREBP-1a {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
1	Smp_132790	Myc-type, basic helix-loop-helix (bHLH) domain	Transcription factor HES-1
1	Smp_332460	Helix-loop-helix DNA-binding domain superfamily	Usf B/HLH domain {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
1	Smp_019440	SUI1 domain	Eukaryotic translation initiation factor eIF-1 (SUI1) {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
1	Smp_019430	Matrin/U1-C-like, C2H2-type zinc finger	RNA binding protein ZFa
1	Smp_331670	Uncharacterised protein family SERF, N-terminal	4F5
1	Smp_301670	Zinc finger, RING-type	E3 ubiquitin-protein ligase parkin (E.C.2.3.2.31)
1	Smp_019380	W2 domain	Translation initiation factor eIF-2B subunit
1	Smp_132750	G-protein alpha subunit, group Q	G-protein alpha subunit Galpha7
1	Smp_019350	NA	ALPHA-BUNGAROTOXIN ISOFORM V31, ALPHA-BUNGAROTOXIN ISOFORM
1	Smp_019340	Tetraspanin	Tetraspanin
1	Smp_132740	Cytochrome b5-like heme/steroid binding domain	FA_desaturase
1	Smp_019310	Dynein light chain, type 1/2	Dynein light chain 1 (DLC1) {Fruit fly ( <i>Drosophila melanogaster</i> ) [TaxId: 7227]}
1	Smp_319270	Dynein light chain, type 1/2	Dynein light chain 1 (DLC1) {Fruit fly ( <i>Drosophila melanogaster</i> ) [TaxId: 7227]}
1	Smp_319260	Dynein light chain, type 1/2	Dynein light chain 1 (DLC1) {Fruit fly ( <i>Drosophila melanogaster</i> ) [TaxId: 7227]}
1	Smp_132730	G protein-coupled receptor, rhodopsin-like	Neurotensin receptor type 1,Endolysin,NEUROTENSIN RECEPTOR

## Annexe 4

1	Smp_319250	NA	Engineered transmembrane domain variant
1	Smp_019280	Cyclic nucleotide-binding domain	Protein kinase, cAMP-dependent, catalytic, alpha
1	Smp_019260	Ankyrin repeat	Myotrophin {Rat ( <i>Rattus norvegicus</i> ) [TaxId: 10116]}
1	Smp_342710	Dual specificity phosphatase, catalytic domain	31 6DT1_A DNA dC->dU-editing enzy 98.1 1.6E-08 1.4E-13 100.3 0.0 261 353-651 171-505 (507)
1	Smp_132700	Immunoglobulin subtype 2	CARCINOEMBRYONIC ANTIGEN
1	Smp_019190	Pre-mRNA-splicing factor SLU7 domain	Protein mago nashi homolog 2
1	Smp_132690	AAA ATPase domain	RuvB-like 2 (E.C.3.6.4.12)
1	Smp_019170	ARID DNA-binding domain	Lysine-specific demethylase 5A (E.C.1.14.11.-)
1	Smp_132670	EF-hand domain	Myosin Essential Chain {Bay scallop ( <i>Aequipecten irradians</i> ) [TaxId: 31199]}
1	Smp_132660	Dbl homology (DH) domain	FYVE, RhoGEF and PH domain-containing
1	Smp_201130	Tumor protein p53-inducible protein 11	p53-inducible11
1	Smp_132640	CDP-alcohol phosphatidyltransferase	CTP L-myo-inositol-1-phosphate cytidyltransferase/CDP-L-myo-inositol myo-inositolphosphotransferase
1	Smp_019130	Pre-mRNA cleavage complex subunit Clp1, C-terminal	Protein clpf-1
1	Smp_301660	NA	NA
1	Smp_132620	Dynamin superfamily	Mitofusin-2,cDNA FLJ57997, highly similar to
1	Smp_019100	Dynamin superfamily	EH DOMAIN-CONTAINING PROTEIN-2
1	Smp_132610	NA	NA
1	Smp_000020	Beta-N-acetylglucosaminidase	Protein O-GlcNAcase (E.C.3.2.1.169,3.2.1.-)
1	Smp_000030	Armadillo-type fold	RPN1_RPN2_N
1	Smp_000040	Kinesin light chain	Kinesin light chain 2, C-Jun-amino-terminal
1	Smp_000050	Ankyrin repeat	nvTRPM2 channel
1	Smp_000070	Phospholipid/glycerol acyltransferase	1-acyl-sn-glycerol-3-phosphate acyltransferase (E.C.2.3.1.51)
1	Smp_000075	Homeobox domain	Engrailed Homeodomain { <i>Drosophila melanogaster</i> [TaxId: 7227]}
1	Smp_319230	Target SNARE coiled-coil homology domain	Syntaxin-1A
1	Smp_000080	Target SNARE coiled-coil homology domain	Syntaxin-1A
1	Smp_335570	Sodium channel and clathrin linker 1	CCCAP
1	Smp_319220	LSM domain, eukaryotic/archaea-type	U6 snRNA-associated Sm-like protein LSm2
1	Smp_346820	Sodium channel and clathrin linker 1	CCCAP
1	Smp_127580	DIS3-like exonuclease 2, C-terminal	Dis3L2/RNA Complex

## Annexe 4

1	Smp_319210	Ribonuclease II/R	Dis3L2/RNA Complex
1	Smp_248110	SANT/Myb domain	DNA-binding protein reb1/DNA Complex
1	Smp_010230	G-protein, beta subunit	beta1-subunit of the signal-transducing G protein heterotrimer {Cow ( <i>Bos taurus</i> ) [TaxId: 9913]}
1	Smp_127600	Cyclin-like	Cyclin-dependent protein kinase PHO85 (E.C.2.7.11.22)
1	Smp_010250	Dbl homology (DH) domain	Triple functional domain protein TRIO {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
1	Smp_010260	XPA	General transcription and DNA repair
1	Smp_248120	NA	HAUS4
1	Smp_010280	Phospholipase/carboxylesterase/thioesterase	Putative serine hydrolase Ydr428c {Baker's yeast ( <i>Saccharomyces cerevisiae</i> ) [TaxId: 4932]}
1	Smp_127450	IQ motif, EF-hand binding site	MYOSIN 2 HEAVY CHAIN STRIATED
1	Smp_127440	Centrosome-associated protein 350	NA
1	Smp_127430	NA	Probable metalloprotease ARX1, rRNA-processing protein
1	Smp_315790	CBS domain	5'-AMP-activated protein kinase catalytic subunit
1	Smp_127380	HAP1, N-terminal	HAP1_N
1	Smp_009830	Translocon-associated protein subunit beta	TRAP_beta
1	Smp_127370	Glycosyl transferase, family 14	beta-1,6-N-acetylglucosaminyltransferase (E.C.2.4.1.102)
1	Smp_009810	NA	beta-1,6-N-acetylglucosaminyltransferase (E.C.2.4.1.102)
1	Smp_319200	C-type lectin fold	NK cell receptor {Mouse ( <i>Mus musculus</i> ), ly49-c [TaxId: 10090]}
1	Smp_127360	NA	NA
1	Smp_009800	Protein kinase domain	Serine/threonine-protein kinase Nek2 {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
1	Smp_337140	NADH-quinone oxidoreductase, subunit D	NADH-ubiquinone oxidoreductase chain 3 (E.C.1.6.5.3)
1	Smp_301620	14-3-3 protein	zeta isoform {Cow ( <i>Bos taurus</i> ) [TaxId: 9913]}
1	Smp_301610	NA	Neutrophil cytosol factor 4
1	Smp_009760	14-3-3 protein	zeta isoform {Cow ( <i>Bos taurus</i> ) [TaxId: 9913]}
1	Smp_127310	G protein-coupled receptor, rhodopsin-like	Chimera of Proteinase-activated receptor 1
1	Smp_121460	NA	NA
1	Smp_341860	50S ribosomal protein L30e-like	Ribosome
1	Smp_121430	Proteasome, subunit alpha/beta	Proteasome beta subunit (catalytic) {Cow ( <i>Bos taurus</i> ) [TaxId: 9913]}
1	Smp_121420	Myristoyl-CoA	GLYCYLPEPTIDE N-TETRADECANOYLTRANSFERASE (E.C.2.3.1.97)
1	Smp_194430	PRC-barrel-like superfamily	RIKEN cDNA 2310057j16 protein (KIAA1543) {Mouse ( <i>Mus musculus</i> ) [TaxId: 10090]}

## Annexe 4

1	Smp_332450	Equilibrative nucleoside transporter	Equilibrative nucleoside transporter 1
1	Smp_337200	Putative homeodomain transcription factor 1/2	NA
1	Smp_173840	AAA ATPase domain	Proteasome subunit alpha type-6 (E.C.3.4.25.1)
1	Smp_094320	Zinc finger, CCCH-type	zf-C3H1
1	Smp_315830	NA	Muscarinic acetylcholine receptor M5, T4
1	Smp_102690	PAZ domain	Argonaute2/RNA Complex
1	Smp_179320	PAZ domain	CRISPR-associated endonuclease Cpf1/RNA Complex
1	Smp_146460	Dynein light chain, type 1/2	Tegumental protein 20.8 kDa
1	Smp_045010	Dynein light chain, type 1/2	CALCIUM BINDING PROTEIN
1	Smp_146470	NA	Rhodopsin
1	Smp_319170	NA	Neurotensin receptor type 1, Endolysin, NEUROTENSIN RECEPTOR
1	Smp_045200	Dynein light chain, type 1/2	Tegumental protein 20.8 kDa
1	Smp_337040	Ribosomal protein L11/L12	28S ribosomal protein S2, mitochondrial
1	Smp_319140	Mediator complex, subunit Med27	Med27
1	Smp_319130	FCH domain	SLIT-ROBO Rho GTPase-activating protein 2
1	Smp_319120	Zinc finger, RING-type	zf-PHD-like
1	Smp_319110	FeS cluster biogenesis	Fe-S scaffold protein IscA (YfhF) { <i>Escherichia coli</i> [TaxId: 562]}
1	Smp_319100	PPM-type phosphatase, divalent cation binding	Protein serine/threonine phosphatase 2C, catalytic domain {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
1	Smp_319090	NA	HIN recombinase (DNA-binding domain) {Synthetic}
1	Smp_319080	NA	NA
1	Smp_319070	NA	DUF2852
1	Smp_319060	Matrin/U1-C-like, C2H2-type zinc finger	zf-U1
1	Smp_045070	WD40 repeat	Groucho/tle1, C-terminal domain {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
1	Smp_045060	WD40 repeat	Cell cycle arrest protein BUB3 {Baker's yeast ( <i>Saccharomyces cerevisiae</i> ) [TaxId: 4932]}
1	Smp_319050	Transmembrane protein 164	TMEM164
1	Smp_316440	Ribosomal RNA methyltransferase FtsJ domain	60S ribosomal protein L13-A, 60S
1	Smp_146480	NA	NA
1	Smp_140250	G protein-coupled receptor, rhodopsin-like	Platelet-activating factor receptor, Flavodoxin
1	Smp_140260	CARD domain	Apaf-1 related killer DARK

## Annexe 4

1	Smp_140270	NAD(P)-binding domain superfamily	Carbonyl reductase/20beta-hydroxysteroid dehydrogenase {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
1	Smp_033530	Short-chain dehydrogenase/reductase SDR	Carbonyl reductase/20beta-hydroxysteroid dehydrogenase {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
1	Smp_033540	Short-chain dehydrogenase/reductase SDR	Carbonyl reductase/20beta-hydroxysteroid dehydrogenase {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
1	Smp_152870	NA	NA
1	Smp_200980	NA	Type-B carboxylesterase/lipase {Fungus ( <i>Geotrichum candidum</i> ), ATCC 34614 [TaxId: 27317]}
1	Smp_319040	NA	NA
1	Smp_152860	Tetratricopeptide-like helical domain superfamily	Protein farnesyltransferase alpha subunit (E.C.2.5.1.-)
1	Smp_057230	Aminoacyl-tRNA synthetase, class II (G/ P/ S/T)	Seryl-tRNA synthetase (SerRS) { <i>Thermus thermophilus</i> , strain hb27 [TaxId: 274]}
1	Smp_152850	Tetratricopeptide-like helical domain superfamily	Putative 70 kda peptidylprolyl isomerase PFL2275c { <i>Plasmodium falciparum</i> [TaxId: 5833]}
1	Smp_057210	Leucine-rich repeat	Slit {Fruit fly ( <i>Drosophila melanogaster</i> ) [TaxId: 7227]}
1	Smp_152840	Helicase, C-terminal	CG9323, isoform A/DNA Complex
1	Smp_142620	Na-Ca exchanger/integrin-beta4	Na/Ca exchange protein
1	Smp_210470	Rab-GTPase-TBC domain	TBC1 domain family member 22A
1	Smp_337800	Bromodomain	CREB-binding protein, CBP {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
1	Smp_319010	NA	Uncharacterized protein BT2961 { <i>Bacteroides thetaiotaomicron</i> [TaxId: 818]}
1	Smp_057120	GYF domain	Hypothetical rotein At5g08430 {Thale cress ( <i>Arabidopsis thaliana</i> ) [TaxId: 3702]}
1	Smp_200960	NA	Guanine nucleotide dissociation inhibitor, GDI {Baker's yeast ( <i>Saccharomyces cerevisiae</i> ) [TaxId: 4932]}
1	Smp_197230	Adenylate kinase/UMP-CMP kinase	Adenylate kinase {Pig ( <i>Sus scrofa</i> ) [TaxId: 9823]}
1	Smp_301570	F-box domain	Protein prenyltransferase alpha subunit repeat-containing
1	Smp_301560	Leucine-rich repeat	Cyclin A/CDK2-associated p19, Skp2 {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
1	Smp_319000	NA	Kunitz-type protease inhibitor 1
1	Smp_057090	Radial spokehead-like protein	Radial_spoke
1	Smp_152800	Armadillo-like helical	Exportin-5/13-mer peptide/GTP-binding nuclear protein Ran/RNA
1	Smp_152790	Gtr1/RagA G protein	Ragulator complex protein LAMTOR3, Ragulator



## Annexe 4

1	Smp_057060	Acetyl-CoA hydrolase/transferase	4-hydroxybutyrate CoA-transferase
1	Smp_301530	NA	NA
1	Smp_057020	ATP-dependent RNA helicase DEAD-box, conserved site	Ribosome biogenesis protein RLP7, Ribosome
1	Smp_344520	NA	Golgin_A5
1	Smp_056990	Ionotropic glutamate receptor	Glutamate receptor ionotropic, delta-2
1	Smp_152760	Cyclin-Q	Cyclin K {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
1	Smp_152750	PTP type protein phosphatase	Receptor-type tyrosine-protein phosphatase zeta
1	Smp_056970	Glyceraldehyde-3-phosphate dehydrogenase, type I	Glyceraldehyde-3-phosphate dehydrogenase (E.C.1.2.1.12)
1	Smp_152740	WD40 repeat	40S ribosomal protein S1-A, 40S
1	Smp_056930	Zinc finger, RING-type	Myeloid/lymphoid or mixed-lineage leukemia protein
1	Smp_193610	G protein-coupled receptor, rhodopsin-like	Lysozyme C (E.C.3.2.1.17)
1	Smp_008450	Leucine-rich repeat	Protein toll, Protein spaetzle C-106
1	Smp_008440	DNA-directed RNA polymerase, subunit N/Rpb10	DNA-directed RNA polymerase subunit A'
1	Smp_008430	NA	Protease Do (DegP, HtrA), C-terminal domains { <i>Escherichia coli</i> [TaxId: 562]}
1	Smp_008470	CTLH, C-terminal LisH motif	Vacuolar import and degradation protein
1	Smp_126720	Pre-mRNA-splicing factor Isy1	Pre-mRNA-processing-splicing factor 8, 116 kDa
1	Smp_318980	BOLA protein	SufE-like protein, chloroplastic
1	Smp_318970	Domain of unknown function DUF1330	uncharacterized conserved protein
1	Smp_008490	Glycosyl transferase, family 8	GLYCOGENIN-1 (E.C.2.4.1.186)
1	Smp_008500	Transcription factor IIS, N-terminal	Transcription factor IWS1
1	Smp_126730	G protein-coupled receptor, rhodopsin-like	7tm_1
1	Smp_094020	ATP-dependent RNA helicase DEAD-box, conserved site	Ribosome biogenesis protein RLP7, Ribosome
1	Smp_173640	NA	TatT
1	Smp_347850	NA	Rap guanine nucleotide exchange factor 5, RapGEF5 {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
1	Smp_246820	NA	hypothetical protein TTHA1254
1	Smp_093980	NA	NA
1	Smp_173620	Transmembrane protein TqsA-like	AI-2E_transport
1	Smp_318960	NA	NA
1	Smp_093950	OTU domain	OTU domain-containing protein 5 (E.C.3.4.19.12)
1	Smp_173610	NA	YuzL
1	Smp_200540	NA	NA
1	Smp_318950	Histone H3/CENP-A	<i>Leishmania</i> histone H3, Histone H4

## Annexe 4

1	Smp_125980	Tetratricopeptide-like helical domain superfamily	PARTNER OF INSCUTEABLE, RE60102P
1	Smp_193150	NA	NA
1	Smp_126000	NA	NA
1	Smp_007070	G protein-coupled receptor, rhodopsin-like	M3-mT4L, Tiotropium
1	Smp_126020	Fe-S hydro-lyase, tartrate dehydratase alpha-type, catalytic domain	Fumarate hydratase 2 (E.C.4.2.1.2)
1	Smp_126030	Homeobox domain	PBC
1	Smp_126050	Cadherin-like	Protocadherin-15
1	Smp_126070	NA	NA
1	Smp_126080	Arrestin	Bovine arrestin-2 (full length)
1	Smp_007170	Armadillo-like helical	Mo25 protein {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
1	Smp_007180	CID domain	Putative RNA-binding protein 16
1	Smp_007190	Lariat debranching enzyme, C-terminal	DBR1
1	Smp_110280	Srp40, C-terminal	SRP40_C
1	Smp_007210	Rab-GTPase-TBC domain	Rab GTPase-activating protein 1
1	Smp_196700	WD40/YVTN repeat-like-containing domain superfamily	Bardet-Biedl syndrome 18 protein, BBS1
1	Smp_318920	NA	Biopolymer transport protein ExbB, TonB
1	Smp_318900	NA	PTS system, cellobiose-specific IIB component
1	Smp_318890	NA	Biopolymer transport protein ExbB, TonB
1	Smp_126110	AAA ATPase domain	Spastin (E.C.5.6.1.1), polyglutamate peptide
1	Smp_007260	P-type ATPase	SARCOPLASMIC/ENDOPLASMIC RETICULUM CALCIUM ATPASE 1
1	Smp_007270	Calponin homology domain	Actin binding domain of plectin {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
1	Smp_007290	NA	UPF0249 protein EF_3048
1	Smp_007300	NA	Coiled-coil-helix-coiled-coil-helix domain-containing protein 7
1	Smp_301430	Phospholipase B-like	Putative phospholipase B-like 2 (E.C.3.1.1.-)
1	Smp_007330	NA	NA
1	Smp_210450	Ribosomal protein S23/S29, mitochondrial	Ribosomal protein L2, 50S ribosomal
1	Smp_210440	Aldehyde dehydrogenase domain	Aldehyde reductase (dehydrogenase), ALDH {Baltic cod ( <i>Gadus callarias</i> ) [TaxId: 8053]}
1	Smp_196480	Small integral membrane protein 15	UPF0542
1	Smp_114160	GTP binding domain	Probable GTPase YlqF { <i>Bacillus subtilis</i> [TaxId: 1423]}
1	Smp_007380	UPF0428 family	DUF2325

## Annexe 4

1	Smp_126140	Integrin alpha chain	Integrin alpha-V, Integrin beta-3
1	Smp_158330	Coiled-coil domain-containing protein 170-like	CT398
1	Smp_158320	NA	NA
1	Smp_066830	Protein N-terminal asparagine amidohydrolase	Protein N-terminal asparagine amidohydrolase (E.C.3.5.1.-)
1	Smp_066820	von Willebrand factor, type A	INT_SG_DDX_CT_C
1	Smp_093580	Helicase, C-terminal	Pre-mRNA-splicing factor 8, Pre-mRNA-splicing factor
1	Smp_173420	Leucine-rich repeat	von Willebrand factor binding domain of glycoprotein Ib alpha {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
1	Smp_173410	Mediator complex, subunit Med30, metazoa	Med30
1	Smp_093540	Ser/Thr protein kinase, TGFB receptor	Type I TGF-beta receptor R4 {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
1	Smp_347110	Ribosomal protein L5 eukaryotic/L18 archaeal	60S ribosomal protein L8, 60S
1	Smp_346200	SANT/Myb domain	DnaJ chaperone, N-terminal (J) domain { <i>Escherichia coli</i> [TaxId: 562]}
1	Smp_173380	Protein kinase domain	Musk tyrosine kinase {Rat ( <i>Rattus norvegicus</i> ) [TaxId: 10116]}
1	Smp_025620	Coiled-coil domain-containing protein 151	Tubulin beta, Tubulin alpha, PACRG
1	Smp_115290	Ran binding domain	Ran-binding protein 1, Ranbp1 {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
1	Smp_136090	Ribosomal protein L30, ferredoxin-like fold domain	60S ribosomal protein L8, 60S
1	Smp_343010	Kelch repeat type 1	S-phase kinase-associated protein 1, F-box/LRR-repeat
1	Smp_136110	NA	Lin-12 and glp-1 phenotype protein
1	Smp_343020	Myc-type, basic helix-loop-helix (bHLH) domain	DNA-binding protein inhibitor ID-1
1	Smp_025670	Zinc finger C2H2-type	Zinc finger protein 568/DNA Complex
1	Smp_025680	CSN8/PSMD8/EIF3K	COP9 signalosome complex subunit 1
1	Smp_331660	NA	26S proteasome regulatory subunit RPN9
1	Smp_025700	Kelch-type beta propeller	Kelch-like ECH-associated protein 1, KEAP1 {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
1	Smp_200910	NA	NA
1	Smp_266750	NA	NA
1	Smp_318870	NA	IIA domain of mannose transporter, IIA-Man { <i>Escherichia coli</i> [TaxId: 562]}
1	Smp_136130	Parathyroid hormone-responsive B1	Bardet-Biedl syndrome 18 protein, BBS1

## Annexe 4

1	Smp_136190	Nuclear protein MDM1	MDM1
1	Smp_025740	Small GTPase	RhoE (RND3) {Mouse ( <i>Mus musculus</i> ) [TaxId: 10090]}
1	Smp_025750	NA	NA
1	Smp_200530	NA	Weak toxin DE-1 homolog 1
1	Smp_305800	Proteasome, subunit alpha/beta	Proteasome beta subunit (catalytic) {Baker's yeast ( <i>Saccharomyces cerevisiae</i> ) [TaxId: 4932]}
1	Smp_213780	Transmembrane protein 223	TMEM223
1	Smp_303770	NA	DNA-directed RNA polymerase subunit, DNA-directed
1	Smp_301340	Proteasome, subunit alpha/beta	Proteasome beta subunit (catalytic) {Baker's yeast ( <i>Saccharomyces cerevisiae</i> ) [TaxId: 4932]}
1	Smp_318850	RNA recognition motif domain	DNA-directed RNA polymerase subunit, DNA-directed
1	Smp_025830	ADP-ribosyl cyclase (CD38/157)	Bone marrow stromal cell antigen 1, BST-1/CD157 (ADP ribosyl cyclase-2) {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
1	Smp_136150	Leucine-rich repeat	Rna1p (RanGAP1), N-terminal domain {Fission yeast ( <i>Schizosaccharomyces pombe</i> ) [TaxId: 4896]}
1	Smp_025860	Beta-thymosin	Actin, alpha skeletal muscle, Thymosin
1	Smp_136160	p53 tumour suppressor family	p53 tumor suppressor, DNA-binding domain {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
1	Smp_025890	Rab-GTPase-TBC domain	TBC1 domain family member 22B
1	Smp_343030	Sterol-sensing domain	Protein dispatched
5	Smp_165490	Calcineurin-like phosphoesterase domain, ApaH type	Protein phosphatase 2A catalytic subunit alpha isoform, PP2A-alpha {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
5	Smp_334460	NA	Lebercilin
5	Smp_165530	Rho GTPase-activating protein domain	p50 RhoGAP domain {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
5	Smp_070600	Transaldolase/Fructose-6-phosphate aldolase	Transaldolase { <i>Escherichia coli</i> [TaxId: 562]}
5	Smp_329930	NA	NA
5	Smp_329940	NA	NA
5	Smp_070610	MOB kinase activator family	Mob1a {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
5	Smp_212250	Short-chain dehydrogenase/reductase SDR	Sepiapterin reductase {Mouse ( <i>Mus musculus</i> ) [TaxId: 10090]}
5	Smp_160530	N2227-like	Carnosine N-methyltransferase (E.C.2.1.1.22)
5	Smp_070640	FERM/acyl-CoA-binding protein superfamily	Fermitin family homolog 3
5	Smp_160540	Primosome PriB/single-strand DNA-binding	Primosomal replication protein n
5	Smp_345210	Dyp-type peroxidase	Melanin biosynthesis protein TyrA { <i>Shewanella oneidensis</i> [TaxId: 70863]}

## Annexe 4

5	Smp_314200	Dyp-type peroxidase	Hypothetical protein BT1219 { <i>Bacteroides thetaiotaomicron</i> [TaxId: 818]}
5	Smp_317150	Dyp-type peroxidase	Hypothetical protein BT1219 { <i>Bacteroides thetaiotaomicron</i> [TaxId: 818]}
5	Smp_346170	Dyp-type peroxidase	Melanin biosynthesis protein TyrA { <i>Shewanella oneidensis</i> [TaxId: 70863]}
5	Smp_314210	Dyp-type peroxidase	Melanin biosynthesis protein TyrA { <i>Shewanella oneidensis</i> [TaxId: 70863]}
5	Smp_329950	NA	NA
5	Smp_154590	NA	Naphthalene 1,2-dioxygenase alpha subunit, N-domain { <i>Rhodococcus sp. ncimb12038</i> [TaxId: 92694]}
5	Smp_201330	NA	NA
5	Smp_201320	NA	NA
5	Smp_329960	NA	Phenol_monoox
5	Smp_102020	NA	NA
5	Smp_086630	NA	NA
5	Smp_329970	NA	Nicotinamide mononucleotide (NMN) adenylyltransferase { <i>Bacillus subtilis</i> [TaxId: 1423]}
5	Smp_192610	NA	NA
5	Smp_329980	NA	Synaptotagmin IV {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
5	Smp_329990	C2 domain	Synaptotagmin IV {Rat ( <i>Rattus norvegicus</i> ) [TaxId: 10116]}
5	Smp_330000	Sin3 associated polypeptide p18	HISTONE DEACETYLASE COMPLEX SUBUNIT SAP18
5	Smp_330010	Alpha-ketoglutarate-dependent dioxygenase alkB homologue 7	Alkylated DNA repair protein AlkB/DNA
5	Smp_330020	NA	NA
5	Smp_314630	NA	RCR
5	Smp_332090	Alpha-ketoglutarate-dependent dioxygenase alkB homologue 7	Alkylated DNA repair protein AlkB/DNA
5	Smp_093070	NA	NA
5	Smp_345030	WD40 repeat	mTOR, Target of rapamycin complex
5	Smp_068000	Ubiquitin-conjugating enzyme E2	Ubiquitin conjugating enzyme, UBC { <i>Caenorhabditis elegans</i> , E2 2 [TaxId: 6239]}
5	Smp_315750	Ubiquitin-conjugating enzyme E2	Ubiquitin conjugating enzyme, UBC { <i>Caenorhabditis elegans</i> , E2 2 [TaxId: 6239]}
5	Smp_158990	EF-hand domain	Frequenin (neuronal calcium sensor 1) {Baker's yeast ( <i>Saccharomyces cerevisiae</i> ) [TaxId: 4932]}
5	Smp_158980	EF-hand domain	Recoverin {Cow ( <i>Bos taurus</i> ) [TaxId: 9913]}
5	Smp_158970	ATP-dependent RNA helicase DEAD-box, conserved site	ATP-dependent RNA helicase MSS116 (E.C.3.6.1.-)/RNA
5	Smp_341090	Zinc finger C2H2-type	Zn-C2H2_12

## Annexe 4

5	Smp_067930	28kDa heat- and acid-stable phosphoprotein	DNA-directed RNA polymerase subunit, DNA-directed
5	Smp_345020	NA	QVR
5	Smp_158950	Protein kinase-like domain superfamily	SERINE/THREONINE-PROTEIN KINASE HASPIN (E.C.2.7.11.1)
5	Smp_067900	Cdc37	HEAT SHOCK PROTEIN HSP 90
5	Smp_067890	Proteasome alpha-subunit, N-terminal domain	Proteasome alpha subunit (non-catalytic) {Baker's yeast ( <i>Saccharomyces cerevisiae</i> ) [TaxId: 4932]}
5	Smp_330050	NA	Strabismus
5	Smp_330060	NA	NA
5	Smp_330070	NA	NA
5	Smp_330080	NA	Eukaryotic translation initiation factor 4E-like
5	Smp_330090	NA	NA
5	Smp_330100	NA	NA
5	Smp_056460	NA	DUF3091
5	Smp_330110	NA	NA
5	Smp_330120	NA	NA
5	Smp_344470	NA	NA
5	Smp_330130	NA	NA
5	Smp_152430	CTF transcription factor/nuclear factor 1	zf-His_Me_endon
5	Smp_056500	Proteasome, subunit alpha/beta	Proteasome subunit alpha type-6 (E.C.3.4.25.1)
5	Smp_152440	Nucleotide-binding alpha-beta plait domain superfamily	Splicing factor 3A subunit 3
5	Smp_266930	NA	Zinc finger protein 295, ZNF295 {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
5	Smp_337370	Enolase	Enolase {Baker's yeast ( <i>Saccharomyces cerevisiae</i> ) [TaxId: 4932]}
5	Smp_342340	NA	NA
5	Smp_128040	NA	OB-fold domains of BRCA2 {Rat ( <i>Rattus norvegicus</i> ) [TaxId: 10116]}
6	Smp_345940	NA	Thiamine transporter
6	Smp_169860	NA	NA
6	Smp_169870	Lysosomal cystine transporter	sugar transporter
6	Smp_169880	Immunoglobulin subtype 2	Kin of IRRE-like protein 1
6	Smp_169890	Immunoglobulin subtype	Defective proboscis extension response 2
6	Smp_341830	P-loop containing nucleoside triphosphate hydrolase	Sulfotransferase
6	Smp_203170	Oligosacaryltransferase	Ost4
6	Smp_176180	Cysteine-rich secretory protein-related	Insect allergen 5 (AG5) {Yellow jacket ( <i>Vespula vulgaris</i> ), Ves v 5 [TaxId: 7454]}
6	Smp_176160	Cysteine-rich secretory protein-related	Insect allergen 5 (AG5) {Yellow jacket ( <i>Vespula vulgaris</i> ), Ves v 5 [TaxId: 7454]}

## Annexe 4

6	Smp_120670	Cysteine-rich secretory protein-related	Insect allergen 5 (AG5) {Yellow jacket ( <i>Vespula vulgaris</i> ), Ves v 5 [TaxId: 7454]}
6	Smp_176170	CAP domain	Insect allergen 5 (AG5) {Yellow jacket ( <i>Vespula vulgaris</i> ), Ves v 5 [TaxId: 7454]}
6	Smp_154260	Cysteine-rich secretory protein-related	Insect allergen 5 (AG5) {Yellow jacket ( <i>Vespula vulgaris</i> ), Ves v 5 [TaxId: 7454]}
6	Smp_154270	Rossmann-like alpha/beta/alpha sandwich fold	ATP sulfurylase catalytic domain {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
6	Smp_317130	ATPase, V0 complex, subunit e1/e2	ATPase H+-transporting V1 subunit A
6	Smp_214020	S-adenosyl-L-methionine-dependent methyltransferase	Fibrillarin homologue { <i>Archaeon Pyrococcus horikoshii</i> [TaxId: 53953]}
6	Smp_214030	Metallo-beta-lactamase	Hypothetical protein TM0894 { <i>Thermotoga maritima</i> [TaxId: 2336]}
6	Smp_059670	Marvel domain	MARVEL
6	Smp_340940	WD40 repeat	F-box/WD-repeat protein 1 (beta-TRCP1) {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
6	Smp_305690	Small GTPase	Rad {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
6	Smp_059710	SAP domain	Protein DEK
6	Smp_154280	Divergent protein kinase domain 2A/B	PIP49_C
6	Smp_191970	NA	Endolysin, Claudin-4 (E.C.3.2.1.17), Heat-labile enterotoxin B
6	Smp_154290	Cysteine-rich secretory protein-related	Insect allergen 5 (AG5) {Yellow jacket ( <i>Vespula vulgaris</i> ), Ves v 5 [TaxId: 7454]}
6	Smp_059600	Tektin	Tektin
6	Smp_317500	NA	PGM_PMM_I
6	Smp_154250	NA	NA
6	Smp_141490	Zinc finger, RING-type	brca1 RING domain {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
6	Smp_203030	Uncharacterised domain UPF0506	UPF0506
6	Smp_141500	Uncharacterised domain UPF0506	UPF0506
6	Smp_317510	Uncharacterised domain UPF0506	UPF0506
6	Smp_316080	BTB/POZ domain	Voltage-gated potassium channel subunit beta-2
6	Smp_316090	Carboxylesterase, type B	Acetylcholinesterase {Pacific electric ray ( <i>Torpedo californica</i> ) [TaxId: 7787]}
6	Smp_141520	PDZ domain	Hypothetical protein KIAA1095 {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
6	Smp_347170	Cysteine alpha-hairpin motif superfamily	28S ribosomal protein S24, mitochondrial
6	Smp_343430	Anoctamin	Anoctamin-1
6	Smp_247980	Deoxynucleoside kinase	Deoxyribonucleoside kinase {Fruit fly ( <i>Drosophila melanogaster</i> ) [TaxId: 7227]}
6	Smp_141540	Leucine-rich repeat	Splicesomal U2A' protein {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}

## Annexe 4

6	Smp_210630	Histone H2A	Histone H2A {Chicken ( <i>Gallus gallus</i> ), erythrocytes [TaxId: 9031]}
6	Smp_347320	Cysteine-rich secretory protein-related	Cysteine-rich secretory protein (SteCRISP) {Chinese green tree viper ( <i>Trimeresurus stejnegeri</i> ) [TaxId: 39682]}
6	Smp_141550	Cysteine-rich secretory protein-related	Cysteine-rich secretory protein (SteCRISP) {Chinese green tree viper ( <i>Trimeresurus stejnegeri</i> ) [TaxId: 39682]}
6	Smp_317520	Cysteine-rich secretory protein-related	Pathogenesis-related protein 1 (PR1) {Tomato ( <i>Lycopersicon esculentum</i> ), P14a [TaxId: 4081]}
6	Smp_141560	Cysteine-rich secretory protein-related	Cysteine-rich secretory protein (SteCRISP) {Chinese green tree viper ( <i>Trimeresurus stejnegeri</i> ) [TaxId: 39682]}
6	Smp_316760	Cysteine-rich secretory protein-related	Insect allergen 5 (AG5) {Yellow jacket ( <i>Vespula vulgaris</i> ), Ves v 5 [TaxId: 7454]}
6	Smp_160250	Cysteine-rich secretory protein-related	Insect allergen 5 (AG5) {Yellow jacket ( <i>Vespula vulgaris</i> ), Ves v 5 [TaxId: 7454]}
6	Smp_313710	Cysteine-rich secretory protein-related	Insect allergen 5 (AG5) {Yellow jacket ( <i>Vespula vulgaris</i> ), Ves v 5 [TaxId: 7454]}
6	Smp_070250	Cysteine-rich secretory protein-related	Insect allergen 5 (AG5) {Yellow jacket ( <i>Vespula vulgaris</i> ), Ves v 5 [TaxId: 7454]}
6	Smp_317530	NA	NA
6	Smp_300070	Cysteine-rich secretory protein-related	Insect allergen 5 (AG5) {Yellow jacket ( <i>Vespula vulgaris</i> ), Ves v 5 [TaxId: 7454]}
6	Smp_336880	Cysteine-rich secretory protein-related	Pathogenesis-related protein 1 (PR1) {Tomato ( <i>Lycopersicon esculentum</i> ), P14a [TaxId: 4081]}
6	Smp_300080	Cysteine-rich secretory protein-related	Insect allergen 5 (AG5) {Yellow jacket ( <i>Vespula vulgaris</i> ), Ves v 5 [TaxId: 7454]}
6	Smp_317540	NA	NA
6	Smp_156240	NA	DUF5588
6	Smp_156230	NA	Tk-SP_N-pro
6	Smp_344760	PTP type protein phosphatase	RPTP Lar {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
6	Smp_063190	Transforming growth factor-beta, C-terminal	Activin A (Inhibin beta A) {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
6	Smp_063180	NA	NA
6	Smp_316770	NA	Cytochrome P450, family 17, subfamily
6	Smp_300090	Terminal nucleotidyltransferase	NTP_transf_7
6	Smp_156180	EF-hand domain	DUF5580
6	Smp_063120	Ham1-like protein	Inosine triphosphate pyrophosphatase, ITPase {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
6	Smp_063110	Protein of unknown function DUF3504	DUF3504
6	Smp_156170	ARID DNA-binding domain	Transcription activator BRG1 (E.C.3.6.4.-), AT-rich



## Annexe 4

6	Smp_063090	Aconitase A/isopropylmalate dehydratase small subunit, swivel domain	ACONITASE
6	Smp_156160	WD40 repeat	Elongator complex protein 1, Elongator
6	Smp_063070	Small GTPase	CDC42 {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
6	Smp_300140	NA	NA
6	Smp_300150	NA	NA
6	Smp_309350	Transport protein particle (TRAPP) component	Trafficking protein particle complex subunit 3, Bet3 homolog {Mouse ( <i>Mus musculus</i> ) [TaxId: 10090]}
6	Smp_050880	CXXC motif containing zinc binding protein, eukaryotic	Hypothetical protein MAL13P1.257 {Malaria parasite ( <i>Plasmodium falciparum</i> ) [TaxId: 5833]}
6	Smp_050890	Hid-1/Ecm30	Dymeclin
6	Smp_149370	Transmembrane protein 181	MIG-14_Wnt-bd
6	Smp_050920	RNA recognition motif domain	RNA-binding region containing protein 1 {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
6	Smp_149380	SET domain	SET domain-containing protein 6, RelA
6	Smp_050940	Ribosomal protein L5	eL18, uL16, uL5, eL13, uL13
6	Smp_347550	C-type lectin-like	Low affinity immunoglobulin epsilon Fc receptor {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
6	Smp_243000	NA	Low density lipoprotein (LDL) receptor, different EGF domains {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
6	Smp_051000	Syndetin, C-terminal	Vps54_N
6	Smp_051010	Ubiquitin-conjugating enzyme E2	Ubiquitin conjugating enzyme, UBC {Human ( <i>Homo sapiens</i> ), ubch10 [TaxId: 9606]}
6	Smp_149400	Peptidase S8/S53 domain	Furin (E.C.3.4.21.75)
6	Smp_149410	NA	NA
6	Smp_177180	NA	Merozoite surface protein 1 (MSP-1) {Malaria parasite ( <i>Plasmodium falciparum</i> ) [TaxId: 5833]}
6	Smp_051070	Gonadal family	DGCR6
6	Smp_051080	Clathrin light chain	Clathrin heavy chain 1, Clathrin
6	Smp_051090	CFA20 domain	DUF667
6	Smp_149420	K Homology domain	Protein quaking/RNA
6	Smp_247990	NA	DUF4806
6	Smp_248000	39S ribosomal protein L52, mitochondrial	Mitochondrial ribosomal protein L27, Mitochondrial
6	Smp_051120	START-like domain superfamily	Uncharacterized protein DR_A0006
6	Smp_051130	Ets domain	Fli-1 {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
6	Smp_149430	Phenylalanyl-tRNA synthetase	Phenyl-tRNA synthetase (PheRS) alpha subunit, PheS { <i>Thermus thermophilus</i> [TaxId: 274]}
6	Smp_344200	Small GTPase	MITOCHONDRIAL RHO GTPASE (E.C.3.6.5.-)
6	Smp_149450	tRNA-dihydrouridine synthase	tRNA-dihydrouridine(20) synthase [NAD(P)+]-like (E.C.1.3.1.-)

## Annexe 4

6	Smp_149460	Protein kinase domain	Tyrosine-protein kinase SYK {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
6	Smp_149480	Leucine-rich repeat	von Willebrand factor binding domain of glycoprotein Ib alpha {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
6	Smp_344210	NA	NA
6	Smp_340780	Pleckstrin homology domain	Actin-binding protein anillin
6	Smp_046290	Marvel domain	MARVEL
6	Smp_147000	PDZ domain	amyloid beta A4 precursor protein-binding
6	Smp_300200	Zinc finger C2H2-type	Transcriptional repressor CTCF/DNA Complex
6	Smp_300210	Zinc finger C2H2-type	Zinc finger protein 568/DNA Complex
6	Smp_300220	Zinc finger C2H2-type	Zinc finger protein 568/DNA Complex
6	Smp_046360	Zinc finger C2H2-type	Transcriptional repressor CTCF/DNA Complex
6	Smp_147020	Sec7 domain	guanine nucleotide exchange protein
6	Smp_147040	Ankyrin repeat	Glutaminase kidney isoform, mitochondrial (E.C.3.5.1.2)
6	Smp_046410	Tektin	Tektin
6	Smp_147050	ATPase, F1/V1/A1 complex, alpha/beta subunit, nucleotide-binding domain	ATPase H <sup>+</sup> -transporting V1 subunit A
6	Smp_046430	Peptidase C19, ubiquitin carboxyl-terminal hydrolase	Ubiquitin carboxyl-terminal hydrolase 7 (USP7, HAUSP) {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
6	Smp_317560	NA	30S ribosomal protein S2, 30S
6	Smp_147060	NA	NA
6	Smp_147070	Amino acid transporter, transmembrane domain	antibody Fab Heavy Chain, antibody
6	Smp_011030	Regulatory factor, effector binding domain superfamily	Heme-binding protein 2, Programmed cell
6	Smp_128010	NA	NA
6	Smp_011010	NA	39S ribosomal protein L2, mitochondrial
6	Smp_011000	Partial AB-hydrolase lipase domain	Gastric lipase {Dog ( <i>Canis familiaris</i> ) [TaxId: 9615]}
6	Smp_128000	High mobility group box domain	ARS-binding factor 2, mitochondrial/DNA Complex
6	Smp_127990	NA	NA
6	Smp_098420	NA	FtsL
6	Smp_317580	NA	GRP
6	Smp_317590	NA	NA
6	Smp_176270	Target SNARE coiled-coil homology domain	Syntaxin-binding protein 1, Syntaxin-1A
6	Smp_098450	NA	NA
6	Smp_203190	NA	NA
6	Smp_154220	NA	Photosystem II protein D1 (E.C.1.10.3.9)

## Annexe 4

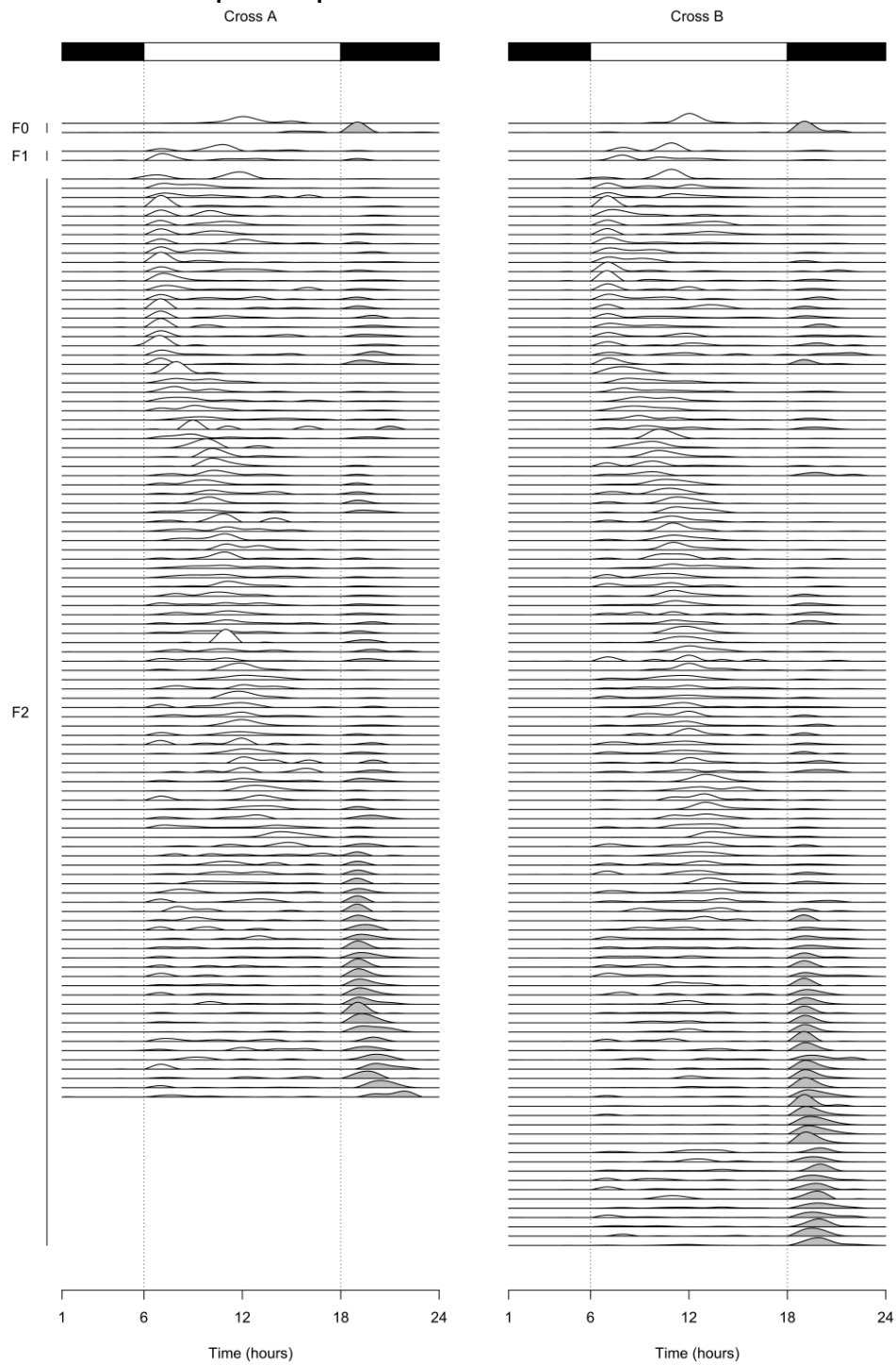
6	Smp_059540	Signal recognition particle, SRP54 subunit, GTPase domain	60S RIBOSOMAL PROTEIN L23, RIBOSOMAL
6	Smp_316100	Domain of unknown function DUF4604	DUF4604
6	Smp_316110	TAP46-like protein	TAP42
6	Smp_059570	Small GTPase	Ras-related protein RalA {Cotton-top tamarin ( <i>Saguinus oedipus</i> ) [TaxId: 9490]}
6	Smp_244170	ARID DNA-binding domain	RSC7, Chromatin structure-remodeling complex protein
6	Smp_136660	EGF-like domain	Neural cell adhesion molecule (NCAM) {Rat ( <i>Rattus norvegicus</i> ) [TaxId: 10116]}
6	Smp_136640	EF-hand domain	Grancalcin {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
6	Smp_026770	LSM domain, eukaryotic/archaea-type	Spliceosomal U1 small nuclear ribonucleoprotein
6	Smp_342570	Diacylglycerol kinase, accessory domain	Diacylglycerol kinase DgkB { <i>Staphylococcus aureus</i> [TaxId: 1280]}
6	Smp_131220	Peptidase S8/S53 domain	Tripeptidyl-peptidase 2 (E.C.3.4.14.10)
6	Smp_016600	Sodium	Neutral amino acid transporter B(0)
6	Smp_131230	WD40 repeat	F-box/WD-repeat protein 1 (beta-TRCP1) {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
6	Smp_016630	Guanine nucleotide binding protein (G-protein), alpha subunit	G-protein alpha subunit Galpha7
6	Smp_317600	NA	Eukaryotic translation initiation factor 4E-binding
6	Smp_131250	NDRG	Protein NDRG3
6	Smp_016670	Pterin 4 alpha carbinolamine dehydratase	DcoH-like protein DCoH2 {Mouse ( <i>Mus musculus</i> ) [TaxId: 10090]}
6	Smp_315800	NA	BETA-CATENIN/TRANSCRIPTION FACTOR XTCF-3
6	Smp_213250	Bcl2-like	Bcl-2 homologous antagonist/killer
6	Smp_213260	Acyl-CoA N-acyltransferase	N-alpha-acetyltransferase 60 (E.C.2.3.1.48,2.3.1.88)
6	Smp_016730	SOUL haem-binding protein	Heme-binding protein 1 {Mouse ( <i>Mus musculus</i> ) [TaxId: 10090]}
6	Smp_131290	NA	SSP160
6	Smp_016750	Zinc finger C2H2-type	ADR1 {Synthetic, based on <i>Saccharomyces cerevisiae</i> sequence}
6	Smp_131320	Zinc finger, PHD-type	Probable histone acetyltransferase MYST1 {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
6	Smp_016780	Tubulin	SPINDLE POLE BODY COMPONENT SPC97
6	Smp_337210	NA	EB
6	Smp_016810	Trafficking protein particle complex subunit 13	DUF974
6	Smp_131350	Bile acid	Transporter, sodium/bile acid symporter family
6	Smp_203200	NA	POLYC9

## Annexe 4

6	Smp_131360	ATP-dependent RNA helicase DEAD-box, conserved site	Splicing factor 3A subunit 3
6	Smp_131370	CAP domain	Ancylostoma secreted protein 2
6	Smp_340240	Armadillo-like helical	Exportin-4, GTP-binding nuclear protein Ran
6	Smp_317620	NA	Keratinocyte growth factor, FGF7 {Rat ( <i>Rattus norvegicus</i> ) [TaxId: 10116]}
6	Smp_131380	NA	Keratinocyte growth factor, FGF7 {Rat ( <i>Rattus norvegicus</i> ) [TaxId: 10116]}
6	Smp_131390	Bromodomain	beta1-subunit of the signal-transducing G protein heterotrimer {Cow ( <i>Bos taurus</i> ) [TaxId: 9913]}
6	Smp_016870	Microtubule-associated protein, MAP65/Ase1/PRC1	Protein regulator of cytokinesis 1
6	Smp_214040	Glycosyl hydrolase family 99	Glycosyl hydrolase family 71
6	Smp_214050	Guanylate kinase-like domain	Guanylate kinase { <i>Escherichia coli</i> [TaxId: 562]}
6	Smp_340250	NA	NA
6	Smp_016900	Protein of unknown function DUF2373	DUF2373
6	Smp_016910	2Fe-2S ferredoxin-type iron-sulfur binding domain	Adrenodoxin {Cow ( <i>Bos taurus</i> ) [TaxId: 9913]}
6	Smp_016920	Zinc finger, RING-type	EL5 RING-H2 domain {Rice ( <i>Oryza sativa</i> ) [TaxId: 4530]}
6	Smp_131400	Mannose-6-phosphate receptor binding domain superfamily	PRKCSH_1
6	Smp_131410	Tmem141 superfamily	Herpes_LMP1
6	Smp_317630	NA	PMP22_Claudin
6	Smp_331630	Uncharacterised domain UPF0506	UPF0506
6	Smp_342590	T-complex 11	Tcp11

Chr. : chromosome.

## Annexe 5 : Répartition des profils chronobiologiques des générations F0, F1 et F2 des croisements interspécifiques



**Figure A5 : Joyplots des profils chronobiologiques des générations F0, F1 et F2 des croisements interspécifiques pour le QTL mapping.** Le croisement A est à gauche et le croisement B à droite. L'aire blanche sous la courbe correspond à une émission diurne des cercaires. L'aire grise sous la courbe correspond à une émission nocturne des cercaires.

## Annexe 6 : Liste des gènes candidats issus du QTL mapping des croisements interspécifiques

Chr.	Gene ID	GFF annotation	HHsearch annotation
3B	Smp_018360	PLC-like phosphodiesterase, TIM beta/alpha-barrel domain superfamily	Crystal Structure of the Y247S/Y251S
3B	Smp_132260	Armadillo	Aurora-related kinase 1 (aurora-2) {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
3B	Smp_132270	K Homology domain	Poly(rC)-binding protein 2
3B	Smp_018400	WD40 repeat	beta1-subunit of the signal-transducing G protein heterotrimer {Cow ( <i>Bos taurus</i> ) [TaxId: 9913]}
3B	Smp_132280	Thrombospondin type-1 (TSP1) repeat	Thrombospondin-1 (TSP-1) {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
3B	Smp_018430	Peptidase S24/S26A/S26B/S26C	Maltose binding protein -
3B	Smp_018440	NA	NA
3B	Smp_018460	Phosphatidylinositol 3-/4-kinase, catalytic domain	Phosphatidylinositol 4-kinase type 2-alpha (E.C.2.7.1.67)
3B	Smp_337780	Coiled-coil alpha-helical rod protein 1	HCR
3B	Smp_018500	Carbohydrate kinase, thermoresistant glucokinase	Gluconate kinase { <i>Escherichia coli</i> [TaxId: 562]}
3B	Smp_132300	Kinesin motor domain	Kinesin { <i>Neurospora crassa</i> [TaxId: 5141]}
3B	Smp_337050		380.41
3B	Smp_325920	Ribosomal protein L15, conserved site	Ribosomal protein L8, putative, Uncharacterized
3B	Smp_018530	Nascent polypeptide-associated complex NAC domain	Ribosomal protein L8, uL3, uL4
3B	Smp_132320	Cadherin-like	Protocadherin-15
3B	Smp_018550	NA	SWI5-dependent HO expression protein 3
3B	Smp_132330	Small GTPase	Rab30 {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
3B	Smp_018570	G protein pathway suppressor 2	G_path_suppress
3B	Smp_018580	WD40 repeat	Pectate_lyase22
3B	Smp_018590	Zinc finger, GATA-type	Trans-acting T-cell-specific transcription factor GATA-3/DNA
3B	Smp_132340	NA	NA
3B	Smp_018610	Mediator complex, subunit Med7	Med7
3B	Smp_018620	Peptidase M41	AFG3-like protein 2 (E.C.3.4.24.-), Substrate
3B	Smp_132350	CCAAT-binding factor	60S ribosomal protein L13-A, 60S
3B	Smp_132360	NA	Activity-regulated cytoskeleton associated protein 1
3B	Smp_018640	RUN domain	Rap2 interacting protein X (RUFY3) {Mouse ( <i>Mus musculus</i> ) [TaxId: 10090]}
3B	Smp_246950	PTP type protein phosphatase	Tyrosine-protein phosphatase, non-receptor type 13 (PTPL1) {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
3B	Smp_132390	NA	NA
3B	Smp_018680	Isocitrate dehydrogenase NAD-dependent	Isocitrate dehydrogenase [NAD] subunit 1

## Annexe 6

3B	Smp_132410	G protein-coupled receptor, rhodopsin-like	Lysozyme,Proteinase-activated receptor 2,Soluble cytochrome b562,Proteinase-activated
3B	Smp_018710	WD40 repeat	Uncharacterized WD repeat-containing protein C4F10.18
3B	Smp_335830	Armadillo-type fold	RTTN_N
3B	Smp_018760	Glycoside hydrolase family 31	Protein disulfide-isomerase (E.C.5.3.4.1), Peptide from
3B	Smp_342640	Tetratricopeptide-like helical domain superfamily	O-GlcNAc transferase p110 subunit, OGT {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
3B	Smp_018790	PPM-type phosphatase domain	Probable protein phosphatase 2C 50
3B	Smp_132450	BTB/POZ domain	No mechanoreceptor potential C isoform
3B	Smp_325960	Tetratricopeptide-like helical domain superfamily	O-GlcNAc transferase p110 subunit, OGT {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
3B	Smp_204880	NA	DUF4999
3B	Smp_335840	NA	NA
3B	Smp_347080	Ion transport domain	nvTRPM2 channel
3B	Smp_331920	LSM domain, eukaryotic/archaea-type	Archaeal homoheptameric Sm protein { <i>Archaeon Methanobacterium thermoautotrophicum</i> [TaxId: 145262]}
3B	Smp_094140	MCM domain	DNA replication licensing factor MCM2
3B	Smp_308310	Helix-turn-helix motif	Homeotic bicoid protein {Fruit fly ( <i>Drosophila melanogaster</i> ) [TaxId: 7227]}
3B	Smp_325970	NA	NA
3B	Smp_126200	Nitrogen permease regulator 2	GATOR complex protein NPRL2, GATOR
3B	Smp_007540	NA	Exosome complex RNA-binding protein 1, ECR1 { <i>Archaeoglobus fulgidus</i> [TaxId: 2234]}
3B	Smp_007550	Peptidase M1, alanine aminopeptidase/leukotriene A4 hydrolase	MGC78867 protein (E.C.3.3.2.6)
3B	Smp_126210	NA	NA
3B	Smp_007570	NA	NA
3B	Smp_126220	NA	NA
3B	Smp_126230	NA	NA
3B	Smp_126240	Protein kinase domain	Twitchin, kinase domain {California sea hare ( <i>Aplysia californica</i> ), twk43 [TaxId: 6500]}
3B	Smp_336670	Tetratricopeptide-like helical domain superfamily	N-terminal acetyltransferase-like protein, Naa10, HypK
3B	Smp_179910	Small GTPase	GTP-binding protein RheB {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
3B	Smp_162520	Cadherin-like	EP-cadherin, ectodomain
3B	Smp_074770	Protein KT112/L-seryl-tRNA(Sec) kinase	KT112
3B	Smp_074780	Ribosomal protein S13	30S ribosomal protein S2, 30S
3B	Smp_162540	Tektin	Tektin
3B	Smp_162550	Cadherin-like	Protocadherin-15
3B	Smp_337640	NA	NA

## Annexe 6

3B	Smp_074830	Tubulin-tyrosine ligase/Tubulin polyglutamylase	Tubulin polyglutamylase TLL7 (E.C.6.-.-)
3B	Smp_162600	U3 small nucleolar RNA-associated protein 10	40S ribosomal protein S1-A, 40S
3B	Smp_074870	tRNA-splicing ligase, RtcB	Hypothetical protein PH1602 { <i>Archaeon Pyrococcus horikoshii</i> [TaxId: 53953]}
3B	Smp_074880	NA	Tim17
3B	Smp_074900	IRS-type PTB domain	Docking protein 1, Dok1 { <i>Mouse (Mus musculus)</i> [TaxId: 10090]}
3B	Smp_162610	Cadherin-like	Protocadherin-15
3B	Smp_162620	rRNA adenine dimethylase-like	SAVED domain-containing protein/RNA Complex
3B	Smp_074930	Palmitoyltransferase, DHHC domain	human DHHC20 palmitoyltransferase (E.C.2.3.1.225)
3B	Smp_074940	Alpha/beta hydrolase fold-3	Uncharacterized protein TM1040_2492 { <i>Silicibacter sp. tm1040</i> [TaxId: 292414]}
3B	Smp_074950	RWD domain	RWD domain-containing protein 2 { <i>Human (Homo sapiens)</i> [TaxId: 9606]}
3B	Smp_162630	Ubiquitin/SUMO-activating enzyme E1	Ubiquitin-activating enzyme E1 1, Uba1
3B	Smp_162640	NA	DUF2798
3B	Smp_075000	Regulator of chromosome condensation 1/beta-lactamase-inhibitor protein II	NA
3B	Smp_074990	Regulator of chromosome condensation, RCC1	Regulator of chromosome condensation RCC1 { <i>Human (Homo sapiens)</i> [TaxId: 9606]}
3B	Smp_162650	Anoctamin	Anoctamin
3B	Smp_075030	Ankyrin repeat	Myotrophin { <i>Rat (Rattus norvegicus)</i> [TaxId: 10116]}
3B	Smp_075040	Uncharacterised protein family UPF0538	DUF2340
3B	Smp_162660	Carbon-nitrogen hydrolase	Glutamine-dependent NAD(+) synthetase (E.C.6.3.5.1)
3B	Smp_075060	Carbon-nitrogen hydrolase	Glutamine-dependent NAD(+) synthetase (E.C.6.3.5.1)
3B	Smp_337650	Sperm-associated antigen 17	PapD-like
3B	Smp_162680	Epithelial sodium channel	Acid-sensing ion channel 1, Pi-theraphotoxin-Pc1a
3B	Smp_075110	Cell cycle checkpoint protein, Rad1	Rad1
3B	Smp_075120	Purine-rich element binding protein family	PurA
3B	Smp_075130	G-protein, beta subunit	beta1-subunit of the signal-transducing G protein heterotrimer { <i>Cow (Bos taurus)</i> [TaxId: 9913]}
3B	Smp_075140	Mitochondria-eating protein	MIEAP
3B	Smp_162710	Protein kinase domain	Aurora-related kinase 1 (aurora-2) { <i>Human (Homo sapiens)</i> [TaxId: 9606]}
3B	Smp_333540	Nucleic acid-binding, OB-fold	CDC24_OB3
3B	Smp_346300	NA	HOOK



## Annexe 6

3B	Smp_174780	P-loop containing nucleoside triphosphate hydrolase	Intraflagellar transport protein 22
3B	Smp_202010	NA	NA
3B	Smp_174800	Zinc finger C2H2-type	Zinc finger protein 568/DNA Complex
3B	Smp_174810	Superoxide dismutase, copper/zinc binding domain	Cu,Zn superoxide dismutase, SOD {Blood fluke ( <i>Schistosoma mansoni</i> ) [TaxId: 6183]}
3B	Smp_095980	Superoxide dismutase, copper/zinc binding domain	Cu,Zn superoxide dismutase, SOD {Blood fluke ( <i>Schistosoma mansoni</i> ) [TaxId: 6183]}
3B	Smp_346310	Protein kinase domain	Pkb kinase (Akt-2) {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
3AB	Smp_096000	NA	NUCLEAR PORE COMPLEX PROTEIN NUP155
3AB	Smp_096010	Heat shock protein DnaJ, cysteine-rich domain	Putative chaperone DnaJ
3AB	Smp_096020	Adenosylhomocysteinase-like	adenosylhomocysteinase (E.C.3.3.1.1)
3AB	Smp_346320	Chromo/chromo shadow domain	Polycomb protein, Pc {Fruit fly ( <i>Drosophila melanogaster</i> ) [TaxId: 7227]}
3AB	Smp_174850	NA	Ragulator complex protein LAMTOR3, Ragulator
3AB	Smp_096050	NA	DNA polymerase III subunit alpha
3AB	Smp_325990	NA	NA
3AB	Smp_212410	Cyclic nucleotide-binding domain	Potassium/sodium hyperpolarization-activated cyclic nucleotide-gated channel
3AB	Smp_078780	Amidohydrolase-related	DIHYDROPYRIMIDINASE-RELATED PROTEIN 5
3AB	Smp_078800	DnaJ domain	DnaJ chaperone, N-terminal (J) domain { <i>Escherichia coli</i> [TaxId: 562]}
3AB	Smp_326000	Helix-turn-helix motif	Engrailed Homeodomain { <i>Drosophila melanogaster</i> [TaxId: 7227]}
3AB	Smp_326010	NA	NA
3AB	Smp_345520	Poly A polymerase, head domain	<i>A.fulgidus</i> CCA-adding enzyme (E.C.2.7.7.72)
3AB	Smp_145890	Thrombospondin type-1 (TSP1) repeat	PROPERDIN
3AB	Smp_197390	Peptidase M12B, ADAM/reprolysin	Disintegrin and metalloproteinase domain-containing protein
3AB	Smp_145910	NA	Protein-tyrosine sulfotransferase 2 (E.C.2.8.2.20), C4
3AB	Smp_308350	Protein-tyrosine sulfotransferase	Protein-tyrosine sulfotransferase 1 (E.C.2.8.2.20), gastrin
3AB	Smp_305920	Protein-tyrosine sulfotransferase	Protein-tyrosine sulfotransferase 1 (E.C.2.8.2.20), gastrin
3AB	Smp_044160	Mitochondrial carrier protein	ADP,ATP carrier protein {Cow ( <i>Bos taurus</i> ), heart isoform t1 [TaxId: 9913]}
3AB	Smp_145930	Hemopexin-like domain	Interstitial collagenase (E.C.3.4.24.7)
3AB	Smp_343900	NA	NA
3AB	Smp_246600	p53-like transcription factor, DNA-binding	Lin-12 and glp-1 phenotype protein
3AB	Smp_145980	Sodium	Neutral amino acid transporter B(0)

## Annexe 6

3AB	Smp_145990	ClpA/B family	ClpA, an Hsp100 chaperone, AAA+ modules { <i>Escherichia coli</i> [TaxId: 562]}
3AB	Smp_044250	JAB1/MPN/MOV34 metalloenzyme domain	Prp8, Brr2, Snu114, Rse1, Cus1
3AB	Smp_146000	Small lysine-rich protein 1	NA
3AB	Smp_044260	Autophagy-related protein 3	Ubiquitin-like-conjugating enzyme ATG10 (E.C.6.3.2.-)
3AB	Smp_267300	NA	NA
3AB	Smp_267310	Thiolase	Thiolase {Baker's yeast ( <i>Saccharomyces cerevisiae</i> ) [TaxId: 4932]}
3AB	Smp_208080	Golgin subfamily A member 7/ERF4	Erf4
3AB	Smp_326040	Up-regulated during skeletal muscle growth protein 5	ATP_synth_reg
3AB	Smp_129340	WD40 repeat	Platelet-activating factor acetylhydrolase IB subunit alpha {Mouse ( <i>Mus musculus</i> ) [TaxId: 10090]}
3AB	Smp_013300	Yip1 domain	YIF1
3AB	Smp_246610	Small GTPase	Rad {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
3AB	Smp_129360	Cyclin, N-terminal	Cell division protein kinase 4
3AB	Smp_342430	NA	NA
3AB	Smp_129380	BTB/POZ domain	Voltage-gated potassium channel subunit beta-2
3AB	Smp_013360	RNA recognition motif domain	U1 small nuclear ribonucleoprotein A
3AB	Smp_013370	Tetraspanin	Tetraspanin
3AB	Smp_347820	Vacuolar protein sorting-associated protein 13, N-terminal domain	Putative vacuolar protein sorting-associated protein
3AB	Smp_129410	NA	NUCLEAR PORE COMPLEX PROTEIN NUP155
3AB	Smp_129420	Ribosomal protein S11	Ribosomal protein S11 { <i>Escherichia coli</i> [TaxId: 562]}
3AB	Smp_129430	Transcription factor, MADS-box	Myocyte enhancer factor Mef2a core {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
3AB	Smp_013440	Zinc/iron permease	Zip
3AB	Smp_129450	Phosphatidic acid phosphatase type 2/haloperoxidase	Phosphatidylglycerophosphatase B (E.C.3.1.3.27)
3AB	Smp_013460	SCAMP	SCAMP
3AB	Smp_013470	Ribosomal protein S5	uS2, eS1, uS5, uS3, eS4
3AB	Smp_246650	Cadherin-like superfamily	DAG1
3AB	Smp_246660	Homeobox domain	LIM domain-binding protein 1, LIM/homeobox
3AB	Smp_013520	Nuclear cap-binding protein subunit 3	NCBP3
3AB	Smp_129500	RNA polymerase, alpha subunit	DNA-DIRECTED RNA POLYMERASES I, II
3AB	Smp_013540	Tyrosinase copper-binding domain	5,6-dihydroxyindole-2-carboxylic acid oxidase (E.C.1.14.18.-)
3AB	Smp_129510	Metallo-dependent phosphatase-like	Metallophosphoesterase MPPED2 (E.C.3.1.-.-)
3AB	Smp_311970	Zinc finger, GATA-type	Erythroid transcription factor GATA-1 { <i>Emericella nidulans</i> [TaxId: 162425]}
3AB	Smp_129520	NA	DASH_Spc34

## Annexe 6

3AB	Smp_013600	Homeobox domain	Homeotic bicoid protein {Fruit fly ( <i>Drosophila melanogaster</i> ) [TaxId: 7227]}
3AB	Smp_013610	Uncharacterised protein family UPF0564	UPF0564
3AB	Smp_347790	Cadherin-like	Protocadherin-15
3AB	Smp_326080	Major facilitator, sugar transporter-like	Folate_carrier
3AB	Smp_326090	Major facilitator, sugar transporter-like	Sugar_tr
3AB	Smp_070680	Repressor of RNA polymerase III transcription Maf1	Repressor of RNA polymerase III
3AB	Smp_336680	Tetratricopeptide-like helical domain superfamily	ANAPHASE-PROMOTING COMPLEX SUBUNIT 1, ANAPHASE-PROMOTING
3AB	Smp_246570	Matrin/U1-C-like, C2H2-type zinc finger	RNA binding protein ZFa
3AB	Smp_070710	Polyprenyl synthetase	farnesyl pyrophosphate synthase (E.C.2.5.1.10)
3AB	Smp_308420	Actin-depolymerising factor homology domain	Adf-H domain of twinfilin isoform-1 {Mouse ( <i>Mus musculus</i> ) [TaxId: 10090]}
3AB	Smp_160590	NA	NA
3AB	Smp_070760	YEATS	Protein AF-9 homolog
3AB	Smp_070770	Ubiquitin-fold modifier-conjugating enzyme 1	Ufm1-conjugating enzyme 1, UFC1 {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
3AB	Smp_070780	UDP-glucose 4-epimerase	Uridine diphosphogalactose-4-epimerase (UDP-galactose 4-epimerase) { <i>Escherichia coli</i> [TaxId: 562]}
3AB	Smp_160600	THO complex, subunit THOC1	efThoc1
3AB	Smp_333550	Pyridoxal phosphate homeostasis protein	Hypothetical protein ybl036c {Baker's yeast ( <i>Saccharomyces cerevisiae</i> ) [TaxId: 4932]}
3AB	Smp_160610	Rho GTPase-activating protein domain	p50 RhoGAP domain {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
3AB	Smp_326120	NA	NA
3AB	Smp_181490	Protein kinase domain	Death-associated protein kinase, Dap {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
3AB	Smp_186670	Aspartate carbamoyltransferase	Human Carbamoyl Phosphate Synthetase I
3AB	Smp_181480	tRNA (uracil-O(2)-)-methyltransferase	AdoMet_MTase
3AB	Smp_106130	Chaperone DnaK	HSP70
3AB	Smp_181470	Bax inhibitor 1-related	Herpes_LMP1
3AB	Smp_106110	EF-hand domain	Kchip1, Kv4 potassium channel-interacting protein {Rat ( <i>Rattus norvegicus</i> ) [TaxId: 10116]}
3AB	Smp_181460	EF-hand domain pair	DUF4205
3AB	Smp_335280	NAD(P)-binding domain superfamily	NADH-ubiquinone oxidoreductase chain 3 (E.C.1.6.5.3)
3AB	Smp_181450	NA	CCDC66
3AB	Smp_326130	Svp26/Tex261	Erv26
3AB	Smp_106080	NAD(P)-binding domain	NADH-ubiquinone oxidoreductase chain 3 (E.C.1.6.5.3)

## Annexe 6

3AB	Smp_181440	Dual specificity phosphatase, catalytic domain	Proline directed phosphatase CDC14b2 {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
3AB	Smp_106060	EF-hand domain	ADP,ATP carrier protein {Cow ( <i>Bos taurus</i> ), heart isoform t1 [TaxId: 9913]}
3AB	Smp_106050	Dim1 family	spliceosomal protein U5-15Kd {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
3AB	Smp_246560	GIY-YIG endonuclease	Structure-specific endonuclease subunit SLX1, Structure-specific
3AB	Smp_181410	NA	NA
3AB	Smp_106010	NA	Mitogen activated protein kinase kinase 5, Map2k5 {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
3AB	Smp_181390	CREB3 regulatory factor	Zta transcription factor/DNA Complex
3AB	Smp_213040	26S Proteasome non-ATPase regulatory subunit 9	Probable 26S proteasome regulatory subunit
3AB	Smp_213050	NA	ComS
3AB	Smp_105970	NA	NUCLEAR PORE COMPLEX PROTEIN NUP155
3AB	Smp_105960	SANT/Myb domain	Zuotin
3AB	Smp_105950	Armadillo	beta-Catenin {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
3AB	Smp_308440	NA	NA
3AB	Smp_105930	S1 domain	ribosome, ribosome biogenesis, rRNA
3AB	Smp_105910	Zinc finger, TAZ-type	Histone acetyltransferase p300 (E.C.2.3.1.48), p53
3AB	Smp_246550	I/LWEQ domain	ENTH DOMAIN OF EPSIN ENT1
3AB	Smp_105860	SKI-interacting protein SKIP, SNW domain	Pre-mRNA-processing-splicing factor 8, 116 kDa
3AB	Smp_246540	Peptidase M14, carboxypeptidase A	METALLO-CARBOXYPEPTIDASE
3AB	Smp_246530	Zinc finger, RING-type	Immediate early protein, IEEHV {Equine herpesvirus 1 [TaxId: 10326]}
3AB	Smp_308470	SEA domain	SEA
3AB	Smp_308480	SEA domain	SEA
3AB	Smp_308490	SEA domain	Transmembrane protease (E.C.3.4.21.-)
3AB	Smp_347300	Ribosomal protein L1, 3-layer alpha/beta-sandwich	Periodic tryptophan protein 2-like protein
3AB	Smp_181290	Ribosomal protein L28/L24	28S ribosomal protein S2, mitochondrial
3AB	Smp_246520	Sel1-like repeat	Cysteine rich protein C (HcpC) { <i>Helicobacter pylori</i> [TaxId: 210]}
3AB	Smp_326150	NA	Cysteine rich protein B (HcpB) { <i>Helicobacter pylori</i> [TaxId: 210]}
3AB	Smp_341430	NA	NA
3AB	Smp_085080	NA	cAMP-dependent protein kinase type II-alpha
3AB	Smp_168330	Trafficking protein particle complex subunit 2	Sedlin (SEDL) {Mouse ( <i>Mus musculus</i> ) [TaxId: 10090]}
3AB	Smp_085110	Amidohydrolase-related	DIHYDROPYRIMIDINASE-RELATED PROTEIN 5

## Annexe 6

3AB	Smp_168340	SANT/Myb domain	REST corepressor 1, CoREST {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
3AB	Smp_168350	CS domain	Suppressor of G2 allele of <i>skp1</i> homolog, <i>gst1</i> {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
3AB	Smp_010160	CS domain	SGS
3AB	Smp_347810	NA	RabGAP/TBC protein
3AB	Smp_010140	G-patch domain	G-patch
3AB	Smp_210480	Peptidase M24	AMINOPEPTIDASE P II (E.C.3.4.11.9)
3AB	Smp_333560	Beta-N-acetylglucosaminidase	Protein O-GlcNAcase (E.C.3.2.1.169,3.2.1.-)
3AB	Smp_010110	Zinc finger, RING-CH-type	IE1B protein (ORF K3), N-terminal domain {Kaposi's sarcoma-associated herpesvirus, KSHV, HHV8 [TaxId: 37296]}
3AB	Smp_010100	Lipoyl synthase	Lipoyl synthase (E.C.2.8.1.8)
3AB	Smp_010090	Snf7 family	Charged multivesicular body protein 3
3AB	Smp_167880	SprT-like	SprT-like domain-containing protein Spartan
3AB	Smp_326160	NA	NA
3AB	Smp_210680	CCR4-NOT transcription complex subunit 9	GENERAL NEGATIVE REGULATOR OF TRANSCRIPTION
3AB	Smp_212540	ST7	ST7
3AB	Smp_316690	Cysteine alpha-hairpin motif superfamily	CHCH
3AB	Smp_127520	PDZ domain	Galpha interacting protein, GalP {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
3AB	Smp_335850	CDK5 regulatory subunit-associated protein 3	DUF773
3AB	Smp_210070	Domain of unknown function DUF3677	DUF3677
3AB	Smp_246450	CDK5 regulatory subunit-associated protein 3	DUF773
3AB	Smp_243720	IQ motif, EF-hand binding site	Myosin S1, motor domain {Bay scallop ( <i>Aequipecten irradians</i> ) [TaxId: 31199]}
3AB	Smp_127500	NA	NA
3AB	Smp_335300	Tim10-like	MITOCHONDRIAL IMPORT INNER MEMBRANE TRANSLOCASE
3AB	Smp_009990	Receptor L-domain	Insulin receptor (E.C.2.7.10.1), Insulin
3AB	Smp_210660	SET domain	Histone-lysine N-methyltransferase 2E (E.C.2.1.1.43)
3AB	Smp_192010	NA	Activity-regulated cytoskeleton associated protein 1
3AB	Smp_210670	LIS1 homology motif	Vacuolar import and degradation protein
3AB	Smp_127490	Helicase, C-terminal	BLOOM'S SYNDROME HELICASE (E.C.3.6.4.12)
3AB	Smp_127480	von Willebrand factor, type A	VWA_3
3AB	Smp_127470	NA	NA
3AB	Smp_127460	Translin-associated factor X-interacting protein 1, N-terminal	TSNAXIP1_N
3AB	Smp_013040	Aspartic peptidase A1 family	Pepsin(ogen) {Pig ( <i>Sus scrofa</i> ) [TaxId: 9823]}

## Annexe 6

3AB	Smp_326170	Aspartic peptidase domain superfamily	Plasmeprin (a hemoglobin-degrading enzyme) { <i>Plasmodium vivax</i> [TaxId: 5855]}
3AB	Smp_326180	NA	NA
3AB	Smp_129160	siRNA-mediated silencing protein NRDE-2	NRDE-2
3AB	Smp_013050	TLDc domain	TLD
3AB	Smp_013060	SMAD domain, Dwarfing-type	SMAD MH1 domain {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
3AB	Smp_326190	NA	NA
3AB	Smp_013070	Homeobox domain	Hepatocyte nuclear factor 6 {Mouse ( <i>Mus musculus</i> ) [TaxId: 10090]}
3AB	Smp_013080	IQ motif, EF-hand binding site	Vacuolar protein sorting-associated protein 4
3AB	Smp_243700	Protein kinase domain	Cyclin-dependent PK, CDK6 {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
3AB	Smp_308550	METTL16/RlmF family	Methyltransferase 10 domain containing protein METT10D {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
3AB	Smp_129130	Helicase, C-terminal	ATP-dependent DNA helicase Q5 (E.C.3.6.4.12)
3AB	Smp_129120	Leucine-rich repeat	BURRH
3AB	Smp_129110	Fanconi-associated nuclease 1-like	Fanconi-associated nuclease 1 (E.C.3.1.21.-, 3.1.4.1)
3AB	Smp_012980	Eukaryotic translation initiation factor 3 subunit J	eIF3_subunit
3AB	Smp_308590	Myotubularin-like phosphatase domain	MYOTUBULARIN-RELATED PROTEIN 6 (E.C.3.1.3.-)
3AB	Smp_308600	FYVE zinc finger	Myotubularin-related protein 2, C-terminal domain {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
3AB	Smp_243690	Cytoplasmic polyadenylation element-binding protein, ZZ domain	Cytoplasmic polyadenylation element-binding protein 1
3AB	Smp_026390	Zinc finger, AN1-type	RIKEN cDNA 2310008M20 protein
3AB	Smp_136400	Cytochrome b561/ferric reductase transmembrane	Cytochrome b reductase 1 (E.C.1.-.-)
3AB	Smp_246440	Immunoglobulin-like fold	PapD-like
3AB	Smp_026340	XRN2-binding (XTBD) domain	5'-3' EXORIBONUCLEASE 2 HOMOLOG (E.C.3.1.13.-)
3AB	Smp_026320	Mediator complex, subunit Med25, von Willebrand factor type A	Med25_VWA
3AB	Smp_332420	GroES-like superfamily	Trans-2-enoyl-CoA reductase (E.C.1.3.1.38)
3AB	Smp_308630	GroES-like superfamily	Trans-2-enoyl-CoA reductase (E.C.1.3.1.38)
3AB	Smp_026300	rRNA biogenesis protein RRP36	RRP36
3AB	Smp_136360	Metallo-beta-lactamase	SGC - Diamond I04-1 fragment
3AB	Smp_026280	NA	NRBF2
3AB	Smp_026270	Zinc finger C2H2-type	Zinc finger protein 297b {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
3AB	Smp_136350	Ribonuclease P/MRP protein subunit	RNase P protein component 2, Rnp2 { <i>Pyrococcus horikoshii</i> [TaxId: 53953]}

## Annexe 6

3AB	Smp_026260	Zinc finger, RING-type	V(D)J recombination activating protein 1 (RAG1), dimerization domain {Mouse ( <i>Mus musculus</i> ) [TaxId: 10090]}
3AB	Smp_026250	Tetratricopeptide-like helical domain superfamily	Uncharacterized protein, Peptide alpha-N-acetyltransferase
3AB	Smp_343040	NA	FAM184
3AB	Smp_026230	NA	Snake coagglutinin alpha chain {Habu snake ( <i>Trimeresurus flavoviridis</i> ), flavocetin-A [TaxId: 88087]}
3AB	Smp_246430	Sterile alpha motif domain	Receptor-type tyrosine-protein phosphatase delta (E.C.3.1.3.48)
3AB	Smp_092810	TMEM33/Pom33 family	UPF0121
3AB	Smp_308680	Protein phosphatase 2A regulatory subunit PR55	Serine/threonine-protein phosphatase 2A 65 kDa
3AB	Smp_335460	NA	NA
3AB	Smp_308690	Protein phosphatase 2A regulatory subunit PR55	NA
3AB	Smp_335470	NA	NA
3AB	Smp_335480	NA	NA
3AB	Smp_308700	Protein phosphatase 2A regulatory subunit PR55	Serine/threonine-protein phosphatase 2A 65 kDa
3AB	Smp_335490	NA	NA
3AB	Smp_326220	NA	NA
3AB	Smp_335500	NA	DUF4183
3AB	Smp_326230	NA	NA
3AB	Smp_335510	NA	NA
3AB	Smp_344300	NA	NA
3AB	Smp_150230	Cation efflux protein	Cation efflux family protein
3AB	Smp_052390	Origin recognition complex, subunit 5	Orc2, Orc3, Orc5, Orc1, Orc6
3AB	Smp_150220	PTP type protein phosphatase	Protein-tyrosine phosphatase alpha {Mouse ( <i>Mus musculus</i> ) [TaxId: 10090]}
3AB	Smp_150210	Zinc finger, RING-type	U1 SNP1-associating protein 1, Degradation
3AB	Smp_150200	Glycosyltransferase, ALG3	ALG3
3AB	Smp_052350	Zinc finger, RanBP2-type	UBIQUITIN THIOESTERASE ZRANB1 (E.C.3.4.19.12)
3AB	Smp_052330	Glycosyl transferase, family 1	DNA alpha-glucosyltransferase (E.C.2.4.1.26)
3AB	Smp_150180	G protein-coupled receptor, rhodopsin-like	Chimera of Proteinase-activated receptor 1
3AB	Smp_084380	Metallo-beta-lactamase	Small nuclear ribonucleoprotein Sm D3
3AB	Smp_167920	Ubiquitin-2	UBN_AB
3AB	Smp_084400	CTLH/CRA C-terminal to LisH motif domain	Vacuolar import and degradation protein
3AB	Smp_007510	Protein kinase domain	Cell division protein kinase 7, CDK7 {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
3AB	Smp_007500	Golgi phosphoprotein 3-like	Sac1, Vps74

## Annexe 6

3AB	Smp_168800	Peptidase C12, ubiquitin carboxyl-terminal hydrolase	Ubiquitin carboxyl-terminal hydrolase UCH-L3 {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
3AB	Smp_085790	NA	NA
3AB	Smp_197540	Lipocalin-interacting membrane receptor	LMBR1
3AB	Smp_326250	NA	30S ribosomal protein S2, 30S
3AB	Smp_085780	Ribosomal protein L10P	28S ribosomal protein S2, mitochondrial
3AB	Smp_168780	NA	NA
3AB	Smp_168770	Mitochondrial import inner membrane translocase subunit Tim16	Pam16
3AB	Smp_085750	Zinc finger protein NOA36	NOA36
3AB	Smp_085740	SH3 domain	Cytoplasmic FMR1-interacting protein 1, Nck-associated
3AB	Smp_076650	Calcium release-activated calcium channel protein	Orai-1
3AB	Smp_168760	Flavin amine oxidase	ferredoxin reductase (E.C.1.18.1.2)
3AB	Smp_131600	Cadherin-like	Protocadherin-15
3AB	Smp_087550	NA	Cytochrome c oxidase subunit 1
3AB	Smp_087560	Dynein regulatory complex protein 1/2, N-terminal	NYD-SP28
3AB	Smp_169780	SMAD domain, Dwarf1-type	Smad4 tumor suppressor C-terminal domain {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
3AB	Smp_087580	BRO1 domain	Crystal Structure of ALIX/AIP1 in
3AB	Smp_243670	DNA topoisomerase, type IA	DNA TOPOISOMERASE 3-ALPHA (E.C.5.99.1.2), RECQ-MEDIATED
3AB	Smp_243660	Cilia- and flagella-associated protein 97	NA
3AB	Smp_087620	BRO1 domain	Crystal Structure of ALIX/AIP1 in
3AB	Smp_169800	Citron homology (CNH) domain	Rho guanyl nucleotide exchange factor
3AB	Smp_169820	Gamma-tubulin complex component protein	Gamma-tubulin complex component 6
3AB	Smp_345930	NA	HNH_5
3AB	Smp_326270	NA	NA
3AB	Smp_210720	Uncharacterised protein family UPF0047	Hypothetical protein TM0723 { <i>Thermotoga maritima</i> [TaxId: 2336]}
3AB	Smp_213030	XPG/Rad2 endonuclease	Flap endonuclease GEN homolog 1/DNA
3AB	Smp_087690	Protease inhibitor I35 (TIMP)	Metalloproteinase inhibitor 1, Stromelysin-2 (E.C.3.4.24.22)
3AB	Smp_156250	NA	Photosystem I iron-sulfur protein PsaC { <i>Synechococcus elongatus</i> [TaxId: 32046]}
3AB	Smp_063230	Leucine-rich repeat	von Willebrand factor binding domain of glycoprotein Ib alpha {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
3AB	Smp_156270	Per1-like	Per1
3AB	Smp_156280	Hyccin	Phosphatidylinositol 4-kinase III alpha (PI4KA)



## Annexe 6

3AB	Smp_063250	Ubiquinol-cytochrome c chaperone, CBP3	Ubiquinol-cytochrome C chaperone
3AB	Smp_063260	FYVE zinc finger	Platelet-activating factor acetylhydrolase IB subunit alpha {Mouse ( <i>Mus musculus</i> ) [TaxId: 10090]}
3AB	Smp_156290	ARID DNA-binding domain	Lysine-specific demethylase 5A (E.C.1.14.11.-)
3AB	Smp_156300	WD40 repeat	Tubulin beta, Tubulin alpha, PACRG
3AB	Smp_063300	Ribosomal protein L36e	Protein translation inhibitor, complex, ribosome
3AB	Smp_156310	NA	NA
3AB	Smp_333570	Sodium	Sodium-dependent serotonin transporter, 8B6 antibody
3AB	Smp_063330	NA	NA
3AB	Smp_340980	Phosphoserine phosphatase, domain 2	Phosphoserine phosphatase {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
3AB	Smp_063350	Ribosomal protein L27e	uL2, uL3, uL4, uL18, eL6
3AB	Smp_308750	NA	NA
3AB	Smp_308760	NA	VASODILATOR-STIMULATED PHOSPHOPROTEIN
3AB	Smp_326300	NA	NA
3AB	Smp_063380	HAD superfamily	Phosphoserine phosphatase {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
3AB	Smp_063390	4'-phosphopantetheinyl transferase domain	HSPC223 (E.C.1.2.1.31)
3AB	Smp_156340	NA	NA
3AB	Smp_156350	WD40 repeat	Actin interacting protein 1 {Nematode ( <i>Caenorhabditis elegans</i> ) [TaxId: 6239]}
3AB	Smp_063420	Chaperonin TCP-1, conserved site	T-COMPLEX PROTEIN 1 SUBUNIT ALPHA
3AB	Smp_241400	Pyruvate kinase	Pyruvate kinase (E.C.2.7.1.40)
3AB	Smp_214390	Homeobox domain	Antennapedia Homeodomain { <i>Drosophila melanogaster</i> [TaxId: 7227]}
3AB	Smp_156370	NA	CCDC144C
3AB	Smp_156390	Ras-associating (RA) domain	Growth factor receptor-bound protein 14
3AB	Smp_243680	AAA ATPase domain	HslU { <i>Haemophilus influenzae</i> [TaxId: 727]}
3AB	Smp_063500	NA	Deltex protein 2 RING-H2 domain {Mouse ( <i>Mus musculus</i> ) [TaxId: 10090]}
3AB	Smp_021340	Myc-type, basic helix-loop-helix (bHLH) domain	Mad protein {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
3AB	Smp_326360		0.34
3AB	Smp_123440	FAD dependent oxidoreductase	D-amino-acid oxidase (E.C.1.4.3.3)
3AB	Smp_123450	Nucleoporin Nup88	Nup88
3AB	Smp_123460	NA	Peptidyl-tRNA hydrolase { <i>Escherichia coli</i> [TaxId: 562]}
3AB	Smp_002510	Transketolase-like, pyrimidine-binding domain	Pyruvate dehydrogenase E1-beta, PdhB, N-terminal domain { <i>Bacillus stearothermophilus</i> [TaxId: 1422]}
3AB	Smp_123470	Leucine-rich repeat	Rna1p (RanGAP1), N-terminal domain {Fission yeast ( <i>Schizosaccharomyces pombe</i> ) [TaxId: 4896]}

## Annexe 6

3AB	Smp_123480	Protein kinase domain	Death-associated protein kinase, Dap {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
3AB	Smp_123490	Cytidyltransferase-like domain	Nicotinamide mononucleotide (NMN) adenylyltransferase {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
3AB	Smp_002550	RNA recognition motif domain	RNA-binding protein 8 {Fruit fly ( <i>Drosophila melanogaster</i> ) [TaxId: 7227]}
3AB	Smp_123500	Helicase, C-terminal	DEAH (Asp-Glu-Ala-His) box polypeptide 37/RNA
3AB	Smp_123510	Vps16, C-terminal	Vps16_C
3AB	Smp_123520	(Uracil-5)-methyltransferase family	rRNA (Uracil-5-)-methyltransferase RumA, catalytic domain { <i>Escherichia coli</i> [TaxId: 562]}
3AB	Smp_123530	tRNA/rRNA methyltransferase, SpoU type	RNA 2'-O-RIBOSE METHYLTRANSFERASE(E.C.2.1.1.-)
3AB	Smp_002600	Peptidase S28	Lysosomal Pro-X carboxypeptidase (E.C.3.4.16.2)
3AB	Smp_202120	Homeobox domain	Msx-1 homeodomain {Mouse ( <i>Mus musculus</i> ) [TaxId: 10090]}
3AB	Smp_002640	Homeobox domain	Antennapedia Homeodomain { <i>Drosophila melanogaster</i> [TaxId: 7227]}
3AB	Smp_123570	Myc-type, basic helix-loop-helix (bHLH) domain	Aryl hydrocarbon receptor nuclear translocator
3AB	Smp_123590	NA	Penicillin-binding protein 2x (pbp-2x), c-terminal domain { <i>Streptococcus pneumoniae</i> [TaxId: 1313]}
3AB	Smp_123600	RNA recognition motif domain	SMART/HDAC1 associated repressor protein, SHARP {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
3AB	Smp_123610	Protein kinase domain	Phosphatidylinositol 3-kinase VPS34 (E.C.2.7.1.137), Serine/threonine-protein
3AB	Smp_123620	OTU domain	OTU domain-containing protein 5 (E.C.3.4.19.12)
3AB	Smp_123630	Peptidase C19, ubiquitin carboxyl-terminal hydrolase	TATA-binding protein, Transcriptional coactivator HFI1/ADA1
3AB	Smp_002740	NA	Arylsulfatase B (4-sulfatase) {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
3AB	Smp_123640	GPCR kinase	G protein-coupled receptor kinase 4
3AB	Smp_308820	Leucine-rich repeat	Myeloperoxidase (E.C.1.11.2.2)
3AB	Smp_308830	NA	NA
3AB	Smp_300810	AAA ATPase domain	Spastin (E.C.5.6.1.1), polyglutamate peptide
3AB	Smp_002820	Zinc finger, CCCH-type	Zinc finger CCCH domain-containing protein C19orf7 (KIAA1064) {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
3AB	Smp_002840	Uncharacterised protein family UPF0489	UPF0489
3AB	Smp_123670	Histidine phosphatase superfamily, clade-2	Prostatic acid phosphatase {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
3AB	Smp_002860	NA	NA
3AB	Smp_002870	TMEM115/Pdh1/Rbl19	DER1

## Annexe 6

3AB	Smp_002880	ATPase, F1/V1/A1 complex, alpha/beta subunit, nucleotide-binding domain	ATP synthase alpha chain, mitochondrial
3AB	Smp_212320	RNA polymerase, alpha subunit	DNA-directed RNA polymerase III subunit
3AB	Smp_347390	Anoctamin	Anoctamin-10
3AB	Smp_123690	tRNA-splicing endonuclease, subunit Sen54, N-terminal	tRNA-splicing endonuclease (E.C.3.1.27.9), NEQ261
3AB	Smp_002920	TRAM/LAG1/CLN8 homology domain	TRAM_LAG1_CLN8
3AB	Smp_337740	Histone H2A	Histone H2A {Chicken ( <i>Gallus gallus</i> ), erythrocytes [TaxId: 9031]}
3AB	Smp_341880	Leucine-rich repeat	von Willebrand factor binding domain of glycoprotein Ib alpha {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
3AB	Smp_084520	SRR1-like domain	SRR1
3AB	Smp_084540	Forkhead-associated (FHA) domain	RNA-binding protein 5, Survival motor
3AB	Smp_165280	WD40 repeat	Transducin-like enhancer protein 1
3AB	Smp_197370	Fibrillar collagen, C-terminal	Collagen alpha-1(I) chain, Collagen alpha-2(I)
3AB	Smp_164590	Fibrillar collagen, C-terminal	Collagen alpha-1(I) chain, Collagen alpha-2(I)
3AB	Smp_078750	Tetraspanin	Tetraspanin
3AB	Smp_345510	P-loop containing nucleoside triphosphate hydrolase	di-Ras2 {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
3AB	Smp_078730	Dihydroorotate dehydrogenase, conserved site	Dihydroorotate dehydrogenase (E.C.1.3.98.1)
3AB	Smp_326390	Dynein light chain, type 1/2	Dynein light chain 1 (DLC1) {Fruit fly ( <i>Drosophila melanogaster</i> ) [TaxId: 7227]}
3AB	Smp_202130	Dynein light chain, type 1/2	Dynein light chain 1 (DLC1) {Fruit fly ( <i>Drosophila melanogaster</i> ) [TaxId: 7227]}
3AB	Smp_164560	NA	NA
3AB	Smp_213150	Pop1, N-terminal	Ribonucleases P/MRP protein subunit POP1
3AB	Smp_168160	HECT domain	Ubiquitin-protein ligase E3a (E6ap) {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
3AB	Smp_084830	UDP-glucose/GDP-mannose dehydrogenase, N-terminal	UDP-glucose 6-dehydrogenase (E.C.1.1.1.22)
3AB	Smp_197480	Abhydrolase domain containing 18	CARBOXYLESTERASE (EST-2) (E.C.3.1.1.1)
3AB	Smp_165180	NA	NA
3AB	Smp_079840	Tetrapyrrole biosynthesis, uroporphyrinogen III synthase	Uroporphyrinogen III synthase (U3S, HemD) {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
3AB	Smp_165170	Ion transport domain	nvTRPM2 channel
3AB	Smp_079810	Peptide chain release factor class I/class II	Probable peptide chain release factor
3AB	Smp_079800	Rab-GTPase-TBC domain	RabGAP/TBC protein
3AB	Smp_165150	Ankyrin repeat	I-kappa-B-alpha {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
3AB	Smp_079770	Disulphide isomerase	Protein disulfide-isomerase A3ERp57 (E.C.5.3.4.1)/Tapasin

## Annexe 6

3AB	Smp_079760	Protein kinase domain	B-Raf kinase {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
3AB	Smp_079750	Ribosomal RNA processing protein 8	Methyltransf_8
3AB	Smp_079740	Cullin, N-terminal	Cullin-5, RING-box protein 2
3AB	Smp_079730	RNA recognition motif domain	mRNA 3'-end-processing protein RNA14, Rna15
3AB	Smp_079710	Ribosomal protein/NADH dehydrogenase domain	NADH-ubiquinone oxidoreductase b8 subunit, CI-B8 {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
3AB	Smp_345590	NA	NA
3AB	Smp_331930	NA	NA
3AB	Smp_183870	NA	NA
3AB	Smp_341780	NA	NA
3AB	Smp_346730	NA	NA
3AB	Smp_185460	NA	NA
3AB	Smp_346550	NA	NA
3AB	Smp_326400	Pseudouridine synthase, RsuA/RluA	Ribosomal large subunit pseudouridine synthase D, RluD { <i>Escherichia coli</i> [TaxId: 562]}
3AB	Smp_333580	Microtubule-associated protein, MAP65/Ase1/PRC1	Protein regulator of cytokinesis 1
3AB	Smp_084940	Small GTPase	Ras-related protein M-Ras (XRas) {Mouse ( <i>Mus musculus</i> ) [TaxId: 10090]}
3AB	Smp_341420	Dual specificity phosphatase, catalytic domain	Mapk phosphatase {Human ( <i>Homo sapiens</i> ), pac-1 [TaxId: 9606]}
3AB	Smp_168200	F-box domain	Cryptochrome-2, F-box/LRR-repeat protein 3, S-phase
3AB	Smp_084910	Protein FMC1 homologue	Complex1_LYR_2
3AB	Smp_084900	Initiation factor 2B-related	Translation initiation factor eIF-2B subunit
3AB	Smp_084890	Type I phosphodiesterase/nucleotide pyrophosphatase/phosphate transferase	PHOSPHONOACETATE HYDROLASE (E.C.3.11.1.2)
3AB	Smp_084870	CID domain	PCF11 protein {Baker's yeast ( <i>Saccharomyces cerevisiae</i> ) [TaxId: 4932]}
3AB	Smp_084860	Domain of unknown function DUF155	DUF155
3AB	Smp_168170	Calcineurin-like phosphoesterase domain, ApaH type	Uncharacterized C17orf48 homolog zgc:64213 {Zebrafish ( <i>Danio rerio</i> ) [TaxId: 7955]}
3AB	Smp_138210	Alpha crystallin/Hsp20 domain	SMALL HEAT SHOCK PROTEIN
3AB	Smp_347470	Homeobox domain	Antennapedia Homeodomain { <i>Drosophila melanogaster</i> [TaxId: 7227]}
3AB	Smp_138220	NA	Enhancer of mRNA-decapping protein 3
3AB	Smp_138230	DM DNA-binding domain	human DmrT1/DNA Complex
3AB	Smp_029670	NA	Methyltransf_21
3AB	Smp_138240	Zinc finger, double-stranded RNA binding	Pre-mRNA-processing-splicing factor 8, 116 kDa
3AB	Smp_029700	NF-kappa-B-activating protein, C-terminal	Nkap_C
3AB	Smp_246360	Uncharacterised domain NUC173	NUC173

## Annexe 6

3AB	Smp_246350	NA	BAR_3
3AB	Smp_029740	Flavinator of succinate dehydrogenase	Succinate dehydrogenase [ubiquinone] flavoprotein subunit
3AB	Smp_029750	Pre-rRNA-processing protein TSR2	Pre-rRNA-processing protein TSR2, 40S ribosomal
3AB	Smp_138280	Ribonuclease III domain	28S ribosomal protein S2, mitochondrial
3AB	Smp_138290	Dynein light chain, type 1/2	DYNEIN LIGHT CHAIN 2, CYTOPLASMIC
3AB	Smp_246340	NA	Tubulin beta, Tubulin alpha, PACRG
3AB	Smp_029800	FERM domain	FERM, RhoGEF and pleckstrin domain
3AB	Smp_138300	Tetraspanin, EC2 domain superfamily	CD81 antigen
3AB	Smp_029820	Ribosomal protein L12 family	Ribosome
3AB	Smp_138310	Dual specificity phosphatase, catalytic domain	Mapk phosphatase {Human ( <i>Homo sapiens</i> ), pac-1 [TaxId: 9606]}
3AB	Smp_029850	Ribonuclease H-like superfamily	3'-5' exonuclease ERI1 (E.C.3.1.-.-)
3AB	Smp_138320	NA	NA
3AB	Smp_029890	Septin	Septin
3AB	Smp_029900	Cyclophilin-type peptidyl-prolyl cis-trans isomerase domain	Peptidyl-prolyl cis-trans isomerase-like 1, PP1L1 {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
3AB	Smp_138330	Tetraspanin, EC2 domain superfamily	CD81 antigen
3AB	Smp_138340	Zinc finger C2H2-type	Transcriptional repressor CTCF {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
3AB	Smp_138350	Zinc finger C2H2-type	NOA36
3AB	Smp_343210	NA	NA
3AB	Smp_340620	NA	NA
3AB	Smp_029940	Ubiquitin-conjugating enzyme E2	Ubiquitin-conjugating enzyme E2 Q2, C-terminal domain {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
3AB	Smp_029970	NA	NA
3AB	Smp_029990	Dedicator of cytokinesis, C-terminal	Dedicator of cytokinesis protein 2
3AB	Smp_030000	Peptidase M17, leucyl aminopeptidase, C-terminal	Cytosol aminopeptidase (E.C.3.4.11.1)
3AB	Smp_030010	CCDC92/74, coiled-coil domain	CCDC92
3AB	Smp_030020	PTIP-associated protein 1	PAXIP1_C
3AB	Smp_030030	TM2 domain	TM2
3AB	Smp_347780	Homeobox domain	Msx-1 homeodomain {Mouse ( <i>Mus musculus</i> ) [TaxId: 10090]}
3AB	Smp_308980	BRCT domain	BRCT-containing protein 1
3AB	Smp_309000	BRCT domain	Regulator of Ty1 transposition protein
3AB	Smp_311980	Zinc finger C2H2-type	LYSINE-SPECIFIC HISTONE DEMETHYLASE 1A (E.C.1.-.-.-)
3AB	Smp_138430	DEP domain	Regulator of G-protein signaling 7
3AB	Smp_030110	NA	Actin, alpha skeletal muscle, Gelsolin
3AB	Smp_316460	Zinc finger, UBR-type	E3 ubiquitin-protein ligase UBR2 (E.C.6.3.2.19)
3AB	Smp_326430	Exoribonuclease, phosphorolytic domain 1	Exosome complex exonuclease RRP46 {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
3AB	Smp_337750	DNA repair protein, Swi5	Swi5-dependent recombination DNA repair protein

## Annexe 6

3AB	Smp_120960	Exoribonuclease, phosphorolytic domain 1	Exosome complex exonuclease RRP46 {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
3AB	Smp_030150	WH1/EVH1 domain	Homer {Rat ( <i>Rattus norvegicus</i> ) [TaxId: 10116]}
3AB	Smp_138470	NA	Type-2 restriction enzyme DpnI (E.C.3.1.21.4)/DNA
3AB	Smp_138480	NEMP family	NEMP
3AB	Smp_030180	Beta-grasp domain superfamily	C9orf74 homolog {Mouse ( <i>Mus musculus</i> ) [TaxId: 10090]}
3AB	Smp_343220	Bardet-Biedl syndrome 5 protein	Bardet-Biedl syndrome 18 protein, BBS1
3AB	Smp_138500	Arf GTPase activating protein	Ras-related protein Rab-35, Iporin
3AB	Smp_030220	HAD-superfamily hydrolase, subfamily IIA	Hypothetical protein TM1742 { <i>Thermotoga maritima</i> [TaxId: 2336]}
3AB	Smp_138510	NA	NA
3AB	Smp_030250	Nuclear transcription factor Y subunit A	NUCLEAR TRANSCRIPTION FACTOR Y SUBUNIT
3AB	Smp_030260	NA	HIG1 domain family member 1B
3AB	Smp_138520	Drought induced 19 protein type, zinc-binding domain	Zinc finger protein 568/DNA Complex
3AB	Smp_030280	OTU domain	OTU domain-containing protein
3AB	Smp_030290	Prefoldin beta-like	Prefoldin beta subunit { <i>Archaeon Methanobacterium thermoautotrophicum</i> [TaxId: 145262]}
3AB	Smp_340630	Heat shock protein Hsp90 family	ATP-DEPENDENT MOLECULAR CHAPERONE HSP82, CO-CHAPERONE
3AB	Smp_030320	WD40 repeat	F-box/WD repeat-containing protein 7, FBXW7 {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
3AB	Smp_030330	Peptidyl-tRNA hydrolase, PTH2	Hypothetical protein TA0108 { <i>Archaeon Thermoplasma acidophilum</i> [TaxId: 2303]}
3AB	Smp_333590	CUB domain	Gastric intrinsic factor, Cubilin
3AB	Smp_030350	CUB domain	Granzyme K {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
3AB	Smp_326450	NA	NA
3AB	Smp_326460	NA	NA
3AB	Smp_326470	NA	NA
3AB	Smp_326480	NA	NA
3AB	Smp_309090	NA	NA
3AB	Smp_309100	NA	NA
3AB	Smp_343230	NA	Lachrymatory-factor synthase
3AB	Smp_341710	NA	Catalytic domain of MutY { <i>Escherichia coli</i> [TaxId: 562]}
3AB	Smp_333600	P2X purinoreceptor	P2X purinoceptor 3
3AB	Smp_309110	NAD(P)-binding domain	TAT-interacting protein TIP30 {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
3AB	Smp_326490	NA	GRP
3AB	Smp_159890	Peptidase M14, carboxypeptidase A	Carboxypeptidase M, catalytic domain {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
3AB	Smp_333610	Nucleolar, Nop52	Nop52

## Annexe 6

3AB	Smp_326500	Dolichol-phosphate mannosyltransferase subunit 3	DPM3
3AB	Smp_159860	G protein-coupled receptor, rhodopsin-like	Cannabinoid receptor 1, Flavodoxin
3AB	Smp_159840	Dynein heavy chain, C-terminal domain	Dynein heavy chain, cytoplasmic
3AB	Smp_159830	NA	CYYR1
3AB	Smp_159800	NA	NA
3AB	Smp_326510	NA	GRP
3AB	Smp_180310	NA	Histidyl-tRNA synthetase (HisRS) { <i>Staphylococcus aureus</i> [TaxId: 1280]}
3AB	Smp_180340	NA	SMK-1
3AB	Smp_180330	NA	HTP reductase { <i>Methanococcus jannaschii</i> [TaxId: 2190]}
3AB	Smp_180320	NA	toxin Hm-3
3AB	Smp_345100	NA	Squamosa-promoter binding-like protein 12, DNA-binding domain {Thale cress ( <i>Arabidopsis thaliana</i> ) [TaxId: 3702]}
3AB	Smp_309120	NA	NA
3AB	Smp_336990	NA	Squamosa-promoter binding-like protein 12, DNA-binding domain {Thale cress ( <i>Arabidopsis thaliana</i> ) [TaxId: 3702]}
3AB	Smp_337000	NA	WAK
3AB	Smp_016640	Histidine phosphatase superfamily, clade-2	Prostatic acid phosphatase {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
3AB	Smp_342580	Myc-type, basic helix-loop-helix (bHLH) domain	Transcription factor E2-alpha, DNA-binding protein
3AB	Smp_179810	Troponin	PROTEIN
3AB	Smp_340020	Small GTP-binding protein domain	ADP-ribosylation factor {Human ( <i>Homo sapiens</i> ), ARF6 [TaxId: 9606]}
3AB	Smp_000720	Protein kinase domain	Dual specificity mitogen-activated protein kinase kinase 1, Mek1 {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
3AB	Smp_000710	Peptidase C19, ubiquitin carboxyl-terminal hydrolase	Ubiquitin carboxyl-terminal hydrolase 7 (USP7, HAUSP) {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
3AB	Smp_000700	SET domain	Smyd3 methyltransferase
3AB	Smp_000690	POU-specific domain	Oct-1 {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
3AB	Smp_000680	Mitochondrial substrate/solute carrier	ADP,ATP carrier protein {Cow ( <i>Bos taurus</i> ), heart isoform t1 [TaxId: 9913]}
3AB	Smp_000670	Arf GTPase activating protein	ARFGAP with coiled-coil, ANK repeat
3AB	Smp_000660	Aminotransferase class-III	Ornithine aminotransferase {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
3AB	Smp_000650	Zinc finger, RING-type	SIAH, seven in absentia homolog {Mouse ( <i>Mus musculus</i> ) [TaxId: 10090]}
3AB	Smp_000640	SH2 domain	v-crk sarcoma virus CT10 oncogene
3AB	Smp_000630	Isochorismatase-like	Ribonuclease MAR1 { <i>Leishmania donovani</i> [TaxId: 5661]}

## Annexe 6

3AB	Smp_000620	Myc-type, basic helix-loop-helix (bHLH) domain	DNA-binding protein inhibitor ID-1
3AB	Smp_000600	WD40 repeat	Rsa4
3AB	Smp_000590	28S ribosomal protein S26	MITORIBOSOMAL PROTEIN US2M, MRPS2, MITORIBOSOMAL
3AB	Smp_000580	NA	Tubulin beta, Tubulin alpha, PACRG
3AB	Smp_122370	Bromodomain	PHD-Bromo of TRIM66 protein, ALA-ARG-THR-LYS-GLN-THR-ALA-ARG-LYS-SER-THR-GLY
3AB	Smp_000560	NA	NA
3AB	Smp_266960	Zinc finger C2H2-type	FIVE-FINGER GLI/DNA COMPLEX
3AB	Smp_333620	Myc-type, basic helix-loop-helix (bHLH) domain	Myc proto-oncogene protein {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
3AB	Smp_185930	NA	Coatomer subunit alpha, Coatomer subunit
3AB	Smp_162070	NA	NA
3AB	Smp_000510	Hexapeptide repeat	Glucose-1-phosphate adenylyltransferase (E.C.2.7.7.27)
3AB	Smp_000500	NA	NA
3AB	Smp_122340	BTB/POZ domain	S-phase kinase-associated protein 1, F-box/LRR-repeat
3AB	Smp_000480	Proteasome component (PCI) domain	40S ribosomal protein SA, 40S
3AB	Smp_000470	ATP-dependent (S)-NAD(P)H-hydrate dehydratase	Hypothetical protein TM0922, C-terminal domain { <i>Thermotoga maritima</i> [TaxId: 2336]}
3AB	Smp_326520	NA	DNA damage-binding protein 1, DDB1-
3AB	Smp_122320	NA	Activity-regulated cytoskeleton associated protein 1
3AB	Smp_309210	UBA-like superfamily	Ubiquitin-associated protein 2-like Uba2l {Mouse ( <i>Mus musculus</i> ) [TaxId: 10090]}
3AB	Smp_309220	Nucleotide-binding alpha-beta plait domain superfamily	Poly(A)-binding protein {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
3AB	Smp_000430	Trematode Eggshell Synthesis	TES
3AB	Smp_000420	NA	NA
3AB	Smp_326530	NA	GRP
3AB	Smp_122280	NA	NA
3AB	Smp_326540	NA	NA
3AB	Smp_191910	Trematode Eggshell Synthesis	TES
3AB	Smp_340010	Trematode Eggshell Synthesis	TES
3AB	Smp_326550	NA	NA
3AB	Smp_326560	NA	NA
3AB	Smp_000390	Trematode Eggshell Synthesis	TES
3AB	Smp_326570	NA	NA
3AB	Smp_326580	NA	NA
3AB	Smp_000370	Selenoprotein SelK/SelG	SelK_SelG
3AB	Smp_312230	NA	CCDC73
3AB	Smp_000340	Nuclear hormone receptor, ligand-binding domain	Retinoic acid receptor RXR-alpha, Retinoic



## Annexe 6

3AB	Smp_000320	NDRG	Protein NDRG3
3AB	Smp_000310	NA	PRA1
3AB	Smp_000300	Tetratricopeptide-like helical domain superfamily	Coatomer subunit alpha, Coatomer subunit
3AB	Smp_112450	Trematode Eggshell Synthesis	TES
3AB	Smp_077890	Trematode Eggshell Synthesis	TES
3AB	Smp_307900	Trematode Eggshell Synthesis	TES
3AB	Smp_332440	Trematode Eggshell Synthesis	TES
3AB	Smp_309250	Trematode Eggshell Synthesis	TES
3AB	Smp_326600	Trematode Eggshell Synthesis	TES
3AB	Smp_326610	Trematode Eggshell Synthesis	TES
3AB	Smp_000280	Trematode Eggshell Synthesis	TES
3AB	Smp_000270	Trematode Eggshell Synthesis	TES
3AB	Smp_000260	Sapoin A-type domain	(Pro)phytepsin {Barley ( <i>Hordeum vulgare</i> ) [TaxId: 4513]}
3AB	Smp_000250	NA	tRNA modification GTPase trmE
3AB	Smp_000240	RNA recognition motif domain	Transportin-1, RNA-binding protein FUS
3AB	Smp_000230	Zinc finger C2H2-type	LYSINE-SPECIFIC HISTONE DEMETHYLASE 1A (E.C.1.-.-.-)
3B	Smp_000200	Smr domain	Polynucleotide kinase, kinase domain {Bacteriophage T4 [TaxId: 10665]}
3B	Smp_122190	NA	p47 {Rat ( <i>Rattus norvegicus</i> ) [TaxId: 10116]}
3B	Smp_326620	Cysteine-rich domain, DPF-motif	C6_DPF
3B	Smp_000190	Short-chain dehydrogenase/reductase SDR	17-beta-hydroxysteroid dehydrogenase type XI {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
3B	Smp_000180	EF-hand domain	Frequenin (neuronal calcium sensor 1) {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
3B	Smp_000160	Leucine-rich repeat N-terminal domain	Slit {Fruit fly ( <i>Drosophila melanogaster</i> ) [TaxId: 7227]}
3B	Smp_122170	NA	Thrombospondin-1 (TSP-1) {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
3B	Smp_000150	Zinc finger, PHD-type	Histone H3.2, Histone H4, Histone
3B	Smp_187940	NAD kinase/diacylglycerol kinase-like domain superfamily	DAGK_cat
3B	Smp_000130	Hormone-sensitive lipase, N-terminal	HSL_N
3B	Smp_000110	Elongation factor EFG, domain V-like	60S RIBOSOMAL PROTEIN L3, 60S
3B	Smp_000100	Filamin/ABP280 repeat	FILAMIN-B
3B	Smp_118830	F-actin-capping protein subunit alpha	Capz alpha-1 subunit {Chicken ( <i>Gallus gallus</i> ) [TaxId: 9031]}
3B	Smp_167630	Zinc/iron permease	Zip
3B	Smp_167640	Peptidase M14, carboxypeptidase A	FAMILY M14 UNASSIGNED PEPTIDASE
3B	Smp_167650	Dbl homology (DH) domain	GEF of TIAM1 (T-Lymphoma invasion and metastasis inducing protein 1) {Mouse ( <i>Mus musculus</i> ) [TaxId: 10090]}
3B	Smp_083870	Peptidase M17, leucyl aminopeptidase, C-terminal	Cytosol aminopeptidase (E.C.3.4.11.1)

## Annexe 6

3B	Smp_083880	G protein-coupled receptor, rhodopsin-like	Neurotensin receptor type 1, Endolysin, NEUROTENSIN RECEPTOR
3B	Smp_167660	CBM21 (carbohydrate binding type-21) domain	Protein phosphatase 1 regulatory subunit
3B	Smp_167670	NA	GTP-sensing transcriptional pleiotropic repressor CodY
3B	Smp_167680	HECT domain	E3 ubiquitin-protein ligase HUWE1 (E.C.6.3.2.-)
3B	Smp_083930	Chromatin modification-related protein Eaf6	Histone acetyltransferase ESA1 (E.C.2.3.1.48), Chromatin
3B	Smp_177720	G protein-coupled receptor, rhodopsin-like	C-X-C chemokine receptor type 4
3B	Smp_083940	G protein-coupled receptor, rhodopsin-like	7TM_GPCR_Srw
3B	Smp_326640	G protein-coupled receptor, rhodopsin-like	7TM_GPCR_Srw
3B	Smp_083950	NA	Soluble cytochrome b562, Thromboxane A2 receptor, Rubredoxin, Thromboxane
3B	Smp_083990	Amino acid/polyamine transporter I	4F2 cell-surface antigen heavy chain
3B	Smp_326650	NA	DUF1998
3B	Smp_083980	Epithelial sodium channel	Amiloride-sensitive sodium channel subunit alpha
3B	Smp_083970	ER membrane protein complex subunit 10	ER membrane protein complex subunit
3B	Smp_336790	FYVE zinc finger	Early Endosomal Autoantigen 1
3B	Smp_246300	WD40 repeat	40S ribosomal protein S1-A, 40S
3B	Smp_144350	NA	BEN
3B	Smp_040710	DNA-directed RNA polymerase Rpb11, 13-16kDa subunit, conserved site	RPB11 {Baker's yeast ( <i>Saccharomyces cerevisiae</i> ) [TaxId: 4932]}
3B	Smp_040720	Protein of unknown function DUF2838	DUF2838
3B	Smp_144360	Myelin proteolipid protein PLP	Myelin_PLP
3B	Smp_144370	Leucine-rich repeat	Ribonuclease inhibitor {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
3B	Smp_144380	Homeobox domain	Msx-1 homeodomain {Mouse ( <i>Mus musculus</i> ) [TaxId: 10090]}
3B	Smp_040770	Aminoacyl-tRNA synthetase, class I, conserved site	Methionyl-tRNA synthetase (E.C.6.1.1.10)
3B	Smp_190740	NA	Coatomer subunit alpha, Coatomer subunit
3B	Smp_040790	Cyclophilin-type peptidyl-prolyl cis-trans isomerase domain	Cyclophilin (eukaryotic) { <i>Caenorhabditis elegans</i> , isoform 5 [TaxId: 6239]}
3B	Smp_040800	WHEP-TRS domain	Glycyl-tRNA synthetase (E.C.6.1.1.14)
3B	Smp_144390	Ser/Thr protein kinase, TGFB receptor	B-Raf kinase {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
3B	Smp_185420	NA	ms48, ms51, ms56, ms59, ms60
3B	Smp_196090	NA	Protein patched homolog 1, Sonic
3B	Smp_205170	NA	NA

## Annexe 6

3B	Smp_144430	Formin, FH3 domain	FMNL1, Cdc42
3B	Smp_144440	NA	NA
3B	Smp_144450	RNA recognition motif domain	Vinculin, Ribonucleoprotein PTB-binding 1
3B	Smp_190940	NA	EGF_3
3B	Smp_172030	Helicase, C-terminal	U1 small nuclear ribonucleoprotein 70
3B	Smp_199800	NA	DNA-directed primase/polymerase protein/DNA Complex
3B	Smp_337590	Matrin/U1-C, C2H2-type zinc finger	Spliceosomal protein U1C {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
3B	Smp_243210	Peptidase C14, p20 domain	Caspase-1 {Fall armyworm ( <i>Spodoptera frugiperda</i> ) [TaxId: 7108]}
3B	Smp_313810	Cell cycle regulator Mat89Bb	Integrator complex subunit 13, Integrator
3B	Smp_309360	Sterol-sensing domain	Protein dispatched homolog 1
3B	Smp_333630	PDZ domain	Synaptic protein PSD-95 {Rat ( <i>Rattus norvegicus</i> ) [TaxId: 10116]}
3B	Smp_326660	NA	LRR_11
3B	Smp_091240	Porin, eukaryotic type	Voltage-dependent anion-selective channel protein 1
3B	Smp_091230	Ubiquitin-fold modifier 1	Ufm1-specific protease (E.C.3.4.22.-), Ubiquitin-fold modifier
3B	Smp_118710	NA	DNA replication licensing factor MCM6
3B	Smp_057130	AdipoR/Haemolysin-III-related	HUMAN ADIPONECTIN RECEPTOR 2, single-chain
3B	Smp_027850	Chromatin SPT2	Histone H3.2, Histone H4, Protein
3B	Smp_027870	NA	NPIP
3B	Smp_027880	Prefoldin beta-like	Prefoldin beta subunit { <i>Archaeon Methanobacterium thermoautotrophicum</i> [TaxId: 145262]}
3B	Smp_347380	Zinc finger, TRAF-type	F-box_4
3B	Smp_137280	Coiled-coil domain-containing protein 87	Protein regulator of cytokinesis 1
3B	Smp_333640	Peptidase S59, nucleoporin	NUCLEOPORIN NUP43, NUCLEAR PORE COMPLEX
3B	Smp_027910	Synembryn	Synembryn-A, G alpha t C
3B	Smp_027920	Tubulin	SPINDLE POLE BODY COMPONENT SPC97
3B	Smp_347950	Myosin head, motor domain	MYOSIN 2 HEAVY CHAIN STRIATED
3B	Smp_241490	G protein-coupled receptor, rhodopsin-like	7TM_GPCR_Srw
3B	Smp_091950	G protein-coupled receptor, rhodopsin-like	7TM_GPCR_Srw
3B	Smp_326670	G protein-coupled receptor, rhodopsin-like	Platelet-activating factor receptor, Flavodoxin
3B	Smp_134960	G protein-coupled receptor, rhodopsin-like	Platelet-activating factor receptor
3B	Smp_023710	G protein-coupled receptor, rhodopsin-like	Platelet-activating factor receptor, Flavodoxin

## Annexe 6

3B	Smp_117340	G protein-coupled receptor, rhodopsin-like	Rhodopsin
3B	Smp_023700	Anticodon-binding domain	AD
3B	Smp_134950	EF-hand domain	Myosin Essential Chain {Bay scallop ( <i>Aequipecten irradians</i> ) [TaxId: 31199]}
3B	Smp_340430	Double-stranded RNA-binding domain	Endoribonuclease Dicer, RISC-loading complex subunit
3B	Smp_023660	U3 small nucleolar ribonucleoprotein complex, subunit Mpp10	Mpp10
3B	Smp_134940	GTP binding domain	YqeH GTPase
3B	Smp_134930	NA	NA
3B	Smp_023620	FERM domain	FERM, RhoGEF and pleckstrin domain
3B	Smp_134910	Protein kinase domain	Aurora-related kinase 1 (aurora-2) {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
3B	Smp_246270	ABC transporter-like	Iron-sulfur clusters transporter ATM1, mitochondrial
3B	Smp_023550	Armadillo	Sys-1 protein, Pop-1 8-residue peptide
3B	Smp_134880	Snurportin-1	35 6DT1_A DNA dC->dU-editing enzy 93.8 0.02 1.7E-07 54.0 0.0 202 99-310 152-399 (507)
3B	Smp_333660	Nucleosome assembly protein (NAP)	Nucleosome assembly protein, NAP {Baker's yeast ( <i>Saccharomyces cerevisiae</i> ) [TaxId: 4932]}
3B	Smp_134870	Zinc finger C2H2-type	Zinc finger protein 297b {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
3B	Smp_134860	HEAT repeat	Constant regulatory domain of protein phosphatase 2a, pr65alpha {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
3B	Smp_309430	RecF/RecN/SMC, N-terminal	Structural maintenance of chromosomes protein
3B	Smp_246290	SAC domain	Sac1, Vps74
3B	Smp_023470	Aminotransferase class V domain	Probable cysteine desulfurase SufS { <i>Synechocystis</i> sp. PCC 6803 [TaxId: 1148]}
3B	Smp_134820	G protein-coupled receptor, rhodopsin-like	Chimera of Proteinase-activated receptor 1
3B	Smp_326690	NA	NA
3B	Smp_134800	Protein kinase domain	Tyrosine-protein kinase ZAP-70 {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
3B	Smp_134790	TAFII28-like protein	TAF(II)28 {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
3B	Smp_342930	Zinc finger, RING-type	Not-4 N-terminal RING finger domain {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
3B	Smp_346950	NA	NA
3B	Smp_342920	NA	NA
3B	Smp_023360	Anticodon-binding	human histidyl-tRNA synthetase, cytoplasmic
3B	Smp_134720	Major facilitator superfamily domain	Melibiose carrier protein
3B	Smp_309500	Major facilitator superfamily domain	Melibiose carrier protein
3B	Smp_309510	Major facilitator superfamily domain	Melibiose carrier protein
3B	Smp_335880	Major facilitator superfamily domain	Melibiose carrier protein

## Annexe 6

3B	Smp_333670	Ionotropic glutamate receptor	N-methyl-D-aspartate receptor subunit NR1-8a, Ionotropic
3B	Smp_026880	Histone H3/CENP-A	Leishmania histone H3, Histone H4
3B	Smp_136710	P-type ATPase	Sarcoplasmic/endoplasmic reticulum calcium ATPase 1
3B	Smp_136730	Aspartic peptidase A1 family	Pepsin(ogen) {Pig ( <i>Sus scrofa</i> ) [TaxId: 9823]}
3B	Smp_026910	Lipase maturation factor	LMF1
3B	Smp_026920	GTP cyclohydrolase I	GTP cyclohydrolase I {Rat ( <i>Rattus norvegicus</i> ) [TaxId: 10116]}
3B	Smp_026930	Glycosyl transferase, family 13	N-acetylglucosaminyltransferase I {Rabbit ( <i>Oryctolagus cuniculus</i> ) [TaxId: 9986]}
3B	Smp_136740	ClpA/B family	ClpB, AAA+ modules { <i>Thermus thermophilus</i> [TaxId: 274]}
3B	Smp_213390	Protein kinase domain	Pkb kinase (Akt-2) {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
3B	Smp_213400	Zinc finger, RING-type	V(D)J recombination activating protein 1 (RAG1), dimerization domain {Mouse ( <i>Mus musculus</i> ) [TaxId: 10090]}
3B	Smp_136760	NA	hypothetical protein YugE { <i>Bacillus stearothermophilus</i> [TaxId: 1422]}
3B	Smp_136770	BCL7	BCL_N
3B	Smp_027000	Serine aminopeptidase, S33	Myristoyl-ACP-specific thioesterase { <i>Vibrio harveyi</i> [TaxId: 669]}
3B	Smp_027010	Clathrin adaptor, mu subunit	ADAPTOR-RELATED PROTEIN COMPLEX 2, ALPHA
3B	Smp_136780	Ubiquitin domain	Ubiquitin-like protein bab25500 (2010008E23Rik) {Mouse ( <i>Mus musculus</i> ) [TaxId: 10090]}
3B	Smp_027040	TB2/DP1/HVA22-related protein	TB2_DP1_HVA22
3B	Smp_333680	Phosphatidylinositol transfer protein	Phosphatidylinositol transfer protein, PITP {Rat ( <i>Rattus norvegicus</i> ) [TaxId: 10116]}
3B	Smp_027080	NA	YIF1
3B	Smp_336330	NA	Myelin_PLP
3B	Smp_335890	Aspartic peptidase A1 family	Pepsin(ogen) {Pig ( <i>Sus scrofa</i> ) [TaxId: 9823]}
3B	Smp_326700	NA	NA
3B	Smp_335900	Aspartic peptidase A1 family	Pepsin(ogen) {Human ( <i>Homo sapiens</i> ), progastricsin (pepsinogen C) [TaxId: 9606]}
3B	Smp_309540	Aspartic peptidase A1 family	Pepsin(ogen) {Human ( <i>Homo sapiens</i> ), progastricsin (pepsinogen C) [TaxId: 9606]}
3B	Smp_018800	Aspartic peptidase A1 family	Pepsin(ogen) {Pig ( <i>Sus scrofa</i> ) [TaxId: 9823]}
3B	Smp_132470	Aspartic peptidase A1 family	Pepsin(ogen) {Human ( <i>Homo sapiens</i> ), progastricsin (pepsinogen C) [TaxId: 9606]}
3B	Smp_132480	Aspartic peptidase A1 family	Pepsin(ogen) {Pig ( <i>Sus scrofa</i> ) [TaxId: 9823]}
3B	Smp_342650	NA	NA
3B	Smp_342660	Small GTPase	Rad {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
3B	Smp_342670	BRCT domain	Breast cancer associated protein, BRCA1 {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
3B	Smp_342680	Sema domain	Macrophage-stimulating protein receptor (E.C.2.7.10.1)

## Annexe 6

3B	Smp_018870	Domain of unknown function DUF4537	DUF4537
3B	Smp_018880	NA	NA
3B	Smp_342690	3-oxo-5-alpha-steroid 4- dehydrogenase, C-terminal	3-oxo-5-alpha-steroid 4-dehydrogenase 2 (E.C.1.3.1.22)
3B	Smp_214060	Phosphoglycerate kinase	Phosphoglycerate kinase {Baker's yeast ( <i>Saccharomyces cerevisiae</i> ) [TaxId: 4932]}
3B	Smp_132540	EGF-type aspartate/asparagine hydroxylation site	Protein toll, Protein spaetzle C-106
3B	Smp_018930	NA	NA
3B	Smp_018940	EF-hand domain	Myosin Essential Chain {Bay scallop ( <i>Aequipecten irradians</i> ) [TaxId: 31199]}
3B	Smp_018950	RNA recognition motif domain	Telomerase reverse transcriptase, Telomerase- associated protein
3B	Smp_132550	Centrosomal protein POC5	NA
4B	Smp_135500	3'5'-cyclic nucleotide phosphodiesterase, catalytic domain	cGMP-specific 3',5'-cyclic phosphodiesterase (E.C.3.1.4.35)
4B	Smp_135510	Glutamate carboxypeptidase 2-like	Glutamate carboxypeptidase III (E.C.3.4.17.21)
4B	Smp_340460	Beta-1,4-galactosyltransferase	beta 1,4 galactosyltransferase (b4GalT1) {Cow ( <i>Bos taurus</i> ) [TaxId: 9913]}
4B	Smp_024660	Immunoglobulin subtype 2	Roundabout homolog 2
4B	Smp_340470	Sirtuin family	NAD-dependent deacetylase sirtuin-6 (E.C.3.5.1.-)
4B	Smp_248070	Dedicator of cytokinesis, C-terminal	Dedicator of cytokinesis protein 10
4B	Smp_211020	Coagulation factor 5/8 C-terminal domain	C2 domain of factor V {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
4B	Smp_247870	NA	NA
4B	Smp_247860	Peptidase M8, leishmanolysin	Peptidase_M8
4B	Smp_247850	Peptidase M8, leishmanolysin	LEISHMANOLYSIN
4B	Smp_135540	NA	Guanine nucleotide-binding protein G(i) subunit
4B	Smp_135550	Leucine-rich repeat	Ribonuclease inhibitor {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
4B	Smp_135560	Fibrillar collagen, C-terminal	Collagen alpha-1(I) chain, Collagen alpha-2(I)
4B	Smp_135570	Actin family	RuvB-like helicase (E.C.3.6.4.12), Ino80, Uncharacterized
4B	Smp_334210	Zinc finger C2H2-type	Cohesin subunit SA-2, Double-strand-break repair
4B	Smp_342970	EGF-type aspartate/asparagine hydroxylation site	DUF4958
4B	Smp_213190	NA	DNA-directed RNA polymerase subunit, DNA- directed
4B	Smp_213200	Amidohydrolase-related	Amidohydrolase 2
4B	Smp_334220	Cadherin-like	Protocadherin-10
4B	Smp_024800	S-adenosyl-L-methionine-dependent methyltransferase	DUF2431
4B	Smp_340480	Tetratricopeptide-like helical domain superfamily	Mitochondrial import receptor subunit tom20-3 {Thale cress ( <i>Arabidopsis thaliana</i> ) [TaxId: 3702]}

## Annexe 6

4B	Smp_024820	Tetraspanin	Tetraspanin
4B	Smp_342980	The Golgi pH regulator/GPCR-type G protein	ABA_GPCR
4B	Smp_343790	Rab-GTPase-TBC domain	RabGAP/TBC protein
4B	Smp_145180	Zinc finger, A20-type	Rab5 GDP/GTP exchange factor, Ras-related
4B	Smp_145190	Protein kinase domain	Calmodulin-dependent protein kinase {Rat ( <i>Rattus norvegicus</i> ) [TaxId: 10116]}
4B	Smp_343800	Homeobox domain engrailed	Engrailed Homeodomain { <i>Drosophila melanogaster</i> [TaxId: 7227]}
4B	Smp_145210	Cadherin-like	Protocadherin-15
4B	Smp_042530	WD40 repeat	Tubulin beta, Tubulin alpha, PACRG
4B	Smp_244220	POU-specific domain	Pituitary-specific positive transcription factor 1/DNA
4B	Smp_244230	WD40 repeat	Platelet-activating factor acetylhydrolase IB subunit alpha {Mouse ( <i>Mus musculus</i> ) [TaxId: 10090]}
4B	Smp_042550	Phox homologous domain	Vacuolar protein sorting-associated protein 26-like
4B	Smp_042560	Porin domain superfamily	Mitochondrial import receptor subunit TOM40
4B	Smp_145230	Transcription factor, T-box	T-box protein 3, tbx3 {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
4B	Smp_042590	NADH-quinone oxidoreductase, chain I	NADH-quinone oxidoreductase chain 9, Nqo9 { <i>Thermus thermophilus</i> [TaxId: 274]}
4B	Smp_343810	GPCR, rhodopsin-like, 7TM	Adenosine receptor A2a, Soluble cytochrome b562, Adenosine
4B	Smp_145250	Cleft lip and palate transmembrane 1	CLPTM1
4B	Smp_042630	Methyltransferase-like	Phenylethanolamine N-methyltransferase, PNMTase {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
4B	Smp_343820	P-loop containing nucleoside triphosphate hydrolase	Apoptotic protease activating factor 1
4B	Smp_042670	DNA mismatch repair protein MutS, C-terminal	DNA mismatch repair protein Msh2/
4B	Smp_042680	Short-chain dehydrogenase/reductase SDR	Tropinone reductase {Jimsonweed ( <i>Datura stramonium</i> ), II [TaxId: 4076]}
4B	Smp_042700	Protein FAM81	Apolipoprotein A-IV
4B	Smp_145290	Alkaline phosphatase	Alkaline phosphatase {Northern shrimp ( <i>Pandalus borealis</i> ) [TaxId: 6703]}
4B	Smp_198170	NA	DUF1664
4B	Smp_307840	Alkaline phosphatase	Alkaline phosphatase {Northern shrimp ( <i>Pandalus borealis</i> ) [TaxId: 6703]}
4B	Smp_145300	Copper type II, ascorbate-dependent monooxygenase, N-terminal	Peptidyl-glycine alpha-amidating monooxygenase (E.C.1.14.17.3,4.3.2.5)
4B	Smp_329490	NA	NA
4B	Smp_042740	Armadillo	Karyopherin alpha {Baker's yeast ( <i>Saccharomyces cerevisiae</i> ) [TaxId: 4932]}
4B	Smp_145310	Reduced growth phenotype protein 1	Rgp1

## Annexe 6

4B	Smp_145320	SLC26A/SuIP transporter	Solute carrier family 26 member
4B	Smp_042770	Rab-GTPase-TBC domain	RabGAP/TBC protein
4B	Smp_042780	Mediator complex, subunit Med20	Mediator of RNA polymerase II
4B	Smp_042790	Glycosyltransferase 2-like	Dolichol monophosphate mannose synthase
4B	Smp_343830	IQ and ubiquitin-like domain-containing protein	Small ubiquitin-related modifier 1, Death
4B	Smp_042810	AAA ATPase domain	Replication factor C subunit 1
4B	Smp_145350	Ubiquitin-conjugating enzyme E2	Ubiquitin conjugating enzyme, UBC {Human ( <i>Homo sapiens</i> ), E2 H [TaxId: 9606]}
4B	Smp_202660	NA	NA
4B	Smp_329500	NA	PFL0690c
4B	Smp_145360	Mitochondria-eating protein	MIEAP
4B	Smp_329510	NA	Anti-silencing protein 1, ASF1 {Baker's yeast ( <i>Saccharomyces cerevisiae</i> ) [TaxId: 4932]}
4B	Smp_145370	NA	NA
4B	Smp_145380	NA	NA
4B	Smp_334230	C-5 cytosine methyltransferase	DNMT2 {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
4B	Smp_042910	Osteopetrosis-associated transmembrane protein 1 precursor	OSTMP1
4B	Smp_042920	Serine-threonine protein kinase 19	Stk19
4B	Smp_329520	Complex 1 LYR protein	Acyl carrier protein, mitochondrial, LYR
4B	Smp_145400	Armadillo-like helical	Rcd1
4B	Smp_303310	NA	NA
4B	Smp_145410	NA	DUF3338
4B	Smp_145420	Sema domain	Semaphorin-like protein 139, Plexin-C1
4B	Smp_042980	GHMP kinase, ATP-binding, conserved site	mevalonate kinase (E.C.2.7.1.36)
4B	Smp_042990	MINDY deubiquitinase	Protein FAM63A
4B	Smp_247830	NA	AF-4
4B	Smp_202680	Ragulator complex protein LAMTOR5	Hepatitis B virus X-interacting protein
4B	Smp_043030	Hexokinase	Hexokinase {Blood fluke ( <i>Schistosoma mansoni</i> ) [TaxId: 6183]}
4B	Smp_340750	NA	DYNC2H1 variant protein
4B	Smp_145440	Zinc finger C2H2-type	Zinc finger protein 568/DNA Complex
4B	Smp_343840	NA	DUF4659
4B	Smp_043070	Phosphorylated adapter RNA export protein, RNA-binding domain	ANKYRIN-REPEAT PROTEIN
4B	Smp_145460	NA	CD225
4B	Smp_145470	Zinc finger C2H2-type	Zinc finger protein 295
4B	Smp_043100	RNA recognition motif domain	Nuclear ribonucleoprotein A1 (RNP A1, UP1) {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
4B	Smp_043110	NA	Phage_AlpA
4B	Smp_043120	Universal stress protein A family	Putative ethylene-responsive protein AT3g01520/F4P13_7 {Thale cress ( <i>Arabidopsis thaliana</i> ) [TaxId: 3702]}



## Annexe 6

4B	Smp_202690	Universal stress protein A family	Putative ethylene-responsive protein AT3g01520/F4P13_7 {Thale cress ( <i>Arabidopsis thaliana</i> ) [TaxId: 3702]}
4B	Smp_043130	Gemin2/Brr1	Small nuclear ribonucleoprotein Sm D1
4B	Smp_043150	Calreticulin/calnexin	calnexin
4B	Smp_145490	K Homology domain	Far upstream binding element, FBP {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
4B	Smp_043170	Transforming acidic coiled-coil-containing protein, C-terminal	TACC_C
4B	Smp_343850	Formin, FH2 domain	Diaphanous protein homolog 1, dia1 {Mouse ( <i>Mus musculus</i> ) [TaxId: 10090]}
4B	Smp_043200	RIO kinase	40S ribosomal protein S27-A, 40S
4B	Smp_043210	Tetraspanin, EC2 domain superfamily	Tetraspanin
4B	Smp_043220	NA	BNIP3
4B	Smp_145510	Leucine-rich repeat	Internalin A { <i>Listeria monocytogenes</i> [TaxId: 1639]}
4B	Smp_043250	NA	GIY-YIG
4B	Smp_043260	G protein-coupled receptor, rhodopsin-like	7tm_1
4B	Smp_043270	G protein-coupled receptor, rhodopsin-like	7tm_1
4B	Smp_336140	G protein-coupled receptor, rhodopsin-like	7tm_1
4B	Smp_348040	G protein-coupled receptor, rhodopsin-like	7tm_1
4B	Smp_043290	G protein-coupled receptor, rhodopsin-like	7tm_1
4B	Smp_043300	G protein-coupled receptor, rhodopsin-like	7tm_1
4B	Smp_348050	G protein-coupled receptor, rhodopsin-like	7tm_1
4B	Smp_332070	G protein-coupled receptor, rhodopsin-like	human Chemokine Receptor 7
4B	Smp_043320	G protein-coupled receptor, rhodopsin-like	7tm_1
4B	Smp_043340	G protein-coupled receptor, rhodopsin-like	7tm_1
4B	Smp_043350	UDPGP family	UDP-N-acetylglucosamine pyrophosphorylase {Human ( <i>Homo sapiens</i> ), AGX1 [TaxId: 9606]}
4B	Smp_043360	NA	SJCHGC06286 protein, Bcl-2 homologous antagonist/killer
4B	Smp_145540	G protein-coupled receptor, rhodopsin-like	Muscarinic acetylcholine receptor M2, Vasopressin V2
4B	Smp_329530	NA	Myelin Oligodendrocyte Glycoprotein

## Annexe 6

4B	Smp_053700	Cytochrome b-c1 complex subunit 8	Ubiquinone-binding protein QP-C of cytochrome bc1 complex (Ubiquinol-cytochrome c reductase) {Cow ( <i>Bos taurus</i> ) [TaxId: 9913]}
4B	Smp_150890	IBR domain	E3 ubiquitin-protein ligase ARIH1 (E.C.2.3.2.-)
4B	Smp_247810	NA	NA
4B	Smp_150870	Leucine-rich repeat	Rab geranylgeranyltransferase alpha-subunit, C-terminal domain {Rat ( <i>Rattus norvegicus</i> ) [TaxId: 10116]}
4B	Smp_329550	NA	NA
4B	Smp_053640	NA	NA
4B	Smp_150860	Protein kinase domain	Cyclin-dependent PK, CDK6 {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
4B	Smp_053610	Aminoacyl-tRNA synthetase, class II (G/ P/ S/T)	Seryl-tRNA synthetase, cytoplasmic (E.C.6.1.1.11)
4B	Smp_150850	Methyl-CpG DNA binding	Histone-lysine N-methyltransferase SETDB1 (E.C.2.1.1.43), Histone
4B	Smp_150840	Zinc finger, C2H2, LYAR-type	60S ribosomal protein L8, 60S
4B	Smp_313300	Small GTPase	Rab21 {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
4B	Smp_313310	EF-hand domain	Sarcoplasmic calcium-binding protein {Amphioxus ( <i>Branchiostoma lanceolatum</i> ) [TaxId: 7740]}
4B	Smp_053560	Protein kinase domain	MAP kinase activated protein kinase 2, mapkap2 {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
4B	Smp_053550	U6 snRNA phosphodiesterase Usb1	U6 snRNA phosphodiesterase (E.C.3.1.4.-)
4B	Smp_213500	HotDog domain superfamily	Acyl-CoA thioesterase
4B	Smp_213510	NA	FA_FANCE
4B	Smp_247800	Homeobox domain	Homeodomain-only protein, Hop {Mouse ( <i>Mus musculus</i> ) [TaxId: 10090]}
4B	Smp_053510	Aspartyl/Asparaginyl-tRNA synthetase, class IIb	Aspartate--tRNA ligase, cytoplasmic (E.C.6.1.1.12)
4B	Smp_313320	NA	Med28
4B	Smp_313330	NA	NA
4B	Smp_329560	NA	conserved hypothetical protein
4B	Smp_344330	Cysteine-rich flanking region, C-terminal	von Willebrand factor binding domain of glycoprotein Ib alpha {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
4B	Smp_334250	Epithelial sodium channel	Acid-sensing ion channel 1
4B	Smp_153890	Post-GPI attachment to proteins factor 2	Frag1
4B	Smp_030760	Flavodoxin-like	NADPH--cytochrome P450 reductase (E.C.1.6.2.4)
4B	Smp_138810	Frag1/DRAM/Sfk1	Frag1
4B	Smp_334260	Zinc finger, PHD-type	Protein AF-10
4B	Smp_153870	Protein kinase domain	Aurora-related kinase 1 (aurora-2) {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
4B	Smp_058990	Meiotic nuclear division protein 1	Mnd1, Putative tbpip family protein
4B	Smp_058980	Oxysterol-binding protein	KLLA0C04147p

## Annexe 6

4B	Smp_058970	Succinate dehydrogenase assembly factor 4	DUF1674
4B	Smp_337560	NA	DNA damage-binding protein 1, DDB1-
4B	Smp_344600	Zinc finger C2H2-type	Transcriptional repressor CTCF/DNA Complex
4B	Smp_153840	BEACH domain	Lipopolysaccharide-responsive and beige-like anchor protein
4B	Smp_153820	SNF2-related, N-terminal domain	DNA-directed RNA polymerase II subunit
4B	Smp_058880	Serine hydrolase FSH	FSH1
4B	Smp_329570	Nicotinamide riboside kinase	Uridine-cytidine kinase 2 {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
4B	Smp_153810	Ankyrin repeat	Asparaginase type II { <i>Vibrio cholerae</i> [TaxId: 666]}
4B	Smp_058850	RAD51 interacting motif	RAD51_interact
4B	Smp_059090	NA	cAMP-dependent protein kinase type II-alpha
4B	Smp_153790	Ribosomal protein L4/L1e	ribosomal protein uL2, ribosomal protein
4B	Smp_059100	PRELI/MSF1 domain	Protein UPS1, mitochondrial, Mitochondrial distribution and
4B	Smp_153920	Cyclin, N-terminal	G1/S-specific cyclin-E1 {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
4B	Smp_153930	Peptidase M8, leishmanolysin	LEISHMANOLYSIN
4B	Smp_059130	CAP Gly-rich domain	Restin {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
4B	Smp_059140	NA	NDUF_B12
4B	Smp_329580	Immunoglobulin-like fold	SIP2 { <i>Saccharomyces cerevisiae</i> [TaxId: 4932]}
4B	Smp_059150	Mitochondrial inner membrane protein Mitofilin	Mitofilin
4B	Smp_153950	Ribosomal protein L24e-related	60S ribosomal protein L2-A, 60S
4B	Smp_059170	Troponin	Troponin C/Troponin T/Troponin I
4B	Smp_153960	Peptidase A22A, presenilin	Presenilin
4B	Smp_153970	JNK/Rab-associated protein-1, N-terminal	RILP
4B	Smp_059200	NA	RIFIN
4B	Smp_059210	Periphilin-1	Periphilin-1, Protein TASOR
4B	Smp_153990	Zinc finger, LIM-type	DNA endonuclease RBBP8, LIM domain
4B	Smp_126460	Protein kinase domain	Serine/threonine-protein kinase Nek2 {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
4B	Smp_126450	ABC transporter-like	ATP-binding cassette sub-family G member
4B	Smp_007970	Ribosome maturation protein Sdo1/SBDS	60S RIBOSOMAL PROTEIN L3, 60S
4B	Smp_007960	Glyoxalase/fosfomycin resistance/dioxygenase domain	4-hydroxyphenylpyruvate dioxygenase, HppD { <i>Streptomyces avermitilis</i> [TaxId: 33903]}
4B	Smp_340160	Beta-1,4-galactosyltransferase	beta 1,4 galactosyltransferase (b4GalT1) {Cow ( <i>Bos taurus</i> ) [TaxId: 9913]}
4B	Smp_071000	DSS1/SEM1	Activated Human 26S Proteasome
4B	Smp_160700	PWWP domain	Histone-lysine N-methyltransferase SETD2 (E.C.2.1.1.43,2.1.1.-), H3.3S31phK36M(29-42)

## Annexe 6

4B	Smp_345220	Nucleic acid-binding, OB-fold	Breast Cancer type 2 susceptibility
4B	Smp_160670	Homeobox domain	Paired protein (prd) {Fruit fly ( <i>Drosophila melanogaster</i> ) [TaxId: 7227]}
4B	Smp_347040	Protein of unknown function DUF2045	DUF2045
4B	Smp_070930	Proteasome alpha-subunit, N-terminal domain	Proteasome alpha subunit (non-catalytic) {Cow ( <i>Bos taurus</i> ) [TaxId: 9913]}
4B	Smp_310920	Forkhead-associated (FHA) domain	Suppressor of IKBKE 1, Sarcolemmal
4B	Smp_212750	G protein-coupled receptor, rhodopsin-like	Lysophosphatidic acid receptor 6a, Endolysin, receptor (E.C.3.2.1.17)
4B	Smp_212740	SMAD domain, Dwarfing-type	Smad3 MH2 domain {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
4B	Smp_337220	AAA ATPase domain	DYNC2H1 variant protein
4B	Smp_070860	Malate dehydrogenase, type 2	MALATE DEHYDROGENASE
4B	Smp_341800	Leucine-rich repeat	von Willebrand factor binding domain of glycoprotein Ib alpha {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
4B	Smp_347030	Dbl homology (DH) domain	GEF of intersectin {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
4B	Smp_347560	DNA recombination and repair protein Rad51-like, C-terminal	DNA REPAIR AND RECOMBINATION PROTEIN
4B	Smp_313450	Peptidase M12B, ADAM/reprolysin	Disintegrin and metalloproteinase domain-containing protein
4B	Smp_348000	NA	NA
4B	Smp_348010	NA	NA
4B	Smp_348020	NA	NA
4B	Smp_348030	NA	NA
4B	Smp_313480	NA	NA
4B	Smp_190540	NA	Coatomer subunit alpha, Coatomer subunit
4B	Smp_303070	Peptidase M8, leishmanolysin	Peptidase_M8
4B	Smp_303760	Peptidase M8, leishmanolysin	Peptidase_M8
4B	Smp_204980	NA	Cyclophilin-like protein, putative, Dipeptide alapro
4B	Smp_093840	NA	Mtp
4B	Smp_346220	NA	NA
4B	Smp_093850	RNA recognition motif domain	Splicing factor 3B subunit 4 {Human ( <i>Homo sapiens</i> ) [TaxId: 9606]}
4B	Smp_167890	Peptide methionine sulphoxide reductase MrsB	C-terminal MsrB domain of peptide methionine sulfoxide reductase PilB { <i>Neisseria gonorrhoeae</i> [TaxId: 485]}
4B	Smp_313490	NA	Transforming protein RhoA, Rho-associated, coiled-coil
4B	Smp_313500	NA	NA
4B	Smp_174580	Legume-like lectin	Emp46p N-terminal domain {Baker's yeast ( <i>Saccharomyces cerevisiae</i> ) [TaxId: 4932]}
4B	Smp_341620	FERM domain	FERM, RhoGEF and pleckstrin domain
4B	Smp_095630	Tetraspanin	CD81 antigen

## Annexe 6

4B	Smp_174610	Zinc finger, LIM-type	Fusion protein of Lmo4 protein
4B	Smp_095650	Selenophosphate synthetase	Putative selenophosphate synthetase (E.C.2.7.9.3)
4B	Smp_332080	NA	28S ribosomal protein S2, mitochondrial
4B	Smp_095670	Zinc finger, CCCH-type	Protein CFT1, mRNA 3'-end-processing protein
4B	Smp_095680	Matrin/U1-C, C2H2-type zinc finger	Pre-mRNA-processing-splicing factor 8, 116 kDa
4B	Smp_174630	Tetraspanin	CD81 antigen
4B	Smp_334280	Sterile alpha motif domain	Stromal interaction molecule 2

Chr. : chromosome.





# Annexe 7 : Article méthodologique sur la technique de ChIPmentation sur échantillons *S. mansoni*



METHOD ARTICLE

## Automated ChIPmentation procedure on limited biological material of the human blood fluke *Schistosoma mansoni*

[version 1; peer review: 1 approved with reservations]

Chrystelle Lasica<sup>1</sup>, Ronaldo de Carvalho Augusto <sup>1,2</sup>, H  l  ne Mon   <sup>1</sup>, Gabriel Mouahid<sup>1</sup>, Cristian Chaparro <sup>1</sup>, Anne-Cl  mence Veillard<sup>3</sup>, Agnieszka Zelisko-Schmidt<sup>3</sup>, Christoph Grunau <sup>1</sup>

<sup>1</sup>IHPE, Univ Perpignan Via Domitia, CNRS, IFREMER, Univ Montpellier, Perpignan, 66860, France

<sup>2</sup>LBMC, Laboratoire de Biologie et Mod  lisation de la Cellule Univ Lyon, ENS de Lyon, Universit   Claude Bernard Lyon 1, CNRS, UMR 5239, INSERM, U1210., Lyon, 69000, France

<sup>3</sup>Diagenode, Seraing, 4100, Belgium

**V1** First published: 11 Apr 2022, 7:133  
<https://doi.org/10.12688/wellcomeopenres.17779.1>  
 Latest published: 11 Apr 2022, 7:133  
<https://doi.org/10.12688/wellcomeopenres.17779.1>

### Abstract

In living cells, the genetic information stored in the DNA sequence is always associated with chromosomal and extra-chromosomal epigenetic information. Chromatin is formed by the DNA and associated proteins, in particular histones. Covalent histone modifications are important bearers of epigenetic information and as such have been increasingly studied since about the year 2000. One of the principal techniques to gather information about the association between DNA and modified histones is chromatin immunoprecipitation (ChIP), also combined with massive sequencing (ChIP-Seq). Automated ChIPmentation procedure is a convenient alternative to native chromatin immunoprecipitation (N-ChIP). It is now routinely used for ChIP-Seq in many model species, using in general roughly  $10^6$  cells per experiment. Such high cell numbers are sometimes difficult to produce. Using the human parasite *Schistosoma mansoni*, whose production requires sacrificing animals and should therefore be kept to a minimum, we show here that automated ChIPmentation is suitable for limited biological material. We define the operational limit as  $\geq 20,000$  *Schistosoma* cells. We also present a streamlined protocol for the preparation of ChIP input libraries.

### Keywords

ChIPmentation, ChIP-seq, N-ChIP, limited biological material, Epigenetics, *Schistosoma mansoni*

### Open Peer Review

#### Approval Status ?


1

#### version 1

11 Apr 2022



view

1. **Nolwenn M. Dheilly** , ENVA, Animal health laboratory, 14 rue Pierre et Marie Curie, France

Any reports and responses or comments on the article can be found at the end of the article.

**Corresponding author:** Christoph Grunau ([christoph.grunau@univ-perp.fr](mailto:christoph.grunau@univ-perp.fr))

**Author roles:** **Lasica C:** Data Curation, Formal Analysis, Investigation, Methodology, Validation, Writing – Original Draft Preparation, Writing – Review & Editing; **de Carvalho Augusto R:** Investigation, Methodology, Writing – Review & Editing; **Moné H:** Conceptualization, Funding Acquisition, Investigation, Project Administration, Supervision, Writing – Review & Editing; **Mouahid G:** Conceptualization, Funding Acquisition, Investigation, Project Administration, Supervision, Writing – Review & Editing; **Chaparro C:** Data Curation, Methodology, Resources, Software, Supervision, Writing – Review & Editing; **Veillard AC:** Methodology, Supervision, Visualization, Writing – Review & Editing; **Zelisko-Schmidt A:** Conceptualization, Methodology, Supervision, Visualization, Writing – Review & Editing; **Grunau C:** Conceptualization, Formal Analysis, Funding Acquisition, Investigation, Methodology, Project Administration, Supervision, Writing – Original Draft Preparation, Writing – Review & Editing

**Competing interests:** Diagenode is the company which commercialized the ChIPmentation technology. ACV and AZS are employees of Diagenode.

**Grant information:** This work was supported by Wellcome [107475/Z/15/Z]. This research was also funded by the French National Agency for Research (ANR) [grant ANR-17-CE12-0005-01] CHRONOGET and the CNRS (French National Centre for Research). CL is a PhD candidate who receives funding from the Occitanie Region (France).

*The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.*

**Copyright:** © 2022 Lasica C *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**How to cite this article:** Lasica C, de Carvalho Augusto R, Moné H *et al.* **Automated ChIPmentation procedure on limited biological material of the human blood fluke *Schistosoma mansoni* [version 1; peer review: 1 approved with reservations]** Wellcome Open Research 2022, 7:133 <https://doi.org/10.12688/wellcomeopenres.17779.1>

**First published:** 11 Apr 2022, 7:133 <https://doi.org/10.12688/wellcomeopenres.17779.1>

## Introduction

*Schistosoma mansoni* is a human parasite with a complex life cycle that shows strong developmental phenotypic plasticity, with intra-molluscal and intra-vertebrate stages, and two free-swimming larval stages (miracidium and cercariae). We had shown by native chromatin immunoprecipitation (N-ChIP) that the different life cycle stages also show strong histone modification plasticity (Augusto *et al.*, 2019; Cosseau *et al.*, 2009; Roquis *et al.*, 2018). While N-ChIP has been successfully used, we found that it is associated with two challenges: one is the high hands-on time with the N-ChIP, and the other is obtaining enough biological material to perform several ChIP experiments with different antibodies. We therefore explored and benchmarked an automated ChIP procedure (Figure 1). ChIPmentation is a ChIP-sequencing (ChIP-seq) technology which uses a transposase to add the sequencing adaptors to the DNA of interest instead of the classical multi-step processing, including end repair, A-tailing, adaptor ligation and size-selection (Schmidl *et al.*, 2015). Thanks to the action of the transposase, loaded with sequencing adaptors, the library preparation is performed in only one step, which reduces hands-on time and material loss. Moreover, in the ChIPmentation approach, this tagmentation process is performed directly on chromatin during the immunoprecipitation process instead of naked DNA after purification. This workflow allows for a more reproducible tagmentation.

The combined facts that ChIPmentation has been automated on Diagenode's IP-Star Compact Automated System and that this technology has been validated on low amounts of human cells (Roels *et al.*, 2020; Schmidl *et al.*, 2015) make it a perfect candidate for ChIP-seq on limited material of other, non-model species. Here we addressed the questions of whether ChIPmentation, which was originally developed for human cell cultures, (i) can be used with schistosomes, (ii) whether it can be automated on a pipetting robot, and finally (iii) what the lowest schistosome cell number would be to obtain robust results with this procedure. We show here that the method is almost as sensitive as N-ChIP, but is about two times faster and can be carried out on the IP-Star pipetting robot, reducing experimenter hands-on and, more importantly, training time.

## Methods

### Production of biological material

*S. mansoni* NMRI eggs were extracted from livers of golden hamsters (donated by ParaDev) 42 days post-infection. Nocturnal *S. mansoni* from Oman (Mouahid *et al.*, 2012) were extracted from livers of two Swiss OF1 mice eight months post-infection. Miracidia were allowed to hatch for two hours in spring water, collected by pipetting and sedimented on ice for 30 min. Miracidia were counted under a microscope, aliquoted and stored at -80°C.

### Ethical considerations

**Experiments on animals permits** The laboratory received the permit number A 66040 for experiments on animals, from both the French *Ministère de l'Agriculture et de la Pêche* (Ministry of Agriculture and Fisheries) and the French *Ministère de l'Éducation Nationale de la Recherche et de la Technologie* (Ministry of Education, Research and Technology, Décret n°

2001-464 du 29 mai 2001 modifiant le décret n° 87-848 du 19 octobre 1987). This includes housing, breeding and care of the mice and hamsters, and animal experimentation. HM holds the official certificate for animal experimentation N° C661101 delivered by the *Direction Départementale de la Protection des Populations* (Articles R 214-87 à R 214-122 du Code Rural et article R 215-10 ; Arrêté du 19 avril 1988).

**Ethical approval number:** we obtained approval from the CEEA - 036 *Comité d'éthique en expérimentation animale Languedoc Roussillon* (CEEA-LR)", which is the registration code of our ethical committee within the *Comité national de réflexion éthique sur l'expérimentation animale* (CNREEA), under the agreement number C66-136-01. The CNREEA is part of the French Ministry of Higher Education, Research and Innovation.

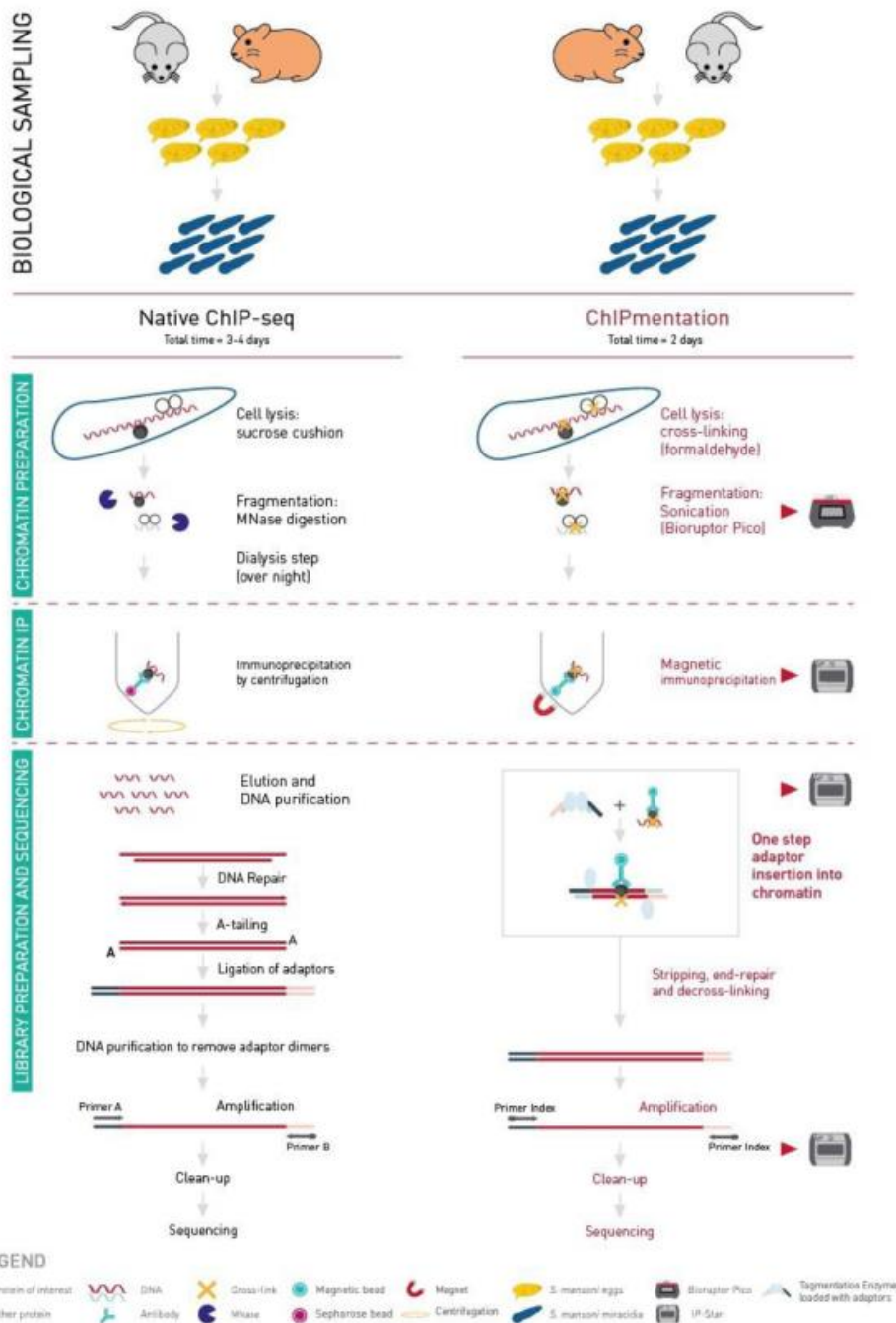
### Cell lysis and chromatin shearing

Chromatin preparation was performed using the Diagenode ChIPmentation Kit for Histones, Cat. No. C01011009 and protocol with minor modifications. A total of 10,000 miracidia (1,000,000 cells based on the observation that one miracidium is composed of 100–120 cells) were resuspended in 1 mL 1x Hank's balanced salt solution (HBSS), split into 2x500 µL and crushed with a plastic pestle in an Eppendorf tube on ice during ~1 min. For cross-linking, 13.5 µL of formaldehyde were added and tubes were incubated for 10 min at room temperature with occasional inversion. To stop cross-linking fixation, 57 µL of glycine were added and samples were incubated for 5 min at room temperature. Samples were centrifuged at 500xg, at 4°C, for 5 min. The pellet was resuspended in 2x1 mL of ice-cold Lysis Buffer iL1, combined and homogenized in a Dounce (pestle A) on ice for 5 min. After another centrifugation (500g, 5 min, 4°C), the supernatant was discarded and the pellet was resuspended in 1 mL of ice-cold Lysis Buffer II.2 and centrifuged (500xg, 5 min, 4°C). The supernatant was discarded and the pellet was resuspended in 100 µL of complete Shearing Buffer iS1 for each tube. Samples were sonicated with the Bioruptor Pico (Diagenode, Cat. No. B01080010) for 5 cycles (30 s ON and 30 s OFF). After transfer into new tubes, samples were centrifuged (16,000g, 10 min, 4°C). The supernatants were transferred into a new single tube (200 µL total) and 20 µL iS1 were added, yielding a total volume of 220 µL. The procedure was done in duplicate (named L and R in the following). Serial dilutions were done in iS1 to produce 100 µL equivalents of 10,000 miracidia (10<sup>6</sup> cells), 1,000 miracidia (10<sup>5</sup> cells), 100 miracidia (10<sup>4</sup> cells), 50 miracidia (5,000 cells), 10 miracidia (10<sup>3</sup> cells), five miracidia (500 cells) and one miracidium (100 cells) or 100 µL iS1 as negative control.

### Magnetic immunoprecipitation and tagmentation

Immunoprecipitation (IP) was performed on the Diagenode IP-Star Compact Automated System (Cat. No. B03000002) according to the ChIPmentation Kit for Histones User Guide and by following the manufacturer's on-screen instructions. Antibody (Ab) coating time was set to 3 h, IP reaction to 13 h, washes to 10 min, and tagmentation to 5 min. For each sample the Ab coating mix was done with 4 µL anti-H3K4me3 (Diagenode, Cat. No. C15410003; mixture of lot A1051D and A1052D; raised in rabbit).





**Figure 1. Comparison of Native-ChIP and ChIPmentation workflows.** Biological sampling time depends on the biological model used. Native-ChIP protocol lasts three to four days. A sucrose cushion is used for cell lysis and MNase digestion for the fragmentation step. The immunoprecipitation is done by centrifugation. The whole process is done manually. ChIPmentation protocol lasts two days. Cross-linking is used for cell lysis and sonication for the fragmentation step. The immunoprecipitation, tagmentation and library cleaning are done with the IP-Star.

Stripping, end repair and reverse cross-linking were done as indicated in the User Guide.

#### Input library tagmentation

In the ChIPmentation Kit for Histones (Diagenode, Cat. No. C01011009) the suggested strategy is to sequence one immunoprecipitated sample with a control immunoglobulin G (IgG) and to use it for sequencing normalization instead of the traditional input, which cannot be treated in exactly the same way as the immunoprecipitated samples. However, IgG are negative control samples, and the generation of such samples in low-amount approaches involves in our experience the use of a high number of amplification cycles that can induce some biases. A protocol for the tagmentation of the input sample was therefore set-up as follows.

For each immunoprecipitated sample, 1  $\mu$ L of sheared chromatin was kept aside before IP in the IP-Star. 1  $\mu$ L of MgCl<sub>2</sub>

(Diagenode ChIPmentation kit for Histones, Cat. No. C01011009), 8  $\mu$ L of molecular biology grade water, 10  $\mu$ L 2xTagment DNA buffer and 1  $\mu$ L of 100-fold in molecular-grade water diluted DNA tagmentation enzyme (Illumina 20034197, lot 20464427) were added to each 1  $\mu$ L input. The tagmentation reaction was performed in a thermocycler for five minutes at 55°C. Then, 25  $\mu$ L of 2xPCR NEB master mix (New England Biolabs M0541L, lot 10067165) was added to each input. The end-repair and de-cross-link were performed in a thermocycler for five minutes at 72°C followed by 10 minutes at 95°C. An aliquot of 2  $\mu$ L was taken from each input and added to 8  $\mu$ L of quantification mix. For each reaction, this quantification mix was composed of 0.3  $\mu$ L of forward and reverse ATAC-seq primers (25  $\mu$ M) (Table 1, (Buenrostro *et al.*, 2015)), 1  $\mu$ L SYBR Green 10X (Diagenode kit), 1.3  $\mu$ L of molecular biology grade water and 5  $\mu$ L of 2xPCR NEB master mix. While not formally tested, leftover primers of the ChIPmentation kit could probably also be used, but the volume must be adjusted

**Table 1. ATAC-seq primer indexes and sequences (Buenrostro *et al.*, 2015).**

Index ID	Sequence
Ad1_noMX:	AATGATACGGCGACCACCGAGATCTACACTCGTCGGCAGCGTCAGATGTG
Ad2.1_TAAGCGCA	CAAGCAGAAGACGGCATAACGAGATTCGCCTTAGTCTCGTGGGCTCGGAGATGT
Ad2.2_CGTAAGTAG	CAAGCAGAAGACGGCATAACGAGATCTAGTACGGTCTCGTGGGCTCGGAGATGT
Ad2.3_AGGCAGAA	CAAGCAGAAGACGGCATAACGAGATTTCTGCCTGTCTCGTGGGCTCGGAGATGT
Ad2.4_TCCTGAGC	CAAGCAGAAGACGGCATAACGAGATGCTCAGGAGTCTCGTGGGCTCGGAGATGT
Ad2.5_GGACTCCT	CAAGCAGAAGACGGCATAACGAGATAGGAGTCCGTCTCGTGGGCTCGGAGATGT
Ad2.6_TAGGCATG	CAAGCAGAAGACGGCATAACGAGATCATGCCTAGTCTCGTGGGCTCGGAGATGT
Ad2.7_CTCTCTAC	CAAGCAGAAGACGGCATAACGAGATGTAGAGAGGTCTCGTGGGCTCGGAGATGT
Ad2.8_CAGAGAGG	CAAGCAGAAGACGGCATAACGAGATCCTCTCTGGTCTCGTGGGCTCGGAGATGT
Ad2.9_GCTACGCT	CAAGCAGAAGACGGCATAACGAGATAGCGTAGCGTCTCGTGGGCTCGGAGATGT
Ad2.10_CGAGGCTG	CAAGCAGAAGACGGCATAACGAGATCAGCCTCGGTCTCGTGGGCTCGGAGATGT
Ad2.11_AAGAGGCA	CAAGCAGAAGACGGCATAACGAGATTGCCTCTTGTCTCGTGGGCTCGGAGATGT
Ad2.12_GTAGAGGA	CAAGCAGAAGACGGCATAACGAGATTCCTCTACGTCTCGTGGGCTCGGAGATGT
Ad2.13_GTCGTGAT	CAAGCAGAAGACGGCATAACGAGATATCACGACGTCTCGTGGGCTCGGAGATGT
Ad2.14_ACCACTGT	CAAGCAGAAGACGGCATAACGAGATACAGTGGTGTCTCGTGGGCTCGGAGATGT
Ad2.15_TGGATCTG	CAAGCAGAAGACGGCATAACGAGATCAGATCCAGTCTCGTGGGCTCGGAGATGT
Ad2.16_CCGTTTGT	CAAGCAGAAGACGGCATAACGAGATACAAACGGGTCTCGTGGGCTCGGAGATGT
Ad2.17_TGCTGGGT	CAAGCAGAAGACGGCATAACGAGATACCCAGCAGTCTCGTGGGCTCGGAGATGT
Ad2.18_GAGGGGTT	CAAGCAGAAGACGGCATAACGAGATAACCCCTCGTCTCGTGGGCTCGGAGATGT
Ad2.19_AGTTGGG	CAAGCAGAAGACGGCATAACGAGATCCCAACCTGTCTCGTGGGCTCGGAGATGT
Ad2.20_GTGTGGTG	CAAGCAGAAGACGGCATAACGAGATCACCACAGTCTCGTGGGCTCGGAGATGT
Ad2.21_TGGGTTTC	CAAGCAGAAGACGGCATAACGAGATGAAACCCAGTCTCGTGGGCTCGGAGATGT
Ad2.22_TGGTCACA	CAAGCAGAAGACGGCATAACGAGATTGTGACCAGTCTCGTGGGCTCGGAGATGT
Ad2.23_TTGACCCT	CAAGCAGAAGACGGCATAACGAGATAGGGTCAAGTCTCGTGGGCTCGGAGATGT
Ad2.24_CCACTCCT	CAAGCAGAAGACGGCATAACGAGATAGGAGTGGGTCTCGTGGGCTCGGAGATGT

as the primer pairs in the kit were 10  $\mu$ M instead of 25  $\mu$ M. Library amplification, purification, quality checking and sequencing steps were performed as for the immunoprecipitated samples (see below).

#### Library amplification

To determine optimal number of library amplification cycles we proceeded as described in steps 5.1 to 5.5 of the ChIPmentation User Guide with modifications detailed below. We determined the number of amplification cycles for each library by using the number of cycles that corresponded to 1/3 of the qPCR amplification curve slope during the exponential phase. Results are shown in [Table 6](#).

Amplification was done in step 5.7. After PCR, 48  $\mu$ l of library amplification mix were AMPure-purified on the IP-Star using 86  $\mu$ l of AMPure beads (1.8x). Instead of a resuspension buffer, we used ChIP grade water for a final volume of 20  $\mu$ l.

#### Library check and sequencing

Library fragment size and concentration was checked on an Agilent 2100 Bioanalyzer with a High Sensitivity DNA Assay v1.03. Paired-end sequencing (2x75 cycles) was performed at the Bio-Environnement platform (University of Perpignan, France) on a NextSeq 550 instrument (Illumina, USA).

#### Bioinformatics analysis

Reads were quality-checked with FastQC, Galaxy version 0.72 and Galaxy 1 ([Smith, 2019](#)). Adapters were detected in less than 4% of reads. Reads were aligned with the *S. mansoni* v7 reference genomes using [Bowtie2](#) Galaxy version 2.3.4.3 ([Langmead & Salzberg, 2012](#)), the default sensitive settings.

Uniquely aligned reads were filtered from the BAM files using the XS: tag of [Bowtie2](#), Galaxy version 2.3.4.3 and Galaxy 0 ([Langmead & Salzberg, 2012](#)). PCR duplicates were removed with [samtools rmdup](#) Galaxy version 2.0.1 ([Li et al., 2009](#)). BAM files were subsampled to 3.8 M uniquely aligned reads with [Picard DownsamplSam](#) Galaxy version 2.18.2.1 (Broad Institute 2022, February 23). Peakcalling was done with [Peakranger](#) (v1.17) Galaxy version 1.0.0 ([Feng et al., 2011](#)), with P value cut-off: 0.0001, false discovery rate (FDR) cut-off: 0.05, Read extension length 100 - 2,000 bp, Smoothing bandwidth: 99, Delta: 0.8, Detection mode: region. [ChromstaR](#), Galaxy version 0.99.0 ([Taudt et al., 2016](#)) was used with a bin size same as [Peakranger](#) extension lengths and a step size of half a bin size. *Miracidia* genomic DNA libraries served as input. In [ChromstaR](#) postprocessing, the maximum posterior probabilities to adjust sensitivity of peak detection was set to 0.999. This kept broad peaks intact. We reasoned that gaps between peaks that were larger than a nucleosome are not biologically meaningful and peaks were merged with [BEDTOOLS](#), Galaxy version 2.29.2 ([Quinlan & Hall, 2010](#)) when they were  $\leq 150$  bp apart (the average length of DNA in a nucleosome).

Enrichment plots over metagenes were produced over 5,073 genes on the positive strand based on the canonical annotation v7 of *S. mansoni*.

## Results

ChIPmentation has a sensitivity that is comparable to N-ChIP

We used *S. mansoni* miracidia which are composed of 100–120 cells as starting material, allowing for a good estimate of cell numbers. After the chromatin fragmentation and library amplification,  $10^6$  to  $10^4$  cell equivalents gave comparable Bioanalyzer profiles with peaks around 1kb. For 5,000 cells equivalents and below, fragments of smaller size became clearly visible. No high-molecular fragments were observed in the negative control without chromatin ([Figure 2](#) and [Table 2](#)).

Alignments to the genome gave expected results (~50% uniquely aligned reads) for  $10^6$  to  $10^4$  cell equivalents, but dropped to ~35% with  $10^3$  cells equivalent and were <20% for 100 cells equivalent. No contaminating DNA was detected ([Table 3](#)).

In order to compare ChIPmentation results to N-ChIP we re-analysed earlier data obtained by N-ChIP ([Augusto et al., 2019](#)) using the same data-cleaning, alignment and peak-calling parameters as for ChIPmentation ([Table 4](#)). For ChIPmentation, peakcalling with [Peakranger](#) was highly robust for  $10^6$  cells and delivered the expected values (based on earlier N-ChIP results). Below this cell equivalent, peak-calling became dependent on bin size ([Table 5](#)).

We iteratively identified 1,000 bp as best-read extension length. Using the HMM-based [ChromstaR](#) improved peak calling for ChIPmentation  $10^4$  cell equivalent, but not for  $10^3$  or 100 cells. We obtained comparable results for ChIPmentation of  $10^6$ ,  $10^4$  cells and N-ChIP ~ $10^6$ ,  $10^4$  and  $10^3$  cells ([Table 6](#)).

[ChromstaR](#) metagene profiles showed consistent profiles for ChIPmentation  $10^6$ ,  $10^4$  and  $10^3$  cells, and all N-ChIP, but not for ChIPmentation on 100 cell equivalents ([Figure 3](#)).

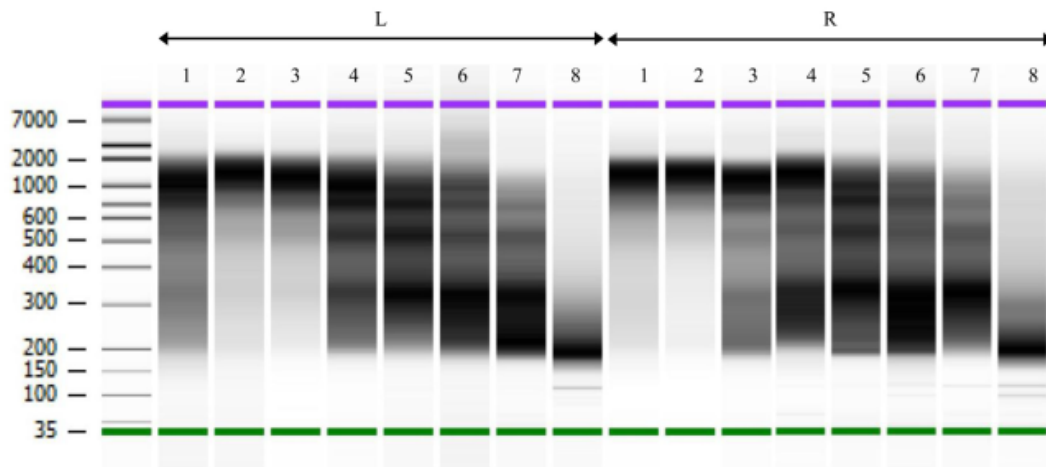
[ChromstaR](#) allows for estimating the correlation of chromatin profiles based on read counts ([Figure 4](#)), and indicates high correlations between ChIPmentation  $10^6$ , and N-ChIP ~ $10^6$ ,  $10^4$ .

All other overall chromatin profiles were below 0.75 correlation coefficient. This is surprising given the high similarity of metagene profiles. Visual inspection of peaks and profiles showed that peaks were actually correctly identified by [ChromstaR](#) (but much less by [Peakranger](#)) in ChIPmentation until  $10^4$  cells, but there was higher background than in N-ChIP which probably decreased the correlation for lower cell equivalents ([Figure 5](#)).

[BEDTools](#) intersect with default parameters identified 5,948 common peaks between the 6,682 ChIPmentation peaks and the 6,186 N-ChIP peaks obtained with  $10^6$  cells. This corresponds to an empirical FDR of 0.89–0.96%. This FDR cannot distinguish between biological and technical variation.

ChIPmentation input library can rapidly be produced in parallel to the automated procedure

After having formally established that automated ChIPmentation had a comparable sensitivity to our routine N-ChIP procedure, we aimed to identify the optimal way to produce input libraries



**Figure 2.** ChIPmentation library profiles for two technical replicates (L and R). Electrophoresis performed on BioAnalyser. Lanes 1–8 from left to right:  $10^6$ ,  $10^5$ ,  $10^4$ ,  $5 \times 10^3$ ,  $10^3$ , 500, 100, and 0 cells. Libraries used in lanes L1, L3, L5 and L7, and R1, R3, R5 and R7 were subsequently sequenced. Size indicated on the left in base pairs.

**Table 2.** Cell number equivalents and library amplification cycles for ChIPmentation.

Position	Equivalent Miracidia	Equivalent cells	Amplification cycles L1-8	Amplification cycles R1-8	Sequenced
1	10,000	1,000,000	17	18	y
2	1,000	100,000	17	17	
3	100	10,000	18	19	y
4	50	5,000	19	18	
5	10	1,000	19	20	y
6	5	500	22	20	
7	1	100	22	22	y
8	0	0	26	25	

**Table 3.** General statistics on ChIPmentation libraries. All libraries were downsampled to 3.8 M uniquely aligned reads.

Pos	Equivalent miracidia	Equivalent cells	Amplification cycles	qBit DNA HS ng/ $\mu$ l	Read pairs	Uniquely aligned	After deduplication	% unique non-dup	% to keep for 3.8 M
L1	10,000	1,000,000	17	23.2	5,728,173	7,969,894	6,775,525	0	0
R1	10,000	1,000,000	18	21.2	8,388,53	13,469,770	10,007,402	0	0
L3	100	10,000	18	25.2	12,537,243	19,157,585	14,417,614	0	0
R3	100	10,000	19	30.4	8,665,837	13,838,684	8,409,793	0	0
L5	10	1,000	19	13.8	11,436,716	17,920,411	8,664,467	0	0
R5	10	1,000	20	19.3	11,109,380	17,854,360	7,214,153	0	0
L7	1	100	22	22.0	11,662,120	15,737,202	4,095,477	0	0
R7	1	100	22	11.9	10,566,906	15,221,913	3,827,034	0	0

**Table 4. General statistics on N-ChIP libraries.** When possible, libraries were downsampled to 3.8 M uniquely aligned reads.

Pos	Equivalent Miracidia	Equivalent cells	Amplification cycles	Read pairs	Uniquely aligned	After deduplication	% unique non-dup	% downsampled
A	8,000	800,000	14	13,129,021	11,299,195	10,763,378	41 %	35 %
B	100	10,000	14	21,415,343	5,409,960	2,477,749	6 %	100 %
C	10	1,000	14	29,857,139	12,170,093	6,718,570	11 %	57 %

**Table 5. Optimization of peakcalling with Peakranger.** Number of peaks identified for each condition. ChIPmentation on top. In bold: 1000 bp was selected as the best extension length and applied to N-ChIP data below. N-ChIP A is for  $0.8 \times 10^6$  cells.

cell equivalents	1,000,000			10,000			1,000			100	
	ChIPmentation			ChIPmentation			ChIPmentation			ChIPmentation	
	L1	R1	N-ChIP A	L3	R3	N-ChIP B	L5	R5	N-ChIP C	L7	R7
100	6,871	6,767		625	424		200	370		4,732	5,683
300	10,369	9,545		2,125	1,211		499	635		3,944	5,264
600	10,581	9,924		3,513	2,823		610	813		3,197	3,816
<b>1,000</b>	<b>9,632</b>	<b>9,835</b>	<b>8,320</b>	<b>4,033</b>	<b>2,749</b>	<b>3,888</b>	<b>695</b>	<b>811</b>	<b>131</b>	<b>3,098</b>	<b>3,416</b>
1,500	9,901	9,426		3,315	2,925		847	837		4,019	3,526
2,000	9,067	8,353		3,052	2,631		626	693		4,336	3,897

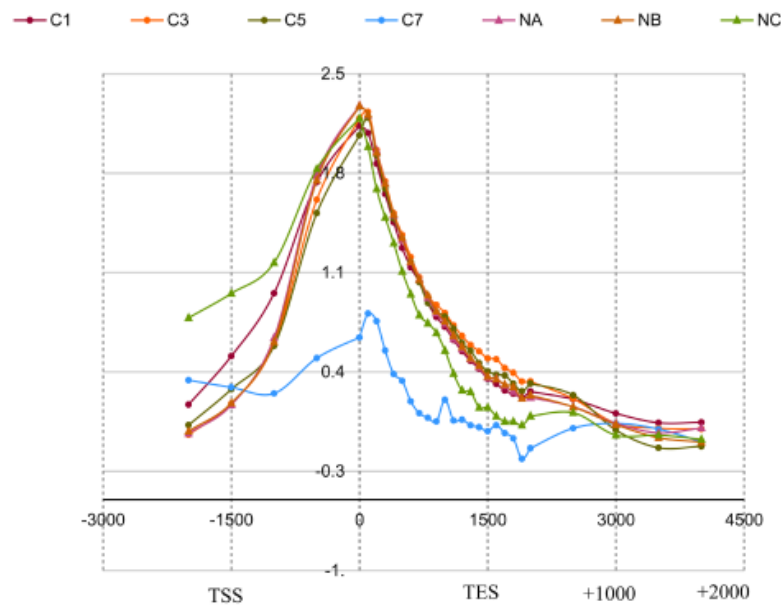
**Table 6. Peakcalling with HMM-based ChromstaR before and after merging adjacent peaks (in bold).**

Cell equivalents	1,000,000		10,000		1,000		100	
ChromstaR bin 1000, step 500, post prob 0.999	All	merged	All	Merged	All	Merged	All	Merged
ChIPmentation	13,565	<b>6,682</b>	7,194	<b>5,504</b>	3,137	<b>2,876</b>	10,665	<b>10,522</b>
N-ChIP	6,262	<b>6,186</b>	7,058	<b>6,922</b>	5,296	<b>5,129</b>		

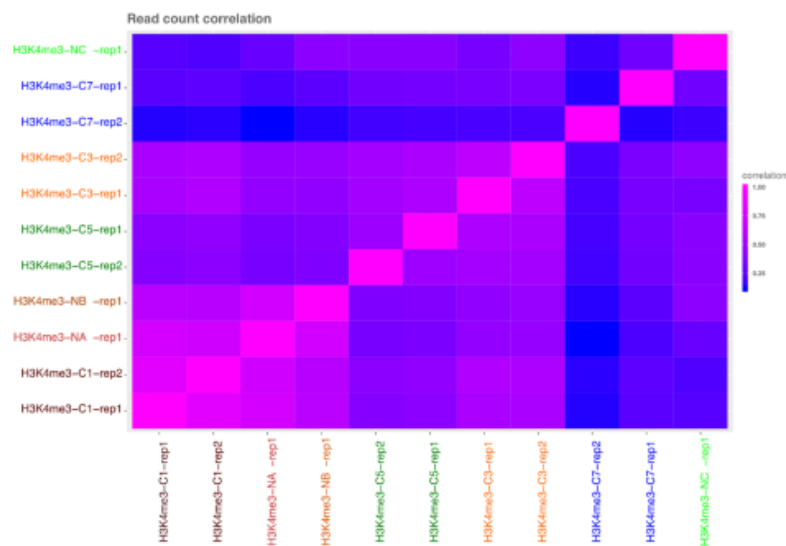
for control of unspecific enrichment. The production of input chromatin is "built-in" the N-ChIP protocol (Cosseau *et al.*, 2009; de Carvalho Augusto *et al.*, 2020; Roquis *et al.*, 2018) and needed to be adapted to the automated ChIPmentation procedure. During a ChIPmentation experiment three types of input can be considered (Figure 6): (i) 1  $\mu$ L of chromatin before immunoprecipitation, (ii) chromatin that binds non-specifically to any support, and (iii) chromatin truly available for IP. In (i), an aliquot of 1  $\mu$ L is taken from the sample before immunoprecipitation. The two other types (input library ii and

iii) need a supplementary sample in which mock IP is done without antibody. After IP, this supplementary sample contains magnetic beads with the non-specifically bound chromatin and the supernatant, which is the available chromatin for IP.

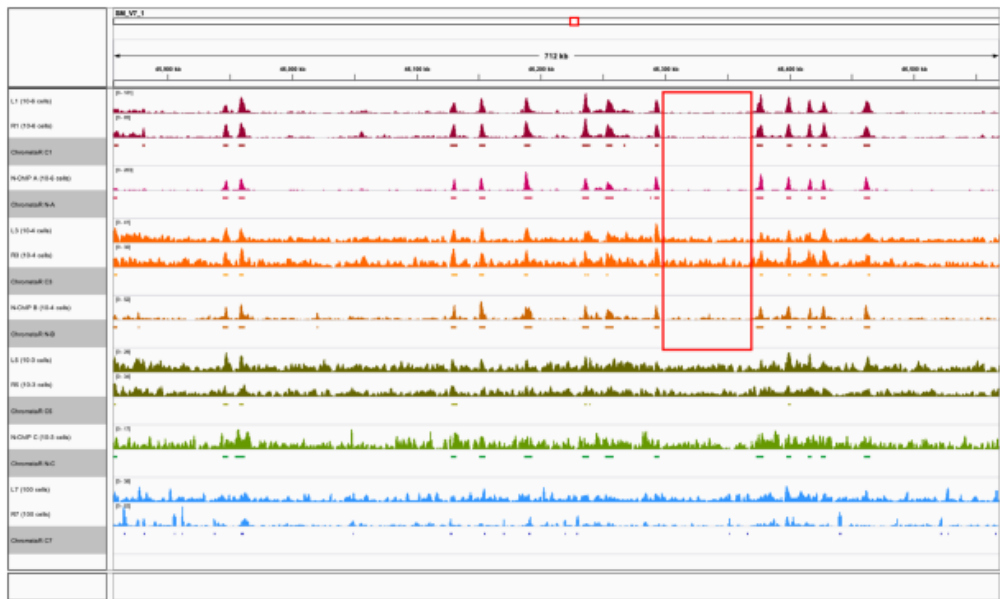
Only (i), *i.e.* before IP chromatin and (iii) *i.e.* free chromatin available for IP give ideal library sizes (Figure 6). We decided to optimize the input protocol for option (i), non-immunoprecipitated chromatin because it does not occupy a slot in the IP-Star. In addition, during the preliminary



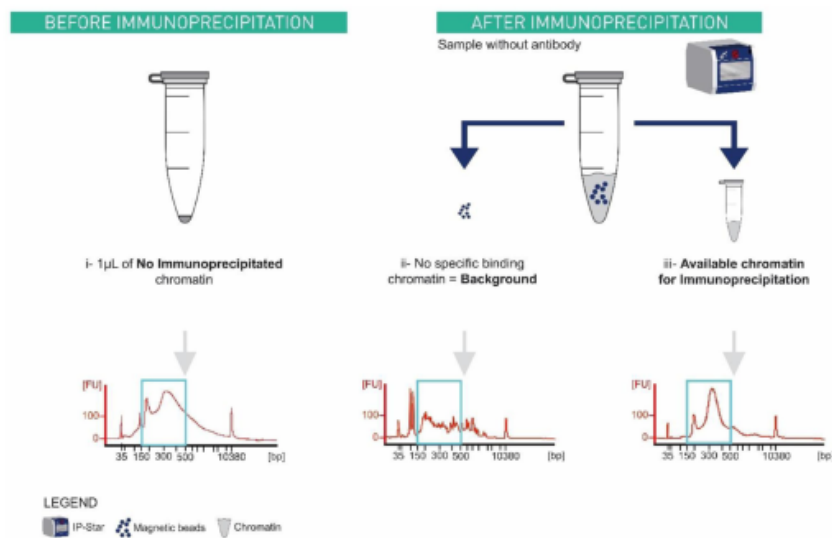
**Figure 3. Average metagene profiles over 5,073 plus strand genes for ChIPmentation and N-ChIP.** ChIPmentation samples were  $10^6$  (C1),  $10^4$  (C3),  $10^3$  (C5) and 100 cell equivalents (C7). N-ChIP samples were  $0.8 \times 10^6$  cells (NA),  $10^4$  (NB) and  $10^3$  (NC). X-axis: bp upstream, within and downstream of genes. TSS/TES for transcription start and end sites. Y-axis: log(observed/expected). Not all genes contributed to the profiles as only roughly half of the genes show a H3K4me3 peak at the TSS.



**Figure 4. ChromstaR read count correlations between libraries (lowest 0, highest 1).** L and R samples were considered as replicates 1 and 2 for ChromstaR analysis. ChIPmentation  $10^6$  (C1),  $10^4$  (C3),  $10^3$  (C5) and 100 cell equivalent (C7). N-ChIP  $0.8 \times 10^6$  cells (NA),  $10^4$  (NB) and  $10^3$  (NC).



**Figure 5. Genome browser screen shot of a typical region of the *S. mansoni* genome with Peakranger chromatin profiles for visual inspection and HMM model-based ChromStaR peak regions (grey underlay).** ChIPmentation  $10^5$  (L1, R1),  $10^4$  (L3, R3),  $10^3$  (L5, R5) and 100 cell equivalent (L7, R7), and N-ChIP  $0.8 \times 10^6$  cells (N-A),  $10^4$  (N-B) and  $10^3$  (N-C). Color codes as in previous figures : ChIPmentation  $10^6$  (L1, R1) in dark red, N-ChIP  $0.8 \times 10^6$  cells (N-A) in magenta, ChIPmentation  $10^4$  (L3, R3) in orange, N-ChIP  $10^4$  (N-B) in dark orange, ChIPmentation  $10^3$  (L5, R5) in green, N-ChIP  $10^3$  (N-C) in light green and ChIPmentation 100 cell equivalent (L7, R7) in light blue. The region circled in red illustrates higher background for  $10^4$  cells equivalent in ChIPmentation C3 (orange, replicates L3, R3) than in N-ChIP B (dark orange).



**Figure 6. Types of inputs generated during ChIPmentation and their associated bioanalyzer profiles.** X axis represents the size in base pair, and Y axis represents the fluorescence intensity. Ideal library size is between 150 bp and 500 bp (blue rectangle). (i)  $1 \mu\text{L}$  aliquot of one sample chromatin taken before its immunoprecipitation. (ii) and (iii) are from a supplementary sample where immunoprecipitation in the IP-Star is done without antibody. After immunoprecipitation, this supplementary sample contains magnetic beads with the (ii) non-specifically binding chromatin and the supernatant which is the (iii) available chromatin for immunoprecipitation. (ii) No specific binding chromatin which did not deliver a usable library.

test, 1  $\mu$ L chromatin input showed a lower Ct compared to option (iii) input, which means that it requires fewer amplification cycles (data on Zenodo).

For the preparation of the 1  $\mu$ L chromatin input libraries, we identified two critical parameters. The first one was the dilution of the tagmentation enzyme. Using undiluted Tn5 caused complete over-tagmentation. Between 10- and 100-fold dilutions in water delivered optimal results (Figure 7A).

Secondly, we found that, in our hands, there was no need to add tagmentation neutralizer (0.2% sodium dodecyl sulfate [SDS]) after tagmentation (Figure 7-B and Figure 8). This actually inhibits the PCR amplification step (Picelli *et al.*, 2014). Interestingly, parameters like tagmentation temperature (37°C-55°C) (Figure 8), tagmentation time (2–10 min) and addition of MgCl<sub>2</sub> did not have a critical effect on input library generation.

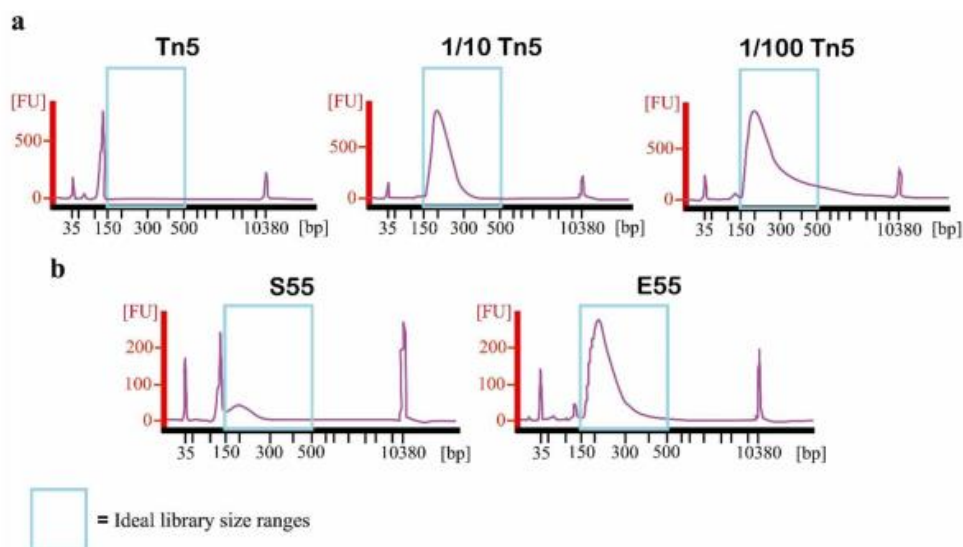
### Discussion

As many methods, ChIPmentation had been developed using readily available but highly artificial cell cultures (Schmidl *et al.*, 2015). The transition from such model systems to non-model species and ecologically realistic conditions is sometimes very difficult or even impossible. Here we showed that the

automated ChIPmentation can be done with a parasitic flatworm and delivers results comparable to N-ChIP, the current method of choice for this species. However, the method is roughly two times faster and requires roughly six times less hands-on time. Since the procedure is done on a pipetting robot with on-screen instructions for the researcher, in our experience, training time was reduced to about a week.

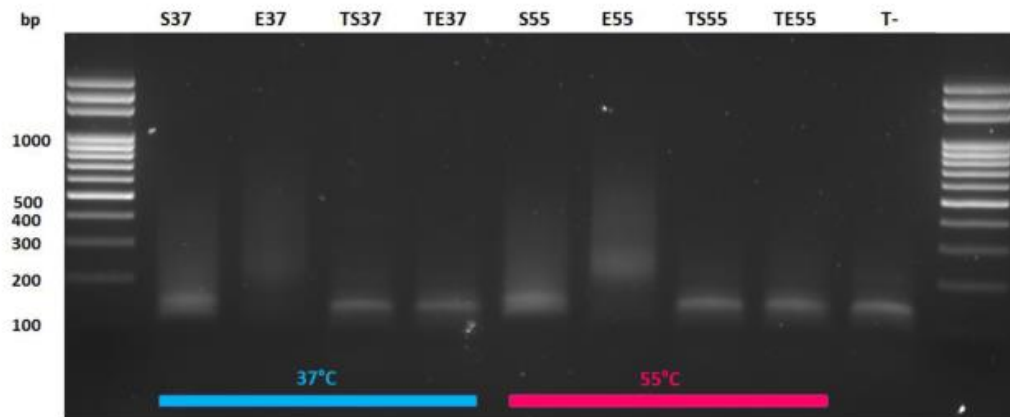
We empirically define the limit for the robust detection of peaks with ChIPmentation to be 100,000 cells equivalent per antibody reaction, with 10,000 being the absolute limit if background is acceptable. Operational limit for N-ChIP is 10,000 cells equivalent, confirming our previous results. To avoid variations that might be introduced by small errors in the estimation of cell numbers, we arbitrarily doubled this lower cell limit and established  $\geq 20,000$  cells as the lower limit for the routine ChIPmentation-seq procedures in *S. mansoni*. This is a little higher than what has been described in Human cells where good results have been obtained with as little as 5,000 cells.

To improve signal to noise ratio and reduce background in ChIPmentation, it could be useful to increase the washing time (currently 10 min) and speed (currently medium) but we recommend increasing cell number rather than to invest in washing optimization.



**Figure 7. Bioanalyzer profiles for input libraries performed with different Tn5 and SDS conditions. a)** with undiluted Tn5 (**Tn5**), 10-fold diluted Tn5 (**1/10 Tn5**) and 100-fold diluted Tn5 (**1/100 Tn5**). **b)** 10-fold diluted Tn5 with 0.2% SDS (**S55**) and 10-fold diluted Tn5 without 0.2% SDS (**E55**). X axis represents the size in base pair, and Y axis represents the fluorescence intensity. Ideal library size is between 150 bp and 500 bp (blue rectangle).





**Figure 8. Picture of size separation of PCR products after qPCR for input library quantification by electrophoresis.** Electrophoresis was performed through an agarose gel stained with ethidium bromide. Tagmentation was performed at 37°C and 55°C with 10-fold diluted Tn5. Sample legend: **(S)** SDS, **(E)** No SDS, **(TS)** no Tn5 enzyme with 0.2% SDS, **(TE)** no Tn5 enzyme and no SDS, **(T-)** qPCR negative control without chromatin. Ideal library size is between 150 bp and 500 bp. Only **E** samples had the right library size.

Using 1  $\mu$ L of non-immunoprecipitated chromatin for the reference input library production is the best compromise to save space in the IP-Star, experiment time and biological materials when one is restricted by quantity.

This ChIPmentation protocol is not limited to miracidia cells. We also performed this protocol on *S. mansoni* adult worms and sporocysts. It should be noted that the number of sonication cycles needs to be experimentally determined and adapted for each sample type before proceeding to ChIPmentation experiments.

A new version of the ChIPmentation solution, called  $\mu$ ChIPmentation for Histones (Diagenode, Cat. No. C01011011), has also been released recently in order to improve the quality for low-amount samples. This relies on a reduced number of steps, especially during chromatin preparation, and reduced number of tube transfers, in order to avoid DNA loss. It also contains a new protocol to process the non-immunoprecipitated chromatin input samples up to the sequencing step. This new version of  $\mu$ ChIPmentation may be a good alternative for experiments on very low cell numbers in the future.

### Data availability

#### Underlying data

NCBI SRA: Automated ChIPmentation procedure on limited biological material of the human blood fluke *Schistosoma mansoni*. Accession number: PRJNA816041, <https://identifiers.org/ncbiprotein:PRJNA816041>

This project contains the Fastq data of the ChIPmentation libraries.

Zenodo: Supporting information for "Automated ChIPmentation procedure on limited biological material of the human blood fluke *Schistosoma mansoni*", <https://doi.org/10.5281/zenodo.6375548> (Grunau, 2022)

This project contains the following underlying data:

- 20200612-1\_Report.pdf (qPCR report for inputs 1 $\mu$ L and available chromatin ("Row7"), see Figure 6)
- 20200919-1\_Report.pdf (qPCR report for testing SDS after Tn5, see Figure 7B and Figure 8)
- 20200921-1\_Report.pdf (qPCR report for comparing inputs with enzyme of Diagenode kit and our protocol with other enzyme Tn5)
- CG\_Ro\_1\_HighSensitivityDNAAssay\_DE13805677\_2019-06-20\_09-10-48.pdf (DNA assay file underlying Figure 2)
- CG\_Ro\_2\_HighSensitivityDNAAssay\_DE13805677\_2019-06-20\_10-12-12.pdf (DNA assay file underlying Figure 2)
- gel qpcr test input sds\_01.Tif

Data are available under the terms of the [Creative Commons Attribution 4.0 International license \(CC-BY 4.0\)](https://creativecommons.org/licenses/by/4.0/).

### Acknowledgements

This study is set within the framework of the *Laboratoire d'Excellence (LabEx) TULIP* (ANR-10-LABX-41), with the support of LabEx CeMEB, an ANR *Investissements d'avenir* program (ANR-10-LABX-04-01) and the Environmental Epigenomics Core Service at IHPE.

We thank S.A.S. *ParaDev* and Julien Portela for livers of infected hamsters. We also thank the Bio-Environment platform (University of Perpignan Via Domitia) and Jean-François Allienne for support in library preparation and sequencing. We have obtained permission from these colleagues to be acknowledged.

## References

- Augusto RdC, Cosseau C, Grunau C: **Histone Methylome of the Human Parasite *Schistosoma Mansoni*. *The DNA RNA, and Histone Methylomes*. S. Jurga and J. Barciszewski, Cham, Springer International Publishing; 2019; 607–624. [Publisher Full Text](#)**
- Broad Institute: **Picard. Broad Institute, GitHub repository**. 2022. [Reference Source](#)
- Buenrostro JD, Wu B, Chang HY, et al.: **ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide**. *Curr Protoc Mol Biol*. 2015; **109**: 21.29.21–21.29.29. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Cosseau C, Azzi A, Smith K, et al.: **Native chromatin immunoprecipitation (N-ChIP) and ChIP-Seq of *Schistosoma mansoni*: Critical experimental parameters**. *Mol Biochem Parasitol*. 2009; **166**(1): 70–76. [PubMed Abstract](#) | [Publisher Full Text](#)
- de Carvalho Augusto R, Roquis D, Al Picard M, et al.: **Measuring Histone Modifications in the Human Parasite *Schistosoma mansoni***. *Methods Mol Biol*. 2020; **2151**: 93–107. [PubMed Abstract](#) | [Publisher Full Text](#)
- Feng X, Grossman R, Stein L: **PeakRanger: a cloud-enabled peak caller for ChIP-seq data**. *BMC Bioinformatics*. 2011; **12**: 139. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Grunau C: **Supporting information for "Automated ChIPmentation procedure on limited biological material of the human blood fluke *Schistosoma mansoni*" (Version 1) [Data set]**. *Zenodo*. 2022. <http://www.doi.org/10.5281/zenodo.6382509>
- Langmead B, Salzberg SL: **Fast gapped-read alignment with Bowtie 2**. *Nat Methods*. 2012; **9**(4): 357–359. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Li H, Handsaker B, Wysoker A, et al.: **The Sequence Alignment/Map format and SAMtools**. *Bioinformatics*. 2009; **25**(16): 2078–2079. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Mouahid G, Idris MA, Verneau O, et al.: **A new chronotype of *Schistosoma mansoni*: adaptive significance**. *Trop Med Int Health*. 2012; **17**(6): 727–732. [PubMed Abstract](#) | [Publisher Full Text](#)
- Picelli S, Björklund AK, Reinius B, et al.: **Tn5 transposase and tagmentation procedures for massively scaled sequencing projects**. *Genome Res*. 2014; **24**(12): 2033–2040. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Quinlan AR, Hall IM: **BEDTools: a flexible suite of utilities for comparing genomic features**. *Bioinformatics*. 2010; **26**(6): 841–842. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Roels J, Kuchmiy A, De Decker M, et al.: **Distinct and temporary-restricted epigenetic mechanisms regulate human  $\alpha\beta$  and  $\gamma\delta$  T cell development**. *Nat Immunol*. 2020; **21**(10): 1280–1292. [PubMed Abstract](#) | [Publisher Full Text](#)
- Roquis D, Taudt A, Geyer KK, et al.: **Histone methylation changes are required for life cycle progression in the human parasite *Schistosoma mansoni***. *PLoS Pathog*. 2018; **14**(5): e1007066. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Schmidl C, Rendeiro AF, Sheffield NC, et al.: **ChIPmentation: fast, robust, low-input ChIP-seq for histones and transcription factors**. *Nat Methods*. 2015; **12**(10): 963–965. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Smith AD: **A quality control tool for high throughput sequence data**. 2019. [Reference Source](#)
- Taudt A, Nguyen MA, Heinig M, et al.: **chromstaR: Tracking combinatorial chromatin state dynamics in space and time**. *bioRxiv*. 2016. [Publisher Full Text](#)

## Annexe 8 : Liste des gènes différentiellement marqués par H3K4me3 issus du DEseq2 entre les chronotypes diurne et nocturne de *S. mansoni*

Chr.	Gene ID	GFF annotation	HHsearch annotation	Surreprésenté	QTL
2	Smp_327240	NA		Diurne	
2	Smp_310040	Alpha 1,6 fucosyltransferase I		Diurne	
2	Smp_122820	NA		Diurne	
2	Smp_327530	NA		Diurne	
2	Smp_327550	NA		Diurne	
2	Smp_303430	previous_stable_id=Smp_001430		Diurne	
2	Smp_001420	Gh regulated tbc protein-1, putative		Diurne	
2	Smp_122810	Putative spfh domain / Band 7 family protein		Diurne	
2	Smp_343980	previous_stable_id=Smp_047850,Smp_147740		Diurne	
2	Smp_337360	Peptidase_M13_N domain-containing protein		Diurne	
2	Smp_021170	VPS13C protein, putative		Diurne	
2	Smp_335380	previous_stable_id=Smp_193370		Diurne	
2	Smp_328700	NA		Diurne	
2	Smp_328710	previous_stable_id=Smp_170370		Diurne	
1	Smp_162010	Non-specific serine/threonine protein kinase		Diurne	
SM _v7 _2H008	Smp_305410	previous_stable_id=Smp_085410		Diurne	
3	Smp_336660	previous_stable_id=Smp_212400,Smp_197760, Smp_055500		Diurne	
3	Smp_098950	TATA-box binding protein, putative		Diurne	
3	Smp_247260	previous_stable_id=Smp_194670,Smp_149660		Diurne	
3	Smp_085780	Ribosomal protein L10P	28S ribosomal protein S2, mitochondrial	Diurne	QTL3AB
3	Smp_000370	Selenoprotein SelK/SelG	SelK_SelG	Diurne	QTL3AB
1	Smp_180260	Amiloride-sensitive sodium channel,putative		Diurne	
5	Smp_334550	previous_stable_id=Smp_156830,Smp_205340		Diurne	

## Annexe 8

5	Smp_334620	previous_stable_id=Smp_210140		Diurne	
6	Smp_155560	Serpin, putative		Diurne	
6	Smp_336290	previous_stable_id=Smp_162320		Diurne	
6	Smp_331250	NA		Diurne	
6	Smp_164130	Dopey-related		Diurne	
ZW	Smp_324090	NA		Diurne	
ZW	Smp_173120	Phosphatidic acid phosphatase type 2-related		Diurne	
1	Smp_319050	Transmembrane protein 164	TMEM164	Diurne	QTL1
1	Smp_127370	Glycosyl transferase, family 14	beta-1,6-N-acetylglucosaminyltransferase (E.C.2.4.1.102)	Nocturne	QTL1
2	Smp_172990	Putative leucine rich repeat-containing		Nocturne	

Chr. : chromosome.

Annexe 9 : Liste des gènes différentiellement marqués par H3K9ac issus du DEseq2 entre les chronotypes diurne et nocturne de *S. mansoni*

Chr.	Gene ID	GFF annotation	HHsearch annotation	Surreprésenté	QTL
6	Smp_336310	NA		Diurne	
6	Smp_147020	Sec7 domain	guanine nucleotide exchange protein	Diurne	QTL6
2	Smp_011220	Putative glucosamine-fructose-6-phosphate aminotransferase		Nocturne	
3	Smp_337380	NA		Nocturne	
3	Smp_205760	Guanine nucleotide-binding protein subunit gamma		Nocturne	
3	Smp_029160	Putative ribonuclease p/mrp subunit		Nocturne	
3	Smp_308470	SEA domain	SEA	Nocturne	QTL3AB
3	Smp_084540	Forkhead-associated (FHA) domain	RNA-binding protein 5, Survival motor	Nocturne	QTL3AB
5	Smp_106390	NA		Nocturne	
5	Smp_327560	Thioredoxin domain-containing protein		Nocturne	
6	Smp_341570	NA		Nocturne	
SM_V7_6H013	Smp_321200	NA		Nocturne	
7	Smp_010380	Isochorismatase domain-containing protein		Nocturne	
SM_V7_MITO	Smp_900070	Cytochrome b (mitochondrion)		Nocturne	
SM_V7_MITO	Smp_900080	NA		Nocturne	
SM_V7_MITO	Smp_900090	NADH dehydrogenase subunit 4 (mitochondrion)		Nocturne	
ZW	Smp_169010	Putative adenylate cyclase		Nocturne	
1	Smp_127370	Glycosyl transferase, family 14	beta-1,6-N-acetylglucosaminyl-transferase (E.C.2.4.1.102)	Nocturne	QTL1

Annexe 9

1	Smp_158570	NA		Nocturne	
---	------------	----	--	----------	--

Chr. : chromosome.

## Annexe 10 : Liste des gènes diurnes rythmiques

Chr.	Gene ID	GFF annotation	HHsearch annotation	Pic (h)
SM_V7_1	Smp_003300	Putative serpin		8
SM_V7_1	Smp_003610	NA		16
SM_V7_1	Smp_003940	NA		4
SM_V7_1	Smp_005180	Transcription factor, GTP-binding domain	uL2, uL3, uL4, 60S ribosomal	24
SM_V7_1	Smp_007260	P-type ATPase	SARCOPLASMIC/ENDOPLASMIC RETICULUM CALCIUM ATPASE 1	8
SM_V7_1	Smp_007270	Putative alpha-actinin		12
SM_V7_1	Smp_008400	Adenosine kinase		4
SM_V7_1	Smp_009600	Protein kinase domain	Aurora-related kinase 1 (aurora-2) {Human (Homo sapiens) [TaxId: 9606]}	4
SM_V7_1	Smp_009830	Translocon-associated protein subunit beta	TRAP_beta	20
SM_V7_1	Smp_012810	Putative four and A half lim domains		12
SM_V7_1	Smp_019190	Pre-mRNA-splicing factor SLU7 domain	Protein mago nashi homolog 2	12
SM_V7_1	Smp_019310	Dynein light chain, type 1/2	Dynein light chain 1 (DLC1) {Fruit fly (Drosophila melanogaster) [TaxId: 7227]}	8
SM_V7_1	Smp_019340	Tetraspanin	Tetraspanin	20
SM_V7_1	Smp_019440	SUI1 domain	Eukaryotic translation initiation factor eIF-1 (SUI1) {Human (Homo sapiens) [TaxId: 9606]}	20
SM_V7_1	Smp_020070	NA		8
SM_V7_1	Smp_020080	Putative gtp-binding protein (I) alpha-1subunit%2C gnai1		12
SM_V7_1	Smp_025750	NA	NA	8
SM_V7_1	Smp_032180	Small GTP-binding protein domain	ADP-ribosylation factor {Human (Homo sapiens), ARF6 [TaxId: 9606]}	24
SM_V7_1	Smp_032230	Tetratricopeptide-like helical domain superfamily	Mitochondria fission protein Fis1 {Baker's yeast (Saccharomyces cerevisiae) [TaxId: 4932]}	20
SM_V7_1	Smp_032250	Beta-lactamase-related	Carboxylesterase (E.C.3.1.1.1)	20
SM_V7_1	Smp_034660	Cytohesin-related guanine nucleotide-exchange protein		8
SM_V7_1	Smp_034670	Tubulin gamma chain		4
SM_V7_1	Smp_035250	NA		12
SM_V7_1	Smp_035280	Adenylyltransferase and sulfurtransferase MOCS3 homolog Adenylyltransferase Sulfurtransferase		12
SM_V7_1	Smp_035470	Dolichyl-diphosphooligosaccharide--protein glycosyltransferase 48 kDa subunit		8
SM_V7_1	Smp_048260	Putative n-acetyltransferase		20
SM_V7_1	Smp_049550	Putative heat shock protein 70 (Hsp70)		8
SM_V7_1	Smp_049600	Putative DNAj (Hsp40) homolog%2C subfamily C%2C member		8
SM_V7_1	Smp_055390	Putative dullard protein		24
SM_V7_1	Smp_055990	Adenylosuccinate synthetase		8

## Annexe 10

SM_V7_1	Smp_056200	Isocitrate dehydrogenase [NAD] subunit%2C mitochondrial		12
SM_V7_1	Smp_057060	Acetyl-CoA deacylase		8
SM_V7_1	Smp_059980	Arginase		8
SM_V7_1	Smp_060570	Putative sarg904		8
SM_V7_1	Smp_061650	26S proteasome subunit S9%2C putative		4
SM_V7_1	Smp_065610	Pyruvate kinase		8
SM_V7_1	Smp_071050	NA		16
SM_V7_1	Smp_071390	Adenylate kinase%3B Putative adenylate kinase		8
SM_V7_1	Smp_072450	Putative rhodopsin-like orphan GPCR		16
SM_V7_1	Smp_072560	NA		4
SM_V7_1	Smp_073560	G beta-like protein gbl		12
SM_V7_1	Smp_073610	Transcriptional co-activator crsp77-related		4
SM_V7_1	Smp_073800	RecQ-mediated genome instability protein 1		12
SM_V7_1	Smp_075770	NA		12
SM_V7_1	Smp_075800	Legumain		8
SM_V7_1	Smp_076660	Putative DNA polymerase delta small subunit		20
SM_V7_1	Smp_077030	NA		20
SM_V7_1	Smp_078690	Calponin		12
SM_V7_1	Smp_083710	Helicase, C-terminal	PROBABLE ATP-DEPENDENT RNA HELICASE DDX58	4
SM_V7_1	Smp_084600	Putative solute carrier family 37 member 2 (Glycerol-3-phosphate transporter)		8
SM_V7_1	Smp_085010	Cathepsin B-like peptidase (C01 family)		8
SM_V7_1	Smp_085840	MEG-4 (10.3) family		8
SM_V7_1	Smp_086480	Antigen Sm21.7		8
SM_V7_1	Smp_086530	Sm 20.8%3B Tegumental antigen Sm20.8%3B Tegumental protein Sm 20.8		8
SM_V7_1	Smp_088950	Hypoxia upregulated 1 (Hyou1)-related		8
SM_V7_1	Smp_090860	Protein-serine/threonine phosphatase		8
SM_V7_1	Smp_091390	Btz domain-containing protein		20
SM_V7_1	Smp_093210	Putative amiloride-sensitive sodium channel		12
SM_V7_1	Smp_093540	Receptor protein serine/threonine kinase		8
SM_V7_1	Smp_093580	Helicase, C-terminal	Pre-mRNA-splicing factor 8, Pre-mRNA-splicing factor	12
SM_V7_1	Smp_094360	Putative hepatitis B virus X associated protein%2C hbxa		8
SM_V7_1	Smp_102920	Putative trna delta(2)-isopentenylpyrophosphate transferase		24
SM_V7_1	Smp_103610	Cathepsin B1 isotype 1		8
SM_V7_1	Smp_104980	NA	PHD finger protein 7 (NYD-SP6) {Mouse (Mus musculus) [TaxId: 10090]}	4
SM_V7_1	Smp_124050	Venom allergen-like (VAL) 6 protein		8
SM_V7_1	Smp_124070	Venom allergen-like (VAL) 16 protein		8



## Annexe 10

SM_V7_1	Smp_124120	DNA polymerase II subunit 2		4
SM_V7_1	Smp_124340	NRF domain-containing protein		8
SM_V7_1	Smp_129020	Innexin		16
SM_V7_1	Smp_129560	C2H2-type domain-containing protein		24
SM_V7_1	Smp_129830	Putative cadherin		4
SM_V7_1	Smp_133020	Mitogen-activated protein kinase		4
SM_V7_1	Smp_136200	NA		12
SM_V7_1	Smp_136700	NA		12
SM_V7_1	Smp_137390	Histone RNA stem-loop-binding protein SLBP1/SLBP2	SLBP/RNA Complex	20
SM_V7_1	Smp_139620	PDZ domain	Glutamate receptor interacting protein 2, GRIP2 (KIAA1719) {Human (Homo sapiens) [TaxId: 9606]}	4
SM_V7_1	Smp_140460	NA		8
SM_V7_1	Smp_140550	Putative 39s ribosomal protein L35%2C mitochondrial		20
SM_V7_1	Smp_141380	Protein kinase		4
SM_V7_1	Smp_144510	VWFA domain-containing protein		12
SM_V7_1	Smp_149700	Heterogeneous nuclear ribonucleoprotein I (Hnrnp I)%2C putative		4
SM_V7_1	Smp_150420	NA		12
SM_V7_1	Smp_151360	NA		16
SM_V7_1	Smp_154600	Carboxylic ester hydrolase		12
SM_V7_1	Smp_159600	Putative collagen alpha chain		8
SM_V7_1	Smp_160490	NA		12
SM_V7_1	Smp_160850	Nucleoprotein TPR		8
SM_V7_1	Smp_160880	Phosphatidylcholine-sterol acyltransferase (Lecithin-cholesterol acyltransferase)/ Phospholipase A		8
SM_V7_1	Smp_160950	RING-type domain-containing protein		24
SM_V7_1	Smp_161360	Putative syndecan		12
SM_V7_1	Smp_161790	NA		12
SM_V7_1	Smp_165050	MEG-7		8
SM_V7_1	Smp_165950	Tektin		12
SM_V7_1	Smp_165960	Dcp1 related		8
SM_V7_1	Smp_167240	Nicastrin		4
SM_V7_1	Smp_167300	SWIM-type domain-containing protein		12
SM_V7_1	Smp_169190	Tegumental allergen-like protein		8
SM_V7_1	Smp_170340	Putative collagen alpha-1(V) chain		8
SM_V7_1	Smp_170570	Pre-mRNA processing protein prp39- related		4
SM_V7_1	Smp_172110	Protein disulfide-isomerase		24
SM_V7_1	Smp_173150	Tetraspanin		12
SM_V7_1	Smp_174520	Dynein light chain		8
SM_V7_1	Smp_176930	Putative cationic amino acid transporter		8
SM_V7_1	Smp_181050	NA		24
SM_V7_1	Smp_194040	NA		8
SM_V7_1	Smp_194770	Taurocyamine kinase		8
SM_V7_1	Smp_195090	Tegumental antigen		8

## Annexe 10

SM_V7_1	Smp_197230	Adenylate kinase/UMP-CMP kinase	Adenylate kinase {Pig ( <i>Sus scrofa</i> ) [TaxId: 9823]}	12
SM_V7_1	Smp_200020	Protein yippee-like		16
SM_V7_1	Smp_200110	NA		12
SM_V7_1	Smp_208090	NA		12
SM_V7_1	Smp_210160	NA		20
SM_V7_1	Smp_212230	Proteasome component (PCI) domain	COP9 signalosome complex subunit 1	20
SM_V7_1	Smp_212860	previous_stable_id=Smp_131860		20
SM_V7_1	Smp_213850	previous_stable_id=Smp_167250		20
SM_V7_1	Smp_213900	Collagen IV NC1 domain-containing protein		8
SM_V7_1	Smp_214180	Calpain catalytic domain-containing protein		8
SM_V7_1	Smp_245600	previous_stable_id=Smp_191220,Smp_180960,Smp_180950		16
SM_V7_1	Smp_246020	NA		16
SM_V7_1	Smp_246400	Glycos_transf_1 domain-containing protein		8
SM_V7_1	Smp_246460	previous_stable_id=Smp_173850		12
SM_V7_1	Smp_247180	SHSP domain-containing protein		12
SM_V7_1	Smp_248110	SANT/Myb domain	DNA-binding protein reb1/DNA Complex	12
SM_V7_1	Smp_302370	previous_stable_id=Smp_148580,Smp_196250		16
SM_V7_1	Smp_307220	Developmentally regulated antigen 10.3		8
SM_V7_1	Smp_310240	Calponin-homology (CH) domain-containing protein		12
SM_V7_1	Smp_314730	previous_stable_id=Smp_073050		4
SM_V7_1	Smp_315890	JmjC domain	PHD FINGER PROTEIN 8	8
SM_V7_1	Smp_318890	NA	Biopolymer transport protein ExbB, TonB	8
SM_V7_1	Smp_318900	NA	PTS system, cellobiose-specific IIB component	8
SM_V7_1	Smp_318920	NA	Biopolymer transport protein ExbB, TonB	8
SM_V7_1	Smp_319140	Mediator complex, subunit Med27	Med27	4
SM_V7_1	Smp_319260	Dynein light chain, type 1/2	Dynein light chain 1 (DLC1) {Fruit fly ( <i>Drosophila melanogaster</i> ) [TaxId: 7227]}	8
SM_V7_1	Smp_319300	NA	NA	8
SM_V7_1	Smp_319490	NA	Nucleoporin p54	8
SM_V7_1	Smp_319720	NA		8
SM_V7_1	Smp_319810	previous_stable_id=Smp_056290		8
SM_V7_1	Smp_319920	NA		8
SM_V7_1	Smp_319930	NA		8
SM_V7_1	Smp_320110	NA		8
SM_V7_1	Smp_320300	Dynein light chain		8
SM_V7_1	Smp_320540	NA		20
SM_V7_1	Smp_329760	N6-L-threonylcarbamoyladenine synthase		12
SM_V7_1	Smp_331730	Poly(U)-specific endoribonuclease		12

## Annexe 10

SM_V7_1	Smp_332610	previous_stable_id=Smp_172100,Smp_172080		12
SM_V7_1	Smp_337250	previous_stable_id=Smp_205800		12
SM_V7_1	Smp_337310	NA		12
SM_V7_1	Smp_340440	previous_stable_id=Smp_023830		12
SM_V7_1	Smp_341700	AMPKBI domain-containing protein		12
SM_V7_1	Smp_342030	previous_stable_id=Smp_124360		12
SM_V7_1	Smp_342240	previous_stable_id=Smp_126630		4
SM_V7_1	Smp_342940	UPF0506 domain-containing protein		20
SM_V7_1	Smp_344020	previous_stable_id=Smp_147860		12
SM_V7_1	Smp_344320	Tubulin-folding cofactor E		8
SM_V7_1	Smp_345320	previous_stable_id=Smp_161990		16
SM_V7_1	Smp_345500	previous_stable_id=Smp_164550		8
SM_V7_1	Smp_345580	previous_stable_id=Smp_165100		12
SM_V7_1	Smp_346350	26S proteasome non-ATPase regulatory subunit 1		8
SM_V7_1H002	Smp_329840	RFX-type winged-helix domain-containing protein		4
SM_V7_2	Smp_001590	NADH dehydrogenase [ubiquinone] 1 alpha subcomplex assembly factor 3		4
SM_V7_2	Smp_004260	NA		4
SM_V7_2	Smp_004680	Interferon-related developmental regulator-related		8
SM_V7_2	Smp_004690	ZZ-type domain-containing protein		8
SM_V7_2	Smp_004810	Serine/threonine-protein phosphatase		24
SM_V7_2	Smp_011150	Taspase-1 (T02 family)		12
SM_V7_2	Smp_014570	Smp_014570.2		8
SM_V7_2	Smp_014620	Putative translation initiation factor eif-2b beta subunit		24
SM_V7_2	Smp_016250	Putative gtp-binding protein (I) alpha-2 subunit%2C gnai2		24
SM_V7_2	Smp_021140	40S ribosomal protein S7		20
SM_V7_2	Smp_021920	Paramyosin		8
SM_V7_2	Smp_022090	Ribose-phosphate diphosphokinase		8
SM_V7_2	Smp_034410	Cathepsin F (C01 family)		12
SM_V7_2	Smp_037540	Putative alpha-amylase		24
SM_V7_2	Smp_046740	Dihydrolipoyl dehydrogenase		8
SM_V7_2	Smp_046910	Histone-like transcription factor ccaat-related		24
SM_V7_2	Smp_047620	Putative cyclin B3		4
SM_V7_2	Smp_047650	Ferritin		16
SM_V7_2	Smp_047680	Ferritin		16
SM_V7_2	Smp_086220	NA		20
SM_V7_2	Smp_099690	Protein RER1		20
SM_V7_2	Smp_102490	NA		12
SM_V7_2	Smp_103340	NA		12
SM_V7_2	Smp_105420	LGG%3B Saposin containing protein		8
SM_V7_2	Smp_123390	Requim%2C req/dpf2%2C putative		8
SM_V7_2	Smp_124430	Coatomer subunit beta		24
SM_V7_2	Smp_125490	NA		16
SM_V7_2	Smp_127200	Glucose transporter		12
SM_V7_2	Smp_128200	Kinesin-like protein		4

## Annexe 10

SM_V7_2	Smp_134040	Putative elf1		8
SM_V7_2	Smp_137720	NA		12
SM_V7_2	Smp_138750	Alpha 1%2C6 fucosyltransferase K		8
SM_V7_2	Smp_140700	Mitogen-activated protein kinase		16
SM_V7_2	Smp_148220	NA		24
SM_V7_2	Smp_154830	NA		8
SM_V7_2	Smp_154880	Delta-tubulin		20
SM_V7_2	Smp_157290	Condensin complex subunit 2		8
SM_V7_2	Smp_157360	RING-type E3 ubiquitin transferase		4
SM_V7_2	Smp_158240	Fumarate hydratase		4
SM_V7_2	Smp_167750	Putative fatty-acid amide hydrolase		12
SM_V7_2	Smp_169090	NA		8
SM_V7_2	Smp_176300	Dynein heavy chain%2C putative		4
SM_V7_2	Smp_210300	Putative 60s ribosomal protein L7a		8
SM_V7_2	Smp_245790	SAM domain-containing protein		12
SM_V7_2	Smp_247360	C2H2-type domain-containing protein		12
SM_V7_2	Smp_310600	previous_stable_id=Smp_161310		12
SM_V7_2	Smp_311170	PDZ domain-containing protein		12
SM_V7_2	Smp_311700	Putative kunitz-type protease inhibitor		8
SM_V7_2	Smp_318800	Metallophos domain-containing protein		8
SM_V7_2	Smp_332350	Putative kunitz-type protease inhibitor		8
SM_V7_2	Smp_337470	RING-type domain-containing protein		24
SM_V7_2	Smp_341520	RFXA_RFXANK_bdg domain-containing protein		20
SM_V7_2	Smp_345700	IPPC domain-containing protein		4
SM_V7_3	Smp_000500	NA	NA	8
SM_V7_3	Smp_000790	Actin binding LIM protein family member 2-related		12
SM_V7_3	Smp_001100	NA		8
SM_V7_3	Smp_002920	TRAM/LAG1/CLN8 homology domain	TRAM_LAG1_CLN8	8
SM_V7_3	Smp_004480	Red protein (Ik factor) (Cytokine ik)%2C putative		12
SM_V7_3	Smp_010160	CS domain	Suppressor of G2 allele of <i>skp1</i> homolog, <i>gst1</i> {Human (Homo sapiens) [TaxId: 9606]}	20
SM_V7_3	Smp_013300	Yip1 domain	YIF1	24
SM_V7_3	Smp_013440	Zinc/iron permease	Zip	8
SM_V7_3	Smp_018610	Mediator complex, subunit Med7	Med7	8
SM_V7_3	Smp_018710	WD40 repeat	Uncharacterized WD repeat-containing protein C4F10.18	24
SM_V7_3	Smp_018880	NA	NA	8
SM_V7_3	Smp_019010	Dipeptidyl peptidase 3		4
SM_V7_3	Smp_027870	NA	NPIP	8
SM_V7_3	Smp_029500	Thimet oligopeptidase (M03 family)		24
SM_V7_3	Smp_030220	HAD-superfamily hydrolase, subfamily IIA	Hypothetical protein TM1742 { <i>Thermotoga maritima</i> [TaxId: 2336]}	24
SM_V7_3	Smp_052350	Zinc finger, RanBP2-type	UBIQUITIN THIOESTERASE ZRANB1 (E.C.3.4.19.12)	8
SM_V7_3	Smp_052390	Origin recognition complex, subunit 5	Orc2, Orc3, Orc5, Orc1, Orc6	4
SM_V7_3	Smp_075040	Uncharacterised protein family UPF0538	DUF2340	16

## Annexe 10

SM_V7_3	Smp_075110	Cell cycle checkpoint protein, Rad1	Rad1	20
SM_V7_3	Smp_085750	Zinc finger protein NOA36	NOA36	8
SM_V7_3	Smp_100130	GYF domain-containing protein		20
SM_V7_3	Smp_105950	Armadillo	beta-Catenin {Human (Homo sapiens) [TaxId: 9606]}	16
SM_V7_3	Smp_106110	EF-hand domain	Kchip1, Kv4 potassium channel-interacting protein {Rat (Rattus norvegicus) [TaxId: 10116]}	12
SM_V7_3	Smp_127490	Helicase, C-terminal	BLOOM'S SYNDROME HELICASE (E.C.3.6.4.12)	4
SM_V7_3	Smp_132210	NA		12
SM_V7_3	Smp_134720	Major facilitator superfamily domain	Melibiose carrier protein	8
SM_V7_3	Smp_134870	Zinc finger C2H2-type	Zinc finger protein 297b {Human (Homo sapiens) [TaxId: 9606]}	16
SM_V7_3	Smp_134880	Snurportin-1	35 6DT1_A DNA dC->dU-editing enzy 93.8 0.02 1.7E-07 54.0 0.0 202 99- 310 152-399 (507)	8
SM_V7_3	Smp_137910	Putative tight junction protein		12
SM_V7_3	Smp_137920	Putative chloride intracellular channel		8
SM_V7_3	Smp_138320	NA	NA	12
SM_V7_3	Smp_145990	ClpA/B family	ClpA, an Hsp100 chaperone, AAA+ modules {Escherichia coli [TaxId: 562]}	8
SM_V7_3	Smp_150210	Zinc finger, RING-type	U1 SNP1-associating protein 1, Degradation	8
SM_V7_3	Smp_160000	WD_REPEATS_REGION domain-containing protein		8
SM_V7_3	Smp_162100	ADA3-like		12
SM_V7_3	Smp_162630	Ubiquitin/SUMO-activating enzyme E1	Ubiquitin-activating enzyme E1 1, Uba1	8
SM_V7_3	Smp_162710	Protein kinase domain	Aurora-related kinase 1 (aurora-2) {Human (Homo sapiens) [TaxId: 9606]}	20
SM_V7_3	Smp_167540	Ris2 protein%2C putative		4
SM_V7_3	Smp_167630	Zinc/iron permease	Zip	16
SM_V7_3	Smp_177040	GPI mannosyltransferase 2		8
SM_V7_3	Smp_177100	NA		8
SM_V7_3	Smp_181460	EF-hand domain pair	DUF4205	12
SM_V7_3	Smp_194420	NA		12
SM_V7_3	Smp_194720	Family S33 non-peptidase homologue (S33 family)		12
SM_V7_3	Smp_196220	RING-type E3 ubiquitin transferase		8
SM_V7_3	Smp_201950	NA		8
SM_V7_3	Smp_202190	NA		20
SM_V7_3	Smp_204770	Citrate synthase		8
SM_V7_3	Smp_210720	Uncharacterised protein family UPF0047	Hypothetical protein TM0723 {Thermotoga maritima [TaxId: 2336]}	20
SM_V7_3	Smp_211070	NEDD8-activating enzyme E1 catalytic subunit		8
SM_V7_3	Smp_211260	Subfamily S1B unassigned peptidase (S01 family)		20
SM_V7_3	Smp_212540	ST7	ST7	4
SM_V7_3	Smp_213360	previous_stable_id=Smp_029150		20
SM_V7_3	Smp_243200	Calpain catalytic domain-containing protein		12

## Annexe 10

SM_V7_3	Smp_243680	AAA ATPase domain	HslU {Haemophilus influenzae [TaxId: 727]}	8
SM_V7_3	Smp_246560	GIY-YIG endonuclease	Structure-specific endonuclease subunit SLX1, Structure-specific	20
SM_V7_3	Smp_266760	NA		12
SM_V7_3	Smp_308750	NA	NA	8
SM_V7_3	Smp_309100	NA	NA	12
SM_V7_3	Smp_309580	previous_stable_id=Smp_019000		8
SM_V7_3	Smp_326250	NA	30S ribosomal protein S2, 30S	12
SM_V7_3	Smp_326390	Dynein light chain, type 1/2	Dynein light chain 1 (DLC1) {Fruit fly (Drosophila melanogaster) [TaxId: 7227]}	8
SM_V7_3	Smp_326460	NA	NA	8
SM_V7_3	Smp_326480	NA	NA	8
SM_V7_3	Smp_326490	NA	GRP	8
SM_V7_3	Smp_326660	NA	LRR_11	8
SM_V7_3	Smp_326730	previous_stable_id=Smp_108550		20
SM_V7_3	Smp_336680	Tetratricopeptide-like helical domain superfamily	ANAPHASE-PROMOTING COMPLEX SUBUNIT 1, ANAPHASE-PROMOTING	20
SM_V7_3	Smp_337650	Sperm-associated antigen 17	PapD-like	8
SM_V7_3	Smp_341710	NA	Catalytic domain of MutY {Escherichia coli [TaxId: 562]}	8
SM_V7_3	Smp_342670	BRCT domain	Breast cancer associated protein, BRCA1 {Human (Homo sapiens) [TaxId: 9606]}	4
SM_V7_3	Smp_343230	NA	Lachrymatory-factor synthase	8
SM_V7_3	Smp_346510	previous_stable_id=Smp_177110		8
SM_V7_3	Smp_346550	NA	NA	8
SM_V7_3	Smp_346960	previous_stable_id=Smp_171410,Smp_197430		8
SM_V7_3	Smp_347290	previous_stable_id=Smp_210210		8
SM_V7_3H001	Smp_326590	previous_stable_id=Smp_000370		12
SM_V7_3H003	Smp_320020	TPR_REGION domain-containing protein		20
SM_V7_4	Smp_008830	Putative helicase		4
SM_V7_4	Smp_009030	Ribonucleoside-diphosphate reductase		8
SM_V7_4	Smp_024820	Tetraspanin	Tetraspanin	8
SM_V7_4	Smp_036130	Phosducin-like protein (Phlp)		12
SM_V7_4	Smp_042160	Fructose-bisphosphate aldolase		8
SM_V7_4	Smp_042440	(Dihydro)ceramide Synthase (LAG1)		8
SM_V7_4	Smp_042560	Porin domain superfamily	Mitochondrial import receptor subunit TOM40	24
SM_V7_4	Smp_043030	Hexokinase	Hexokinase {Blood fluke (Schistosoma mansoni) [TaxId: 6183]}	8
SM_V7_4	Smp_043350	UDPGP family	UDP-N-acetylglucosamine pyrophosphorylase {Human (Homo sapiens), AGX1 [TaxId: 9606]}	8
SM_V7_4	Smp_049890	WD-repeat protein%2C putative		12
SM_V7_4	Smp_050080	NAD dehydrogenase%2C putative		20
SM_V7_4	Smp_050740	NA		8
SM_V7_4	Smp_050830	Putative synaptojanin		8

## Annexe 10

SM_V7_4	Smp_053140	Histone acetyltransferase		4
SM_V7_4	Smp_053510	Aspartyl/Asparaginyl-tRNA synthetase, class IIb	Aspartate--tRNA ligase, cytoplasmic (E.C.6.1.1.12)	20
SM_V7_4	Smp_053820	Putative preprotein translocase secy subunit (Sec61)		8
SM_V7_4	Smp_055220	Putative surface protein PspC		12
SM_V7_4	Smp_080100	Putative synaptophysin/synaptoporin		24
SM_V7_4	Smp_080180	Putative carcinoma associated protein hoj-1		12
SM_V7_4	Smp_080210	Putative lipid-binding protein		8
SM_V7_4	Smp_089200	Putative multidrug resistance protein 1%2C 2%2C 3 (P glycoprotein 1%2C 2%2C 3)		8
SM_V7_4	Smp_094930	Early growth response protein%2C putative		16
SM_V7_4	Smp_096750	40S ribosomal protein S13		20
SM_V7_4	Smp_100380	DNA replication complex GINS protein PSF2		4
SM_V7_4	Smp_104860	Ribonuclease p-related		12
SM_V7_4	Smp_136950	NA		20
SM_V7_4	Smp_142780	Putative retinoblastoma binding protein		4
SM_V7_4	Smp_145370	NA	NA	16
SM_V7_4	Smp_150020	NA		16
SM_V7_4	Smp_150160	Geranylgeranyl transferase type-2 subunit beta		8
SM_V7_4	Smp_151070	DRB sensitivity-inducing factor large subunit		12
SM_V7_4	Smp_169640	Nup54 domain-containing protein		24
SM_V7_4	Smp_170770	NA		12
SM_V7_4	Smp_211030	Protein phosphatase methylesterase 1		8
SM_V7_4	Smp_211040	Protein phosphatase methylesterase 1 (S33 family)		12
SM_V7_4	Smp_213200	Amidohydrolase-related	Amidohydrolase 2	4
SM_V7_4	Smp_213500	HotDog domain superfamily	Acyl-CoA thioesterase	20
SM_V7_4	Smp_247460	ACB domain-containing protein		24
SM_V7_4	Smp_247630	ARMC6-like protein		20
SM_V7_4	Smp_247860	Peptidase M8, leishmanolysin	Peptidase_M8	8
SM_V7_4	Smp_312440	Aldehyde dehydrogenase%2Cputative		24
SM_V7_4	Smp_313300	Small GTPase	Rab21 {Human (Homo sapiens) [TaxId: 9606]}	8
SM_V7_4	Smp_317120	NA		8
SM_V7_4	Smp_328940	NA		12
SM_V7_4	Smp_329280	NA		12
SM_V7_4	Smp_329700	Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit 2		12
SM_V7_4	Smp_334090	previous_stable_id=Smp_182770,Smp_038730		16
SM_V7_4	Smp_334160	Kinesin motor domain-containing protein		4
SM_V7_4	Smp_334180	Hydrogen voltage-gated channel 1		16
SM_V7_4	Smp_336100	previous_stable_id=Smp_170820		8

## Annexe 10

SM_V7_4	Smp_341080	previous_stable_id=Smp_067800		24
SM_V7_4	Smp_345010	previous_stable_id=Smp_158930		4
SM_V7_4	Smp_345910	previous_stable_id=Smp_169620		12
SM_V7_4	Smp_346530	previous_stable_id=Smp_177320		12
SM_V7_4H0 21	Smp_303590	Putative mitochondrial 39S ribosomal protein L45		20
SM_V7_4H0 34	Smp_325270	NA		16
SM_V7_5	Smp_006440	Heterogeneous nuclear ribonucleoprotein I (Hnrnp I)%2C putative		8
SM_V7_5	Smp_053040	Snf7-related		24
SM_V7_5	Smp_056760	Protein disulfide-isomerase		4
SM_V7_5	Smp_058140	Protein-L-isoaspartate O-methyltransferase		20
SM_V7_5	Smp_061570	Mediator of RNA polymerase II transcription subunit 19		12
SM_V7_5	Smp_064350	Putative vacuolar ATP synthase proteolipid subunit		4
SM_V7_5	Smp_064400	NA		12
SM_V7_5	Smp_094630	RRM domain-containing protein		8
SM_V7_5	Smp_146290	Protein kinase		24
SM_V7_5	Smp_160530	N2227-like	Carnosine N-methyltransferase (E.C.2.1.1.22)	8
SM_V7_5	Smp_165490	Calcineurin-like phosphoesterase domain, ApaH type	Protein phosphatase 2A catalytic subunit alpha isoform, PP2A-alpha {Human (Homo sapiens) [TaxId: 9606]}	4
SM_V7_5	Smp_247660	NA		8
SM_V7_5	Smp_266890	previous_stable_id=Smp_133440		24
SM_V7_5	Smp_314310	CN hydrolase domain-containing protein		20
SM_V7_5	Smp_314640	Peroxiredoxin%2C Prx1		16
SM_V7_5	Smp_315750	Ubiquitin-conjugating enzyme E2	Ubiquitin conjugating enzyme, UBC {Caenorhabditis elegans, E2 2 [TaxId: 6239]}	12
SM_V7_5	Smp_330180	NA		8
SM_V7_5	Smp_330240	NA		8
SM_V7_5	Smp_330380	NA		20
SM_V7_5	Smp_333340	previous_stable_id=Smp_205630,Smp_121070		8
SM_V7_5	Smp_334600	Phosphodiesterase		12
SM_V7_5	Smp_336210	previous_stable_id=Smp_174500		8
SM_V7_5	Smp_340920	Heme O synthase		8
SM_V7_5	Smp_342420	previous_stable_id=Smp_129220		8
SM_V7_6	Smp_001890	Venom allergen-like (VAL) 18 protein		8
SM_V7_6	Smp_002060	Venom allergen-like (VAL) 10 protein%3B Venom allergen-like protein 10		8
SM_V7_6	Smp_002070	Venom allergen-like protein 4		8
SM_V7_6	Smp_005290	Cytochrome C1%2C putative		8
SM_V7_6	Smp_009340	KRR1 small subunit processome component		4



## Annexe 10

SM_V7_6	Smp_016920	Zinc finger, RING-type	EL5 RING-H2 domain {Rice ( <i>Oryza sativa</i> ) [TaxId: 4530]}	16
SM_V7_6	Smp_024900	Putative retinoblastoma-like protein		4
SM_V7_6	Smp_052740	Putative gamma-soluble nsf attachment protein (Snap)		8
SM_V7_6	Smp_062080	Serpin%2C putative		8
SM_V7_6	Smp_063090	Aconitase A/isopropylmalate dehydratase small subunit, swivel domain	ACONITASE	12
SM_V7_6	Smp_067420	Splicing factor 3b%2C subunit 4%2Cputative		24
SM_V7_6	Smp_067430	Basic helix-loop-helix protein%2C putative		4
SM_V7_6	Smp_074330	PEST proteolytic signal-containing nuclear protein		12
SM_V7_6	Smp_074370	Putative pre-mRNA splicing factor		4
SM_V7_6	Smp_091750	NA		16
SM_V7_6	Smp_123090	Venom allergen-like (VAL) 19 protein		8
SM_V7_6	Smp_138850	Putative sh2a		8
SM_V7_6	Smp_141550	Cysteine-rich secretory protein-related	Cysteine-rich secretory protein (SteCRISP) {Chinese green tree viper ( <i>Trimeresurus stejnegeri</i> ) [TaxId: 39682]}	8
SM_V7_6	Smp_147110	Putative ankyrin repeat-containing		8
SM_V7_6	Smp_147160	Protein quiver		24
SM_V7_6	Smp_149400	Peptidase S8/S53 domain	Furin (E.C.3.4.21.75)	8
SM_V7_6	Smp_158600	Dolichyl-diphosphooligosaccharide--protein glycotransferase		4
SM_V7_6	Smp_162230	Microtubule-associated protein		16
SM_V7_6	Smp_165660	PKD_channel domain-containing protein		12
SM_V7_6	Smp_168730	Carbonic anhydrase		8
SM_V7_6	Smp_172190	Putative methyltransferase		16
SM_V7_6	Smp_175310	NA		12
SM_V7_6	Smp_179350	previous_stable_id=Smp_036760		12
SM_V7_6	Smp_202970	Smp_202970		8
SM_V7_6	Smp_267040	KH_dom_type_1 domain-containing protein		8
SM_V7_6	Smp_300070	Cysteine-rich secretory protein-related	Insect allergen 5 (AG5) {Yellow jacket ( <i>Vespula vulgaris</i> ), Ves v 5 [TaxId: 7454]}	8
SM_V7_6	Smp_305550	Serpin%2C putative		8
SM_V7_6	Smp_331270	NA		8
SM_V7_6	Smp_340060	Peptidase S1 domain-containing protein		12
SM_V7_6	Smp_341900	EF-hand domain-containing protein		12
SM_V7_6	Smp_341930	previous_stable_id=Smp_123250		8
SM_V7_6	Smp_347320	Cysteine-rich secretory protein-related	Cysteine-rich secretory protein (SteCRISP) {Chinese green tree viper ( <i>Trimeresurus stejnegeri</i> ) [TaxId: 39682]}	8
SM_V7_7	Smp_005720	Putative aquaporin-3		8

## Annexe 10

SM_V7_7	Smp_005910	TYR_PHOSPHATASE_2 domain-containing protein	8
SM_V7_7	Smp_010820	Peptidylamidoglycolate lyase	12
SM_V7_7	Smp_015020	Sodium/potassium-transporting ATPase subunit alpha	8
SM_V7_7	Smp_023010	ATPase inhibitor-like protein trans-spliced	24
SM_V7_7	Smp_033260	Family C48 unassigned peptidase (C48 family)	24
SM_V7_7	Smp_033380	NA	8
SM_V7_7	Smp_041430	14-3-3 protein%2C putative	8
SM_V7_7	Smp_041480	Rna recognition motif (Rrm) domain containing protein%2C putative	4
SM_V7_7	Smp_041650	Putative 40s ribosomal protein S27	20
SM_V7_7	Smp_140100	Putative double-stranded rna-binding protein zn72d	4
SM_V7_7	Smp_152960	previous_stable_id=Smp_057410,Smp_152970	12
SM_V7_7	Smp_153100	Putative potassium/sodium hyperpolarization-activated cyclic nucleotide-gated channel	12
SM_V7_7	Smp_166680	Putative sorting and assembly machinery (Sam50) protein	4
SM_V7_7	Smp_166690	NA	12
SM_V7_7	Smp_198360	Putative ankyrin repeat-containing	16
SM_V7_7	Smp_210370	Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit 2	8
SM_V7_7	Smp_212190	Acetyltransferase component of pyruvate dehydrogenase complex	8
SM_V7_7	Smp_267240	Palmitoyltransferase	8
SM_V7_7	Smp_318210	NA	8
SM_V7_7	Smp_318300	NA	8
SM_V7_7	Smp_318310	NA	8
SM_V7_7	Smp_332980	previous_stable_id=Smp_125220,Smp_005740	8
SM_V7_7	Smp_335530	previous_stable_id=Smp_194470	12
SM_V7_7	Smp_341850	previous_stable_id=Smp_121240	8
SM_V7_U005	Smp_317840	NA	12
SM_V7_ZW	Smp_005020	E3 ubiquitin-protein ligase CBL	24
SM_V7_ZW	Smp_006340	ER membrane protein complex subunit 4	20
SM_V7_ZW	Smp_007730	ER lumen protein-retaining receptor	20
SM_V7_ZW	Smp_011700	Receptor protein-tyrosine kinase	16
SM_V7_ZW	Smp_011720	NA	20
SM_V7_ZW	Smp_012200	Putative zinc finger protein	12
SM_V7_ZW	Smp_015390	Putative wd-repeat protein	24
SM_V7_ZW	Smp_015530	NA	20
SM_V7_ZW	Smp_015840	Putative zinc finger protein	4
SM_V7_ZW	Smp_017710	Putative rab	16
SM_V7_ZW	Smp_019630	N-acyl-L-amino-acid amidohydrolase	20
SM_V7_ZW	Smp_019820	NA	12

## Annexe 10

SM_V7_ZW	Smp_019840	DNA ligase	20
SM_V7_ZW	Smp_020550	Low-density lipoprotein receptor (Ldl)	8
SM_V7_ZW	Smp_021390	Subfamily M23B non-peptidase homologue (M23 family)	12
SM_V7_ZW	Smp_021460	Glutamine synthetase bacteria%2C putative	8
SM_V7_ZW	Smp_022340	Putative pdz and lim domain protein	8
SM_V7_ZW	Smp_022400	Glucose-6-phosphate isomerase	8
SM_V7_ZW	Smp_022470	Putative excision repair cross-complementing 1 ercc1	12
SM_V7_ZW	Smp_022500	NA	20
SM_V7_ZW	Smp_022700	Very-long-chain (3R)-3-hydroxyacyl-CoA dehydratase	12
SM_V7_ZW	Smp_025150	Sugar nucleotide epimerase related	4
SM_V7_ZW	Smp_025260	Prolyl-tRNA synthetase	20
SM_V7_ZW	Smp_026400	Thyrotroph embryonic factor related	12
SM_V7_ZW	Smp_028670	Putative carbonic anhydrase II (Carbonate dehydratase II)	8
SM_V7_ZW	Smp_030370	Calreticulin	24
SM_V7_ZW	Smp_030440	NA	12
SM_V7_ZW	Smp_031010	Putative dead box ATP-dependent RNA helicase	20
SM_V7_ZW	Smp_031310	40S ribosomal protein S26	20
SM_V7_ZW	Smp_032370	Putative short chain dehydrogenase	8
SM_V7_ZW	Smp_032420	Rhomboid-like protein	20
SM_V7_ZW	Smp_032920	Putative ob fold nucleic acid binding protein	20
SM_V7_ZW	Smp_032930	NA	12
SM_V7_ZW	Smp_032950	Putative calmodulin (CaM)	16
SM_V7_ZW	Smp_032970	Putative calmodulin	8
SM_V7_ZW	Smp_032980	NA	8
SM_V7_ZW	Smp_032990	Putative calmodulin-4 (Calcium-binding protein Dd112)	8
SM_V7_ZW	Smp_033000	Calcium-binding protein	8
SM_V7_ZW	Smp_033010	Similar to 16 kDa calcium-binding protein	8
SM_V7_ZW	Smp_033020	Putative anti-silencing protein	20
SM_V7_ZW	Smp_036550	NA	20
SM_V7_ZW	Smp_037700	Nuclear transport factor 2-like protein%3B Nuclear transport factor%2C putative	20
SM_V7_ZW	Smp_038410	Protein SYS1 homolog	24
SM_V7_ZW	Smp_044470	Autophagy_act_C domain-containing protein	20
SM_V7_ZW	Smp_044520	Putative vmac	12
SM_V7_ZW	Smp_044560	NA	20
SM_V7_ZW	Smp_046030	U6 snRNA-associated Sm-like protein LSm1	12
SM_V7_ZW	Smp_046980	Pre-mRNA-splicing factor 38	20
SM_V7_ZW	Smp_053430	NA	12
SM_V7_ZW	Smp_054120	Ribosome biogenesis protein NSA2 homologue	16

## Annexe 10

SM_V7_ZW	Smp_054410	Adenine phosphoribosyltransferase	20
SM_V7_ZW	Smp_057620	Polypeptide N-acetylgalactosaminyltransferase	16
SM_V7_ZW	Smp_057650	IST1 homolog	12
SM_V7_ZW	Smp_058170	Agenet-like domain-containing protein	4
SM_V7_ZW	Smp_058270	Putative amiloride-sensitive sodium channel	16
SM_V7_ZW	Smp_058380	Putative long-N-terminal adhesion GPCR	8
SM_V7_ZW	Smp_058690	Glutathione peroxidase	12
SM_V7_ZW	Smp_058700	Glutathione peroxidase	12
SM_V7_ZW	Smp_058780	Cir_N domain-containing protein	20
SM_V7_ZW	Smp_059290	Putative dual-specificity phosphatase	16
SM_V7_ZW	Smp_059920	NA	20
SM_V7_ZW	Smp_060210	Putative lipopolysaccharide-induced transcription factor regulating tumor necrosis factor alpha	8
SM_V7_ZW	Smp_061940	Putative adenylate kinase 1	8
SM_V7_ZW	Smp_061950	Putative u3 small nucleolar ribonucleoprotein	20
SM_V7_ZW	Smp_064800	Alba domain-containing protein	20
SM_V7_ZW	Smp_067510	NA	16
SM_V7_ZW	Smp_067590	NA	20
SM_V7_ZW	Smp_068300	Putative zinc finger protein	12
SM_V7_ZW	Smp_071770	Putative rnase h (70)	8
SM_V7_ZW	Smp_074450	NA	8
SM_V7_ZW	Smp_074560	NA	8
SM_V7_ZW	Smp_074570	NA	8
SM_V7_ZW	Smp_075370	NA	8
SM_V7_ZW	Smp_075400	NA	8
SM_V7_ZW	Smp_075500	NA	8
SM_V7_ZW	Smp_076010	NA	8
SM_V7_ZW	Smp_076390	Glycolipid transfer protein (Gltp)%2C putative	24
SM_V7_ZW	Smp_076600	Sox transcription factor	4
SM_V7_ZW	Smp_077280	NA	20
SM_V7_ZW	Smp_079960	Tubulin beta chain	24
SM_V7_ZW	Smp_080370	NA	16
SM_V7_ZW	Smp_080530	Putative replication protein A	20
SM_V7_ZW	Smp_080980	NA	16
SM_V7_ZW	Smp_082250	Peptidase_M14 domain-containing protein	12
SM_V7_ZW	Smp_084060	NA	8
SM_V7_ZW	Smp_085240	Diacylglycerol pyrophosphate (Phosphatidic acid) phosphatase 1	16
SM_V7_ZW	Smp_091010	Putative glyoxalase II (Hydroxyacylglutathione hydrolase)	16
SM_V7_ZW	Smp_091170	Ribosomal protein	12
SM_V7_ZW	Smp_091840	Bestrophin homolog	20
SM_V7_ZW	Smp_091850	Glucosamine-6-phosphate isomerase	20
SM_V7_ZW	Smp_092530	AP-1 complex subunit gamma	12
SM_V7_ZW	Smp_093790	CAI-2 protein%2C putative	12

## Annexe 10

SM_V7_ZW	Smp_093930	Receptor protein-tyrosine kinase	16
SM_V7_ZW	Smp_095130	Hydroxymethylbilane synthase	20
SM_V7_ZW	Smp_095150	Protein-synthesizing GTPase	16
SM_V7_ZW	Smp_096160	ADP-ribosylation factor-like 13%2C arl13	12
SM_V7_ZW	Smp_100350	NA	12
SM_V7_ZW	Smp_102240	Upf3 regulator of nonsense transcripts-like protein	12
SM_V7_ZW	Smp_103920	Axonemal dynein intermediate chain%2C putative	12
SM_V7_ZW	Smp_104450	NA	12
SM_V7_ZW	Smp_119110	NA	12
SM_V7_ZW	Smp_120380	Testis specific leucine rich repeat protein%2Cputative	12
SM_V7_ZW	Smp_121130	Mitochondrial carrier protein%2C putative	16
SM_V7_ZW	Smp_121160	Mitochondrial carrier protein%2C putative	20
SM_V7_ZW	Smp_125550	Protein RIC1 homolog	16
SM_V7_ZW	Smp_126370	RPB5 homolog	12
SM_V7_ZW	Smp_128430	NA	8
SM_V7_ZW	Smp_128650	Nuclear movement protein related	12
SM_V7_ZW	Smp_132020	Cell division cycle 20 (Fizzy)-related	24
SM_V7_ZW	Smp_132090	Putative wd40 protein	8
SM_V7_ZW	Smp_133770	Putative glutamine synthetase bacteria	8
SM_V7_ZW	Smp_133800	NA	12
SM_V7_ZW	Smp_133830	NA	12
SM_V7_ZW	Smp_133900	NA	12
SM_V7_ZW	Smp_134140	Phosphodiesterase	12
SM_V7_ZW	Smp_134390	Acyl-CoA: cholesterol acyltransferase protein isotype 1.1%3B O-acyltransferase	16
SM_V7_ZW	Smp_135250	NA	12
SM_V7_ZW	Smp_135270	NA	12
SM_V7_ZW	Smp_135840	NA	20
SM_V7_ZW	Smp_135900	Putative dna double-strand break repair rad50 ATPase	12
SM_V7_ZW	Smp_135960	NA	8
SM_V7_ZW	Smp_136010	NA	4
SM_V7_ZW	Smp_136530	Palmitoyltransferase	24
SM_V7_ZW	Smp_137610	Tyrosine kinase	16
SM_V7_ZW	Smp_139030	NA	12
SM_V7_ZW	Smp_141830	Transcriptional enhancer factor (Tef) related	16
SM_V7_ZW	Smp_142150	NA	8
SM_V7_ZW	Smp_142160	NA	8
SM_V7_ZW	Smp_143130	Putative four and A half lim domains	12
SM_V7_ZW	Smp_143730	Putative carboxypeptidase regulatory region-containing	24
SM_V7_ZW	Smp_145550	Sister chromatid cohesion protein DCC1	20

## Annexe 10

SM_V7_ZW	Smp_145580	Putative progesterone-induced-blocking factor	16
SM_V7_ZW	Smp_145640	Putative forkhead protein/ forkhead protein domain	12
SM_V7_ZW	Smp_145810	Poly [ADP-ribose] polymerase	12
SM_V7_ZW	Smp_147420	EGF-like domain-containing protein	16
SM_V7_ZW	Smp_147520	Putative tripartite motif protein	12
SM_V7_ZW	Smp_150340	Monocarboxylate transporter%2C putative	8
SM_V7_ZW	Smp_152370	Putative arf gtpase-activating protein	24
SM_V7_ZW	Smp_153660	Cyclin d%2C putative	4
SM_V7_ZW	Smp_157100	Sphingoid long chain base kinase	8
SM_V7_ZW	Smp_157910	NA	12
SM_V7_ZW	Smp_157930	Putative leukocyte receptor cluster (Lrc) member 4 protein	8
SM_V7_ZW	Smp_158030	Intu_longin_1 domain-containing protein	20
SM_V7_ZW	Smp_158720	Protein kinase	12
SM_V7_ZW	Smp_162380	Isoamyl acetate-hydrolyzing esterase 1	12
SM_V7_ZW	Smp_163300	m7GpppX diphosphatase	4
SM_V7_ZW	Smp_163820	Phospholipid-transporting ATPase	12
SM_V7_ZW	Smp_164760	NA	12
SM_V7_ZW	Smp_165230	Polycystin 1-related	8
SM_V7_ZW	Smp_165470	Receptor protein-tyrosine kinase	16
SM_V7_ZW	Smp_165550	Afg3-like protein 2 (M41 family)	20
SM_V7_ZW	Smp_165580	General transcription and DNA repair factor IIH helicase subunit XPB	12
SM_V7_ZW	Smp_165780	Glycine cleavage system H protein	20
SM_V7_ZW	Smp_165820	NA	12
SM_V7_ZW	Smp_166500	Phosphatidylcholine-sterol acyltransferase (Lecithin-cholesterol acyltransferase)/ Phospholipase A	12
SM_V7_ZW	Smp_166600	Ubiquitin conjugating enzyme variant%3B Ubiquitin conjugating enzyme%2C putative	20
SM_V7_ZW	Smp_166610	previous_stable_id=Smp_166620,Smp_166630	8
SM_V7_ZW	Smp_171640	previous_stable_id=Smp_171650	12
SM_V7_ZW	Smp_171670	Putative wd-repeat protein	20
SM_V7_ZW	Smp_171770	NA	20
SM_V7_ZW	Smp_171860	Kh-domain rna binding protein-related	16
SM_V7_ZW	Smp_172280	Serine/threonine protein phosphatase 2a regulatory subunit A%2C putative	8
SM_V7_ZW	Smp_173910	Nuclear movement protein related	12
SM_V7_ZW	Smp_174870	NA	12
SM_V7_ZW	Smp_174950	Putative 40s ribosomal protein S19	20
SM_V7_ZW	Smp_175090	Long-chain-fatty-acid--CoA ligase	12
SM_V7_ZW	Smp_176090	Fbxl4%2C putative	12
SM_V7_ZW	Smp_177790	Iron-sulfur cluster scaffold protein nfu-related	12
SM_V7_ZW	Smp_178750	Transcription initiation factor IIE subunit alpha	20

## Annexe 10

SM_V7_ZW	Smp_179110	Purine nucleoside phosphorylase	8
SM_V7_ZW	Smp_179720	Oxysterol-binding protein	12
SM_V7_ZW	Smp_180000	40S ribosomal protein S9	16
SM_V7_ZW	Smp_180360	NA	20
SM_V7_ZW	Smp_190020	NA	12
SM_V7_ZW	Smp_194240	SPRY domain-containing protein 7	12
SM_V7_ZW	Smp_194610	Serine/threonine kinase	8
SM_V7_ZW	Smp_194620	tRNA (adenine(58)-N(1))-methyltransferase catalytic subunit TRMT61A	4
SM_V7_ZW	Smp_195060	NA	8
SM_V7_ZW	Smp_195070	NA	8
SM_V7_ZW	Smp_196110	Ceramidase	12
SM_V7_ZW	Smp_198850	ESCRT-II complex subunit VPS22	12
SM_V7_ZW	Smp_198890	previous_stable_id=Smp_150330	8
SM_V7_ZW	Smp_201220	Non-specific serine/threonine protein kinase	12
SM_V7_ZW	Smp_203500	G_PROTEIN_RECEP_F1_2 domain-containing protein	8
SM_V7_ZW	Smp_203630	NA	16
SM_V7_ZW	Smp_203830	NA	20
SM_V7_ZW	Smp_203970	Calcium load-activated calcium channel	20
SM_V7_ZW	Smp_204120	NA	12
SM_V7_ZW	Smp_204380	Smp_204380	16
SM_V7_ZW	Smp_204390	Smp_204390	8
SM_V7_ZW	Smp_209040	Long-chain-fatty-acid--CoA ligase	8
SM_V7_ZW	Smp_210520	Putative tyrosyl-DNA phosphodiesterase	4
SM_V7_ZW	Smp_210860	Putative dystrobrevin	12
SM_V7_ZW	Smp_210900	NA	12
SM_V7_ZW	Smp_214240	previous_stable_id=Smp_011600,Smp_128400	12
SM_V7_ZW	Smp_241570	previous_stable_id=Smp_154950	8
SM_V7_ZW	Smp_242500	previous_stable_id=Smp_155410	8
SM_V7_ZW	Smp_244130	Elongation factor G%2C mitochondrial	12
SM_V7_ZW	Smp_244310	previous_stable_id=Smp_168630,Smp_168640	8
SM_V7_ZW	Smp_244350	Signal recognition particle 14 kDa protein	20
SM_V7_ZW	Smp_244640	previous_stable_id=Smp_094500	8
SM_V7_ZW	Smp_244650	previous_stable_id=Smp_094500	8
SM_V7_ZW	Smp_244920	previous_stable_id=Smp_179020,Smp_179030	12
SM_V7_ZW	Smp_244960	previous_stable_id=Smp_171960	8
SM_V7_ZW	Smp_245040	previous_stable_id=Smp_174910	16
SM_V7_ZW	Smp_245140	MFS domain-containing protein	8
SM_V7_ZW	Smp_245150	MFS domain-containing protein	12
SM_V7_ZW	Smp_245160	previous_stable_id=Smp_134990	8
SM_V7_ZW	Smp_245370	NA	16
SM_V7_ZW	Smp_245400	NA	12
SM_V7_ZW	Smp_246330	previous_stable_id=Smp_166520	8
SM_V7_ZW	Smp_266790	NUDE_C domain-containing protein	16

## Annexe 10

SM_V7_ZW	Smp_267110	Ribosomal RNA-processing protein 40	4
SM_V7_ZW	Smp_267120	previous_stable_id=Smp_194600	4
SM_V7_ZW	Smp_303980	previous_stable_id=Smp_153230,Smp_057790,Smp_153240	8
SM_V7_ZW	Smp_304250	Purine nucleoside phosphorylase	4
SM_V7_ZW	Smp_304260	Purine nucleoside phosphorylase	24
SM_V7_ZW	Smp_304710	previous_stable_id=Smp_167940	12
SM_V7_ZW	Smp_304800	Mannosyltransferase	12
SM_V7_ZW	Smp_305120	G domain-containing protein	12
SM_V7_ZW	Smp_306610	previous_stable_id=Smp_130420,Smp_015190	8
SM_V7_ZW	Smp_307260	Protein kinase domain-containing protein	8
SM_V7_ZW	Smp_314290	previous_stable_id=Smp_176530	8
SM_V7_ZW	Smp_315980	previous_stable_id=Smp_171890	12
SM_V7_ZW	Smp_321810	NA	8
SM_V7_ZW	Smp_322220	60S ribosomal protein L18a	20
SM_V7_ZW	Smp_322240	NA	12
SM_V7_ZW	Smp_322560	NA	20
SM_V7_ZW	Smp_322590	NA	12
SM_V7_ZW	Smp_322650	NA	12
SM_V7_ZW	Smp_322820	LIM zinc-binding domain-containing protein	20
SM_V7_ZW	Smp_322830	LIM zinc-binding domain-containing protein	8
SM_V7_ZW	Smp_322850	NA	24
SM_V7_ZW	Smp_322970	Endo/exonuclease/phosphatase domain-containing protein	12
SM_V7_ZW	Smp_323230	previous_stable_id=Smp_136600	8
SM_V7_ZW	Smp_323280	NA	12
SM_V7_ZW	Smp_323440	NA	8
SM_V7_ZW	Smp_323900	NA	8
SM_V7_ZW	Smp_324080	ATP-binding domain-containing protein 4	12
SM_V7_ZW	Smp_324230	NA	8
SM_V7_ZW	Smp_332890	previous_stable_id=Smp_137550,Smp_137560	16
SM_V7_ZW	Smp_332910	previous_stable_id=Smp_198850,Smp_130400,Smp_198650,Smp_139110	16
SM_V7_ZW	Smp_332930	G_PROTEIN_RECEP_F2_4 domain-containing protein	12
SM_V7_ZW	Smp_333080	RRM domain-containing protein	12
SM_V7_ZW	Smp_335130	NA	8
SM_V7_ZW	Smp_337330	60S ribosomal protein L21	20
SM_V7_ZW	Smp_337390	Ribosome biogenesis regulatory protein	12
SM_V7_ZW	Smp_340410	SMP-LTD domain-containing protein	16
SM_V7_ZW	Smp_340960	previous_stable_id=Smp_062720	8
SM_V7_ZW	Smp_341550	Guanine nucleotide exchange factor mss4%2Cputative	12
SM_V7_ZW	Smp_342910	Ribosome assembly factor mrt4	16



## Annexe 10

SM_V7_ZW	Smp_343350	Ubiquitin-like domain-containing protein		12
SM_V7_ZW	Smp_343530	previous_stable_id=Smp_142140		8
SM_V7_ZW	Smp_343890	previous_stable_id=Smp_145870,Smp_203570		20
SM_V7_ZW	Smp_343970	Tektin		12
SM_V7_ZW	Smp_344840	previous_stable_id=Smp_157130		8
SM_V7_ZW	Smp_345200	previous_stable_id=Smp_160350		4
SM_V7_ZW	Smp_345560	Autophagy protein 5		12
SM_V7_ZW	Smp_345770	Phosphatase and actin regulator%2C putative		12
SM_V7_ZW	Smp_346050	zf-CHCC domain-containing protein		16
SM_V7_ZW	Smp_346340	RRM domain-containing protein		20
SM_V7_ZW	Smp_346810	previous_stable_id=Smp_192180		12
SM_V7_ZW	Smp_347710	previous_stable_id=Smp_153180,Smp_057710		12
SM_V7_ZW H005	Smp_330930	previous_stable_id=Smp_190330		8
SM_V7_ZW H005	Smp_330950	NA		8
SM_V7_ZW H005	Smp_330960	NA		8
SM_V7_ZW H006	Smp_330990	NA		8
SM_V7_ZW H010	Smp_316820	previous_stable_id=Smp_075500		8
SM_V7_ZW H019	Smp_321350	NA		24
SM_V7_ZW H041	Smp_325450	NA		12

Chr. : chromosome.

## Annexe 11 : Liste des gènes nocturnes rythmiques

Chr.	Gene ID	GFF annotation	HHsearch annotation	H
SM_V7_1	Smp_003970	Putative elongation factor-1 beta%2Cdelta		24
SM_V7_1	Smp_006860	PDZ domain-containing protein		4
SM_V7_1	Smp_019280	Cyclic nucleotide-binding domain	Protein kinase, cAMP-dependent, catalytic, alpha	4
SM_V7_1	Smp_032260	Ribosomal protein L15e	uL2, uS2, uL15, eS26, uL3	24
SM_V7_1	Smp_033870	NA		16
SM_V7_1	Smp_034550	Putative alpha-actinin		4
SM_V7_1	Smp_034860	Putative nuclear receptor		4
SM_V7_1	Smp_035180	Putative voltage-gated potassium channel		4
SM_V7_1	Smp_035210	Myosin_tail_1 domain-containing protein		4
SM_V7_1	Smp_036180	Diacylglycerol kinase	NA	4
SM_V7_1	Smp_040360	Serologically defined breast cancer antigen ny-br-84-related		12
SM_V7_1	Smp_040970	V-type proton ATPase subunit a		4
SM_V7_1	Smp_049270	Putative heat shock protein		4
SM_V7_1	Smp_054620	NA		12
SM_V7_1	Smp_055450	Late endosomal/lysosomal adaptor and MAPK and MTOR activator 4		12
SM_V7_1	Smp_056280	NA		16
SM_V7_1	Smp_059780	Putative rab6-interacting protein 2 (ERC protein 1)		4
SM_V7_1	Smp_060030	NA		4
SM_V7_1	Smp_060480	C2 domain	Copine	4
SM_V7_1	Smp_073080	Putative rRNA processing protein (Rrp20)		24
SM_V7_1	Smp_073270	C-terminal peptide		4
SM_V7_1	Smp_079400	Putative goliath E3 ubiquitin ligase		4
SM_V7_1	Smp_081190	Putative rho3 GTPase%3B Rho3 GTPase		4
SM_V7_1	Smp_083650	Cys/Met metabolism, pyridoxal phosphate-dependent enzyme	Methionine gamma-lyase, MGL {Pseudomonas putida [TaxId: 303]}	24
SM_V7_1	Smp_083680	Ribosome biogenesis protein Nop53/GLTSCR2	Nop53	24
SM_V7_1	Smp_086830	Putative spfh domain / Band 7 family protein		4
SM_V7_1	Smp_087760	Ferritin-1 heavy chain		24
SM_V7_1	Smp_094880	1%2C4-alpha-glucan branching enzyme		4
SM_V7_1	Smp_094890	Rho GTPase-activating protein domain	Beta-chimaerin, C-terminal domain {Human (Homo sapiens) [TaxId: 9606]}	4
SM_V7_1	Smp_097800	Y-box binding protein%3B Y-box binding protein homolog		24
SM_V7_1	Smp_097940	Serine/threonine kinase		4
SM_V7_1	Smp_099850	Putative divalent metal transporter DMT1B		4
SM_V7_1	Smp_104660	Ribonucleoprotein		24
SM_V7_1	Smp_105020	Immunoglobulin subtype 2	Titin	4
SM_V7_1	Smp_123860	Putative neurexin		4
SM_V7_1	Smp_123910	Putative ift20		4
SM_V7_1	Smp_126050	Cadherin-like	Protocadherin-15	4
SM_V7_1	Smp_126140	Integrin alpha chain	Integrin alpha-V, Integrin beta-3	4

## Annexe 11

SM_V7_1	Smp_126540	Endostatin domain-containing protein		4
SM_V7_1	Smp_126620	NA		4
SM_V7_1	Smp_128920	Putative spectrin beta chain		4
SM_V7_1	Smp_129060	Putative solute carrier family 17%2C member 7 (Vesicular glutamate transporter)		4
SM_V7_1	Smp_129840	DUF5735 domain-containing protein		4
SM_V7_1	Smp_131870	Putative syntaxin binding protein-1%2C%2C3		4
SM_V7_1	Smp_132700	Immunoglobulin subtype 2	CARCINOEMBRYONIC ANTIGEN	4
SM_V7_1	Smp_139480	Protein kinase domain	ephb2 receptor tyrosine kinase {Mouse (Mus musculus) [TaxId: 10090]}	4
SM_V7_1	Smp_139520	SH3 domain	Hypothetical protein FLJ21935	4
SM_V7_1	Smp_140250	G protein-coupled receptor, rhodopsin-like	Platelet-activating factor receptor, Flavodoxin	4
SM_V7_1	Smp_140510	Putative transmembrane protein 56		16
SM_V7_1	Smp_140520	Putative ank repeat-containing		4
SM_V7_1	Smp_144210	Putative death associated protein kinase		4
SM_V7_1	Smp_144240	NA		4
SM_V7_1	Smp_149660	DRBM domain-containing protein		4
SM_V7_1	Smp_152060	NA		4
SM_V7_1	Smp_152830	Putative sodium/calcium exchanger		4
SM_V7_1	Smp_154670	Glycosyl transferase, family 14	beta-1,6-N-acetylglucosaminyltransferase (E.C.2.4.1.102)	12
SM_V7_1	Smp_155290	Rlla domain-containing protein		4
SM_V7_1	Smp_157620	NA		4
SM_V7_1	Smp_158450	BTB/POZ domain	S-phase kinase-associated protein 1, F-box/LRR-repeat	4
SM_V7_1	Smp_158480	AMP-dependent synthetase/ligase	Acetyl-CoA synthetase {Baker's yeast (Saccharomyces cerevisiae) [TaxId: 4932]}	4
SM_V7_1	Smp_159730	Protein kinase		4
SM_V7_1	Smp_159780	NA		4
SM_V7_1	Smp_159790	Putative kif1		4
SM_V7_1	Smp_161510	NA		4
SM_V7_1	Smp_161680	Phosphorylase b kinase regulatory subunit		4
SM_V7_1	Smp_163080	Diacylglycerol kinase		4
SM_V7_1	Smp_163420	Protein kinase		4
SM_V7_1	Smp_163430	C2H2-type domain-containing protein		24
SM_V7_1	Smp_163500	SH2 domain-containing protein		4
SM_V7_1	Smp_163570	Ryanodine receptor related		4
SM_V7_1	Smp_165940	DH domain-containing protein		4
SM_V7_1	Smp_166310	Putative ran gtpase-activating protein		20
SM_V7_1	Smp_167200	Putative ATPase		24
SM_V7_1	Smp_170290	Neuronal differentiation protein		4
SM_V7_1	Smp_171010	Putative ef-hand domain (C-terminal) containing protein		16
SM_V7_1	Smp_171020	NA		4
SM_V7_1	Smp_172040	Putative unc-13 (Munc13)		4
SM_V7_1	Smp_180140	G protein-coupled receptor, rhodopsin-like	P2Y purinoceptor 1, Rubredoxin	12

## Annexe 11

SM_V7_1	Smp_181150	Putative abc transporter		4
SM_V7_1	Smp_194780	Neurotracting/Isamp/neurotrimin/obcam related cell adhesion molecule		4
SM_V7_1	Smp_196300	Putative rhogef and pleckstrin domain protein		4
SM_V7_1	Smp_196400	Putative sphingosine phosphate lyase		4
SM_V7_1	Smp_198930	previous_stable_id=Smp_125000		4
SM_V7_1	Smp_200450	Hexosyltransferase		4
SM_V7_1	Smp_201130	Tumor protein p53-inducible protein 11	p53-inducible11	12
SM_V7_1	Smp_210820	GPCR kinase	Beta-adrenergic receptor kinase 1 (E.C.2.7.11.15)	4
SM_V7_1	Smp_213180	Putative 12-lipoxygenase		4
SM_V7_1	Smp_213740	previous_stable_id=Smp_162050		4
SM_V7_1	Smp_241950	HMG-CoA synthase		4
SM_V7_1	Smp_245820	Kettin/titin-related protein		4
SM_V7_1	Smp_246030	tRNA-intron lyase		12
SM_V7_1	Smp_246240	previous_stable_id=Smp_160900,Smp_160910,Smp_160890		4
SM_V7_1	Smp_246380	previous_stable_id=Smp_012680		4
SM_V7_1	Smp_247030	Putative protocadherin alpha		4
SM_V7_1	Smp_301990	previous_stable_id=Smp_180750,Smp_180760		4
SM_V7_1	Smp_303260	previous_stable_id=Smp_142230		24
SM_V7_1	Smp_305430	G_PROTEIN_RECEP_F3_4 domain-containing protein		4
SM_V7_1	Smp_310210	previous_stable_id=Smp_031530,Smp_139210		4
SM_V7_1	Smp_315690	previous_stable_id=Smp_120050,Smp_199230		4
SM_V7_1	Smp_315760	C1q domain-containing protein	Nucleoid-associated protein Rv3716c	4
SM_V7_1	Smp_319130	FCH domain	SLIT-ROBO Rho GTPase-activating protein 2	4
SM_V7_1	Smp_319330	SAM-dependent methyltransferase RsmB/NOP2-type	60S ribosomal protein L13-A, 60S	12
SM_V7_1	Smp_323010	NA		4
SM_V7_1	Smp_323430	NA		24
SM_V7_1	Smp_328690	NA		4
SM_V7_1	Smp_332470	B-box-type zinc finger	E3 UBIQUITIN/ISG15 LIGASE TRIM25 (E.C.6.3.2.19)	4
SM_V7_1	Smp_332520	RIB43A	Tubulin beta, Tubulin alpha, PACRG	4
SM_V7_1	Smp_332600	MAM domain-containing protein		4
SM_V7_1	Smp_333140	previous_stable_id=Smp_180980,Smp_168870		4
SM_V7_1	Smp_333210	previous_stable_id=Smp_175280,Smp_179700		4
SM_V7_1	Smp_335630	Tetraspanin		4
SM_V7_1	Smp_337040	Ribosomal protein L11/L12	28S ribosomal protein S2, mitochondrial	24
SM_V7_1	Smp_340760	Tropomyosin-1		4
SM_V7_1	Smp_341660	Thiol oxidase		4

## Annexe 11

SM_V7_1	Smp_342080	Stonin homology	Second domain of Mu2 adaptin subunit (ap50) of ap2 adaptor {Rat (Rattus norvegicus) [TaxId: 10116]}	4
SM_V7_1	Smp_342950	ARID domain-containing protein		4
SM_V7_1	Smp_343540	I/LWEQ domain	I_LWEQ	4
SM_V7_1	Smp_343700	SH2 domain-containing protein		4
SM_V7_1	Smp_346630	Membrane associated guanylate kinase inverted related		4
SM_V7_1003	Smp_165590	NA		12
SM_V7_2	Smp_004190	Putative cationic amino acid transporter		4
SM_V7_2	Smp_009560	BTB domain-containing protein		4
SM_V7_2	Smp_016340	Mediator of RNA polymerase II transcription subunit 8		12
SM_V7_2	Smp_021070	Putative gtp-binding protein-animal		24
SM_V7_2	Smp_021100	SAM domain-containing protein		4
SM_V7_2	Smp_024520	Rab GDP dissociation inhibitor		4
SM_V7_2	Smp_030690	Elongation factor 1-beta%2C putative		24
SM_V7_2	Smp_045360	NA		4
SM_V7_2	Smp_047700	Putative ring finger protein 11 (Sid 1669) (Nedd4 ww domain-binding protein 2)		4
SM_V7_2	Smp_048980	Putative multiple pdz domain protein		4
SM_V7_2	Smp_063680	NA		4
SM_V7_2	Smp_065180	Eh domain containing/past-1-related		4
SM_V7_2	Smp_089430	60S ribosomal protein L40		20
SM_V7_2	Smp_122810	Putative spfh domain / Band 7 family protein		4
SM_V7_2	Smp_122950	NA		4
SM_V7_2	Smp_125480	PDZ domain-containing protein		4
SM_V7_2	Smp_125510	Putative cadherin		4
SM_V7_2	Smp_128170	NA		4
SM_V7_2	Smp_128750	Putative titin		4
SM_V7_2	Smp_129670	Apoptosis stimulating of P53%2C putative		4
SM_V7_2	Smp_130160	Sema domain-containing protein		4
SM_V7_2	Smp_131050	Putative camp-dependent protein kinase type II regulatory subunit		4
SM_V7_2	Smp_131690	MAGUK homolog		4
SM_V7_2	Smp_133580	NA		4
SM_V7_2	Smp_142520	Palmitoyltransferase		12
SM_V7_2	Smp_146580	Putative homeobox protein knotted-1		4
SM_V7_2	Smp_147260	Ras-like family		4
SM_V7_2	Smp_148400	Putative calmodulin-binding transcription activator (Camta)		4
SM_V7_2	Smp_148460	Putative neurofibromin		4
SM_V7_2	Smp_154850	Putative rap gtpase-activating protein		4
SM_V7_2	Smp_156400	Putative cytochrome P450		4
SM_V7_2	Smp_156720	Putative o-linked n-acetylglucosamine transferase%2C ogt		4
SM_V7_2	Smp_171580	Aromatic-L-amino-acid decarboxylase (AADC) (DOPA decarboxylase) (DDC)%2Cputative		4
SM_V7_2	Smp_207080	Proteasome subunit alpha type-6		24
SM_V7_2	Smp_213010	DFP domain-containing protein		12

## Annexe 11

SM_V7_2	Smp_242110	Putative sam/hd domain protein		4
SM_V7_2	Smp_245580	Putative cadherin		4
SM_V7_2	Smp_245680	Nephrin related		4
SM_V7_2	Smp_245840	previous_stable_id=Smp_148310,Smp_148300		4
SM_V7_2	Smp_245970	previous_stable_id=Smp_140660,Smp_140670		4
SM_V7_2	Smp_246050	Vinculin		4
SM_V7_2	Smp_310160	previous_stable_id=Smp_208030,Smp_146950		4
SM_V7_2	Smp_310170	previous_stable_id=Smp_208030,Smp_146950		4
SM_V7_2	Smp_310430	previous_stable_id=Smp_016230,Smp_131010		4
SM_V7_2	Smp_310550	previous_stable_id=Smp_143620,Smp_143610		4
SM_V7_2	Smp_310620	previous_stable_id=Smp_139300,Smp_139310,Smp_072030		4
SM_V7_2	Smp_310650	previous_stable_id=Smp_161180		4
SM_V7_2	Smp_310770	Serine/threonine kinase		8
SM_V7_2	Smp_327250	NA		20
SM_V7_2	Smp_328730	NA		4
SM_V7_2	Smp_333920	previous_stable_id=Smp_147550		4
SM_V7_2	Smp_335360	BPTI/Kunitz inhibitor domain-containing protein		4
SM_V7_2	Smp_337610	previous_stable_id=Smp_206120		4
SM_V7_2	Smp_340270	previous_stable_id=Smp_017350		4
SM_V7_2	Smp_340770	Putative zinc finger protein		12
SM_V7_2	Smp_344000	previous_stable_id=Smp_147810		4
SM_V7_2	Smp_346250	Putative anosmin-1 (Kallmann syndrome protein) (Adhesion molecule-like X-linked)		4
SM_V7_2	Smp_346270	alpha-1%2C2-Mannosidase		4
SM_V7_2	Smp_347740	Coiled-coil domain-containing protein 39		4
SM_V7_3	Smp_000100	Filamin/ABP280 repeat	FILAMIN-B	4
SM_V7_3	Smp_000620	Myc-type, basic helix-loop-helix (bHLH) domain	DNA-binding protein inhibitor ID-1	4
SM_V7_3	Smp_007890	DNA-directed RNA polymerase III subunit RPC9		4
SM_V7_3	Smp_009990	Receptor L-domain	Insulin receptor (E.C.2.7.10.1), Insulin	4
SM_V7_3	Smp_018260	Glycogen [starch] synthase		4
SM_V7_3	Smp_018790	PPM-type phosphatase domain	Probable protein phosphatase 2C 50	4
SM_V7_3	Smp_021340	Myc-type, basic helix-loop-helix (bHLH) domain	Mad protein {Human (Homo sapiens) [TaxId: 9606]}	4
SM_V7_3	Smp_023620	FERM domain	FERM, RhoGEF and pleckstrin domain	4
SM_V7_3	Smp_026230	NA	Snake coagglutinin alpha chain {Habu snake (Trimeresurus flavoviridis), flavocetin-A [TaxId: 88087]}	4
SM_V7_3	Smp_030010	CCDC92/74, coiled-coil domain	CCDC92	4

## Annexe 11

SM_V7_3	Smp_063250	Ubiquinol-cytochrome c chaperone, CBP3	Ubiquinol-cytochrome C chaperone	24
SM_V7_3	Smp_072900	Hsp90 co-chaperone (Tebp)%2C putative		24
SM_V7_3	Smp_079840	Tetrapyrrole biosynthesis, uroporphyrinogen III synthase	Uroporphyrinogen III synthase (U3S, HemD) {Human (Homo sapiens) [TaxId: 9606]}	4
SM_V7_3	Smp_087690	Protease inhibitor I35 (TIMP)	Metalloproteinase inhibitor 1, Stromelysin-2 (E.C.3.4.24.22)	4
SM_V7_3	Smp_098960	60S ribosomal protein L26%2C putative%3B 60S ribosomal protein L26-like protein		24
SM_V7_3	Smp_123630	Peptidase C19, ubiquitin carboxyl-terminal hydrolase	TATA-binding protein, Transcriptional coactivator HFI1/ADA1	4
SM_V7_3	Smp_126240	Protein kinase domain	Twitchin, kinase domain {California sea hare (Aplysia californica), twk43 [TaxId: 6500]}	4
SM_V7_3	Smp_129360	Cyclin, N-terminal	Cell division protein kinase 4	4
SM_V7_3	Smp_132320	Cadherin-like	Protocadherin-15	4
SM_V7_3	Smp_132410	G protein-coupled receptor, rhodopsin-like	Lysozyme,Proteinase-activated receptor 2,Soluble cytochrome b562,Proteinase-activated	4
SM_V7_3	Smp_134910	Protein kinase domain	Aurora-related kinase 1 (aurora-2) {Human (Homo sapiens) [TaxId: 9606]}	4
SM_V7_3	Smp_138430	DEP domain	Regulator of G-protein signaling 7	4
SM_V7_3	Smp_138510	NA	NA	4
SM_V7_3	Smp_145890	Thrombospondin type-1 (TSP1) repeat	PROPERDIN	4
SM_V7_3	Smp_150220	PTP type protein phosphatase	Protein-tyrosine phosphatase alpha {Mouse (Mus musculus) [TaxId: 10090]}	4
SM_V7_3	Smp_156280	Hyccin	Phosphatidylinositol 4-kinase III alpha (PI4KA)	4
SM_V7_3	Smp_159840	Dynein heavy chain, C-terminal domain	Dynein heavy chain, cytoplasmic	4
SM_V7_3	Smp_159990	Voltage-dependent L-type calcium channel subunit alpha		4
SM_V7_3	Smp_162520	Cadherin-like	EP-cadherin, ectodomain	4
SM_V7_3	Smp_162610	Cadherin-like	Protocadherin-15	12
SM_V7_3	Smp_162660	Carbon-nitrogen hydrolase	Glutamine-dependent NAD(+) synthetase (E.C.6.3.5.1)	12
SM_V7_3	Smp_169700	LIM zinc-binding domain-containing protein		4
SM_V7_3	Smp_175510	NA		4
SM_V7_3	Smp_176980	Putative beta-arrestin 1		12
SM_V7_3	Smp_196090	NA	Protein patched homolog 1, Sonic	4
SM_V7_3	Smp_196810	Dihydropyrimidinase related protein-3 (M38 family)		4
SM_V7_3	Smp_197640	Voltage-dependent calcium channel		4
SM_V7_3	Smp_212410	Cyclic nucleotide-binding domain	Potassium/sodium hyperpolarization-activated cyclic nucleotide-gated channel	4

## Annexe 11

SM_V7_3	Smp_212710	Gynecophoral canal protein		4
SM_V7_3	Smp_246890	previous_stable_id=Smp_137830,Smp_137820		4
SM_V7_3	Smp_246950	PTP type protein phosphatase	Tyrosine-protein phosphatase, non-receptor type 13 (PTPL1) {Human (Homo sapiens) [TaxId: 9606]}	4
SM_V7_3	Smp_247230	Putative single-stranded DNA-binding protein mssp-1		4
SM_V7_3	Smp_247260	IPT/TIG domain-containing protein		4
SM_V7_3	Smp_307780	previous_stable_id=Smp_193230,Smp_176470,Smp_199910		12
SM_V7_3	Smp_307880	VWFA domain-containing protein		4
SM_V7_3	Smp_307960	MAP kinase-activating death domain protein		4
SM_V7_3	Smp_308440	NA	NA	4
SM_V7_3	Smp_308820	Leucine-rich repeat	Myeloperoxidase (E.C.1.11.2.2)	4
SM_V7_3	Smp_309510	Major facilitator superfamily domain	Melibiose carrier protein	12
SM_V7_3	Smp_311920	FHA domain-containing protein		4
SM_V7_3	Smp_315360	previous_stable_id=Smp_159360		4
SM_V7_3	Smp_316660	previous_stable_id=Smp_176510		24
SM_V7_3	Smp_331880	COesterase domain-containing protein		4
SM_V7_3	Smp_333520	previous_stable_id=Smp_204450,Smp_167460		4
SM_V7_3	Smp_333540	Nucleic acid-binding, OB-fold	CDC24_OB3	4
SM_V7_3	Smp_333660	Nucleosome assembly protein (NAP)	Nucleosome assembly protein, NAP {Baker's yeast (Saccharomyces cerevisiae) [TaxId: 4932]}	24
SM_V7_3	Smp_333720	previous_stable_id=Smp_197770,Smp_165890		4
SM_V7_3	Smp_333740	previous_stable_id=Smp_151620,Smp_151630		4
SM_V7_3	Smp_337590	Matrin/U1-C, C2H2-type zinc finger	Spliceosomal protein U1C {Human (Homo sapiens) [TaxId: 9606]}	20
SM_V7_3	Smp_337640	NA	NA	12
SM_V7_3	Smp_340630	Heat shock protein Hsp90 family	ATP-DEPENDENT MOLECULAR CHAPERONE HSP82, CO-CHAPERONE	4
SM_V7_3	Smp_340860	previous_stable_id=Smp_052880		4
SM_V7_3	Smp_341890	Tr-type G domain-containing protein		4
SM_V7_3	Smp_346670	RNase H domain-containing protein		4
SM_V7_3	Smp_347130	previous_stable_id=Smp_202290		4
SM_V7_3	Smp_347790	Cadherin-like	Protocadherin-15	4
SM_V7_4	Smp_024660	Immunoglobulin subtype 2	Roundabout homolog 2	4
SM_V7_4	Smp_041740	T-complex protein 1 subunit gamma		24
SM_V7_4	Smp_041980	Putative 60s acidic ribosomal protein P1		24
SM_V7_4	Smp_042340	Catenin and plakophilin%2C putative		16
SM_V7_4	Smp_042530	WD40 repeat	Tubulin beta, Tubulin alpha, PACRG	4
SM_V7_4	Smp_050220	Complexin%2C putative		4
SM_V7_4	Smp_053900	Beta-hexosaminidase		12
SM_V7_4	Smp_062970	WD-repeat protein%2C putative		4



## Annexe 11

SM_V7_4	Smp_070030	Snf7-related		24
SM_V7_4	Smp_126990	Putative synapsin		4
SM_V7_4	Smp_140990	NA		4
SM_V7_4	Smp_145040	MFS domain-containing protein		4
SM_V7_4	Smp_145210	Cadherin-like	Protocadherin-15	4
SM_V7_4	Smp_145540	G protein-coupled receptor, rhodopsin-like	Muscarinic acetylcholine receptor M2,Vasopressin V2	4
SM_V7_4	Smp_148790	Laminin beta chain-related		4
SM_V7_4	Smp_150140	Putative pi3kinase		4
SM_V7_4	Smp_150870	G protein-coupled receptor, rhodopsin-like	Muscarinic acetylcholine receptor M2,Vasopressin V2	4
SM_V7_4	Smp_156630	NA		4
SM_V7_4	Smp_169660	NA		4
SM_V7_4	Smp_170700	Putative wd-repeat protein		4
SM_V7_4	Smp_173100	Axon guidance protein		4
SM_V7_4	Smp_180240	F-spondin%2C putative		4
SM_V7_4	Smp_200520	Smp_200520		12
SM_V7_4	Smp_241510	Protein kinase domain-containing protein		24
SM_V7_4	Smp_247280	C2H2-type domain-containing protein		4
SM_V7_4	Smp_247620	previous_stable_id=Smp_081700,Smp_166190		4
SM_V7_4	Smp_248170	Nop domain-containing protein		24
SM_V7_4	Smp_313100	F5/8 type C domain-containing protein		4
SM_V7_4	Smp_313110	previous_stable_id=Smp_126880		4
SM_V7_4	Smp_313210	previous_stable_id=Smp_063930,Smp_156640		4
SM_V7_4	Smp_315730	TGc domain-containing protein		4
SM_V7_4	Smp_315840	previous_stable_id=Smp_149280,Smp_149290		4
SM_V7_4	Smp_317070	SAC domain-containing protein		24
SM_V7_4	Smp_329150	NA		16
SM_V7_4	Smp_336730	previous_stable_id=Smp_126950,Smp_126940		4
SM_V7_4	Smp_336900	previous_stable_id=Smp_151860,Smp_199540,Smp_151850		4
SM_V7_4	Smp_343120	Calcium-transporting ATPase		4
SM_V7_4	Smp_343130	previous_stable_id=Smp_137200		4
SM_V7_4	Smp_344270	previous_stable_id=Smp_149910		4
SM_V7_4	Smp_344900	previous_stable_id=Smp_158090		4
SM_V7_4	Smp_345160	P/Homo B domain-containing protein		4
SM_V7_4	Smp_347500	previous_stable_id=Smp_212450,Smp_198230,Smp_168510		4
SM_V7_4H004	Smp_318540	Autophagy-related protein 101		12
SM_V7_4H030	Smp_332730	Camp-dependent protein kinase type II-alpha regulatory subunit%2C putative		4
SM_V7_4H037	Smp_325410	NA		8
SM_V7_5	Smp_067890	Proteasome alpha-subunit, N-terminal domain	Proteasome alpha subunit (non-catalytic) {Baker's yeast (Saccharomyces cerevisiae) [TaxId: 4932]}	24

## Annexe 11

SM_V7_5	Smp_085540	Myosin heavy chain		4
SM_V7_5	Smp_088270	Subfamily M14A unassigned peptidase (M14 family)		4
SM_V7_5	Smp_101450	60S ribosomal protein L7%2C putative		24
SM_V7_5	Smp_129240	SBF2 domain-containing protein		4
SM_V7_5	Smp_143940	Dynein light intermediate chain		4
SM_V7_5	Smp_145700	Putative ankyrin 2%2C3/unc44		4
SM_V7_5	Smp_153440	Neuropilin (Nrp) and tolloid (Tll)-like		4
SM_V7_5	Smp_154520	previous_stable_id=Smp_154510		24
SM_V7_5	Smp_166230	NA		12
SM_V7_5	Smp_170210	Tektin		4
SM_V7_5	Smp_314670	previous_stable_id=Smp_199030		12
SM_V7_5	Smp_329980	NA	Synaptotagmin IV {Human (Homo sapiens) [TaxId: 9606]}	12
SM_V7_5	Smp_337500	Peptidase S1 domain-containing protein		4
SM_V7_5	Smp_343630	previous_stable_id=Smp_143410		4
SM_V7_5	Smp_343680	Protein-tyrosine-phosphatase		4
SM_V7_5H022	Smp_307520	NA		24
SM_V7_6	Smp_001900	Septate junction protein		12
SM_V7_6	Smp_025030	Glutamate-cysteine ligase		4
SM_V7_6	Smp_051130	Ets domain	Fli-1 {Human (Homo sapiens) [TaxId: 9606]}	16
SM_V7_6	Smp_077860	Putative actin		4
SM_V7_6	Smp_078230	cGMP-dependent protein kinase		4
SM_V7_6	Smp_082300	acidPPc domain-containing protein		4
SM_V7_6	Smp_102960	Nebula%2C putative		4
SM_V7_6	Smp_123100	MEG-32.1 protein		12
SM_V7_6	Smp_127130	Venom allergen-like (VAL) 20 protein		4
SM_V7_6	Smp_154080	NA		24
SM_V7_6	Smp_163950	NA		4
SM_V7_6	Smp_315510	previous_stable_id=Smp_179570		12
SM_V7_6	Smp_317810	NA		4
SM_V7_6	Smp_331200	NA		8
SM_V7_6	Smp_342090	CUB domain-containing protein		4
SM_V7_6	Smp_342990	previous_stable_id=Smp_135670		12
SM_V7_7	Smp_006000	Eukaryotic translation initiation factor 3 subunit I		24
SM_V7_7	Smp_041190	NA		4
SM_V7_7	Smp_041370	NA		4
SM_V7_7	Smp_041460	Tetraspanin%2C putative		12
SM_V7_7	Smp_057360	Serine/threonine kinase		4
SM_V7_7	Smp_057530	Subfamily S9B unassigned peptidase (S09 family)		4
SM_V7_7	Smp_097380	10 kDa heat shock protein%2C mitochondrial		24
SM_V7_7	Smp_125350	BC026374 protein (S09 family)		4
SM_V7_7	Smp_125440	Putative 4.1 G protein		4
SM_V7_7	Smp_127940	Anion exchange protein		4
SM_V7_7	Smp_130260	Calponin-homology (CH) domain-containing protein		4
SM_V7_7	Smp_134590	Putative unc-13 (Munc13)		4
SM_V7_7	Smp_140040	Guanylate cyclase domain-containing protein		4

## Annexe 11

SM_V7_7	Smp_214210	previous_stable_id=Smp_175620	8
SM_V7_7	Smp_246140	V-type proton ATPase subunit H	4
SM_V7_7	Smp_246160	previous_stable_id=Smp_210730	4
SM_V7_7	Smp_300560	BTB domain-containing protein	4
SM_V7_7	Smp_300840	Bruno-like rna binding protein	4
SM_V7_7	Smp_318260	EF-hand domain-containing protein	4
SM_V7_7	Smp_332270	previous_stable_id=Smp_127800,Smp_127790,Smp_205400	4
SM_V7_7	Smp_332280	previous_stable_id=Smp_140170	4
SM_V7_7	Smp_332360	Erythrocyte membrane protein%2C putative	4
SM_V7_7	Smp_340660	previous_stable_id=Smp_140090,Smp_033200	4
SM_V7_7	Smp_343310	previous_stable_id=Smp_139400	4
SM_V7_7	Smp_344530	previous_stable_id=Smp_152890	4
SM_V7_ZW	Smp_007650	previous_stable_id=Smp_007640	4
SM_V7_ZW	Smp_011860	Putative transcription factor LCR-F1	8
SM_V7_ZW	Smp_011960	Inositol-pentakisphosphate 2-kinase	16
SM_V7_ZW	Smp_015630	Putative cys-loop ligand gated ion channel subunit	4
SM_V7_ZW	Smp_028650	Rho-GAP domain-containing protein	16
SM_V7_ZW	Smp_031050	BSD domain-containing protein	12
SM_V7_ZW	Smp_039640	Putative retinoblastoma binding protein	8
SM_V7_ZW	Smp_039810	Muscleblind-like protein	4
SM_V7_ZW	Smp_045860	Inositol hexakisphosphate and diphosphoinositol-pentakisphosphate kinase	4
SM_V7_ZW	Smp_046130	Putative alpha-actinin	4
SM_V7_ZW	Smp_052630	Amiloride-sensitive sodium channel%2Cputative	4
SM_V7_ZW	Smp_074500	Proteasome subunit beta	20
SM_V7_ZW	Smp_075350	NA	4
SM_V7_ZW	Smp_080520	RING-type domain-containing protein	4
SM_V7_ZW	Smp_084090	Putative junctophilin	4
SM_V7_ZW	Smp_094190	Serine/threonine kinase	4
SM_V7_ZW	Smp_099090	Putative alpha(1%2C3)fucosyltransferase	4
SM_V7_ZW	Smp_135210	Egf-like domain protein	4
SM_V7_ZW	Smp_135300	NA	4
SM_V7_ZW	Smp_136070	previous_stable_id=Smp_118100,Smp_157650	4
SM_V7_ZW	Smp_139760	Kinesin light chain	4
SM_V7_ZW	Smp_142740	Rho-GAP domain-containing protein	4
SM_V7_ZW	Smp_151280	Camp-response element binding protein-related	12
SM_V7_ZW	Smp_152390	NA	4
SM_V7_ZW	Smp_155050	Agrin%2C putative	4
SM_V7_ZW	Smp_156990	Serine/threonine kinase	4
SM_V7_ZW	Smp_157830	Putative erythrocyte membrane protein	4
SM_V7_ZW	Smp_161120	Actin bundling/missing in metastasis-related	4
SM_V7_ZW	Smp_161140	Putative voltage-gated potassium channel	4
SM_V7_ZW	Smp_163750	Putative otoferlin	4
SM_V7_ZW	Smp_164680	Synaptic ras gtpase activating protein%2C syngap%2Cputative	4
SM_V7_ZW	Smp_170060	previous_stable_id=Smp_118020	4

## Annexe 11

SM_V7_ZW	Smp_175900	Synaptotagmin		4
SM_V7_ZW	Smp_176200	Superoxide dismutase [Cu-Zn]		24
SM_V7_ZW	Smp_176540	Putative cadherin		4
SM_V7_ZW	Smp_182620	previous_stable_id=Smp_142190		4
SM_V7_ZW	Smp_210580	Sad1/unc-84-like protein		12
SM_V7_ZW	Smp_212180	Putative glucose-methanol-choline (Gmc) oxidoreductase		12
SM_V7_ZW	Smp_212240	NA		16
SM_V7_ZW	Smp_241420	AGC-kinase C-terminal domain-containing protein		4
SM_V7_ZW	Smp_241470	Metalloendopeptidase		4
SM_V7_ZW	Smp_243900	Innexin		4
SM_V7_ZW	Smp_244080	Protein-tyrosine-phosphatase		4
SM_V7_ZW	Smp_244190	ZP domain-containing protein		4
SM_V7_ZW	Smp_244250	previous_stable_id=Smp_198430,Smp_167820		4
SM_V7_ZW	Smp_244400	FERM domain-containing protein		4
SM_V7_ZW	Smp_244950	previous_stable_id=Smp_171960		12
SM_V7_ZW	Smp_245310	previous_stable_id=Smp_126350		12
SM_V7_ZW	Smp_266820	previous_stable_id=Smp_096370,Smp_173270,Smp_175050		4
SM_V7_ZW	Smp_303910	NA		4
SM_V7_ZW	Smp_305740	Putative nachr subunit		4
SM_V7_ZW	Smp_306120	NA		4
SM_V7_ZW	Smp_306200	FERM domain-containing protein		4
SM_V7_ZW	Smp_306240	Acetylcholine-gated chloride channel subunit		4
SM_V7_ZW	Smp_306270	previous_stable_id=Smp_171450,Smp_171440		4
SM_V7_ZW	Smp_307000	SH3_10 domain-containing protein		4
SM_V7_ZW	Smp_315810	previous_stable_id=Smp_168650,Smp_198820,Smp_138620,Smp_138630,Smp_187350		4
SM_V7_ZW	Smp_333320	previous_stable_id=Smp_076180,Smp_076170,Smp_173170,Smp_163200		4
SM_V7_ZW	Smp_337260	previous_stable_id=Smp_054170,Smp_198960		4
SM_V7_ZW	Smp_337410	SB domain-containing protein		8
SM_V7_ZW	Smp_345440	previous_stable_id=Smp_188590,Smp_163920		4
SM_V7_ZW	Smp_345530	previous_stable_id=Smp_164670		24
SM_V7_ZW H003	Smp_314110	Alpha-mann_mid domain-containing protein		12

Chr. : chromosome.

## Annexe 12

## Annexe 12 : Liste des gènes candidats sélectionnés pour être potentiellement impliqués dans le rythme d'émission des cercaires

Chr.	Gene ID	GFF annotation	HHsearch annotation	Moyenne Reads D	Moyenne Reads N	QTL	H3K4me3	H3K9ac	Rythmique	Sélection
1	Smp_132810	Myc-type, basic helix-loop-helix (bHLH) domain	Myc proto-oncogene protein [Human (Homo sapiens) [Taxid: 9606]]	53,6037965	82,6678821	Intraspé.				Motif bHLH
1	Smp_132800	Myc-type, basic helix-loop-helix (bHLH) domain	Aryl hydrocarbon receptor nuclear translocator-like	91,0373972	4830,18226	Intraspé.				Motif bHLH
1	Smp_019460	Myc-type, basic helix-loop-helix (bHLH) domain	SREBP-1a [Human (Homo sapiens) [Taxid: 9606]]	53,4203059	408,194317	Intraspé.				Motif bHLH
1	Smp_132790	Myc-type, basic helix-loop-helix (bHLH) domain	Transcription factor HES-1	16,8412608	637,812492	Intraspé.				Motif bHLH
1	Smp_343020	Myc-type, basic helix-loop-helix (bHLH) domain	DNA-binding protein inhibitor ID-1	3,16691957	38,930902	Intraspé.				Motif bHLH
1	Smp_180140	G protein-coupled receptor, rhodopsin-like	P2Y purinoceptor 1, Rubredoxin	0,15105096	156,233733	Intraspé.			N (L-O et O-L)	Rhodospine et rhodospine like
1	Smp_316850	G protein-coupled receptor, rhodopsin-like	C-C chemokine receptor type 2	3,61511644	64,6441694	Intraspé.				Rhodospine et rhodospine like
1	Smp_204580	G protein-coupled receptor, rhodopsin-like	Neurotensin receptor type 1, Neurotensin	2,54493418	486,176692	Intraspé.				Rhodospine et rhodospine like
1	Smp_118040	G protein-coupled receptor, rhodopsin-like	Chimera of Proteinase-activated receptor 1	4,64630648	782,09813	Intraspé.				Rhodospine et rhodospine like
1	Smp_137320	G protein-coupled receptor, rhodopsin-like	Srg	0,46920763	4273,7631	Intraspé.				Rhodospine et rhodospine like
1	Smp_027940	G protein-coupled receptor, rhodopsin-like	ADENOSINE RECEPTOR A2A	0,19641185	632,323537	Intraspé.				Rhodospine et rhodospine like
1	Smp_137310	G protein-coupled receptor, rhodopsin-like	7TM_GPCR_Srw	1,63386995	4103,87566	Intraspé.				Rhodospine et rhodospine like
1	Smp_137300	G protein-coupled receptor, rhodopsin-like	7TM_GPCR_Srw	2,31619242	2573,30923	Intraspé.				Rhodospine et rhodospine like
1	Smp_319310	NA	7TM_GPCR_Srab	0,99318232	93,7737103	Intraspé.				Rhodospine et rhodospine like
1	Smp_178420	GPCR, rhodopsin-like, 7TM	ADENOSINE RECEPTOR A2A	1,04431464	939,306078	Intraspé.				Rhodospine et rhodospine like
1	Smp_132730	G protein-coupled receptor, rhodopsin-like	Neurotensin receptor type 1, Endolysin, NEUROTENSIN RECEPTOR	10,5536811	2022,0823	Intraspé.	x			Rhodospine et rhodospine like
1	Smp_127310	G protein-coupled receptor, rhodopsin-like	Chimera of Proteinase-activated receptor 1	55,3426429	6,11219416	Intraspé.	x			Rhodospine et rhodospine like
1	Smp_146470	NA	Rhodopsin	1,24781665	1021,28325	Intraspé.				Rhodospine et rhodospine like
1	Smp_140250	G protein-coupled receptor, rhodopsin-like	Platelet-activating factor receptor, Flavodoxin	23,2810068	1485,52478	Intraspé.			N (L-O et O-L)	Rhodospine et rhodospine like
1	Smp_193610	G protein-coupled receptor, rhodopsin-like	Lysozyme C (E.C.3.2.1.17)	13,4895674	2,94755447	Intraspé.				Rhodospine et rhodospine like
1	Smp_007070	G protein-coupled receptor, rhodopsin-like	M3-mT4L, Tiotropium	30,7991451	760,084209	Intraspé.				Rhodospine et rhodospine like
1	Smp_094890	Rho GTPase-activating protein domain	Beta-chimaerin, C-terminal domain [Human (Homo sapiens) [Taxid: 9606]]	488,132151	1006,31934	Intraspé.	x		N (L-O)	Acteurs phototransduction
1	Smp_028030	Rho GTPase-activating protein domain	SLIT-ROBO Rho GTPase-activating protein 2	1841,29072	1749,04695	Intraspé.	x			Acteurs phototransduction
1	Smp_180150	Rho GTPase-activating protein domain	Beta-chimaerin, C-terminal domain [Human (Homo sapiens) [Taxid: 9606]]	1037,28971	230,457733	Intraspé.				Acteurs phototransduction
1	Smp_180160	Rho GTPase-activating protein domain	Graf [Chicken (Gallus gallus) [Taxid: 9031]]	604,819963	4030,17202	Intraspé.	x			Acteurs phototransduction
1	Smp_319130	FCH domain	SLIT-ROBO Rho GTPase-activating protein 2	131,032691	10,4521916	Intraspé.	x		N (L-O)	Acteurs phototransduction
1	Smp_007210	Rab-GTPase-TBC domain	Rab GTPase-activating protein 1	1264,20337	6834,72269	Intraspé.	x			Acteurs phototransduction
5	Smp_165530	Rho GTPase-activating protein domain	p50 RhoGAP domain [Human (Homo sapiens) [Taxid: 9606]]	122,950283	809,709826	Intraspé.	x			Acteurs phototransduction
1	Smp_211230	Adenylyl cyclase class-3/4/guanylyl cyclase	Soluble guanylyl cyclase alpha-1 subunit	4,46877352	9,47380194	Intraspé.				Acteurs phototransduction
1	Smp_097470	Adenylyl cyclase class-3/4/guanylyl cyclase	Soluble guanylyl cyclase alpha-1 subunit	2,37668024	299,613768	Intraspé.				Acteurs phototransduction
5	Smp_158980	EF-hand domain	Recoverin [Cow (Bos taurus) [Taxid: 9913]]	6,57180603	415,080824	Intraspé.				Acteurs phototransduction
5	Smp_158990	EF-hand domain	Frequenin (neuronal calcium sensor 1) [Baker's yeast (Saccharomyces cerevisiae) [Taxid: 4932]]	35,1542874	394,752873	Intraspé.				Acteurs phototransduction
6	Smp_156180	EF-hand domain	DUF5580	0	416,684558	Intraspé.				Acteurs phototransduction
6	Smp_136640	EF-hand domain	Grancalcin [Human (Homo sapiens) [Taxid: 9606]]	1433,53103	2523,6077	Intraspé.	x			Acteurs phototransduction
1	Smp_083740	EF-hand domain	Myosin Essential Chain [Bay scallop (Aequipecten irradians) [Taxid: 31199]]	739,409005	1224,24346	Intraspé.	x			Acteurs phototransduction
1	Smp_005230	EF-hand domain	Calcium-dependent protein kinase sk5 CLD [Soybean (Glycine max) [Taxid: 3847]]	3763,84069	173,141661	Intraspé.	x			Acteurs phototransduction
1	Smp_137410	EF-hand domain pair	M-CALPAIN LARGE AND SMALL SUBUNITS	128,358601	548,023528	Intraspé.	x			Acteurs phototransduction
1	Smp_132670	EF-hand domain	Myosin Essential Chain [Bay scallop (Aequipecten irradians) [Taxid: 31199]]	6010,69651	5332,5733	Intraspé.				Acteurs phototransduction
1	Smp_068060	Protein kinase domain	Serine/threonine protein kinase TAO2 [Rat (Rattus norvegicus) [Taxid: 10116]]	3107,58026	631,614881	Intraspé.	x			Acteurs phototransduction
1	Smp_005190	Protein kinase domain	STE20-like serine/threonine-protein kinase, SLK [Human (Homo sapiens) [Taxid: 9606]]	1236,88844	165,090057	Intraspé.	x			Acteurs phototransduction
1	Smp_009800	Protein kinase domain	Serine/threonine-protein kinase Nek2 [Human (Homo sapiens) [Taxid: 9606]]	19,0607243	1,90069089	Intraspé.				Acteurs phototransduction
5	Smp_158950	Protein kinase-like domain superfamily	SERINE/THREONINE-PROTEIN KINASE HASPIN (E.C.2.7.11.1)	172,964297	1651,38688	Intraspé.				Acteurs phototransduction
5	Smp_165490	Calcineurin-like phosphoesterase domain, ApaH type	Protein phosphatase 2A catalytic subunit alpha isoform, PP2A-alpha [Human (Homo sapiens) [Taxid: 9606]]	4779,87312	1606,50822	Intraspé.	x			Acteurs phototransduction
1	Smp_067140	Small GTPase	RhoA [Human (Homo sapiens) [Taxid: 9606]]	687,376028	144,309619	Intraspé.				Acteurs phototransduction
1	Smp_025740	Small GTPase	RhoE [RND3] [Mouse (Mus musculus) [Taxid: 10090]]	267,897857	545,542907	Intraspé.	x			Acteurs phototransduction
1	Smp_246990	Transcription factor CBF/NF-Y/archaeal histone domain	Histone H3.1, Histone H4, Histone	583,452168	1375,80951	Intraspé.				Histone
1	Smp_036220	Histone H2B	Histone H2B [Chicken (Gallus gallus), erythrocytes [Taxid: 9031]]	1995,08293	543,166586	Intraspé.	x			Histone
1	Smp_124840	Zinc finger, UBR-type	HISTONE H3.2, Histone H4, Histone	2078,18437	447,202343	Intraspé.				Histone
1	Smp_318950	Histone H3/CENP-A	Leishmania histone H3, Histone H4	0,27740305	719,694697	Intraspé.				Histone
6	Smp_210630	Histone H2A	Histone H2A [Chicken (Gallus gallus), erythrocytes [Taxid: 9031]]	3662,2237	2098,24735	Intraspé.	x			Histone
6	Smp_131320	Zinc finger, PHD-type	Probable histone acetyltransferase MYST1 [Human (Homo sapiens) [Taxid: 9606]]	1153,00181	75,4036476	Intraspé.	x			Histone
1	Smp_005210	Histone deacetylase family	METASTASIS-ASSOCIATED PROTEIN MTA1, HISTONE DEACETYLASE	3274,91636	10,6869863	Intraspé.	x			Histone
5	Smp_330000	Sin3 associated polypeptide p18	HISTONE DEACETYLASE COMPLEX SUBUNIT SAP18	866,154826	168,49628	Intraspé.				Histone
1	Smp_126730	G protein-coupled receptor, rhodopsin-like	7tm_1	29,1329131	1223,40209	Intraspé.				
5	Smp_329990	C2 domain	Synaptotagmin IV [Rat (Rattus norvegicus) [Taxid: 10116]]	22,3182533	457,016358	Intraspé.				
5	Smp_314630	NA	RCR	42,4353226	35,6081083	Intraspé.	x			
1	Smp_105040	Endonuclease/exonuclease/phosphatase	2',5'-phosphodiesterase 12 (E.C.3.1.4.-, 3.1.13.4)	291,88514	236,676729	Intraspé.	x			
5	Smp_345020	NA	QVR	25,4178812	170,969739	Intraspé.	x			GOsleep
1	Smp_009630	Homeobox domain	SIX1_SD	148,366449	5349,04638	Intraspé.				GO developmental maturation
1	Smp_315790	CBS domain	5'-AMP-activated protein kinase catalytic subunit	255,672243	420,816428	Intraspé.	x			K07200
1	Smp_010230	G-protein, beta subunit	beta1-subunit of the signal-transducing G protein heterotrimer [Cow (Bos taurus) [Taxid: 9913]]	680,225115	1,01257237	Intraspé.	x			
6	Smp_016630	Guanine nucleotide binding protein (G-protein), alpha subunit	G-protein alpha subunit Galpha7	1132,64942	0,08891906	Intraspé.	x			
1	Smp_132750	G-protein alpha subunit, group Q	G-protein alpha subunit Galpha7	1,24197493	596,446731	Intraspé.				
1	Smp_210820	GPCR kinase	Beta-adrenergic receptor kinase 1 (E.C.2.7.11.15)	594,96695	579,415435	Intraspé.			N (L-O)	K00910

## Annexe 12

1	Smp_126080	Arrestin	Bovine arrestin-2 (full length)	601,479348	2159,11532	Intraspé.	x			
1	Smp_152830	Na-Ca exchanger/integrin-beta4	Na/Ca exchange protein	335,193874	13,2323608	Intraspé.			N (L-O)	K05849
3	Smp_021340	Myc-type, basic helix-loop-helix (bHLH) domain	Mad protein (Human (Homo sapiens) [Taxid: 9606])	3171,81347	763,33068	Interspé.			N (L-O)	Motif bHLH
3	Smp_123570	Myc-type, basic helix-loop-helix (bHLH) domain	Aryl hydrocarbon receptor nuclear translocator	262,402038	1031,30756	Interspé.	x			Motif bHLH
3	Smp_342580	Myc-type, basic helix-loop-helix (bHLH) domain	Transcription factor E2-alpha, DNA-binding protein	10,7735232	547,83698	Interspé.				Motif bHLH
3	Smp_000620	Myc-type, basic helix-loop-helix (bHLH) domain	DNA-binding protein inhibitor ID-1	15,2282684	0,60010372	Interspé.			N (L-O)	Motif bHLH
3	Smp_333620	Myc-type, basic helix-loop-helix (bHLH) domain	Myc proto-oncogene protein (Human (Homo sapiens) [Taxid: 9606])	2934,35195	270,433032	Interspé.	x			Motif bHLH
3	Smp_150180	G protein-coupled receptor, rhodopsin-like	Chimera of Proteinase-activated receptor 1	3,32989335	1065,25224	Interspé.				Rhodospine et rhodospine like
3	Smp_159860	G protein-coupled receptor, rhodopsin-like	Cannabinoid receptor 1, Flavodoxin	3,96183552	262,487843	Interspé.				Rhodospine et rhodospine like
3	Smp_132410	G protein-coupled receptor, rhodopsin-like	Lysozyme, Proteinase-activated receptor 2, Soluble cytochrome b562, Proteinase-activated	19,2022749	33,1519434	Interspé.			N (L-O)	Rhodospine et rhodospine like
3	Smp_083880	G protein-coupled receptor, rhodopsin-like	Neurotensin receptor type 1, Endolysin, NEUROTENSIN RECEPTOR	0,09078652	1420,40407	Interspé.	x			Rhodospine et rhodospine like
3	Smp_177720	G protein-coupled receptor, rhodopsin-like	C-X-C chemokine receptor type 4	0,09730639	42,6576636	Interspé.				Rhodospine et rhodospine like
3	Smp_083940	G protein-coupled receptor, rhodopsin-like	7TM_GPCR_Srw	9,3390145	0	Interspé.				Rhodospine et rhodospine like
3	Smp_326640	G protein-coupled receptor, rhodopsin-like	7TM_GPCR_Srw	0	427,467123	Interspé.	x			Rhodospine et rhodospine like
3	Smp_241490	G protein-coupled receptor, rhodopsin-like	7TM_GPCR_Srw	0,69401572	89,9435101	Interspé.				Rhodospine et rhodospine like
3	Smp_091950	G protein-coupled receptor, rhodopsin-like	7TM_GPCR_Srw	0,19361732	392,952619	Interspé.				Rhodospine et rhodospine like
3	Smp_326670	G protein-coupled receptor, rhodopsin-like	Platelet-activating factor receptor, Flavodoxin	4,24614636	1545,2751	Interspé.				Rhodospine et rhodospine like
3	Smp_134960	G protein-coupled receptor, rhodopsin-like	Platelet-activating factor receptor	6,8856344	6,67864644	Interspé.				Rhodospine et rhodospine like
3	Smp_023710	G protein-coupled receptor, rhodopsin-like	Platelet-activating factor receptor, Flavodoxin	1,37522946	14,8816496	Interspé.				Rhodospine et rhodospine like
3	Smp_117340	G protein-coupled receptor, rhodopsin-like	Rhodopsin	1,65972958	4949,25023	Interspé.	x			Rhodospine et rhodospine like
4	Smp_343810	GPCR, rhodopsin-like, 7TM	Adenosine receptor A2a, Soluble cytochrome b562, Adenosine	37,7981714	781,939417	Interspé.				Rhodospine et rhodospine like
4	Smp_043260	G protein-coupled receptor, rhodopsin-like	7tm_1	4,52025031	2235,39202	Interspé.				Rhodospine et rhodospine like
4	Smp_043270	G protein-coupled receptor, rhodopsin-like	7tm_1	2,31687872	654,548813	Interspé.				Rhodospine et rhodospine like
4	Smp_336140	G protein-coupled receptor, rhodopsin-like	7tm_1	0,91088034	1621,13261	Interspé.				Rhodospine et rhodospine like
4	Smp_348040	G protein-coupled receptor, rhodopsin-like	7tm_1	0,36462583	1038,44216	Interspé.				Rhodospine et rhodospine like
4	Smp_043290	G protein-coupled receptor, rhodopsin-like	7tm_1	0,47471714	118,097675	Interspé.				Rhodospine et rhodospine like
4	Smp_043300	G protein-coupled receptor, rhodopsin-like	7tm_1	0,21959669	165,906881	Interspé.				Rhodospine et rhodospine like
4	Smp_348050	G protein-coupled receptor, rhodopsin-like	7tm_1	0,79699199	1269,34159	Interspé.				Rhodospine et rhodospine like
4	Smp_332070	G protein-coupled receptor, rhodopsin-like	human Chemokine Receptor 7	1,48945005	2706,94684	Interspé.				Rhodospine et rhodospine like
4	Smp_043320	G protein-coupled receptor, rhodopsin-like	7tm_1	0,09880883	223,716815	Interspé.				Rhodospine et rhodospine like
4	Smp_043340	G protein-coupled receptor, rhodopsin-like	7tm_1	0,23576484	2103,46195	Interspé.				Rhodospine et rhodospine like
4	Smp_212750	G protein-coupled receptor, rhodopsin-like	Lysophosphatidic acid receptor 6a, Endolysin, receptor (E.C.3.2.1.17)	2,44988415	212,848143	Interspé.				Rhodospine et rhodospine like
3	Smp_138500	Arf GTPase activating protein	Ras-related protein Rab-35, Iporin	632,036028	628,922412	Interspé.	x			Acteurs phototransduction
3	Smp_000670	Arf GTPase activating protein	ARFGAP with coiled-coil, ANK repeat	787,647933	0,8046734	Interspé.	x			Acteurs phototransduction
3	Smp_160610	Rho GTPase-activating protein domain	p50 RhoGAP domain (Human (Homo sapiens) [Taxid: 9606])	194,197882	4334,59485	Interspé.				Acteurs phototransduction
3	Smp_138430	DEP domain	Regulator of G-protein signaling 7	238,951866	366,55032	Interspé.	x			Acteurs phototransduction
3	Smp_084890	Type I phosphodiesterase/nucleotide pyrophosphatase/phosphate transferase	PHOSPHONACETATE HYDROLASE (E.C.3.11.1.2)	926,177763	287,758284	Interspé.				Acteurs phototransduction
3	Smp_018360	PLC-like phosphodiesterase, TIM beta/alpha-barrel domain superfamily	Crystal Structure of the Y2475/Y2515	943,10973	0,21454483	Interspé.				Acteurs phototransduction
4	Smp_053550	U6 snRNA phosphodiesterase Usb1	U6 snRNA phosphodiesterase (E.C.3.1.4.-)	246,183457	395,867207	Interspé.				Acteurs phototransduction
3	Smp_106110	EF-hand domain	Kchip1, Kv4 potassium channel-interacting protein [Rat (Rattus norvegicus) [Taxid: 10116]]	23,7359547	192,04885	Interspé.			D (L-O)	Acteurs phototransduction
3	Smp_181460	EF-hand domain pair	DUF4205	419,705945	22,2483905	Interspé.	x		D (L-O)	Acteurs phototransduction
3	Smp_106060	EF-hand domain	ADP, ATP carrier protein (Cow (Bos taurus), heart isoform t1 [Taxid: 9913])	398,839106	88,6154242	Interspé.	x			Acteurs phototransduction
3	Smp_000180	EF-hand domain	Frequenin (neuronal calcium sensor 1) (Human (Homo sapiens) [Taxid: 9606])	4,50921799	1351,85154	Interspé.				Acteurs phototransduction
3	Smp_134950	EF-hand domain	Myosin Essential Chain [Bay scallop (Aequipecten irradians) [Taxid: 31199]]	449,311197	1162,46489	Interspé.	x			Acteurs phototransduction
4	Smp_313310	EF-hand domain	Sarcoplasmic calcium-binding protein [Amphioxus (Branchiostoma lanceolatum) [Taxid: 7740]]	8,22979325	584,346012	Interspé.			DN	Acteurs phototransduction
3	Smp_018400	WD40 repeat	beta1-subunit of the signal-transducing G protein heterotrimer [Cow (Bos taurus) [Taxid: 9913]]	438,761194	529,772316	Interspé.	x			Acteurs phototransduction
3	Smp_123610	Protein kinase domain	Phosphatidylinositol 3-kinase VPS34 (E.C.2.7.1.137), Serine/threonine-protein	506,853188	55,8904665	Interspé.				Acteurs phototransduction
3	Smp_123640	GPCR kinase	G protein-coupled receptor kinase 4	508,110359	562,613409	Interspé.	x			Acteurs phototransduction
3	Smp_144390	Ser/Thr protein kinase, TGFbeta receptor	B-Raf kinase (Human (Homo sapiens) [Taxid: 9606])	830,269412	613,673974	Interspé.	x			Acteurs phototransduction
4	Smp_042920	Serine-threonine protein kinase 19	Stk19	273,061349	40,7021666	Interspé.	x			Acteurs phototransduction
4	Smp_308680	Protein kinase domain	Serine/threonine-protein kinase Nek2 (Human (Homo sapiens) [Taxid: 9606])	1065,27069	848,171547	Interspé.	x			Acteurs phototransduction
3	Smp_308690	Protein phosphatase 2A regulatory subunit PR55	NA	276,900956	763,532311	Interspé.				Acteurs phototransduction
3	Smp_308700	Protein phosphatase 2A regulatory subunit PR55	Serine/threonine-protein phosphatase 2A 65 kDa	7,76765622	878,795614	Interspé.				Acteurs phototransduction
3	Smp_246610	Small GTPase	Rad (Human (Homo sapiens) [Taxid: 9606])	88,0819514	133,394512	Interspé.				Acteurs phototransduction
3	Smp_084940	Small GTPase	Ras-related protein M-Ras (XRas) [Mouse (Mus musculus) [Taxid: 10090]]	928,541401	126,391329	Interspé.	x			Acteurs phototransduction
3	Smp_132330	Small GTPase	Rab30 (Human (Homo sapiens) [Taxid: 9606])	257,926321	26,892619	Interspé.				Acteurs phototransduction
3	Smp_134940	GTP binding domain	Yqeh GTPase	1096,94943	130,919352	Interspé.				Acteurs phototransduction
3	Smp_342660	Small GTPase	Rad (Human (Homo sapiens) [Taxid: 9606])	15,3716692	844,395068	Interspé.				Acteurs phototransduction
4	Smp_313300	Small GTPase	Rab21 (Human (Homo sapiens) [Taxid: 9606])	38,8927175	3108,98173	Interspé.			D (L-O)	Acteurs phototransduction
3	Smp_337740	Histone H2A	Histone H2A [Chicken (Gallus gallus), erythrocytes [Taxid: 9031]]	215,345187	1022,78704	Interspé.				Histone
3	Smp_000150	Zinc finger, PHD-type	Histone H3.2, Histone H4, Histone	905,606672	6505,47622	Interspé.				Histone
3	Smp_027850	Chromatin SPT2	Histone H3.2, Histone H4, Protein	609,945723	289,167188	Interspé.				Histone
3	Smp_026880	Histone H3/CENP-A	Leishmania histone H3, Histone H4	0	852,292191	Interspé.				Histone
3	Smp_105910	Zinc finger, TAZ-type	Histone acetyltransferase p300 (E.C.2.3.1.48), p53	4566,00741	22,4878089	Interspé.	x			Histone
3	Smp_083930	Chromatin modification-related protein Eaf6	Histone acetyltransferase ESA1 (E.C.2.3.1.48), Chromatin	1002,70645	801,874546	Interspé.	x			Histone
3	Smp_210660	SET domain	Histone-lysine N-methyltransferase 2E (E.C.2.1.1.43)	1211,11036	243,241395	Interspé.	x			Histone
4	Smp_150850	Methyl-CpG DNA binding	Histone-lysine N-methyltransferase SETDB1 (E.C.2.1.1.43), Histone	1163,0633	2588,48552	Interspé.	x			Histone
4	Smp_160700	PWWP domain	Histone-lysine N-methyltransferase SETD2 (E.C.2.1.1.43, 2.1.1.-), H3.3531phK36M(29-42)	935,275395	312,146006	Interspé.				Histone



Annexe 12

	1	Smp_086860	Histone H2A		3300,68917	312,58784		x		Histone
	7	Smp_031720	Histone H2A		2964,8147	3326,42021		x		Histone
ZW		Smp_082240	Histone H3		12656,1237	555,023314		x		Histone
	1	Smp_178700	Putative histone acetyltransferase type B catalytic subunit		884,324648	2481,69561		x		Histone
	2	Smp_194520	Histone acetyltransferase		280,988236	2866,68794		x		Histone
	4	Smp_053140	Histone acetyltransferase		877,773134	1549,20364		x	D (L-O)	Histone
ZW		Smp_198670	Histone acetyltransferase		562,985347	490,186844		x		Histone
	1	Smp_069380	Putative histone deacetylase 4, 5		510,978471	767,937322		x		Histone
	1	Smp_091990	Putative histone deacetylase 1, 2, 3		884,409109	260,502211		x		Histone
	2	Smp_138770	Histone deacetylase, putative		347,31463	0,73535486		x		Histone
	5	Smp_093280	Histone deacetylase		946,536548	84,8427208		x		Histone
	1	Smp_140390	Histone-lysine N-methyltransferase		1824,48666	42,0710763		x		Histone
	3	Smp_138030	Histone-lysine N-methyltransferase		2812,50744	200,22716		x		Histone
	4	Smp_070170	Histone-lysine N-methyltransferase		686,10412	182,576826		x		Histone
	4	Smp_027300	Histone-lysine N-methyltransferase		703,555319	18,25879		x		Histone
	5	Smp_043580	Putative histone-lysine n-methyltransferase, setb1		128,089362	11,4564123		x		Histone
	6	Smp_165000	Histone-lysine N-methyltransferase, H3 lysine-79 specific		1562,7874	1843,36117		x		Histone
ZW		Smp_210650	Histone-lysine N-methyltransferase		1437,73889	206,108752		x		Histone
	5	Smp_150560	Putative lysine-specific histone demethylase 1		817,609421	508,265591		x		Histone
ZW		Smp_095130	Putative porphobilinogen deaminase		246,204347	10906,4062		x	D (L-O)	
	1	Smp_140790	Ferrochelatase		764,085121	275,424078		x		
	2	Smp_045260	5-aminolevulinatase synthase		989,112985	260,830481		x		DN
	3	Smp_079840	Uroporphyrinogen-III synthase		646,090086	1245,87292		x		N (L-O)
	4	Smp_137190	Frataxin homolog, mitochondrial, putative		337,46529	1325,19166		x		
	5	Smp_340920	previous stable id=Smp_057930		1091,20744	538,10452		x		D (L-O)
	6	Smp_162280	Coproporphyrinogen oxidase		329,346133	984,709566		x		
ZW		Smp_143740	Uroporphyrinogen decarboxylase		329,668652	988,238438		x		
ZW		Smp_333090	previous stable id=Smp_199730,Smp_199720,Smp_147510,Smp_147500		2039,06417	127,498461		x		
ZW		Smp_045810	Delta-aminolevulinic acid dehydratase		446,08514	826,029862		x		
ZW		Smp_065890	NA		272,318596	221,620058		x		
ZW		Smp_004910	Rab-2,4,14, putative	Ras-related protein Rab-14	563,831914	846,458679		x		
	1	Smp_089630	Putative polyglutamine binding protein		1045,18712	670,702612		x		
	1	Smp_041050		G-patch domain-containing protein	1057,34496	10,4344192		x		
	1	Smp_012780	Putative sarm1	ADP-ribosyl cyclase/cyclic ADP-ribose hydrolase	327,400059	399,231208		x		
	3	Smp_001030	5'-amp-activated protein kinase gamma-2 non-catalytic subunit transcript variant 2		234,514885	617,069314		x		K07200
	4	Smp_334010	NA		62,7964225	2557,78588		x		K07200
ZW		Smp_170160	5-AMP-activated protein kinase, beta subunit, putative		787,691557	7895,60296		x		K07199
	1	Smp_341700	AMPKBI domain-containing protein		785,238006	737,721057		x		D (L-O) K07199
	3	Smp_052910	Skp1-related		1288,33863	262,189434		x		K03094
	3	Smp_099930	Putative f-box and wd40 domain protein		982,672389	668,221374		x		K03362
ZW		Smp_214100	Non-specific serine/threonine protein kinase		1417,09423	226,920302		x		K07198
ZW		Smp_245320	ANF receptor domain-containing protein		168,68939	19316,7996		x		K05208
ZW		Smp_026560	Calmodulin 2		5362,10614	191,414259		x		K02183
ZW		Smp_011660	Protein kinase		1926,81391	600,521307		x		K04515
	1	Smp_020080	Putative gtp-binding protein (I) alpha-1subunit, gna1		36,9592898	0		x		D (L-O) K04630
	2	Smp_016250	Putative gtp-binding protein (I) alpha-2 subunit, gna2		692,742701	0,64443613		x		D (L-O) K04630
	7	Smp_005790	Putative gtp-binding protein (Q) alpha-11 subunit, gna11		632,10564	625,292788		x		K04634
	2	Smp_063640	Putative g-protein, beta subunit		2209,21385	64,2475109		x		K04536
	4	Smp_142050	Mitogen-activated protein kinase		58,5492814	301,452631		x		K04371
	4	Smp_336900	Inositol 1,4,5-trisphosphate receptor		15,5334374	602,730612		x		N (L-O) K04958
	1	Smp_163570	Ryanodine receptor related		2051,34732	405,215777		x		N (L-O) K04962
	4	Smp_102340	Adenylate cyclase type IX, putative		2538,44659	2203,03812		x		K08049
ZW		Smp_152330	Serine/threonine kinase		479,726111	67,9349543		x		
ZW		Smp_267270	NA		35,5357316	775,58214		x		K04345
	1	Smp_020270	Voltage-dependent L-type calcium channel subunit alpha		225,997532	43,3999267		x		K04851
	3	Smp_159990	Voltage-dependent L-type calcium channel subunit alpha		144,880033	31,1977222		x		N (L-O) K04851
	4	Smp_172640	Putative gtp-binding protein alpha subunit		31,5407288	138,300805		x		K04632
	2	Smp_245900	Phosphoinositide phospholipase C		488,040101	33,2561482		x		K05858
ZW		Smp_241420	AGC-kinase C-terminal domain-containing protein		276,304707	2843,16684		x		N (L-O) K02677
ZW		Smp_099030	Protein kinase		5020,49088	262,542507		x		K03097
	6	Smp_025010	Casein kinase II subunit beta		2154,70155	58,1874919		x		K03115
	2	Smp_125450	Guanylate cyclase		6,37015434	1428,30966		x		K12322
ZW		Smp_128680	Lipase		384,287943	649,158714		x		K13806
	2	Smp_140770	Serine/threonine-protein phosphatase with EF-hands		58,8780932	2504,30171		x		K13807
	2	Smp_122900	Putative beta-arrestin 1		386,425187	2452,69684		x		
	1	Smp_152290	Putative beta-arrestin 1		37,5075607	276,983829		x		
	4	Smp_149770	Putative g-protein coupled receptor		18,1073535	2630,64798		x		K04153
	1	Smp_319050	Transmembrane protein 164	TMEM164	859,953736	1734,3278		x		
	3	Smp_085780	Ribosomal protein L10P	28S ribosomal protein S2, mitochondrial	213,641686	217,671147		x		
	3	Smp_000370	Selenoprotein SelK/SelG	SelK SelG	340,549469	909,805039		x		
	1	Smp_127370	Glycosyl transferase, family 14	beta-1,6-N-acetylglucosaminyltransferase (E.C.2.4.1.102)	3,82520644	1552,4241		x	x	
	3	Smp_134100	G-protein coupled receptor, putative		1,2698184	6952,2787		x		Rhodospine et rhodospine like
	3	Smp_205760	Guanine nucleotide-binding protein subunit gamma		137,021604	794,43005		x	x	
		Smp_147020	Putative arf6 guanine nucleotide exchange factor	Putative arf6 guanine nucleotide exchange factor	558,598225	2788,14535		x	x	
		Smp_124660	Putative guanyl-nucleotide exchange factor	Putative guanyl-nucleotide exchange factor	129,09676	214,305466		x		
SM_V 7_MIT O		Smp_900070	Cytochrome b (mitochondrion)		7632,78497	400,256562		x	x	
SM_V 7_MIT O		Smp_900090	NADH dehydrogenase subunit 4 (mitochondrion)		8919,57763	1728,24977		x	x	
	3	Smp_308470	SEA domain	SEA	6,95190946	931,539179		x		
		Smp_084540	Forkhead-associated (FHA) domain	RNA-binding protein 5, Survival motor	1120,47018	11086,4612		x	x	
	1	Smp_158570	NA		57,0753046	674,663091		x	x	
	2	Smp_011220	Putative glucosamine-fructose-6-phosphate aminotransferase		90,3317331	502,715624		x	x	
	5	Smp_327560	Thioredoxin domain-containing protein		189,457861	11814,2048		x	x	
	6	Smp_341570	NA		195,28918	987,958037		x	x	
	7	Smp_010380	Isochorismatase domain-containing protein		168,208013	112,126106		x	x	
		Smp_900080	NA		2003,66364	1792,77786		x	x	
	6	Smp_336310		NA	17,2420897	1430,48575		x	x	
	1	Smp_072450	Putative rhodopsin-like orphan GPCR		29,4953463	536,823608				D (L-O) Rhodospine et rhodospine like
ZW		Smp_203500	G PROTEIN RECEPTOR F1 2 domain-containing protein		14,7720279	1067,93035				D (L-O)
ZW		Smp_332930	G PROTEIN RECEPTOR F2 4 domain-containing protein		1442,33196	204,827051				D (L-O)
ZW		Smp_152370	Putative arf gtpase-activating protein		1072,70429	200,986886				D (L-O) Acteurs phototransduction
	5	Smp_334600	Phosphodiesterase		10,6580078	19779,3702				D (L-O) Acteurs phototransduction
ZW		Smp_134140	Phosphodiesterase		197,662779	690,972234				D (L-O) Acteurs phototransduction
	6	Smp_341900	EF-hand domain-containing protein		126,8371	15,4622564				D (L-O) Acteurs phototransduction
	1	Smp_009600	Serine/threonine-protein kinase PLK		4487,81793	0				D (L-O) Acteurs phototransduction
	1	Smp_342030	NA		218,613756	355,567515				D (L-O) GOrsponse to oxidative stress
ZW		Smp_058690	Glutathione peroxidase		1280,34346	34,9837174				D (L-O) GOrsponse to oxidative stress
ZW		Smp_058700	Glutathione peroxidase		981,580833	11292,787				D (L-O) GOrsponse to oxidative stress
	2	Smp_047680	Ferritin		666,787097	40,2929118				D (L-O) GO homeostatic process
	2	Smp_047650	Ferritin		1814,59787	62,062005				D (L-O) GO homeostatic process
ZW		Smp_203970	Calcium load-activated calcium channel		307,763535	22293,9025				D (L-O) GO homeostatic process
	4	Smp_043030	Hexokinase		2333,69548	136,44615				D (L-O) GO homeostatic process
	2	Smp_046740	Dihydropyridyl dehydrogenase		2950,53301	254,749746				D (L-O) GO homeostatic process
ZW		Smp_201220	Non-specific serine/threonine protein kinase		1744,94912	41,3215573				D (L-O et O-L) K08959
ZW		Smp_033010	Similar to 16 kDa calcium-binding protein		21,4012449	123,838448				D (L-O) K02183
ZW		Smp_307260	Protein kinase domain-containing protein		59,4142632	536,299063				D (L-O) K04345
ZW		Smp_026400	Thyrotroph embryonic factor related		499,831745	569,006512				D (L-O) K09057



## Annexe 12

1	Smp_344020	NA		63,9769703	651,793823				D (L-O)	K04967
2	Smp_154850	Putative rap gtpase-activating protein		98,4199718	717,242079				N (L-O)	Acteurs phototransduction
ZW	Smp_164680	Synaptic ras gtpase activating protein%2C syngap%2Cputative		488,414977	3,94498225				N (L-O)	Acteurs phototransduction
1	Smp_166310	Putative ran gtpase-activating protein		1620,42997	19,017753				N (L-O)	Acteurs phototransduction
1	Smp_081190	Putative rho3 GTPase%3B Rho3 GTPase		69,6709537	0,76051515				N (L-O)	Acteurs phototransduction
7	Smp_140040	Guanylate cyclase domain-containing protein		1573,99644	1046,00776				N (L-O)	Acteurs phototransduction
1	Smp_305430	G_PROTEIN_RECEP_F3_4 domain-containing protein		30,769832	1188,47448				N (L-O)	Acteurs phototransduction
2	Smp_310770	Serine/threonine kinase		531,288616	250,577256				N (L-O et O-L)	Acteurs phototransduction
7	Smp_318260	EF-hand domain-containing protein		18,0261485	43,2845419				N (L-O)	Acteurs phototransduction
ZW	Smp_244250	NA		668,270923	6275,04501				N (L-O)	
2	Smp_316380	NA		41,8562197	8,5592052				N (L-O)	
4	Smp_126990	Putative synapsin		366,903621	1,05419522				N (L-O)	
ZW	Smp_175900	Synaptotagmin		256,045556	1577,79207				N (L-O)	
ZW	Smp_245310	NA	glutamate receptor ionotropic, NMDA 1	70,0364074	17055,1783				N (L-O)	K05208
3	Smp_176980	Putative beta-arrestin 1		4,54627442	1025,24067				N (L-O)	
ZW	Smp_180350	Rhodopsin		7,18443175	197,656822				N (L-O)	
ZW	Smp_162980	Putative rhodopsin-like orphan GPCR		10,780007	1368,605					
1	Smp_056060	Putative rhodopsin-like orphan GPCR		10,4721482	76,8694442					
ZW	Smp_167870	Rhodopsin-like orphan GPCR,putative [		0,83224902	575,984625					
ZW	Smp_172810	Putative rhodopsin-like orphan GPCR		12,4494133	194,214496					
ZW	Smp_129810	Putative rhodopsin-like orphan GPCR		3,31127787	5833,2722					
ZW	Smp_153200	Putative rhodopsin-like orphan GPCR		2,69906486	1116,50964					
2	Smp_041700	Rhodopsin-like orphan GPCR,putative		7,20934203	741,222966					
ZW	Smp_157050	Rhodopsin-like orphan GPCR,putative		6,05112192	162,565997					
3	Smp_137980	Putative rhodopsin-like orphan GPCR		6,09011859	2369,18839					
ZW	Smp_084270	Rhodopsin-like orphan GPCR,putative		1,08020081	337,75706					
1	Smp_161500	Rhodopsin-like orphan GPCR, putative		13,739393	437,280594					
1	Smp_125890	Putative DNA photolyase		655,264768	477,045066					
1	Smp_033700	DNA photolyase, putative		1287,58474	0					
4	Smp_027210	Flavoprotein domain-containing protein		23,1675611	109,08272					
2	Smp_062270	12.5 kDa retinal tissue protein (Rtp12.5),putative		267,87092	591,628504					
ZW	Smp_306970	Putative retinal dehydrogenase		8,88763091	1116,24643					
ZW	Smp_104210	Putative opsin-like receptor		6,69824823	20666,3499					
ZW	Smp_180030	Putative opsin-like receptor		0,50439852	411,97546					
ZW	Smp_170020	Putative neuropeptide receptor		9,25295687	261,767147					
ZW	Smp_141880	Putative neuropeptide receptor		4,42406751	3076,17155					
4	Smp_041880	Putative peptide (Allatostatin/somatostatin)-like receptor		0,25739638	8016,51462					
1	Smp_149580	Peptide (FMRFamide/somatostatin)-like receptor%2Cputative		0,94640122	1724,46934					
ZW	Smp_204230	G_PROTEIN_RECEP_F1_2 domain-containing protein		5,46647038	240,445056					
2	Smp_173010	Putative dro/myosuppressin receptor		11,6425195	882,661491					
2	Smp_140620	Putative g-protein coupled receptor		3,46108488	967,058522					
ZW	Smp_131980	G_PROTEIN_RECEP_F1_2 domain-containing protein		0,82338607	612,149383					
1	Smp_049330	GpcrRhopsn4 domain-containing protein		603,449227	3112,54544					

Les nombres moyens de reads indiqués en rouge sont inférieurs à 10 reads.

## Annexe 13 : Liste des gènes candidats différentiellement exprimés

Sur exprimé	C hr	Gene ID	GFF annotation	HHsearch annotation	Moyenne Reads D	Moyenne Reads N	QTL	H3K4me3	H3K9ac	Rythmique	Sélection
N	1	Smp_137300	G protein-coupled receptor, rhodopsin-like	7TM_GPCR_Srw	2	2573	Intras pé.				Rhodospine et rhodospine like
N	1	Smp_132670	EF-hand domain	Myosin Essential Chain {Bay scallop (Aequipecten irradians) [TaxId: 31199]}	6011	5333	Intras pé.				Acteurs phototransduction
N	1	Smp_152830	Na-Ca exchanger/integrin-beta4	Na/Ca exchange protein	335	13	Intras pé.			N (L-O)	K05849
N	3	Smp_326670	G protein-coupled receptor, rhodopsin-like	Platelet-activating factor receptor, Flavodoxin	4	1545	Inters pé.				Rhodospine et rhodospine like
N	3	Smp_308700	Protein phosphatase 2A regulatory subunit PR55	Serine/threonine-protein phosphatase 2A 65 kDa	8	879	Inters pé.				Acteurs phototransduction
N	4	Smp_313300	Small GTPase	Rab21 {Human (Homo sapiens) [TaxId: 9606]}	39	3109	Inters pé.			D (L-O)	Acteurs phototransduction
N	4	Smp_145540	G protein-coupled receptor, rhodopsin-like	Muscarinic acetylcholine receptor M2, Vasopressin V2	63	1438	Inters pé.			N (L-O)	K04131 + rhodospine
N	1	Smp_170610	Putative rhodopsin-like orphan GPCR		2	516		x			Rhodospine et rhodospine like
N	1	Smp_159160	Putative ef-hand domain (C-terminal) containing protein		31	7342		x			Acteurs phototransduction
N	4	Smp_336900	Inositol 1,4,5-trisphosphate receptor		16	603		x		N (L-O)	K04958
N	4	Smp_172640	Putative gtp-binding protein alpha subunit		32	138		x			K04632
N	2	Smp_125450	Guanylate cyclase		6	1428		x			K12322
N	1	Smp_127370	Glycosyl transferase, family 14	beta-1,6-N-acetylglucosaminyltransferase (E.C.2.4.1.102)	4	1552		x	x		H3K4me3 et H3K9ac surreprésentés chez nocturne
N	3	Smp_134100	G-protein coupled receptor, putative		1	6952			x		Rhodospine et rhodospine like
N	S M - V 7 - M I T O	Smp_900090	NADH dehydrogenase subunit 4 (mitochondrion)		8920	1728		x	x		GO Respiration cellulaire (N H3K9ac -1 et K4)
N	1	Smp_158570	NA		57	675		x	x		H3K9ac surreprésentés chez nocturne
N	Z W	Smp_203500	G_PROTEIN_RECEP_F1_2 domain-containing protein		15	1068				D (L-O)	Rhodospine et rhodospine like (GPR)
N	1	Smp_342030	NA		219	356				D (L-O)	GOresponse to oxidative stress

## Annexe 13

N	Z W	Smp_ 03301 0	Similar to 16 kDa calcium-binding protein		21	124				D (L- O)	K02183
N	Z W	Smp_ 30726 0	Protein kinase domain-containing protein		59	536				D (L- O)	K04345
N	Z W	Smp_ 18035 0	Rhodopsin		7	198					Rhodospine et rhodospine like (annotation)
N	Z W	Smp_ 10421 0	Putative opsin-like receptor		7	20666					Rhodospine et rhodospine like (Liang 2016 + Hoffman 2000 + Cosseau 2009)
N	Z W	Smp_ 17002 0	Putative neuropeptide receptor		9	262					Rhodospine et rhodospine like (Liang 2016)
N	Z W	Smp_ 20423 0	G_PROTEIN_RECEP_F1_2 domain-containing protein		5	240					Rhodospine et rhodospine like (Liang 2016)
D	1	Smp_ 13280 0	Myc-type, basic helix- loop-helix (bHLH) domain	Aryl hydrocarbon receptor nuclear translocator-like	91	4830	Intras pé.				Motif bHLH
D	1	Smp_ 00721 0	Rab-GTPase-TBC domain	Rab GTPase-activating protein 1	1264	6835	Intras pé.	x			Acteurs phototransduction
D	5	Smp_ 15898 0	EF-hand domain	Recoverin {Cow (Bos taurus) [TaxId: 9913]}	7	415	Intras pé.				Acteurs phototransduction
D	6	Smp_ 21063 0	Histone H2A	Histone H2A {Chicken (Gallus gallus), erythrocytes [TaxId: 9031]}	3662	2098	Intras pé.	x			Histone
D	3	Smp_ 30869 0	Protein phosphatase 2A regulatory subunit PR55	NA	277	764	Inters pé.				Acteurs phototransduction
D	1	Smp_ 31905 0	Transmembrane protein 164	TMEM164	860	1734		x			H3K4me3 surreprésenté chez diurne
D	3	Smp_ 00037 0	Selenoprotein SelK/SelG	SelK_SelG	341	910		x			H3K4me3 surreprésenté chez diurne
D	7	Smp_ 01038 0	Isochorismatase domain-containing protein		168	112		x	x		H3K9ac surreprésentés chez nocturne
D	Z W	Smp_ 30697 0	Putative retinal dehydrogenase		9	1116					Rhodospine et rhodospine like (annotation)

Les nombres moyens de reads indiqués en rouge sont inférieurs à 10 reads.



## **Emission rythmique des cercaires de *Schistosoma mansoni* : du phénotype au(x) gène(s)**

*Schistosoma mansoni* est un parasite responsable de la bilharziose. Cette maladie tropicale négligée sévit aussi en Corse depuis 2013. Le cycle de vie du parasite est divisé entre un hôte mammifère où il fait sa reproduction sexuée et un hôte mollusque d'eau douce où il fait une multiplication clonale. La forme du parasite qui infeste le mammifère est la cercaire. L'émission de ces cercaires suit un rythme de 24h dont le pic est calqué sur le pic d'activité aquatique de l'hôte mammifère. Afin de mieux comprendre la transmission des schistosomes vers l'hôte mammifère, les travaux de cette thèse s'intéressent au rythme d'émission des cercaires des chronotypes diurne et nocturne de *Schistosoma mansoni*. Cette étude est transversale en étudiant le phénomène de rythme des cercaires du phénotype (chronobiologie, histologie) jusqu'aux gènes impliqués (génétique, épigénétique et transcriptomique). Les approches de chronobiologie ont révélé que le rythme d'émission des cercaires s'adapte à l'inversion de photopériode et disparaît en l'absence d'alternance jour/nuit (lumière ou obscurité continue) sans pour autant statuer sur la nature endogène ou exogène de ce rythme. L'approche histologique a dévoilé que, en absence d'alternance jour/nuit, il y a une accumulation de cercaires matures en attente à l'intérieur des sporocystes et un arrêt de la cercariogenèse mettant ainsi en lumière le sporocyste comme autre acteur dans le phénotype du rythme d'émission des cercaires. Au niveau moléculaire, l'association des approches génétique (linkage mapping), épigénétique (ChIPmentation) et transcriptomique (RNA-seq) a montré qu'aucun gène canonique de l'horloge (clock) n'est impliqué. Toutefois les trois approches ont identifié des gènes impliqués dans le mécanisme de phototransduction, avec la rhodopsine comme photorécepteur.

**Mots-clés :** *Schistosoma mansoni*, Rythme d'émission des cercaires, Chronobiologie, Génétique, Epigénétique, Transcriptomique.

### ***Schistosoma mansoni* cercariae shedding rhythm: From phenotype to gene(s)**

*Schistosoma mansoni* is a parasite responsible for bilharzia. This neglected tropical disease has also been prevalent in Corsica since 2013. The life cycle of the parasite is divided between a mammalian host where it does its sexual reproduction and a freshwater snail host where it does clonal multiplication. Cercariae is the parasite form that infests the mammal. The cercariae shedding follows a 24-hour rhythm based on the water activity peak of the mammalian host. In order to enhance the understanding of schistosome transmission to the mammalian host, the work of this thesis focuses on the cercariae shedding rhythm of the diurnal and nocturnal chronotypes of *Schistosoma mansoni*. This study is transversal by studying the phenomenon of cercariae shedding rhythm from the phenotype (chronobiology, histology) to the genes involved (genetics, epigenetics and transcriptomics). Chronobiology approaches revealed that the cercariae shedding rhythm adapts to photoperiod inversion and disappears in the absence of day/night alternation (continuous light or darkness) without however ruling on the endogenous or exogenous nature of this rhythm. The histological approach revealed that, in the absence of day/night alternation, there is an accumulation of mature cercariae waiting inside the sporocysts and an arrest of cercariogenesis thus highlighting the sporocyst as another actor in the phenotype of the cercarial emission rhythm. At the molecular level, the combination of genetic (linkage mapping), epigenetic (ChIPmentation) and transcriptomic (RNA-seq) approaches showed that no canonical clock gene is involved. However, all three approaches identified genes involved in the phototransduction mechanism with rhodopsin as the photoreceptor.

**Keywords:** *Schistosoma mansoni*, Cercariae shedding rhythm, Chronobiology, Genetics, Epigenetics, Transcriptomics.