

Université de Montréal

L'influence de mitochondries exogènes : changements phénotypiques chez Chrosomus eos

par Vincent Chapdelaine

Départements des Sciences Biologiques,
Faculté des arts et sciences

Mémoire présenté
en vue de l'obtention du grade de Maîtrise
en Science Biologiques

Juillet 2016

© Vincent Chapdelaine, 2016

Résumé

Les interactions mito-nucléaire sont au centre des fonctions de la mitochondrie, telles que la production d'énergie. Or, peu d'informations sont connues sur comment le génome mitochondrial peut influencer la réponse nucléaire. Les cybrides (hybrides cytoplasmiques) peuvent présenter des modifications phénotypiques (adaptatives ou non) dûe à ces interactions. Dans cette étude, les différents mitotypes (haplotype mitochondriale) de l'espèce *Chrosomus eos* furent utilisés comme modèle afin d'étudier plus en profondeur les interactions entre les différents génomes d'une cellule. Ce complexe présente en plus du type sauvage, deux types de cybrides, dont les mitochondries proviennent de deux refuges glaciaires différents : Mississippien et Atlantique. Cette étude fut effectuée sur des individus en sympatrie afin de prendre en compte l'influence environnementale. Des populations en allopatrie furent également mesurés séparément afin de compléter les études précédentes. Afin de répondre à ces objectifs, plusieurs approches furent employées, adressant ainsi divers niveaux d'intégration : la méthylation, la transcription, la protéomique et l'activité enzymatique. Des différences entre les divers mitotypes furent observées, mais de magnitude inférieure à l'influence environnementale. Ce résultat suggère donc que les différences associées au mitotype en sympatrie affecte leur répartition et leur habileté à coloniser différents types d'environnements, créant ainsi la similarité intra-mitotype inter-lac observé dans les études précédentes. Pour adresser l'analyse transcriptomique, une référence a été produite. Cette référence offre beaucoup d'informations pour des applications futures.

Mots-clés : Transcriptomique, Cybride, *Chrosomus eos*, RNA-seq, méthylome, protéome, Enzymatique

Abstract

The mito-nuclear interactions are at the center of the mitochondrial functions, such as energy production. Despite their importance, little is known about how the influence of the mitochondrial genome on the nuclear genes expressions. The cybrids (cytoplasmic hybrids) can present phenotype modification (adaptive or not) caused by the mito-nuclear interactions. In this study, different mitotype of Cybrids of the *Chrosomus eos* species were used as model to further study the interaction between a cell's genomes. This complex of species has two types of cybrids, the mitochondria of which originates from two different glacial refugia : Atlantic and Mississippian. This set up of this study was a sympatric lake, to minimise the environmental factor. Allopatric population were also separately analysed to supplement past studies. To accomplish these goals, multiple methods were used, assessing also multiple levels of gene expression: Methylation, transcription, protein and enzymatic activity. Differences between mitotypes were observed, but of lesser magnitude then the environmental factor. Thus, this result suggest that the differences associated to the sympatric mitotypes drive a deference in environment colonization, which would create the inter-lake intra-mitotype similarity observed in past studies. To assess the transcription of mitotype, a transcriptomic reference was produced. This reference offers a lot of information and future applications.

Keywords : Transcriptomic, Cybrid, *Chrosomus eos*, RNA-seq, methylome, proteome, enzymatic activity

Table des matières

Résumé.....	ii
Abstract.....	iii
Table des matières.....	iv
Liste des tableaux.....	vi
Liste des figures	vii
Liste des abréviations.....	viii
Remerciements.....	ix
Chapitre 1- Introduction.....	1
Plasticité phénotypique	1
Épigénétique	1
Environnement intracellulaire.....	2
Mitochondrie.....	2
Communication rétrograde	4
Cybride.....	4
Objectif	6
Hypothèse	7
Modèle biologique : Les cybrides <i>Chrosomus eos</i>	7
Chapitre 2 - Genomic Resources Notes	10
Chapitre 3 - To what extent can foreign mitochondria impact the phenotype, a holistic study of the red belly dace <i>Chrosomus eos</i>	13
Contribution	13
Résumé.....	13
Abstract.....	14
3.1 Introduction.....	15
3.2 Materials and methods	19
Sampling and genotyping	19
Epigenetic	20

Transcriptomics.....	21
Proteomic	22
Enzymatic activity	23
Statistical analyses	23
3.3 Results.....	24
Epigenetic	24
Transcriptomics.....	26
Proteomics.....	28
Enzymatic activity	30
3.4 Discussion	31
3.5 Acknowledgments.....	36
Supplementary Material – Transcriptomics.....	37
Chapitre 4 - Conclusion	40
Les cybrides : le modèle parfait.....	40
Influence des mitochondries exogènes	41
Retour sur les objectifs.....	41
Distribution géographique	44
Hypothèse des changements post-traductionnels.....	46
Liens entre les approches	47
Conclusion	48
Bibliographie.....	i
Annexes.....	vii
Annexe 1 Article Gene identification	vii

Liste des tableaux

Table 4.1: Number and origin of the <i>C. eos</i> wild type (WT), Mississippian (MC) and Atlantic cybrids (AC) used for the epigenetic, the transcriptomic, the proteomic and the enzymatic activity analyses.	20
Supplementary table SI. The significantly differentially expressed and characterized transcripts in white muscle, when <i>C eos</i> wild type and Mississippian cybrids from Lake Desjardins are compared. Transcripts associated to foreign mitochondria effects (14) alone are indicated by an asterisk. mRNAs homologous to proteins attributed to mitochondria (3) are in bold. Only matches with an e-value below 10^{-5} were considered. Significance was set as $P_{adj} < 0.05$	vii
Table supplémentaire SII. Analyse des termes <i>GO</i> surreprésentés entre WT Desjardins et WT Saad. Analysé par le package GOSEQ de R, seulement les termes significativement surreprésentés à un $P_{adj} < 0,05$ sont présentés.	Erreur ! Signet non défini.
Table supplémentaire SIII. Analyse des termes <i>GO</i> surreprésentés entre CM et WT. Analysé par le package GOSEQ de R, seulement les termes significativement surreprésentés à un $P_{adj} < 0,05$ sont présentés.....	Erreur ! Signet non défini.

Liste des figures

Figure 1.1. Les éléments de la chaîne OXPHOS illustrant la contribution du génome mitochondrial (en couleur). Figure partiel de Schon et al 2012 (Schon <i>et al.</i> 2012).....	4
Figure 1.2. Mécanisme proposé de reproduction du complexe <i>Chrosomus eos-neogaeus</i>. Figure de Angers et al. (2012).....	8
Figure 3.1: RDA of the different epigenetic profiles obtained from the caudal fin of the <i>C. eos</i> wild type, Mississippian and Atlantic cybrids. Circles and triangles refer to individuals from Desjardins Lake and Beaver Lake respectively. Black circles correspond to the wild type, grey circles to Atlantic cybrids and blank circles and triangles to Mississippian cybrids.....	25
Figure 3.2: Smear plot visualization of differential expression analyses of <i>C. eos</i> wild type and Mississippian cybrid transcriptomes. (A) Transcriptome comparisons between wild types from different lakes (Desjardins and Saad). (B) Transcriptome comparisons between wild type and Mississippian cybrids from the same lake (Desjardins). Black dots represent the mRNAs considered as differentially expressed ($P_{adj} < 0.05$). Triangles refer to transcripts that could not be represented in the logFC interval. CPM = counts per million; FC = fold change.	Erreur ! Signet non défini.
Figure 3.3: Protein expression comparisons in <i>C. eos</i> wild type and cybrids. Among the 132 studied proteins, only the ones presenting a more than two-fold difference in expression were considered. The comparisons were performed with two protein gels for each <i>C. eos</i> form. The blank boxplot represents the intra-form comparisons while the dotted boxplots represent the inter-form comparisons. Intra-from comparisons were pooled. WT = wild type; AC = Atlantic cybrid; MC = Mississippian cybrid.	29
Figure 3.4: Complex IV activity in the white muscle of <i>C. eos</i> wild type (WT), Mississippian (MC) and Atlantic (AC) cybrids. Significance was set as $P < 0.05$. Letters denote differences between <i>C. eos</i> forms: a is significantly different from b.....	30
Figure 4.1 : Distribution des cybrides (en noir) et des types sauvages (en rouge) du complexe <i>Chrosomus eos-neogaeus</i>. Tiré de Mee et Taylor 2012.....	45

Liste des abréviations

AC : Atlantic Cybrids

ADN\DNA : Acide Désoxyribonucléique \ Deoxyribonucleic Acid

ARN \ RNA : Acide ribonucléique \ ribonucleic Acid

ARNm \ mRNA : ARN messenger \ messenger RNA

ADNmt \ mtDNA : ADN mitochondrial \ mitochondrial DNA

ARNt : ARN de transport

ARNr : ARN ribosomique

bp : Base pair

cDNA : complementary DNA

CPM : Count per million

FC: fold-change

IRIC : Institut de Recherche en Immunologie et Cancérologie

MC : Mississippian Cybrids

MSAP : Methylation sensitive amplified polymorphism

nucDNA : nuclear DNA

NR : non-redundant

NUMTs : *nuclear mitochondrial DNA segments*, Segment d'ADN mitochondrial encode dans le génome nucléaire.

N50 : the shortest sequence length at 50% of the assembly

OXPPOS : Oxidative Phosphorylation

RDA: Redundancy Analysis

RNA-seq : RNA sequencing

ROS: Reactive Oxygen Species

WT : Wild Type

%vol : pourcentage volume

Remerciements

Tout d'abord, j'aimerais remercier mon directeur, Bernard Angers, sans qui ce projet ne se serait jamais réalisé. Merci pour tous les conseils, les encouragements et tous les bons moments.

Merci à tous les gens qui ont été de passage dans le laboratoire de Bernard, ainsi que tous ceux qui ont, ne serait-ce que pour un plus ou moins long moment, gravités aux alentours. Vous m'avez beaucoup aidé, que vous l'ayez réalisé ou non. Tant d'entre vous sont devenus des amis proches. Je ne vous oublierai jamais.

Une pensée pour ma famille, qui m'a toujours supporté inconditionnellement, Merci !

Chapitre 1- Introduction

Plasticité phénotypique

L'expression des gènes du génome n'est pas seulement déterminée par la séquence qui le compose ; l'environnement joue également un grand rôle. En effet, la panoplie de phénotypes possibles est la résultante de l'interaction du génome avec l'environnement. Ceci correspond à la plasticité phénotypique, c'est-à-dire, la capacité d'un génome à moduler son expression pour que divers phénotypes puissent en résulter en fonction de l'environnement (West-Eberhard 1989). Il ne s'agit pas d'adaptation : mécanisme qui augmente la valeur adaptative par sélection naturelle, mais d'acclimatation à l'environnement (Futuyma 2013). C'est le processus par lequel le phénotype d'un individu est altéré en fonction de signaux perçus de l'environnement. Le génome dicte néanmoins les potentiels de réponse à l'environnement. L'environnement quant à lui, agit sur les divers mécanismes dont certains vont altérer l'expression du génome en se liant et en modifiant le contexte d'expression (West-Eberhard 1989). Les sensibilités d'un génome aux stimuli de l'environnement sont donc elles aussi dictées par le génome.

Épigénétique

Certains mécanismes impliqués dans la plasticité phénotypique sont de nature épigénétique ; un ensemble de changements de la chromatine, de l'expression génique et une altération de longue durée du potentiel de transcription de la cellule qui ne sont pas nécessairement héréditaires. Il est à noter que cette définition englobe autant les réponses à l'environnement externe que les différenciations cellulaires programmées. Selon les auteurs, la définition d'épigénétique peut varier (Holliday 2006).

Les mécanismes épigénétiques sont nombreux. Les éléments les mieux connus de ce domaine d'étude sont la méthylation de l'ADN et les modifications causant un remodelage de la chromatine. Ces deux mécanismes affectent l'expression du génome de façon similaire et donc offre une cible d'étude idéale pour étudier les variations épigénétiques globale d'un organisme. La densification de la chromatine affecte indirectement l'expression selon la disponibilité des sites d'interactions avec la machinerie liée à l'expression (Tse *et al.* 1998). En général, la

méthylation des queues d'histone comprime ou décomprime la chromatine et donc régule les gènes associés, selon l'acide aminé méthylé. (Jenuwein et Allis 2001).

Également, l'acétylation des histones est associée, en général, à la décompression la chromatine, et permet indirectement la régulation des gènes associés et potentiellement leurs transcriptions (Jenuwein et Allis 2001). La méthylation de l'ADN régit l'expression des gènes, sauf les gènes constitutifs, par l'ajout de groupement méthyles sur certaines cytosines et ainsi est souvent associée à une diminution de l'expression du gène en question (Jaenisch et Bird 2003).

Un modèle emblématique est celui des abeilles. Selon leur alimentation, les individus d'une même colonie peuvent se développer en ouvrière ou en reine. En effet, la consommation de gelée royale aura comme conséquence de produire un phénotype de reine. Cette différenciation de phénotype est associée très fortement à une différence dans les profils de méthylation de l'ADN (Maleszka 2008, Lyko *et al.* 2010).

Environnement intracellulaire

La plasticité phénotypique dans les populations naturelles a pu être observée avec des signaux extrinsèques provenant de l'environnement de l'individu, d'autres études se focalise sur les signaux intrinsèques de nature développementale, mais peu d'études se sont focalisées sur des signaux à la fois interne à la cellule et extrinsèque au génome nucléaire : l'environnement intracellulaire. Cet aspect d'un stimulus environnemental interne a été suggéré comme facteur pouvant influençant l'épigénome de l'individu puisque l'environnement du noyau change (Angers *et al.* 2012). Certaines organelles possèdent leurs propres génomes co-évoluant avec le génome nucléaire (Rand *et al.* 2004). Des études portant sur les réponses nucléaires de ces variations demeurent alors importantes, afin de comprendre ces processus, d'autant plus si cette organelle occupe une fonction importante pour la cellule.

Mitochondrie

Un élément de l'environnement cellulaire dont la fonction est cruciale au fonctionnement de la cellule est la mitochondrie : une organelle d'origine procaryote fonctionnant maintenant en tant

qu'endosymbionte de la cellule (Andersson *et al.* 1998). Dans le règne animal, cette organelle est généralement héritée de la mère uniquement. Seul un nombre restreint est transmis dans les œufs, il y aura ainsi un effet de dérive lors du développement de l'enfant (Stewart *et al.* 2008). Les mitochondries sont des organelles complexes. En effet, il y en a plusieurs dans le cytoplasme d'une cellule et elles ne doivent pas être perçues comme des entités indépendantes l'une de l'autre, puisque ces dernières fusionnent et fissionnent lors de leur fonctionnement. Ensemble, celles-ci forment un réseau s'occupant de tâches très importantes à la cellule et à l'organisme. Un des rôles les plus connus est celui de la production d'énergie *via* la respiration cellulaire, mais d'autres demeurent moins connus, entre autres : la communication cellulaire, l'apoptose, la fécondation, le vieillissement, la thermogénèse, les fonctions de synthèse et la régulation calcique (Breton *et al.* 2014).

Cette organelle a une propriété particulière, tout comme le chloroplaste, reflétant leurs origines procaryotes : elle possède son propre génome. Le génome mitochondrial encode pour des protéines indispensables au fonctionnement de la mitochondrie. Dans le règne animal, les mitochondries possèdent 13 protéines dans leur génome, ainsi que 22 ARNt et 2 ARNr (Ballard et Whitlock 2004). Bien sûr ce génome seul n'est pas suffisant pour la viabilité de la mitochondrie et de ses fonctions ; il est estimé que 99 % du protéome de cette organelle soit d'origine nucléaire (Bolender *et al.* 2008). Cette coopération est le résultat de la coévolution nécessaire au bon fonctionnement de l'organisme. Dans cette coévolution beaucoup de gènes ont été transférés de la mitochondrie vers le noyau, ce qui a eu comme avantage de réduire le coût de division de la mitochondrie, profiter de la machinerie du noyau (Martin et Herrmann 1998). De plus, par la présence de reproduction sexuée au noyau, évolutivement les NUMTs (*nuclear mitochondrial DNA segments*) évitent le cliquet de Muller (Martin et Herrmann 1998). L'élément le plus représentatif de cette coévolution étroite entre ces deux génomes est la chaîne de protéines permettant l'oxydoréduction des nutriments. Celle-ci occupe une fonction intégrale à la production d'énergie. Dans cette chaîne, quatre des six complexes impliqués ont des sous-unités encodées par les mitochondries (Fig. 1), s'assemblant avec des sous-unités provenant du noyau (Saraste 1999). Pour cette fonction capitale, une balance de l'expression entre ces diverses sous-unités afin d'optimiser les ressources est attendue.

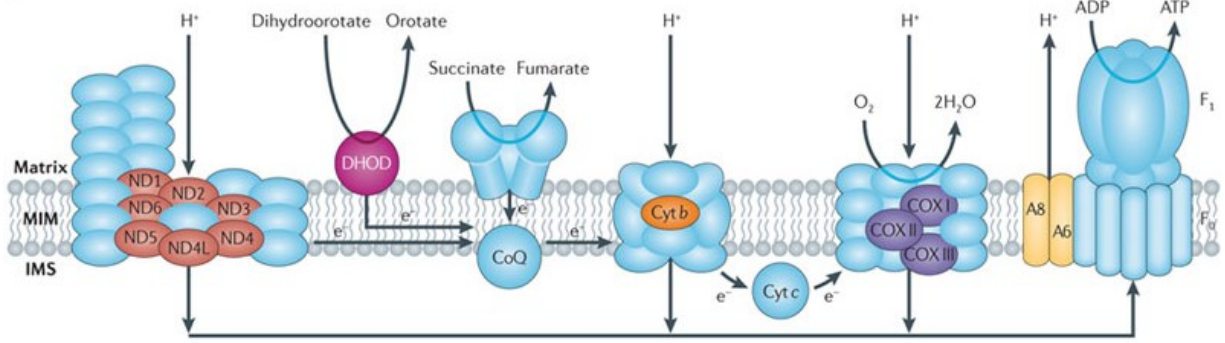


Figure 1.1. Les éléments de la chaîne OXPHOS illustrant la contribution du génome mitochondrial (en couleur). Figure partiel de Schon et al 2012 (Schon *et al.* 2012)

Communication rétrograde

Le génome nucléaire peut altérer son expression pour répondre aux divers stress de la mitochondrie, et ce par divers mécanismes. Cependant très peu de mécanismes sont encodés par le génome mitochondrial, on parle dans ce cas de communication rétrograde (Ryan et Hoogenraad 2007).

Divers états de stress engagent la mitochondrie dans une communication avec le noyau (Ryan et Hoogenraad 2007) tel que : des défauts dans la chaîne OXPHOS, un état d'hypoxie, des mutations dans l'ADNmt et un repliement incorrect des protéines. Les mitochondries communiquent certaines informations au noyau par des moyens tels que : le Ca^{2+} , une variation dans la production d'ATP et de ROS (*Reactive Oxygen Species*) (Ryan et Hoogenraad 2007). Certaines études ont également observé la présence de très petits peptides (ie. MOTS-C (Lee *et al.* 2015)) comme élément signalétique du contenu génomique de la mitochondrie.

Cybride

Les cybrides (hybrides cytoplasmique) représentent un modèle de choix pour l'étude de la relation nucléaire-mitochondriale. La définition fonctionnelle d'un cybride est un organisme avec un génome présent dans le cytoplasme (ie. mitochondrie ou chloroplaste pour les végétaux) ayant une généalogie différente de celle des haplomes du noyau. Les cybrides proviennent soit de manipulations cellulaires en laboratoire (Bunn *et al.* 1974), d'évènements d'hybridation

introgressive, comme chez la paruline à croupion jaune (*Setophaga spp.*) (Toews *et al.* 2014), ou de processus méiotiques particuliers, tel que chez *Chrosomus eos-neogaeus* (Angers et Schlosser 2007) en milieu naturel.

Dans une étude des lignées cellulaires de souris *Mus musculus domesticus*, la production de cybrides a permis de conclure que pour ces diverses lignées, la compatibilité des sous-unités nucléaires avec les sous-unités mitochondriales diminuait avec le temps de divergence entre le noyau et les mitochondries (McKenzie *et al.* 2003). Les cybrides utilisés étaient des lignées nucléaires *Mus musculus domesticus* sans ADN mitochondrial (mtDNA) dans lesquelles ont été introduites des mitochondries de plus en plus divergentes. Ce résultat a été inféré par la présence croissante de lactate, un indicateur de la décroissance de la production d'ATP par la chaîne respiratoire. Cet exemple montre un cas où l'aspect cybride de cette lignée cellulaire ne représentait pas un avantage évolutif.

Contrairement aux cybrides artificiels, la persistance de certains cybrides naturels suggère que ceux-ci peuvent offrir un avantage évolutif. Un exemple est la paruline à croupion jaune (*Setophaga spp.*), un complexe dans lequel le changement de mitochondrie par le biais d'une introgression serait important dans la migration saisonnière (Toews, Mandic *et al.* 2014). En effet, un croisement entre deux populations suivies de nombreux croisements avec une des espèces parentales a donné une forme introgressée de paruline à croupion jaune avec la mitochondrie d'une autre espèce. La forme introgressée est capable de migration saisonnière, tout comme l'espèce donneuse de la mitochondrie. La capacité d'avoir un meilleur contrôle sur leur ratio d'accepteur dans la mitochondrie pourrait en être la cause. Cette étude souligne le rôle de la mitochondrie dans l'écologie, l'évolution et le phénotype des Eucaryotes.

Ce changement de mitochondrie d'un mitotype à l'autre a été lié à un changement d'expression nucléaire et au niveau du protéome. Une étude sur l'épigénome des *Chrosomus eos* à mitochondrie *Chrosomus neogaeus*, et la forme sauvage (Angers, Dallaire *et al.* 2012), par MSAP, a révélé un patron de méthylation de l'ADN similaire intra mitotypes et différent inter mitotypes, et ce pour divers tissus. Ceci suggère ainsi que l'épigénétique est différente entre mitotypes, ce qui serait causé par la mitochondrie différente du nucléaire. Il est possible d'inférer un changement dans l'expression nucléaire par ce changement épigénétique, puisque

la méthylation bien que seulement une partie de l'épigénome régule l'expression génique (Jaenisch et Bird 2003).

Pour supporter cet argument, une autre étude a montré que pour ce même modèle, une expression protéique est différente entre mitotypes. Une faible correspondance entre les protéines des cybrides et les protéines des *C. eos* (R^2 : 0,2046 et 0,204) a été observée. Au contraire, entre les individus d'un mitotype donné, la correspondance demeure forte (R^2 : 0,8718 et 0,9515). Également, cette dernière étude a montré qu'il y a un changement d'abondance de l'expression protéique marquée de certaines protéines entre les deux mitotypes. (Deremiens *et al.* 2015).

Ces deux études sont consistantes l'une avec l'autre. Les protéines sont produites par l'ARNm et leurs abondances sont souvent corrélées avec l'abondance de leurs contigs correspondants, cette abondance étant régulée par entre autres des mécanismes épigénétiques. Une différence au niveau épigénétique est perceptible, toutefois, ceci pourrait être due à des éléments transposables, et que des différences protéiques pourraient être dues à des modifications post-traductionnelles. Une analyse du transcriptome pourrait lier définitivement la conclusion de ces deux études.

Objectif

L'objectif de ce projet fût de quantifier les changements à plusieurs niveaux d'organisation du phénotype (méthylome, transcriptome, protéome et activité enzymatique), en fonction de la nature de la mitochondrie, soit entre les cybrides et la forme sauvage, et ce en sympatrie. Cet objectif se divisant en deux sous-objectifs : évaluer l'influence de la mitochondrie sur l'organisme, globalement et spécifiquement, ainsi qu'évaluer la magnitude de l'influence environnementale. La présente étude portera sur les trois mitotypes connus avec le génome nucléaire de l'espèce *Chrosomus eos* et différentes mitochondries.

Dans le but d'examiner les différences phénotypiques causées par la mitochondrie, une analyse du protéome, du transcriptome, du méthylome et de l'activité enzymatique ont été effectuées. Pour percevoir correctement les changements de l'ensemble du transcriptome, la technologie du

séquençage de nouvelle génération sera employée, spécifiquement le RNA-seq. Celle-ci permet d'identifier et quantifier les ARNm. Pour le méthylome, la technique du MSAP a été employée, pour visualiser les changements dans la méthylation de l'ADN. La photospectrométrie a été utilisée pour l'activité enzymatique et la migration sur gel 2D pour le protéome. La migration sur gel protéique sépare les protéines en points isoélectriques distincts et permet ainsi de visualiser les changements d'expression protéique entre individus (Angers et Schlosser 2007).

Le second sous-objectif vise à observer l'effet environnemental puisque que par l'aspect allopatrique des études précédentes (Angers *et al.* 2012, Deremiens *et al.* 2015) ceux-ci peuvent présenter des facteurs confondants. Pour évaluer l'influence de l'environnement, les analyse du méthylome et transcriptome des *C. eos* d'un autre lac seront pris comme contrôle sur l'effet lac et permettre de quantifier cet effet par rapport à l'effet de la mitochondrie. De ce fait, un but alternatif de cette étude est d'évaluer l'effet environnemental des études précédentes (Angers, Dallaire *et al.* 2012, Deremiens, Schwartz *et al.* 2015). Ces lacs sont à proximité des individus du lac principal.

Hypothèse

Lorsqu'on explore les différences inter-mitotype, notre hypothèse nulle est qu'il n'y a pas de différence entre mitotypes. L'hypothèse alternative est que nos divers mitotypes sont tous différents les uns des autres ou seul un mitotype est différent. Lorsque l'on adresse l'influence environnementale, l'hypothèse nulle est qu'il n'y ait pas d'effet environnemental ou moindre que l'effet du mitotype, supportant ainsi les conclusions des études précédentes. L'hypothèse alternative est que l'effet environnemental soit égal ou supérieur à l'effet de la mitochondrie, apportant ainsi un nouveau contexte aux études précédentes en allopatrie.

Modèle biologique : Les cybrides *Chrosomus eos*

Les cybrides d'origine naturelle comme *C. eos* à mitochondrie *C. neogaeus*, offrent une occasion rare d'observer l'acclimatation des génomes en réponse à des modifications dans le cytoplasme et dans ce cas au niveau de la mitochondrie. En effet, la mitochondrie étrangère de ces organismes est intégrée de façon subite lors de leur formation. Ce modèle offre la possibilité d'étudier le rôle de la plasticité dans la coévolution d'une espèce avec sa mitochondrie, une relation primordiale à la vie d'organismes complexes.

Lorsque les espèces *C. eos* et *C. neogaeus* s'hybrident, il y a création d'une femelle hybride (2n) possédant un haplome *C. eos* et un haplome *C. neogaeus*. Cet hybride se reproduit par gynogenèse, une forme de reproduction clonale. Ce mode de reproduction implique la formation d'œufs non réduits, nécessitant le spermatozoïde d'une des deux espèces parentales pour en déclencher le développement de l'œuf (Goddard *et al.* 1998). Le génome du spermatozoïde ne sera en général pas incorporé. La progéniture issue de cette gynogenèse sera donc constituée de clones hybrides identiques à la mère.

Il peut toutefois survenir des incidents où la gynogenèse sera imparfaite et le génome du spermatozoïde sera intégré dans l'œuf. Lorsque ce génome est intégré, un triploïde est alors créé.

Puisque le complexe à l'étude est observé uniquement avec *C. neogaeus* comme espèce maternelle et *C. eos* comme espèce paternelle, l'hybride diploïde ainsi que triploïde possède une mitochondrie de l'espèce *C. neogaeus*. Il y a un retour à un génome *C. eos* après la reproduction d'un individu triploïde avec un *C.eos* (Goddard et Schultz 1993). Puisque le génome mitochondrial est transmis de façon maternelle, l'individu final sera cybride.

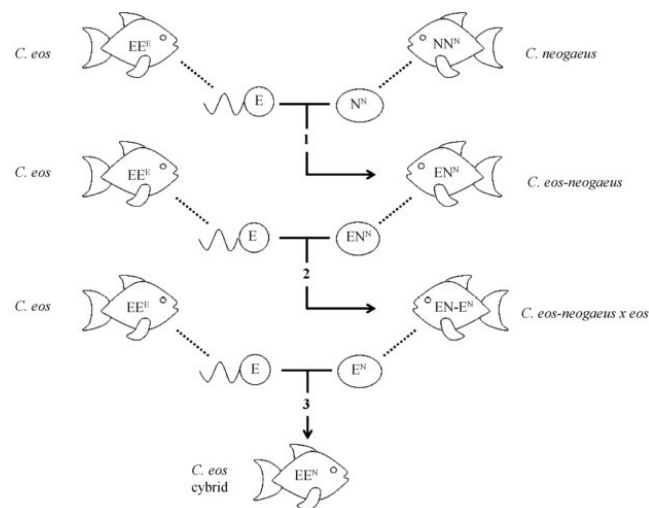


Figure 1.2. Mécanisme proposé de reproduction du complexe *Chrosomus eos-neogaeus*. Figure de Angers et al. (2012)

Le modèle *Chrosomus eos* présente plusieurs haplotypes mitochondriaux (mitotype) différents possible. En effet, un de ces mitotypes est la mitochondrie *C. eos*, la forme sauvage, et deux mitotypes héritées de *Chrosomus neogaeus*, des cybrides. Ces cybrides sont différents puisque les individus ayant transmis leur mitochondrie sont descendants de population de deux refuges glaciaires différents, soient le refuge mississippien et le refuge atlantique. Ces mitotype présente des divergences de près de 10% des nucléotides d'une séquence lorsque comparés à l'haplotype sauvage (9,6% pour l'haplotype mississippien et 9,14% pour l'atlantique). Entre les formes de cybrides, une divergence de 1 million d'années est présente, soit 1,35% de divergence d'entre ces deux haplotypes mitochondriaux (Deremien et al., en préparation).

Les génomes nucléaires des divers mitotypes (mississippien, atlantique et sauvage) peuvent être très génétiquement similaire lorsqu'en sympatrie. Une analyse des microsatellites de ces trois mitotypes à l'intérieur d'un lac a révélé que ceux-ci semblent être très proches génétiquement sans être une seule et même population, suggérant un flux génétique considérable entre ces populations, contrôlant ainsi pour une influence du génome nucléaire (Deremiens et al., en préparation)

Chapitre 2 - Genomic Resources Notes

This article details the availability and transcriptome sequence data, assembled contigs, annotated contigs of the *Chrosomus eos*.

The northern redbelly dace *Chrosomus eos*, (formerly *Phoxinus eos*) is a fresh water cyprinid common to north eastern North-America. Hybridization events between this species and the finescale dace *Chrosomus neogaeus* result in the production of the clonal hybrid *Chrosomus eos-neogaeus*. In absence of genetic variation, these fish represent a useful biological model to test evolutionary and ecological hypotheses in vertebrates.

An additional interest of this species complex is the spontaneous and complete introgression of mitochondrial genome of *C. neogaeus* in the *C. eos* leading to the production of cybrids. These cybrids occurred naturally and are widespread in the freshwater lakes of North America. Assuming a 1.3% divergence per million years, *C. eos* and *C. neogaeus* mtDNA would have diverged for about six million years. According to the important metabolic and physiologic differences observed in allopatry, coexistence of both forms provides an ideal system to investigate the effects of mitochondrial introgression on evolutionary processes, such as genomes coadaptation and sympatric speciation.

Interestingly, both *C. eos* wild type and cybrids can also be found in sympatry. In addition, a divergence of approximately one million years has been observed among mitotypes of *C. neogaeus* in cybrids. Such divergence suggests these mitochondria remained isolated in the distant Mississippian and Atlantic glacial refuge during the Pleistocene. The lack of reproductive isolation between wild type and cybrid leads to the homogenization of the genetic nuclear genome. Altogether, these characteristics make this species an invaluable model organism to assess the effect of allospecific mitochondria on the expression of the nuclear genome.

However, information of neither genomic nor transcriptomic survey is available in the literature. This transcriptome will serve as a reference to explore the effect of cytoplasmic hybridization in this natural complex or how nuclear genome can accommodate different mitochondrial genomes in cybrids, in addition, the reference transcriptome of *Chrosomus eos* will allow the study of hybrids and their intrinsic properties. As a genomic resources the raw or filtered data

can be used as annotation in a future genomic assembly project. This transcriptome can serve as a reference for the development of further genomic resources by the discovery of SNPs, with little further transcriptomic studies...

Meta-information

- Sequencing center: IRIC (“Institut de recherche en Immunologie et cancerologie”)
- Platform and model: Hiseq 2000
- Design description: Goals were to (i) generate a novel gene atlas for *Chrosomus eos*, (ii) to functionally annotate genes and (iii) to identify potential differentiation between cybrids and Wild Type.
- Analysis type: mRNA sequencing
- Run date: 2015-06-22

Library

- Strategy : non-normalized cDNA libraries
- Taxon: *Chrosomus eos*
- Sex: Adult-Female
- Tissus : White muscle
- Location: Fish were sampled during the summer of 2014 in Lake Desjardins and Lake Saad, southern Quebec, Canada (45°54'58.00"N, 74°04'23.00"W) (45°54'53.6"N, 74°01'40.2"W), respectively.
- Sample handling: Fresh material was immediately frozen in dry-ice after sampling and stored at – 80°C until RNA extraction.
- Total RNA extraction :
- Layout: Illumina as paired-end reads of length 101 bp
- Library construction protocol: KAPA RNA Stranded PolyA according to manufacturer instruction.

Processing

- Assembly and annotation- Sequenced reads were filtered by the filter option on the Trinity software along with custom features for poly A tails (Ax20 as adaptor), then de novo assembled using the same package at default settings. The transcriptome assemblies were blasted against a selection of the non-redundant (NR) NCBI protein database (Cypriniform) using the BLASTx program of the BLAST suite (Altschul 1997) using an E value threshold of 10^{-5} . The transcriptome was also annotated using The Trinotate V3.0 and Transdecoder software, with the Swissprot, Uniref90 and Pfam database. Therefore, the gene ontology was inferred by the Swissprot identification. The Bed File was generated by the Trinotate software.

Chapitre 3 - To what extent can foreign mitochondria impact the phenotype, a holistic study of the red belly dace *Chrosomus eos*

Léo Deremiens, Vincent Chapdelaine, Logan Schwartz, Annie Angers, H el ene Gl emet, Bernard Angers

Contribution

L.D. performed the enzymatic, epigenetic analyses and the sampling/genotyping, L.S. performed the proteomic analysis, V.C. performed the transcriptomic assembly and analysis. All project although complementary in nature, their methodologies were thought independently by each party under the supervision of B.A. LD and VC contributed equally to the production and writing of this article.

R esum e

Le transfert d'ADNmt entre esp eces affecte g en eralement les fonctions mitochondriales ainsi que la valeur adaptative des individus. Toutefois, diverses  etudes d emontrent que de tels transferts peuvent  egalement  etre b en efiques. Bien que les introgressions mitochondriales soient connues pour induire des modifications ph enotypiques, de nombreuses questions persistent encore  a ce jour concernant l' etendue de cette influence sur les diff erents niveaux d'organisation du ph enotype. Le complexe *Chrosomus eos* repr esente un syst eme unique pour r epondre  a ces questions. Au sein de ce complexe, trois formes sympatriques diff erentes peuvent  etre distingu ees dans un m eme lac : la forme sauvage et deux formes cybrides arborant l'ADNmt de

Chrosomus neogaeus issus des refuges glaciaires Mississippien et Atlantique. Dans cette étude, nous évaluons les variations épigénétiques, transcriptomiques, protéomiques ainsi que l'activité du complexe IV de la chaîne respiratoire mitochondriale parmi les trois formes détectées. Tel qu'escompter, l'amplitude de ces variations était plus prononcée entre la forme sauvage et les cybrides qu'entre les cybrides Mississipiens et Atlantiques ; ces deux derniers étant caractérisés par des ADNmt moins divergents. Qui plus est, lorsque comparée à une étude similaire mais considérant des formes allopatriques, notre approche holistique révèle que les modifications phénotypiques détectées entre des mitotype sympatriques étaient moins importantes. Ce contraste pourrait provenir d'effets environnementaux et nucléaires dérivant de l'utilisation de poissons allopatriques. De façon générale, nos résultats démontrent que les mitochondries allospécifiques peuvent modifier le phénotype de *C. eos*, à différents niveaux d'organisation phénotypique, lorsque la forme sauvage et les cybrides sont retrouvés en sympatrie.

Abstract

Between-species transfers of mtDNA usually have deleterious effects on mitochondrial function and individual fitness. However, various studies demonstrate that such transfers can also be beneficial. Although mitochondrial introgressions are known to induce phenotypic changes, many questions remain regarding how much these events can impact the phenotype, at various levels. The *Chrosomus eos* complex is a unique system allowing to address these questions. In this complex, we find three distinct forms living in sympatry in a single lake: a wild type and two different cybrids harboring the mtDNA of *Chrosomus neogaeus* derived from the Mississippian and the Atlantic glacial refuges. Here, we assessed epigenetic, transcriptomic, proteomic and enzymatic activity changes among these three forms. According to our results, each of these phenotypic levels displayed variations. As expected, because of the higher mtDNA

divergence detected between wild type and cybrids, the magnitude of this variation was more important between these two than between Mississippian and Atlantic cybrids. Additionally, this holistic survey demonstrated that the phenotypic modifications observed between the sympatric wild type and cybrids were less pronounced than the ones detected from allopatric *C. eos* mitotypes, analysed in a previous study. This discrepancy might be the result of environmental and nuclear effects introduced by the use of allopatric fish. Ultimately, our results show that allospecific mitochondria are able to impact *C. eos* phenotype, at various biological levels, when wild type and cybrids are found in sympatry.

3.1 Introduction

Mitochondrial energy production, which affects every aspect of cellular and individual fitness, requires the interaction of two genomes. The nuclear genome (nucDNA) encodes approximately 1500 proteins that are imported to mitochondria where they constitute the majority of the organelle proteome (Calvo *et al.* 2006). The mitochondrial genome (mtDNA), on the other hand, generally encodes for 22 transfer RNAs, 2 ribosomal RNAs and 13 protein subunits. The 13 proteins are all involved in the formation of most of the mitochondrial respiratory chain complexes (Blier *et al.* 2001, Rand, Haney *et al.* 2004). The ~74 other subunits of these complexes are coded by the nucDNA (Rand, Haney *et al.* 2004). Specific interactions between the products of both genomes are required at many levels; including mtDNA replication, repair, transcription, translation and assembly of the respiratory chain complexes (Ellison and Burton 2010, Burton *et al.* 2013). The critical importance of maintaining the specificity of these interactions suggests the necessity of a strong coevolution between both genomes (Blier, Dufresne *et al.* 2001, Rand, Haney *et al.* 2004, Ballard and Melvin 2010, Burton, Pereira *et al.*

2013, Wolff *et al.* 2014).

Despite this strong constraint, an increasing number of natural organisms characterised by a mtDNA and a nucDNA derived from divergent species have been discovered (Glémet *et al.* 1998, Senjo *et al.* 1999, Boratynski *et al.* 2011, Du *et al.* 2011, Toews *et al.* 2014, Deremiens, Schwartz *et al.* 2015). These cybrids may suffer from a disruption of the mito-nuclear interaction specificity, resulting in an alteration of the mitochondrial function and individual fitness (Kenyon and Moraes 1997, McKenzie, Chiotis *et al.* 2003, Ellison and Burton 2006, Arnqvist *et al.* 2010). Nevertheless, various studies demonstrate that between-species transfers of mtDNA can also be neutral (McKenzie, Chiotis *et al.* 2003, Blier *et al.* 2006, Boratynski, Alves *et al.* 2011) or even have beneficial effects (Pichaud *et al.* 2012, Toews, Mandic *et al.* 2014, Deremiens, Schwartz *et al.* 2015). As an example, the artificial introgression of the fly *Drosophila simulans* *siII* mitochondria into a *Drosophila simulans* *siIII* nuclear background induces an increase of the respiratory chain catalytic capacity, eventually leading to a better fitness of the individuals (Pichaud, Ballard *et al.* 2012). Although these mitochondrial introgressions are known to induce phenotypic changes, many questions still remain about how much allospecific mitochondria can impact the various phenotypic levels (e.g. epigenome, transcriptome, proteome, *etc.*).

The *Chrosomus eos-neogaeus* fish complex (Cyprinidae, Pisces) is a particularly interesting model to address these questions. Among the various biotypes found in this complex, a spontaneous cybrid bearing the nucDNA of *Chrosomus eos* and the mtDNA of *Chrosomus neogaeus* can be distinguished (Dawley *et al.* 1987, Goddard, Megwinoff *et al.* 1998). These

organisms originate from the crossfertilization of *C. eos-neogaeus* \times *eos* triploid hybrid females by *C. eos* males. During the gametogenesis of the triploid individuals, it is assumed that the unmatched set of chromosomes is discarded and meiosis solely occurs on the *C. eos* diploid genome, producing haploid eggs (Goddard, Megwinoff et al. 1998). When fertilized by a *C. eos* male, these eggs give birth to individuals with a pure *C. eos* nuclear diploid genome and a female inherited *C. neogaeus* mitochondrial genome (Binet and Angers 2005, Angers and Schlosser 2007, Angers, Dallaire et al. 2012, Deremiens, Schwartz et al. 2015) . As does *C. eos* wild type, these cybrids reproduce sexually.

Although generally allopatric, *C. eos* wild type and cybrids have been found in sympatry (Deremiens et al. in prep; Mee and Taylor, 2012). A previous study (Deremiens et al. in prep) have revealed the existence of three different forms of *C. eos* in a same lake of the Laurentian region (Quebec, Canada): the wild type and two different cybrids harboring the mitochondrial genome of *C. neogaeus* from the Mississippian and the Atlantic glacial refuges. Based on *nd3-nd4l* and *cox1* gene sequences, Mississippian and Atlantic *C. neogaeus* mtDNA of cybrids are characterised by a divergence of 1.20%, while a difference of 8.38% and 8.44% is found between wild type and Mississippian cybrids and between wild type and Atlantic cybrids, respectively. Assuming a 1.3% divergence per million years (Bermingham *et al.* 1997), these results imply that the mitochondria, harbored by the two cybrids, have diverged for approximately one million years while *C. eos* and *C. neogaeus* mtDNA would have diverged for about seven million years. This difference represents a unique opportunity to study the constraints of mito-nuclear genome interactions at different time scales, in naturally occurring organisms.

According to Angers et al. (2012) and Deremiens et al. (2015), allopatric *C. eos* wild type and cybrids present methylome, enzymatic activity, proteome and swimming performance differences. However, because allopatric individuals were used, these studies could not entirely account for possible environmental and nucDNA effects (Deremiens et al., 2015). The discovery of *C. eos* forms that do not show notable nuclear divergences (Deremiens et al., in prep), provide an ideal system to address these issues and perform a holistic study of allospecific mitochondria impact upon the various biological organization levels of the phenotype.

This study aims at assessing the influence of allospecific mitochondria on *C. eos* nuclear genome considering various levels of biological integration. To address this, wild type, Mississippian and Atlantic cybrid fish living in sympatry were compared through a holistic approach. We first analysed the epigenetic and transcriptomic variations associated with the nuclear gene expression. Afterwards, higher organisation levels were assessed through proteomic and enzymatic analyses. Because cell viability and mitochondrial metabolism decrease as the genetic distance between mtDNA and nucDNA increases (McKenzie, Chiotis et al. 2003, Bayona-Bafaluy *et al.* 2005), we expected a greater divergence between *C. eos* wild type and cybrids than between Mississippian and Atlantic cybrids. Moreover, since no significant difference is measured among the nuclear genomes of the sympatric *C. eos* forms (Deremiens et al. in prep), and because environmental effects are minimized compared to previous study (Deremiens, Schwartz et al. 2015), the phenotypic modifications detected among sympatric fish are expected to be less pronounced than the ones measured between allopatric wild type and cybrids.

3.2 Materials and methods

Sampling and genotyping

The sympatric forms of *C. eos* were sampled in Desjardins Lake, southern Quebec, Canada (45°54'58.00"N, 74°04'23.00"W). Regarding the epigenetic and transcriptomic surveys, individuals were also sampled in Beaver Lake (45°55'29.6"N, 74°03'52.5"W) and Saad Lake (45°54'53.6"N, 74°01'40.2"W) respectively. These fishes were used as reference group to compare the magnitude of environmental/nuclear effects to mitochondrial effects. The number and the origin of wild type, Mississippian and Atlantic cybrids used in each analysis are found in Table I.

All Sampled individuals were genetically identified according to Binet and Angers (2005). Briefly, total DNA extractions were carried out by proteinase K digestion followed by phenol-chloroform purification and ethanol precipitation (Sambrook et al., 1989). An intron of the mesoderm specific transcript gene (PEG/MST1) was amplified to differentiate *C. eos* from hybrids. Afterwards, the amplification of a cytochrome c oxidase subunit I gene segment allowed the discrimination of wild type from cybrids. To differentiate Mississippian cybrids from Atlantic cybrids, an additional step was conducted: a segment of the mitochondrial *nd5* gene was amplified (5' to 3' primer sequences: GTGACGGTCCCCCTAATTGG and TGGCATCCTTTGAGAAGAACC) and then digested with the restriction enzyme *NcoI*. The Mississippian haplotype was characterised by a single 559 bp fragment whereas the Atlantic form had two 129 and 428 bp fragments.

Lake	<i>C. eos</i> form	Epigenetic	Transcriptomic	Proteomic	Enzymatic activity
Desjardins	WT	7	2	4	5
	MC	8	2	4	5
	AC	8	2	4	7
Beaver	MC	8			
Saad	WT		2		

Table I: Number and origin of the *C. eos* wild type (WT), Mississippian (MC) and Atlantic cybrids (AC) used for the epigenetic, the transcriptomic, the proteomic and the enzymatic activity analyses.

Epigenetic

A survey of the DNA methylation variation was performed on female fish sampled on the same day, at the same time. DNA from the caudal fin, was used for the analysis. The epigenetic variation was investigated by assessing the methylation polymorphism at CCGG motifs, using a methylation-sensitive amplified polymorphism analysis (MSAP; Xiong et al., 1999). Fragments displaying a methylation polymorphism with the methylation sensitive treatment (*HpaII*) were considered only when the non-sensitive (*MspI*) treatment exhibited a different banding pattern across individuals. Selective amplifications were performed using ten primer combinations (HpaII-CAA; HpaII-CAC; HpaII-CAG; HpaII-CAT; HpaII-CCA; HpaII-CCC; HpaII-CCG; HpaII-CCT; HpaII-CGT; HpaII-CTG). Amplification products were separated on a 6% denaturing polyacrylamide gel (19:1 acrylamide:bis-acrylamide) and visualized using silver nitrate staining (Bassam *et al.* 1991).

Transcriptomics

A transcriptomic survey was performed to assess the magnitude of *C. neogaeus* mitochondria impact on the *C. eos* nuclear genome expression. Female fishes were euthanized in the field by eugenol overdose, immediately placed on dry ice, brought back to the laboratory and then stored at -80°C. Total RNA was extracted from the white muscle of these fish, as described by Deremiens et al. (2015). To prepare samples for RNA sequencing on the Illumina platform, mRNA was separated from the extracted total RNA and then retro transcribed into cDNA using a Truseq Stranded mRNA sample prep kit. cDNA libraries were constructed with the same kit. Ultimately, samples were barcoded, multiplexed and sequenced (paired-end, 100bp) in a single lane on the Illumina HiSeq2000 platform at IRIC (Université de Montréal, Quebec, Canada).

De novo assembly of the transcriptome was conducted in *Trinity* version 2.1.1 (Grabherr et al. 2011). This was performed with reads previously filtered using the default parameters of the program Trimmomatic within Trinity with the addition of a 25*A adapter for poly-A tails. Following assembly, contigs from the obtained transcriptome (~192,000) were clustered at 90% identity with CD-HIT-EST using the CD-HIT program. Subsequently, filtered reads were aligned on the assembled reference transcriptome using Bowtie2 (Langmead and Salzberg 2012). Only contigs with one count per million uniquely aligned reads for at least two individuals were considered. Transcripts were identified using a BLASTx search against a NCBI non-redundant (NR) sub-database only composed of Cypriniform fishes (TaxID 7952). Transcripts assigned to the mitochondrial genome were removed from all analyses (E-value < 10⁻²⁰).

Differences in gene expression for all transcripts were analysed with pairwise comparisons using the edgeR package (Robinson *et al.* 2010). With this package, *P*-values were adjusted using the Benjamini and Hochberg (1995) method for false discovery rate control. The significance level was set at $FDR < 5\%$. Ultimately, differentially expressed genes "among groups within site" were further identified using a BLASTx search against the whole NR database only the matches with an Evalue $> 10e-5$.

Proteomic

A quantitative proteomic analysis was achieved to assess if protein expression differences could emerge among the sympatric wild type, Mississippian and Atlantic cybrids. Protein homogenates were prepared as described by Angers *et al.* (2012) and total protein separation and revelation were performed according to Deremiens *et al.* (2015). For each 2D-gel, two individuals belonging to the same haplotype were pooled. Gels were analysed using the MELANIE program (version 7.05, Swiss Institute of Bioinformatics). Only spots with good resolution and contrast were selected. To improve the analysis resolution, spots missing in a gel were manually duplicated at the same (x,y) coordinates. Protein expression among the *C. eos* forms was compared using the percentage volume (%Vol). The %Vol of a spot is a normalized value referring to the ratio between its own volume and the total volume of all the spots in the gel.

Enzymatic activity

According to Deremiens et al. (2015), the cytochrome c oxidase (complex IV) was the only mitochondrial respiratory chain complex displaying a significant difference in enzymatic activity between allopatric wild type and cybrids. To assess whether this difference was maintained among the three sympatric *C. eos* forms, the entire white muscle was sampled from one side of these fish and red muscle, found in very small amount, was removed. Homogenates were prepared as described by Deremiens et al. (2015). Complex IV activity was measured using a UV/VIS spectrophotometer (Shimadzu, UV-1800) equipped with a thermoelectrically temperature-controlled cell holder (CPS-240A). All assays were performed in duplicate at 24°C. Conditions for the cytochrome c oxidase enzymatic assay were as described by Deremiens et al. (2015). Total soluble protein content of each homogenate was determined by the Bradford method (Bradford 1976) using bovine serum albumin (BSA) as standard.

Statistical analyses

Redundancy analyses (RDA) were performed to compare the epigenetic and proteomic profiles among the wild type, Mississippian and Atlantic cybrids from Desjardins Lake. For the first two surveys, Mississippian cybrids from Beaver Lake and wild types from Saad Lake were respectively added to the analyses. The epigenetic and proteomic profiles assessments were carried out using the presence-absence matrix of MSAP bands and a matrix of the %Vol of each protein, respectively. Each of these matrices were combined to a matrix grouping of individuals according to their mitochondrial haplotype or the lake of origin. The significance of the results was tested with an ANOVA like permutation test for Constrained Correspondence Analysis with 999 permutations. Regarding enzymatic analyses, because of the very small sample data sets,

non-parametric Mann-Whitney *U*-tests were applied. All statistical analyses were achieved with the R program (version 3.2.1), using the *vegan* package (Oksanen *et al.* 2007). The significance level was set at $P < 0.05$

3.3 Results

Epigenetic

Out of a total of 367 loci screened for epigenetic changes, 128 (34.9%) displayed variation. Among the variable loci, 32 (25%) methylated bands differentiated individuals from Desjardins Lake and Beaver Lake; more precisely 23 of these bands were found in every individual from Desjardins Lake and 9 of them characterised every fish from Beaver Lake.

The first RDA, which considered the Mississippian cybrids from Desjardins Lake and Beaver Lake, revealed a difference between sites. This "between sites within group" difference was significant ($P = 0.002$). In this case, the first axis of the RDA explained 50.4% of the variation. The second RDA, which considered the three *C. eos* forms from Desjardins Lake, demonstrated a difference among groups. This "among groups within site" difference was significant between the wild type and both cybrids ($P < 0.015$) but not between Mississippian and Atlantic cybrids ($P = 0.566$). Here, 12.7% of the variance of the epigenetic profiles was explained by the first two RDA axes. The third RDA (Figure 3.1), combining the three *C. eos* forms from Desjardins Lake with Mississippian cybrids from Beaver Lake ($P = 0.003$), showed that the difference observed "between sites within group" was much more important than the one measured "among groups within site". The first two axes of this RDA explained 19.1% of the variance.

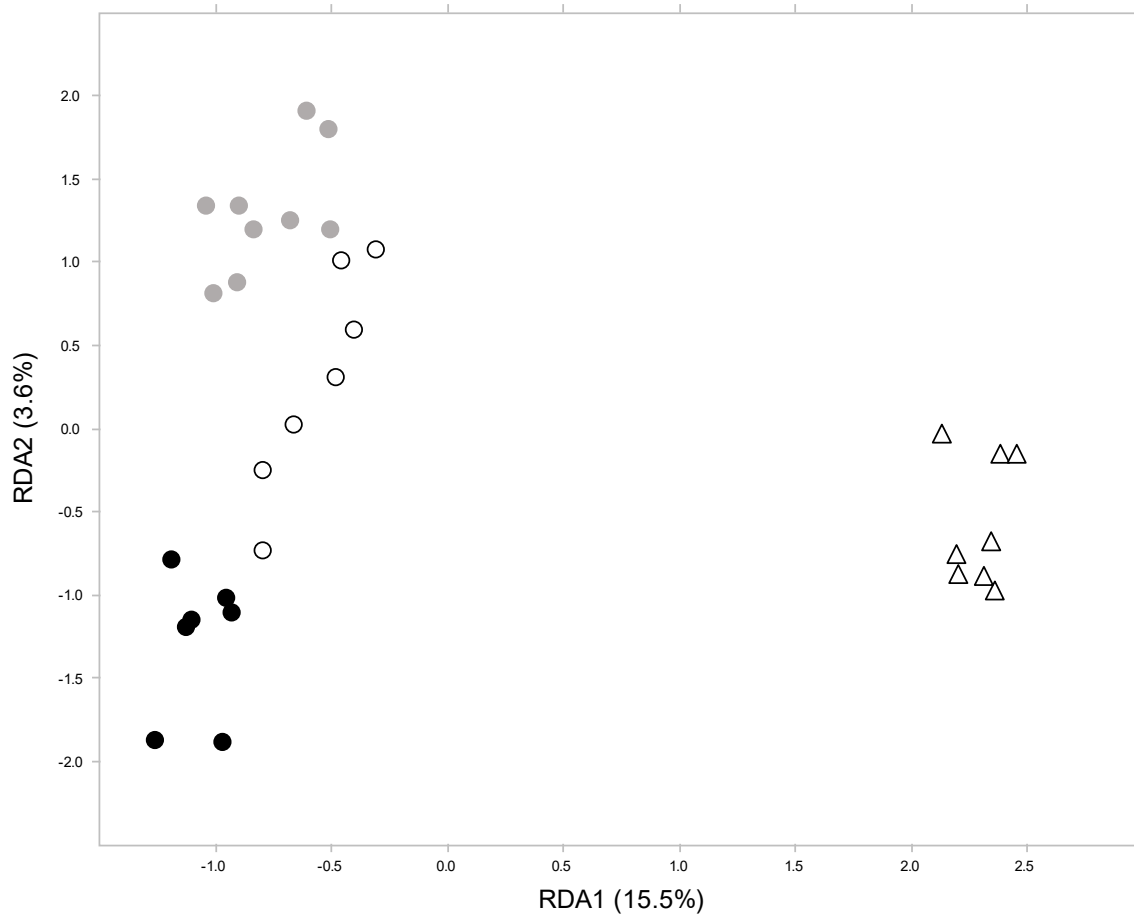


Figure 3.1: RDA of the different epigenetic profiles obtained from the caudal fin of the *C. eos* wild type, Mississippiian and Atlantic cybrids. Circles and triangles refer to individuals from Desjardins Lake and Beaver Lake respectively. Black circles correspond to the wild type, grey circles to Atlantic cybrids and blank circles and triangles to Mississippiian cybrids.

Transcriptomics

Our transcriptomic inquiries initially yielded 216 million paired-end fragments (2×100 bp), averaging 54 million reads of high quality per sample: phred33+ > 30 for 92%, except for one Atlantic cybrid individual at 90.46%. The quality of the clustered assembly (~172k contig) proved to be high; with an N50 of 1324 bases, with over 90% overall realignment for every samples . Of the 30,671 transcripts analysed for differential expression, 21,140 (69%) found an homology.

Transcript expression variations were compared “between sites within group”, but also “among groups within site”. Out of the 30,671 transcripts, 2,392 (7.80%) were considered as differentially expressed between the wild types from Desjardins Lake and Saad Lake (Figure 3.2A). On the other hand, when the wild type and Mississippian cybrids from Desjardins Lake were compared, only 104/30,671 (0.34%) transcripts were characterised as differentially expressed (Figure 3.2B). Based on the whole NR database, 72 of these mRNAs were homologous to 63 proteins. Among those 63 proteins, 31 were characterised (Table A1) while 32 had unknown functions; the latter were not considered in this study. Among the characterised ones, three were nucDNA-encoded proteins associated to mitochondria. When “between sites within group” and “among groups within site” results were compared, 53/104 transcripts were common to both analyses. These mRNAs corresponded to 34 of the 63 homologies previously detected, among which 17 were characterised. Accordingly, the 51/104 differentially expressed transcripts left could be attributed to foreign mitochondria effects alone. These mRNAs

corresponded to 29 of the 63 detected homologies, among which 14 were characterised (Table A1).

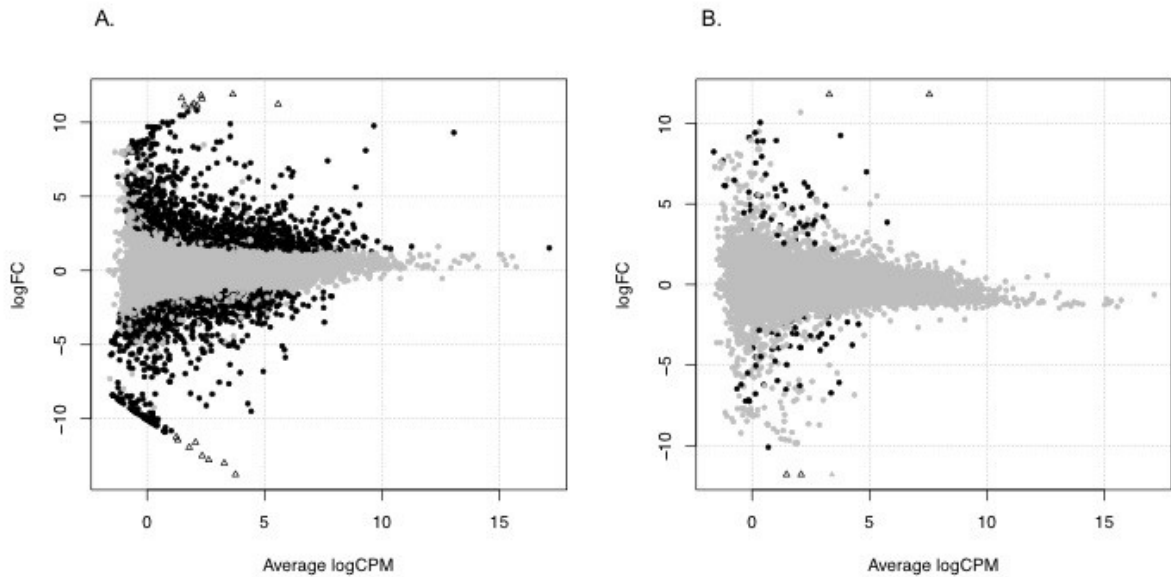


Figure 3.2: Smear plot visualization of differential expression analyses of *C. eos* wild type and Mississippian cybrid transcriptomes. (A) Transcriptome comparisons between wild types from different lakes (Desjardins and Saad). (B) Transcriptome comparisons between wild type and Mississippian cybrids from the same lake (Desjardins). Black dots represent the mRNAs considered as differentially expressed ($P_{\text{adj}} < 0.05$). Triangles refer to transcripts that could not be represented in the logFC interval. CPM = counts per million; FC = fold change.

Proteomics

To infer mitochondria influence on the protein expression of the three *C. eos* forms, the %Vol of a subsample of 132 unambiguous protein spots was analysed. Although the RDA revealed a difference among the wild type, Mississippian and Atlantic cybrids from Desjardins Lake (variance explained by the first two axes = 66.2%), this difference was not statistically significant ($P = 0.067$)

Despite this result, inter-form comparisons indicated protein expression variations (Figure 3.4). Considering an expression ratio difference greater than or equal to 2×, the proportion of differentially expressed proteins within a given *C. eos* form was low; varying from 3.8 to 12%. When inter-form comparisons were tested, a higher percentage of differentially expressed proteins was detected: (i) wild type *versus* Atlantic cybrids, 15.1 to 23.5%; (ii) wild type *versus* Mississippian cybrids, 27.3 to 31.8%; (iii) Atlantic cybrids *versus* Mississippian cybrids, 13.6 to 20.4%.

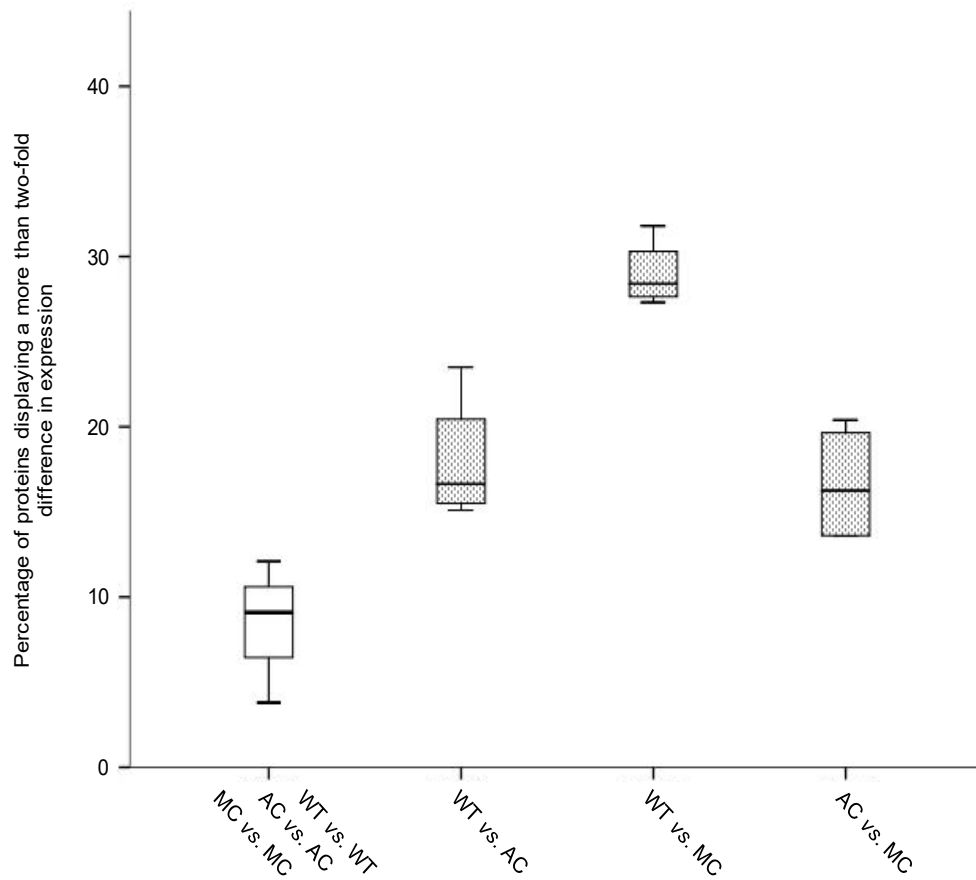


Figure 3.3: Protein expression comparisons in *C. eos* wild type and cybrids. Among the 132 studied proteins, only the ones presenting a more than two-fold difference in expression were considered. The comparisons were performed with two protein gels for each *C. eos* form. The blank boxplot represents the intra-form comparisons while the dotted boxplots represent the inter-form comparisons. Intra-form comparisons were pooled. WT = wild type; AC = Atlantic cybrid; MC = Mississippian cybrid.

Enzymatic activity

A significant difference of complex IV activity was observed in both Mississippian and Atlantic cybrids when compared to their sympatric wild type counterparts (Figure 3.4; Mann-Whitney, $P < 0.02$). These cybrids exhibited a nearly, twice higher activity than the wild type. No significant difference was detected between the two cybrid forms (Mann-Whitney, $P = 0.861$).

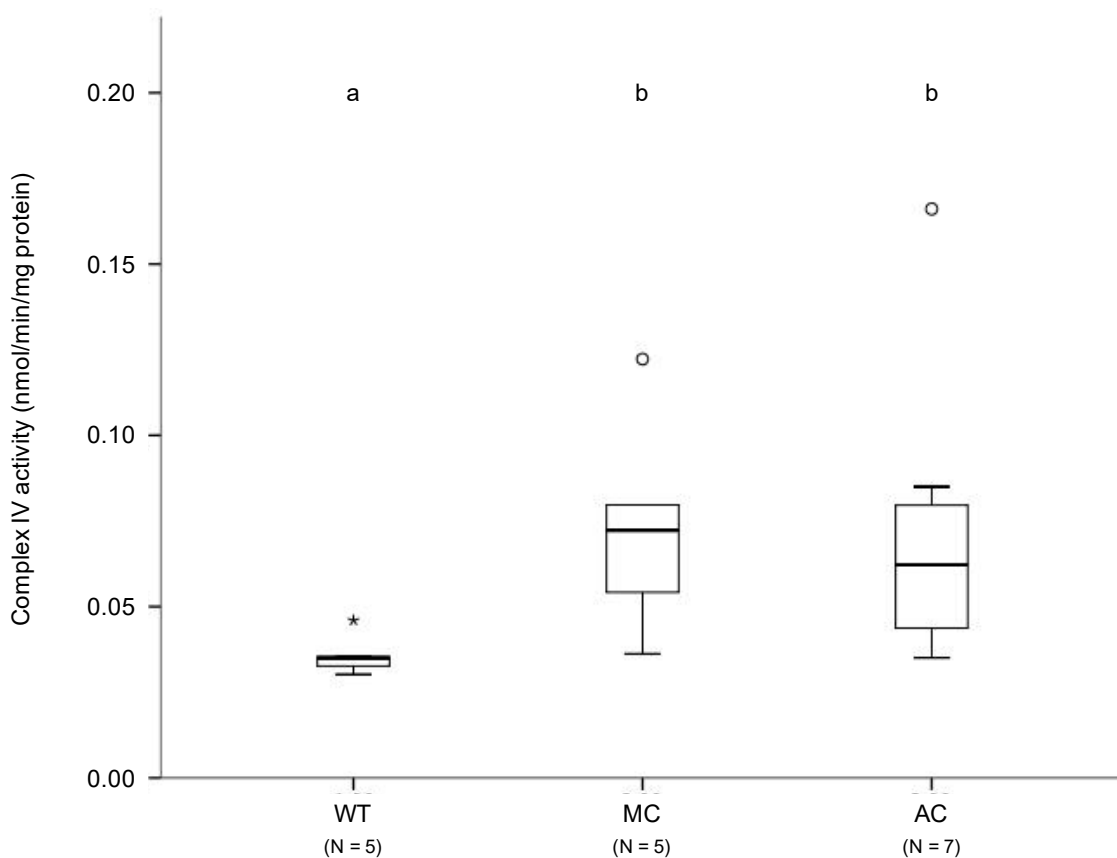


Figure 3.4: Complex IV activity in the white muscle of *C. eos* wild type (WT), Mississippian (MC) and Atlantic (AC) cybrids. Significance was set as $P < 0.05$. Letters denote differences between *C. eos* forms: a is significantly different from b.

3.4 Discussion

As previously described, Mississippian and Atlantic cybrids are characterised by a 1.20% divergence of their mtDNA (Deremiens, Schwartz et al. 2015) while a difference of 8.38% and 8.44% is detected between the wild type and Mississippian cybrids and between the wild type and Atlantic cybrids, respectively. While a divergence of 1.20% would not trigger a wide phenotypic modulation, the 8.38-8.44% difference observed between the wild type and the two cybrids could potentially do so. To test this hypothesis, a holistic approach, going from nuclear genome expression to proteome and enzymatic comparisons, was considered.

In the present study, the comparison of the methylation profiles among the wild type, Mississippian and Atlantic cybrids revealed that *C. neogaeus* mitochondria are able to trigger a reprogramming of the *C. eos* nuclear genome expression. However, this comparison showed that the environment has a stronger effect on DNA methylation than allospecific mitochondria. Presently, the link between mitochondria and epigenetic changes is not yet fully understood. One of the possible processes could imply the production of reactive oxygen species (ROS (Afanas'ev 2015, Wu and Ni 2015)), a by-product of the mitochondrial respiration (Murphy 2009, Bleier and Drose 2013). More precisely, mitochondrial respiration involves the transfer of electrons through a succession of redox centers, from an electron donor (NADH) to a terminal acceptor (O₂). Following the integration of foreign mitochondria, a reduction of this electron flux is generally observed due to mismatches between nucDNA- and mtDNA-encoded subunits of the respiratory chain complexes. Ultimately, this phenomenon leads to a more highly reduced state of these enzymes, which finally induces a higher production of ROS (Moreno-Loshuertos *et al.* 2006, Barja 2007, Gusdon *et al.* 2007). These molecules would impact

epigenetic profiles through both DNA methylation modifications and changes of histone acetylation/deacetylation state (Afanas'ev, 2015; Wu and Ni, 2015). Contrarily to the expected effects of foreign mitochondria integration on respiratory chain complex activity, this study revealed a significant increase of complex IV activity in cybrids. This higher activity would allow complexes I, II and III to be in a more highly oxidized state (Blair and Lemieux 2001), thereby inducing a lower production of ROS compared to *C. eos* wild type, which would eventually modify cybrid epigenetic profiles.

The limited transcriptomics expressions variations between sympatric wild type and Mississippian cybrids might be related to the detected epigenetic variations and/or could derive from post-transcriptional modifications. As revealed by the epigenetic analysis, environmental effects were more pronounced than those of allospecific mitochondria. Although most of the 104 differentially expressed mRNAs detected between the wild type and Mississippian cybrids were mainly unidentified or associated to uncharacterised functions, three of them were homologous to nucDNA-encoded mitochondrial proteins. Among those three, the “1,25-dihydroxyvitamin D(3) 24-hydroxylase” was found in both “between sites within group” and “between groups within site” comparisons. On the other hand, the “glutaminase kidney isoform” and the “pyruvate dehydrogenase kinase isozyme 3” were only found in the “between groups within site” comparison. These two transcript expression modifications can therefore be attributed to foreign mitochondria effects alone. While the "glutaminase kidney isoform" is an enzyme associated to the maintenance of acid-base homeostasis, the "pyruvate dehydrogenase kinase isozyme 3" regulates the pyruvate dehydrogenase complex activity, an enzyme directly involved in glucose metabolism and aerobic respiration. According to their function, discrete

metabolic differences could thus happen between the wild type and Mississippian cybrids. These mRNAs, therefore, reveal that phenotypic modifications might occur at a biological organization level not yet investigated, the metabolome.

Proteomic profiles of the sympatric *C. eos* forms were not significantly differentiated but some protein expression variations were observed among them. These variations might result from the detected transcriptomic changes but could also come from post-translational modifications. Although apparently high, the magnitude of differentially expressed proteins was actually less pronounced when intra-form comparisons were considered. One particularly interesting feature of this analysis come from a similar number of differentially expressed proteins between Mississippian and Atlantic cybrids (i.e. 13.6-20.4%) and between the wild type and Atlantic cybrids (i.e. 15.1-23.5%). This similarity suggests that the 1.20% mtDNA divergence measured between the two cybrids can impact the proteomic phenotype as much as the 8.44% mtDNA difference detected between the wild type and Atlantic cybrids. In the present study, 15.1 to 31.8% of the assessed proteins showed a more than two-fold difference in expression between sympatric wild type and cybrids. On the other hand, according to Deremiens et al. (2015), allopatric wild type and cybrids were characterised by a proteomic difference ranging from 29.0 to 37.0%. These higher values would mainly be caused by environmental and possibly nuclear effects resulting from the use of allopatric biotypes. Ultimately, it is important to mention that part of the inter-form variations could come from proteins encoded by the mtDNA itself. Therefore, the magnitude of allospecific mitochondria influence on protein expression could be less pronounced than what is actually measured.

As mentioned earlier, our comparative analysis demonstrated that *C. neogaeus* mitochondria are able to increase the complex IV activity of *C. eos*. According to Deremiens et al. (2015), similar results were obtained with allopatric *C. eos* wild type and cybrids, and could result from one amino acid change on the *COX3* mitochondrial subunit. This enzyme being of importance in the respiratory chain regulation (Blier and Lemieux 2001, Arnold 2012), its activity change might impact the aerobic metabolism. Moreover, since the sympatric *C. eos* forms do not present notable nuclear differentiation (Deremiens et al. in prep) and because no transcript involved in the mitochondrial respiration *per se* was differentially expressed among the wild type, Mississippian and Atlantic cybrids, the modification of complex IV activity can be assumed to only come from the allospecific mitochondria. Altogether, these observations suggest that mitochondria can occasionally confer specific functional differences, which might ultimately impact mitochondrial metabolism.

According to Mee and Taylor (2012), the wild type form is mainly found in the southern part of the *C. eos* distribution area whereas cybrids, as well as *C. neogaeus*, predominantly occur in the north. This distribution suggests that cybrids can support colder conditions than wild type. In the Laurentian region, where their distributions overlaps, wild type and cybrids are usually found in allopatry and sometimes in sympatry (Mee and Taylor, 2012). To explain this particularity, one assumption is that under certain environmental conditions, the phenotypic differences observed between wild type and cybrids are such that they promote the establishment of one of the two biotypes. However, under less common conditions, these phenotypic variations would be less pronounced and too low to allow the prevalence of any biotypes. On a larger geographical scale, in Laurentian lakes, the influence of *C. neogaeus*

mitochondria on *C. eos* nucDNA expression would not be enough to prevent wild type and cybrids to live together or next to each other. On the other hand, further north, in cold environments, one hypothesis is that the nuclear reprogramming induced by these foreign mitochondria is more pronounced; inducing the expression of genes usually repressed in wild type, eventually leading to cybrids better adapted to cold environments and therefore explaining the peculiar distribution of this biotype. If the higher complex IV activity in cybrids is extrapolated to the metabolic activity, we can also assume that these fish would be metabolically more active than wild types at low temperature; advantage promoting their dominance in cold environments. Both hypotheses are mutually non-exclusive.

To conclude, the holistic approach applied in this study enabled to demonstrate that allospecific mitochondria are able to impact *C. eos* phenotype, at various biological levels, when wild type and cybrids are found in sympatry. As expected, the phenotypic differences detected between these biotypes were more important than the ones measured between Mississippian and Atlantic cybrids. Additionally, this survey revealed that the magnitude of foreign mitochondria influence was more pronounced between allopatric wild type and cybrids than between sympatric wild type and cybrids. This discrepancy might be the result of environmental and nuclear effects resulting from the use of allopatric fishes. Ultimately, all the results obtained through this holistic research significantly contribute to the disentangling of the links existing between mitochondria and the different organisation levels of the phenotype. Furthermore, this approach participates of the age-old question of how genotypic variations affect the phenotype of an organism.

3.5 Acknowledgments

The authors are grateful to L. Baillon, K. Duclos, C. Leung and E. L'Heureux for their precious assistance. LD was partly supported by a PhD scholarship from the GRIL. This research was supported by a generous contribution from S. Lumbroso and by a team grant from the *Fonds Québécois pour la Recherche sur la Nature et les Technologies* (FQRNT) to BA and AA.

Supplementary Material – Transcriptomics

The goal of this study is to identify the functionalities affected by the transcriptomic changes between wild type *Chrosomus eos* and those with the mitochondrial genome of *Chrosomus neogaeus* from the Mississippian glacial refugia.

To address this goal, further transcriptomic analyses were performed to complement those presented in Deremiens *et al.* (in prep). The completeness of the assembly, along with the GO term overrepresentation within the previously identified differentially expressed genes, was assessed. Further details on the transcriptomic section of the article are also provided.

Methods

Within the study of Deremiens *et al.* (in prep) the mRNA of two individuals of each category, i.e. *C. eos* WT, *C. eos* Cybrid and *C. eos* from the lake Saad, were sequenced. The sequencing was performed at the Institute for Research in Immunology and Cancer (IRIC) on a HiSeq 2000 platform. The demultiplexing with adapter masking was also performed at the IRIC. Once recuperated, the quality of the reads was analysed. Before Assembly, the reads were processed further with Trimmomatic using the program Trinity version 2.1.1 (Grabherr *et al.* 2011) at the following settings: trimming of TruSeq3-PE adapter (with the addition of a 25*A and T sequence to filter Poly-A tail), A sliding window of 4 base-pair width, cutting when the quality average per base drops beneath 5, removing the base-pairs of quality below 5 at each extremity, also a minimum of 25 base-pair per reads. The assembly was performed with Trinity with all combined samples at default settings except for the addition of Trimmomatic, resulting in 192 K contigs. The contigs obtained were clustered with the use of cd-hit-est V4.6.4 (Fu *et al.* 2012; Li and Godzik 2006) at 90% identity with a word of 8. The refined Assembly of 172 K contigs was annotated (see below).

The refined assembly was annotated with Trinotate (version: 2.0.2; <https://trinotate.github.io/>). At this point, the assembly was annotated with the Swissprot database (Trinotate provided version) and NR database (<ftp://zhanglab.ccmb.med.umich.edu/download/nr.tar.gz>). Transdecoder version: 2.0.1 was used to provide likely ORF based on a minimum length of 100 amino acids, however only the ORFs with a P-fam domain hit or a Swissprot homology (min 1e-5) were retained. The processed reads were mapped with Bowtie2 (Langmead and Salzberg

2012) on the post CD-hit-est transcriptome. However, only the uniquely mapped reads were kept.

The differential expression analysis was performed with the edger package (Robinson *et al.* 2010) in R V3.2.3, only the genes with a minimum count per million of one for at least 2 individuals were considered for this analysis and further analyses. The FDR significance threshold was set at 0.05 with the Benjamin Hodge method (1995). The categories compared were Cybrids vs C. eos WT and C. eos WT vs C. eos from lake Saad.

In this study, the result of the differential expression analysis of Deremiens *et al* (in prep) was further explored using the GSeq (Young *et al.* 2010) pipeline present in Trinity, a functional analysis at the gene level (as identified by Trinity) was performed for all the previously identified differentially expressed genes analysis. The gene length was obtained from the longest isoform of each gene. The FDR significance threshold was 0.05, corrected with the Benjamin–Hochberg (1995) method. In this analysis, the functions were obtained from the previously identified Swissprot homology (min 1e-5). To assess the completeness of the presented assembly, a BUSCO analysis (Simão *et al.* 2015) was performed with the refined assembly. This analysis identifies Universal Single-Copy Orthologs present within the assembly from a database. The vertebrate database was used for this analysis.

Results

Overall the quality of the cd-hit-est processed assembly proved to be high as analysed with BUSCO. Indeed, it shows that 76 % of the 3023 singles copied genes within the database were found complete, 12% of which were duplicated, another 5% were found fragmented and 18% were missing from the assembly.

Table SI. Analysis assessing the assembly and annotation completeness with Benchmarking Universal Single-Copy Orthologs (BUSCO) with the BUSCO vertebrate database.

Assembly	Complete	Duplicate	Fragmented	Missing
Clustered Assembly	76%	12%	5,0%	18%

The overrepresented GO terms found within the differentially expressed genes common to the “between mitotype within site” and “between site within mitotype” analysis were DNA related: “DNA integration” and “DNA metabolic process”. However, only one overrepresented function is specific to the analysis “between mitotype within site”. This function is “MF oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen”.

Table SII. Analysis of Gene Ontology term overrepresented between Mississippian hybrid and wild type. Analysed at the gene level (as defined by Trinity) by the GOseq R pipeline through Trinity.

Category	over represented pvalue	numDEInCat	over represented FDR	go term
GO:0015074	4.10431264349177e-07	7	0.00118573592270477	BP DNA integration
GO:0006259	1.76670800038812e-05	7	0.0229145014835303	BP DNA metabolic process
GO:0016705	2.37949132746939e-05	4	0.0229145014835303	MF oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen

The overrepresented GO terms for “between site within mitotype” are more numerous and varied in origin. From “carbohydrate binding” to “insulin receptor binding”, seven GO terms are only overrepresented within the “between site within mitotype” analysis. Some of these terms are logically linked such as “immune response” and “immune system process”, “DNA integration”, “metabolic process” and “sequence-specific DNA binding”. Four terms indicate a binding function: “insulin receptor binding”, “sequence-specific DNA binding”, “protein complex binding”, “carbohydrate binding”,

Table SIII. Analysis of Gene Ontology term overrepresented between Desjardins and Saad wild type. Analysed at the gene level (as defined by Trinity) by the GOseq R pipeline through Trinity. Only FDR <0,05 are represented.

Category	over_represented_pvalue	numDEInCat	over_represented_FDR	go_term
GO:0015074	1.93713124389463e-09	30	5.59637216361158e-06	BP DNA integration
GO:0006259	6.76549841306433e-06	39	0.00977276245767143	BP DNA metabolic process
GO:0005158	1.79298985528251e-05	7	0.0172664923063706	MF insulin receptor binding
GO:0002376	3.1927689002579e-05	15	0.0230597733821127	BP immune system process
GO:0008716	5.05966473475023e-05	5	0.0292347428373868	MF D-alanine-D-alanine ligase activity
GO:0043565	9.1886210896032e-05	29	0.0333212571081637	MF sequence-specific DNA binding
GO:0006955	9.22707015802386e-05	14	0.0333212571081637	BP immune response
GO:0032403	0.000114485157529539	18	0.0367497355669819	MF protein complex binding
GO:0030246	0.000152489485760051	11	0.0440542124360788	MF carbohydrate binding

Chapitre 4 - Conclusion

Les cybrides : le modèle parfait

L'objectif de cette maîtrise fut de discriminer l'influence spécifique et globale de la mitochondrie exogène et l'effet environnemental potentiel des études précédentes. À cet effet, une collaboration a été réalisée entre différents étudiants pour regrouper plusieurs projets connexes, étendant le spectre d'analyse original de l'étude. Cette collaboration fut fructueuse puisque celle-ci fut capable de répondre aux deux sous-objectifs de ce projet. Premièrement, il fut établi que l'effet environnemental est supérieur à l'effet du mitotype. Cette observation semble montrer que les études précédentes réalisées en allopatrie (Angers, Dallaire et al. 2012, Deremiens, Schwartz et al. 2015) présentaient potentiellement des effets environnementaux confondants. Deuxièmement, avec certaines approches une différence entre mitotypes fut observable.

Le complexe *Chrosomus eos-neogaeus* présente un modèle peu commun dans l'analyse de l'effet de mitochondrie exogène. La présence de ce modèle en milieu naturel aide à son étude, et en fait donc un modèle avec peu d'analogues. Cette occurrence naturelle de cybrides n'est pas unique, mais certains aspects de la génération de ces individus sont avantageux. Contrairement au cybride naturel obtenu par rétrocroisement tel que la paruline à croupion jaune, ce modèle, par le mode de création, empêche la transmission de gènes nucléaires de l'espèce parentale transmettant la mitochondrie (Toews, Mandic et al. 2014). L'absence de transmission des gènes nucléaires avec la mitochondrie exogène présente un grand avantage sur les autres modèles, éliminant ainsi l'influence potentielle de ce gène intrus sur le phénotype de

l'individu. Cette méthode de reproduction du complexe est donc un attrait unique du complexe pour notre objectif.

Influence des mitochondries exogènes

Retour sur les objectifs

Un parallèle pourrait-être établi entre l'étude de Deremiens, Schwartz et al. (2015), puisque celle-ci a étudié le même phénomène dans certains lacs communs à l'échantillonnage de la présente étude mais en allopatrie. En effet, l'étude de l'effet environnemental ajoute à la littérature au sujet de cette influence qui ne fut pas directement mesurée.

L'effet environnemental mesuré à la fois par l'analyse MSAP sur la méthylation de l'ADN et l'analyse de la transcriptomique, a permis de percevoir que l'effet environnemental est supérieur à l'effet du mitotype. En effet, dans les deux cas, le pourcentage de variation attribué à l'effet observé lorsqu'on compare des lacs différents est significativement supérieur à celui mesuré pour les variations liées au mitotype. Également au niveau transcriptomique, un nombre de gènes d'un ordre supérieur fut observé dans l'analyse d'expression différentielle entre les individus de même mitotype, mais de lac différent, comparée au même type d'analyse comparant les individus d'un même lac et d'un mitotype différent.

Au cours de ce projet, nous avons dû créer une base de données des contigs inférée de nos données transcriptomiques. Cette banque de données ajoute beaucoup d'information sur le complexe, qui avant avait peu d'information génomique et transcriptomique de cette nature. Ceci permettra ainsi de futures études des d'ARNm de *Chrosomus eos* ainsi que certains des autres membres du complexe. Ainsi que de poursuivre une analyse d'expression différentielle sur des données de RNA-seq.

Grâce aux avancées récentes en séquençage de nouvelle génération, plusieurs espèces représentant une bonne diversité de ce taxon ont été séquencées, tant en transcriptomique qu'en génomique. Ces derniers emploient souvent le séquençage de nouvelle génération, tout comme notre assemblage, ce qui offre beaucoup d'information à prix modique. Par contre cette méthode offre seulement de courtes séquences consécutives d'ADN et donc est l'assemblage finale moins certain.

Les bases de données auxquelles notre assemblage fût comparé offre une variété de certitude dans les taxons représentés. De ce fait, l'abondance d'un taxon parmi les homologues n'est pas expliquée uniquement par la proximité génétique des individus, mais également avec la représentation des taxons à l'intérieur de celle-ci. Dans le cas d'une base de données telle que Nr (<ftp://zhanglab.ccmb.med.umich.edu/download/nr.tar.gz>), les taxons ont d'amples différences de représentation et de certitudes. À l'opposé, la base de données Swissprot (<http://www.uniprot.org/uniprot/>), beaucoup plus succincte, se limite aux protéines ayant une fonction prouvée par des expériences biologiques. L'identification des divers contigs aide à établir la fiabilité de l'assemblage. Cependant, la qualité d'un assemblage ne se limite pas seulement à l'identification, puisque celle-ci se limite à l'existence d'information génomique ou protéique d'homologue, informations parfois indisponibles dans les bases de données. Une autre approche, celle du N50, la longueur du contig contenant le milieu nucléotidique d'un assemblage lorsqu'en ordre croissant de longueur, est une statistique analogue à la médiane. Cette mesure illustre la longueur représentative d'un assemblage.

Cette méthode d'évaluation est perçue comme incomplète : aucune évaluation de la redondance des contigs ne lui est intégrée, ainsi qu'une évaluation de l'intégralité des contigs et de l'intégralité de la présence des séquences d'ARNm dans l'assemblage n'est offerte. BUSCO est

une méthode d'évaluation visant à combler ces gaps en évaluant la base de données par rapport à un ensemble de gènes universelles en copie unique (Simao *et al.* 2015). Nous avons évalué notre transcriptome contre la base de données BUSCO des vertébrés 76,1% des gènes orthologues des vertébrés étaient présents et complets (Tableau SI), cependant 12% en présents, complets et répétés ce qui est potentiellement dû à des haplotypes différents ou des isoformes. Ce résultat est comparable à la littérature (Hara *et al.* 2015).

Un avantage d'identifier chaque gène individuellement et d'annoter ceux-ci est l'analyse fonctionnelle des résultats. Cette perspective fonctionnelle de l'analyse de nos données ne fut pas intégrée à l'article actuel. Celle-ci permet de tester l'abondance des fonctions des contigs perçues comme exprimée de façon différente, soit la probabilité que la fonction analysée soit surreprésentée parmi les contigs d'intérêt (Tableau SII et SIII). Cette analyse fût effectuée avec *GSeq* sur *R* (Young *et al.* 2010) application fourni avec *Trinity*.

Plusieurs fonctions attribuées aux gènes différemment exprimés sont surreprésentées (Tableau SIII). Plusieurs des éléments de cette liste sont liés au métabolisme énergétique (ie *Insuline receptor binding* et plausiblement *carbohydrate binding*) ou liés au métabolisme de l'ADN. De plus, les termes liés au système immunitaire peuvent être facilement expliqués par la possibilité que les environnements différents causent des réactions immunitaires différentes.

Puisque le premier objectif évalue l'influence environnementale potentiellement présente dans les études précédentes en allopatrie et que celle-ci fût supérieure à l'effet de la mitochondrie, une analyse des mitotypes en sympatrie représente un élément nouveau dans la littérature sur ce modèle. Le second objectif évalue donc de nouveau l'effet de la mitochondrie. Les résultats ont varié en fonction de la méthode. Tous ont permis de distinguer les cybrides du type sauvage, mais seule l'analyse protéique a pu trouver une différence entre tous les mitotypes.

L'analyse du transcriptome du point de vue fonctionnel (Tableau SII) indique que seulement 3 termes ontologiques sont perçus comme surabondant, et un seul d'entre eux peut présenter un lien avec l'activité mitochondriale (*oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen*). Ce terme est défini comme un catalyseur d'une réaction d'oxydoréduction, dans laquelle un ou plusieurs électrons sont transférés de deux donneurs et qu'un oxygène est réduit ou incorporé. Cette dernière est très pertinente pour son lien avec la respiration cellulaire.

Distribution géographique

Puisque les changements phénotypiques interlacs sont de plus grande amplitude que les changements intra-lac, et que selon la littérature les individus ont des phénotypes plus similaires lorsqu'analysés en allopatrie, ceci suggère que l'influence environnementale représente un effet confondant avec l'effet du mitotype. Ce résultat suggère l'hypothèse que le mitotype ou ses effets sur le phénotype en général cause une répartition géographique différente avec peu d'endroit commun (tel le lac Desjardins). En effet, selon l'hypothèse où les individus d'un même mitotype sont plus apte à coloniser certains environnements, des analyses en allopatrie seraient influencé par un facteur environnemental confondant, et montreraient la similarité inter-lacs et intra-mitotype telle qu'observée dans la littérature sur ce modèle (Angers, Dallaire et al. 2012, Deremiens, Schwartz et al. 2015), malgré que l'effet environnemental soit plus grand que l'effet des mitotypes. Dans Deremiens, Schwartz et al. (2015), une différence dans l'endurance des individus évalués dans un test de nage a été montrée, et tout comme dans la présente étude, des changements caractérisant les différents mitotypes ont été observés. Ces différences pourraient

se refléter dans leur capacité à coloniser divers environnements, engendrant ainsi la similarité inter-mitotype mentionnée précédemment.

Une différence dans la capacité de coloniser divers environnements engendrerait une différence sur la distribution géographique de cette espèce en fonction des conditions environnementales. Cette hypothèse est appuyée par la distribution des membres cybrides/sauvage de ce complexe (voir figure 4.1) (Mee et Taylor 2012). Celle-ci suggère que les individus cybrides correspondraient plus à un environnement nordiques, que les individus de type sauvage, avec

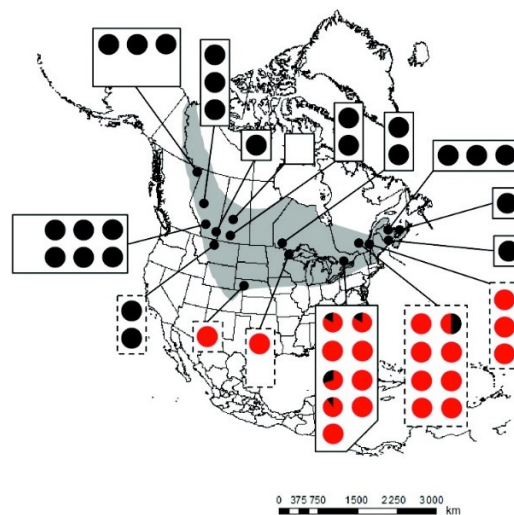


Figure 4.1 : Distribution des cybrides (en noir) et des types sauvages (en rouge) du complexe *Chromosus eos-neogaeus*. Tiré de Mee et Taylor 2012.

peu de territoires intermédiaires où les deux mitotypes peuvent cohabiter. Un changement de répartition en latitude est souvent associée avec des changements biotiques et abiotiques (ie. climat). L'activité mitochondriale du complexe IV, montrait une différence significative à 24°C en sympatrie tout comme en allopatrie. Néanmoins dans la littérature, ce complexe ne présente

aucune différence significative en allopatrie à des températures plus faibles suggérant donc une certaine influence de la température sur l'efficacité mitochondriale de ce complexe.

Selon l'hypothèse précédente, les changements observés entre les individus du lac Desjardins peuvent être dus à un changement d'habitat à l'intérieur du lac. Un exemple de changement de microhabitats pourrait venir de la température, puisqu'à l'intérieur d'un lac la température varie en fonction de la profondeur. Cependant, une analyse fonctionnelle ne suggère pas de facteurs clairs dictant une niche particulière à l'intérieur du lac tel que la température..

Hypothèse des changements post-traductionnels

L'analyse protéique indique un nombre élevé de protéines ayant un changement d'expression entre mitotypes sur le sous-ensemble isoélectrique de protéines analysée. 15% à 30% d'un échantillon de 137 protéines montrent un changement d'expression ce qui est supérieur au résultat obtenu en transcriptomique. Ces résultats différents entre transcriptomique et protéiques pourraient être expliqués par les changements post-traductionnels. En effet, de nombreux changements post-traductionnel sont perceptibles dans cette méthode et ces changements ne se reflèteraient pas directement sur les autres niveaux d'organisation phénotypique analysés. Par exemple, la phosphorylation est un ajout post-traductionnel par une kinase gouvernant l'activation de beaucoup de protéines chez les eucaryotes (Scheeff et Bourne 2005). Parmi les gènes identifiés comme différemment exprimés dans notre étude, un contig ayant une fonction kinase (*Predicted: Pyruvate dehydrogenase (acetyl-transferring) kinase isozyme 3**) pourrait causer un changement au niveau post-traductionnel. Ceci est l'exemple le plus évident, mais de nombreux exemples sont possibles illustrant différentes causalités liant un changement d'expression à des

changements post-traductionnels. Ces résultats peuvent être également expliqués par les différentes étapes méthodologiques.

Liens entre les approches

Lors de la discussion de l'hypothèse des changements post-traductionnel, nous avons exploré un exemple du lien entre les différents niveaux d'intégration analysés et les différentes approches. Chaque approche réalisée par plusieurs parties indépendantes a des avantages et des inconvénients. Bien qu'au niveau de l'organisation phénotypique, il est risqué de faire des liens entre les différents niveaux d'intégration, puisque ces diverses analyses n'ont pas le même niveau de sensibilité ni les mêmes contrôles. Par exemple, les analyses protéiques sont sur 132 points isoélectriques distincts et non identifiés ; or plus de protéines composent le protéome de ces individus. À l'opposé, la méthodologie de la transcriptomique permet l'analyse de l'expression de l'ensemble du génome. Aucune méthode n'est parfaite ; le séquençage de l'ARNm, bien que très précis, est plus susceptible au changement génomique que l'approche MSAP. L'approche holistique et intégrative de ce projet liant plusieurs étudiants a permis d'identifier le signal du changement phénotypique dû à la mitochondrie exogène à de multiples niveaux d'intégration. Ce type d'approche permet une meilleure compréhension d'un phénomène naturel, mais également une meilleure compréhension de l'importance de l'intégration de différents niveaux d'organisation du phénotype. Chaque niveau d'intégration du phénotype analysé nous donne des informations différentes sur le phénotype et lorsque regrouper confirme appuis l'existence d'un signal de changements phénotypiques. Des projets explorant ainsi plusieurs approches face à une question commune nous mènent à plus de

questions et explorent plus d'aspects d'un même phénomène menant ainsi à une réponse plus complète.

Conclusion

En conclusion, les objectifs de ce projet furent atteints en comparant de multiples niveaux d'intégration influençant l'expression génique. Le changement d'expression génique identifié dans cette étude représente un changement phénotypique pouvant entraîner un changement dans la répartition géographique des individus. L'influence de la température suggérée par la répartition nord-sud et le rôle du signal identifié dans l'article ci présent, bien que suggéré dans la littérature demeurée peu étudié.

Bibliographie

Afanas'ev, I. (2015). "Mechanisms of superoxide signaling in epigenetic processes: relation to aging and cancer." Aging Dis **6**(3): 216-227.

Altschul, S. (1997). "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs." Nucleic Acids Research **25**(17): 3389-3402.

Andersson, S. G., et al. (1998). "The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria." Nature **396**(6707): 133-140.

Angers, B., et al. (2012). "The influence of mitochondria in epigenetics revealed through naturally occurring fish cybrids." Current Zoology **58**(1): 138.

Angers, B. and I. J. Schlosser (2007). "The origin of *Phoxinus eos-neogaeus* unisexual hybrids." Mol Ecol **16**(21): 4562-4571.

Angers, B. and I. J. Schlosser (2007). "The origin of *Phoxinus eos-neogaeus* unisexual hybrids." Molecular Ecology **16**: 4562-4571.

Arnold, S. (2012). "The power of life--cytochrome c oxidase takes center stage in metabolic control, cell signalling and survival." Mitochondrion **12**(1): 46-56.

Arnqvist, G., et al. (2010). "Genetic architecture of metabolic rate: environment specific epistasis between mitochondrial and nuclear genes in an insect." Evolution **64**(12): 3354-3363.

Ballard, J. W. and R. G. Melvin (2010). "Linking the mitochondrial genotype to the organismal phenotype." Mol Ecol **19**(8): 1523-1539.

Ballard, J. W. O. and M. C. Whitlock (2004). "The incomplete natural history of mitochondria." Molecular Ecology **13**(4): 729-744.

Barja, G. (2007). "Mitochondrial oxygen consumption and reactive oxygen species production are independently modulated: implications for aging studies." Rejuvenation Res **10**(2): 215-224.

Bassam, B. J., et al. (1991). "Fast and sensitive silver staining of DNA in polyacrylamide gels." Analytical biochemistry **196**(1): 80-83.

Bayona-Bafaluy, M. P., et al. (2005). "Fast adaptive coevolution of nuclear and mitochondrial subunits of ATP synthetase in orangutan." Mol Biol Evol **22**(3): 716-724.

Benjamini, Y. and Y. Hochberg (1995). "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing." Journal of the Royal Statistical Society. Series B (Methodological) **57**(1): 289-300.

Bermingham, E., et al. (1997). "Fish biogeography and molecular clocks: perspectives from the Panamanian Isthmus." Molecular systematics of fishes(7): 21.

Binet, M. C. and B. Angers (2005). "Genetic identification of members of the Phoxinus eos-neogaeus hybrid complex." Journal of Fish Biology **67**(4): 1169-1177.

Bleier, L. and S. Drose (2013). "Superoxide generation by complex III: from mechanistic rationales to functional consequences." Biochim Biophys Acta **1827**(11-12): 1320-1331.

Blier, P. U., et al. (2006). "Functional conservatism in mitochondrial evolution: insight from hybridization of arctic and brook charrs." J Exp Zool B Mol Dev Evol **306**(5): 425-432.

Blier, P. U., et al. (2001). "Natural selection and the evolution of mtDNA-encoded peptides: evidence for intergenomic co-adaptation." Trends Genet **17**(7): 400-406.

Blier, P. U. and H. Lemieux (2001). "The impact of the thermal sensitivity of cytochrome c oxidase on the respiration rate of Arctic charr red muscle mitochondria." J Comp Physiol B **171**(3): 247-253.

Bolender, N., et al. (2008). "Multiple pathways for sorting mitochondrial precursor proteins." EMBO Rep **9**(1): 42-49.

Boratynski, Z., et al. (2011). "Introgression of mitochondrial DNA among *Myodes voles*: consequences for energetics?" BMC Evol Biol **11**: 355.

Bradford, M. M. (1976). "A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding." Analytical biochemistry **72**(1-2): 248-254.

Breton, S., et al. (2014). "A resourceful genome: updating the functional repertoire and evolutionary role of animal mitochondrial DNAs." Trends in genetics : TIG **30**(12): 555-564.

Bunn, C. L., et al. (1974). "Cytoplasmic inheritance of chloramphenicol resistance in mouse tissue culture cells." Proceedings of the National Academy of Sciences of the United States of America **71**(5): 1681-1685.

Burton, R. S., et al. (2013). "Cytonuclear Genomic Interactions and Hybrid Breakdown." Annual Review of Ecology, Evolution, and Systematics **44**(1): 281-302.

Calvo, S., et al. (2006). "Systematic identification of human mitochondrial disease genes through integrative genomics." Nat Genet **38**(5): 576-582.

Dawley, R. M., et al. (1987). "Clonal Reproduction and Polyploidy in Unisexual Hybrids of *Phoxinus eos* and *Phoxinus neogaeus* (Pisces; Cyprinidae)." Copeia **1987**(2): 275.

Deremiens, L., et al. (2015). "Interactions between nuclear genes and a foreign mitochondrial genome in the redbelly dace *Chrosomus eos*." Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology **189**: 80-86.

Du, F. K., et al. (2011). "Direction and extent of organelle DNA introgression between two spruce species in the Qinghai-Tibetan Plateau." New Phytol **192**(4): 1024-1033.

Ellison, C. K. and R. S. Burton (2006). "Disruption of Mitochondrial Function in Interpopulation Hybrids of *Tigriopus Californicus*." Evolution **60**(7): 1382-1391.

Ellison, C. K. and R. S. Burton (2010). "Cytonuclear conflict in interpopulation hybrids: the role of RNA polymerase in mtDNA transcription and replication." J Evol Biol **23**(3): 528-538.

Futuyma, D. J. (2013). Evolution, Macmillan Education.

Glémet, H., et al. (1998). "Geographical extent of arctic char (*Salvelinus alpinus*) mtDNA introgression in brook char populations (*S. fontinalis*) from eastern Quebec, Canada." Molecular Ecology **7**: 1655-1662.

Goddard, K. a., et al. (1998). "Confirmation of gynogenesis in *Phoxinus eos-neogaeus* (Pisces: Cyprinidae)." Journal of Heredity **89**(2): 151-157.

Goddard, K. A. and R. J. Schultz (1993). "Aclonal Reproduction by Polyploid Members of the Clonal Hybrid Species *Phoxinus eos-neogaeus* (Cyprinidae)." Copeia **1993**(3): 650.

Grabherr, M. G., et al. (2011). "Full-length transcriptome assembly from RNA-Seq data without a reference genome." Nature biotechnology **29**(7): 644-652.

Gusdon, A. M., et al. (2007). "Nuclear and mitochondrial interaction involving mt-Nd2 leads to increased mitochondrial reactive oxygen species production." J Biol Chem **282**(8): 5171-5179.

Hara, Y., et al. (2015). "Optimizing and benchmarking de novo transcriptome sequencing: from library preparation to assembly evaluation." BMC Genomics **16**: 977.

Holliday, R. (2006). "Epigenetics: a historical overview." Epigenetics **1**(2): 76-80.

Jaenisch, R. and A. Bird (2003). "Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals." Nature genetics **33 Suppl**(march): 245-254.

Jenuwein, T. and C. D. Allis (2001). "Translating the histone code." Science (New York, N.Y.) **293**(5532): 1074-1080.

Kenyon, L. and C. T. Moraes (1997). "Expanding the functional human mitochondrial DNA database by the establishment of primate xenomitochondrial cybrids." Proc Natl Acad Sci U S A **94**(17): 9131-9135.

Langmead, B. and S. L. Salzberg (2012). "Fast gapped-read alignment with Bowtie 2." Nature Methods **9**(4): 357-359.

Lee, C., et al. (2015). "The Mitochondrial-Derived Peptide MOTS-c Promotes Metabolic Homeostasis and Reduces Obesity and Insulin Resistance." Cell Metabolism **21**(3): 443-454.

Lyko, F., et al. (2010). "The honey bee epigenomes: Differential methylation of brain DNA in queens and workers." PLoS Biology **8**(11).

Maleszka, R. (2008). "Epigenetic integration of environmental and genomic signals in honey bees. The critical interplay of nutritional, brain and reproductive networks." Epigenetics **3**(4): 188-192.

Martin, W. and R. G. Herrmann (1998). "Gene Transfer from Organelles to the Nucleus: How Much, What Happens, and Why?" Plant physiology **118**(1): 9-17.

McKenzie, M., et al. (2003). "Functional respiratory chain analyses in murid xenomitochondrial cybrids expose coevolutionary constraints of cytochrome b and nuclear subunits of complex III." Molecular Biology and Evolution **20**(7): 1117-1124.

Mee, J. A. and E. B. Taylor (2012). "The cybrid invasion : widespread postglacial dispersal by Phoxinus (Pisces : Cyprinidae) cytoplasmic hybrids." **584**: 577-584.

Moreno-Loshuertos, R., et al. (2006). "Differences in reactive oxygen species production explain the phenotypes associated with common mouse mitochondrial DNA variants." Nat Genet **38**(11): 1261-1268.

Murphy, M. P. (2009). "How mitochondria produce reactive oxygen species." Biochem J **417**(1): 1-13.

Oksanen, J., et al. (2007). "The vegan package." Community ecology package **10**.

Pichaud, N., et al. (2012). "Naturally occurring mitochondrial DNA haplotypes exhibit metabolic differences: insight into functional properties of mitochondria." Evolution **66**(10): 3189-3197.

Rand, D. M., et al. (2004). "Cytonuclear coevolution: the genomics of cooperation." Trends in Ecology & Evolution **19**(12): 645-653.

Robinson, M. D., et al. (2010). "edgeR: a Bioconductor package for differential expression analysis of digital gene expression data." Bioinformatics **26**(1): 139-140.

Ryan, M. T. and N. J. Hoogenraad (2007). "Mitochondrial-nuclear communications." Annual review of biochemistry **76**: 701-722.

Saraste, M. (1999). Oxidative Phosphorylation at the fin de siècle. **283**: 1488-1493.

Scheeff, E. D. and P. E. Bourne (2005). "Structural evolution of the protein kinase-like superfamily." PLoS Comput Biol **1**(5): e49.

Schon, E. A., et al. (2012). "Human mitochondrial DNA: roles of inherited and somatic mutations." Nat Rev Genet **13**(12): 878-890.

Senjo, M., et al. (1999). "Extensive Mitochondrial Introgression from *Pinus pumila* to *P. parviflora* var. *pentaphylla* (Pinaceae)." Journal of Plant Research **112**(1): 97-105.

Simao, F. A., et al. (2015). "BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs." Bioinformatics **31**(19): 3210-3212.

Stewart, J. B., et al. (2008). "Purifying selection of mtDNA and its implications for understanding evolution and mitochondrial disease." Nature reviews. Genetics **9**(9): 657-662.

Toews, D. P., et al. (2014). "Migration, mitochondria, and the yellow-rumped warbler." Evolution **68**(1): 241-255.

Toews, D. P. L., et al. (2014). "Migration, mitochondria, and the yellow-rumped warbler." Evolution **68**(1): 241-255.

Tse, C., et al. (1998). "Disruption of Higher-Order Folding by Core Histone Acetylation Dramatically Enhances Transcription of Nucleosomal Arrays by RNA Polymerase III." Molecular and Cellular Biology **18**(8): 4629-4638.

West-Eberhard, M. (1989). "Phenotypic Plasticity And The Origins Of Diversity." Annual Review of Ecology and Systematics **20**(1): 249-278.

Wolff, J. N., et al. (2014). "Mitonuclear interactions: evolutionary consequences over multiple biological scales." Philosophical transactions of the Royal Society of London. Series B, Biological sciences **369**(1646): 20130443-20130443.

Wu, Q. and X. Ni (2015). "ROS-mediated DNA methylation pattern alterations in carcinogenesis." Curr Drug Targets **16**(1): 13-19.

Young, M. D., et al. (2010). "Gene ontology analysis for RNA-seq: accounting for selection bias." Genome biology **11**(2): R14-R14.

Annexes

Annexe 1 Article Gene identification

Supplementary table SI. The significantly differentially expressed and characterized transcripts in white muscle, when *C eos* wild type and Mississippian cybrids from Lake Desjardins are compared. Transcripts associated to foreign mitochondria effects (14) alone are indicated by an asterisk. mRNAs homologous to proteins attributed to mitochondria (3) are in bold. Only matches with an e-value below 10^{-5} were considered. Significance was set as $P_{adj} < 0.05$.

Transcript	Protein ID	Description	E-value
1	XP_691804.2	Predicted: Transmembrane glycoprotein NMB isoform X2*	0
2	NP_998105.2	Ligand of numb-protein X 2b	0
3	XP_007241229.1	Predicted: Zinc finger BED domain-containing protein	0
4	NP_001018510.1	Guanine deaminase*	0
5	XP_007250496.1	Predicted: Carboxypeptidase Q	0
6	XP_007259625.1	Predicted: Pyruvate dehydrogenase (acetyl-transferring) kinase isozyme 3*	0
7	Q08CS6.2	DBH-like monooxygenase protein 2 homolog	0
8	XP_010792626.1	Predicted: Zinc finger protein 862-like*	0
9	XP_003200076.2	Predicted: Integrin alpha-E-like*	0
10	XP_001346132.4	Predicted: Interferon-induced very large GTPase 1-like*	0
11	NP_956704.1	Solute carrier family 16, member 9a	0
12	XP_005465275.1	Predicted: Protein NYNRIN-like	0
13	KKF10883.1	Retrovirus-related Pol polyprotein from transposon 297	0
14	XP_005167957.1	Predicted: Glutaminase kidney isoform*	6E-176
15	XP_002664750.1	Predicted: Sphingosine-1-phosphate phosphatase 2	8E-152
16	XP_009297040.1	Predicted: Kin of IRRE like b isoform X1*	1E-124
17	KRX16276.1	Transposon Ty3-G Gag-Pol polyprotein*	1E-108
18	XP_009291085.1	Predicted: Afadin-like*	1E-100
19	NP_997932.1	Transducin beta-like protein 2 precursor*	1E-89
20	XP_695841.5	Predicted: G protein-coupled receptor kinase 5-like isoform X2	6E-87
21	XP_007242179.1	Predicted: Dihydropyrimidinase*	5E-83
22	XP_005167479.1	Predicted: nectin-1 isoform X2*	1E-75
23	XP_002663014.2	Predicted: Macrophage mannose receptor 1	2E-69
24	KKF13591.1	Zinc finger protein 235	1E-64
25	NP_001082927.1	1,25-dihydroxyvitamin D(3) 24-hydroxylase	9E-62
26	XP_014914825.1	Predicted: Endogenous retrovirus group K member 19 Pol protein-like	2E-61
27	CBN81309.1	Transposable element Tcb1 transposase	3E-51
28	XP_014065326.1	Predicted: Fibroblast growth factor receptor 3-like isoform X1	1E-28
29	AKL71657.1	Hepcidin	8E-27
30	NP_001035458.1	Ankyrin repeat and SOCS box protein 16*	8E-21
31	XP_011492157.1	Predicted: RNA-directed DNA polymerase from mobile element jockey-like	2E-18

