



HAL
open science

Gestion des ressources halieutiques du Haut-Maroni : impacts des pratiques de pêche et des traits d'histoire de vie des espèces sur la diversité génétique intra et inter-populations

Chrystelle Delord

► To cite this version:

Chrystelle Delord. Gestion des ressources halieutiques du Haut-Maroni : impacts des pratiques de pêche et des traits d'histoire de vie des espèces sur la diversité génétique intra et inter-populations. Biodiversité et Ecologie. Agrocampus Ouest, 2019. Français. NNT : 2019NSARH107 . tel-02966951

HAL Id: tel-02966951

<https://theses.hal.science/tel-02966951v1>

Submitted on 14 Oct 2020

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

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

THESE DE DOCTORAT DE

AGROCAMPUS OUEST RENNES
COMUE UNIVERSITE BRETAGNE LOIRE

ECOLE DOCTORALE N° 600
Ecole doctorale Ecologie, Géosciences, Agronomie et Alimentation
Spécialité : *Ecologie et évolution*

Par

Chrystelle DELORD

Gestion des ressources halieutiques du Haut-Maroni :

Impact de la pression de pêche et des traits d'histoire de vie des espèces sur la diversité génétique intra- et interpopulations

Thèse présentée et soutenue à Rennes, le 13 février 2019

Unité de recherche : UMR Ecologie et Santé des Ecosystèmes, INRA/Agrocampus Ouest, Rennes

Thèse N° : 2019-3 H-107

Rapporteurs avant soutenance :

Carole Kerdelhue Directrice de recherche – Centre de Biologie pour la Gestion des Populations (INRA)
Simon Blanchet Directeur de recherche – Laboratoire d'Ecologie Théorique et Expérimentale (CNRS)

Composition du Jury :

Président : Yannick Outreman Professeur – Institut de génétique environnement et protection des plantes (Agrocampus Ouest)
Examineurs : Etienne Bezault Chargé de recherche – Laboratoire de biologie marine (Université des Antilles)
Dir. de thèse : Pierre-Yves Le Bail Directeur de recherche – Laboratoire de physiologie et de génomique des poissons (INRA)
Co-dir. de thèse : Sophie Launey Chargée de recherche – UMR Ecologie et santé des écosystèmes (INRA)
Jean-Marc Roussel Directeur de recherche – UMR Ecologie et santé des écosystèmes (INRA)

Invité(s)

Régis Vigouroux Directeur adjoint du Laboratoire Environnement de Petit-Saut, Hydreco Guyane SARL

Agrocampus Ouest
Université Bretagne Loire

ED 600 : Écologie, Géosciences, Agronomie, Alimentation

THESE

Pour le grade de
Docteur de l'Institut Supérieur des Sciences Agronomiques,
Agro-alimentaires, Horticoles et du Paysage

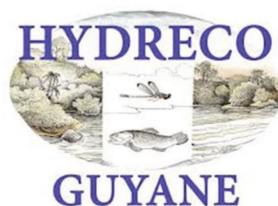
**GESTION DES RESSOURCES HALIEUTIQUES DU HAUT-MARONI :
IMPACT DE LA PRESSION DE PECHE ET DES TRAITS D'HISTOIRE DE VIE DES
ESPECES SUR LA DIVERSITE GENETIQUE INTRA ET INTER-POPULATIONS**

Préparée par :

Chrystelle Delord

à l'Unité Mixte de Recherche INRA-Agrocampus Ouest 0985 ESE
Ecologie et Santé des Ecosystèmes

Soutenue sur le site Agrocampus Ouest de Rennes
Le 13 février 2019



Remerciements

On entend souvent dire que la thèse est une entreprise solitaire. Pourtant, je me suis rarement sentie moins seule que durant ces trois années (et des poussières !) passées aux côtés de toutes les personnes qui ont mûri, porté et soutenu ce projet depuis ses débuts. Ces personnes sont nombreuses et j'espère, à la lecture de ces remerciements, que chacune d'entre elles pourra mesurer l'ampleur de ma reconnaissance.

Avant toute chose, je tiens à remercier les rapporteurs et examinateurs de mon jury : Carole Kerdelhue, Simon Blanchet, Etienne Bezault et Yannick Outreman, d'avoir accepté d'évaluer ce travail. Merci à Carole et Simon d'avoir passé une partie de leurs congés de fin d'année à lire le présent manuscrit ! Merci, à vous tous, de vos conseils et suggestions qui me seront d'une grande aide pour le travail de valorisation à venir.

A présent, je souhaite remercier vivement mes encadrants pour leur présence à mes côtés jusqu'à l'aboutissement de cette thèse. J'ai eu la chance de bénéficier d'un encadrement bi-partite dans le cadre de cette thèse CIFRE, qui s'est conduite en collaboration entre l'INRA et le bureau d'étude Hydreco Guyane (Laboratoire Environnement de Petit Saut à Kourou). A ce titre, je remercie mes encadrants scientifiques Pierre-Yves Le Bail, Sophie Launey et Jean-Marc Roussel de m'avoir accordé leur confiance, jusqu'au bout malgré les dernières semaines un peu difficiles : Pierre-Yves, merci pour ton calme et ta patience indéfectibles qui ont permis aux deux esprits têtus que nous sommes de trouver un terrain d'entente à chacune de nos réunions animées. Sophie, merci pour ta douceur et ta bienveillance qui m'ont toujours rappelé mon droit à l'erreur et permis d'assumer ainsi mes moments de doute. Jean-Marc, merci pour ta rigueur et ta perspicacité qui m'ont régulièrement aidée à prendre du recul et rester réaliste. J'adresse un grand merci également à mon encadrant technique, Régis Vigouroux. Régis, j'ai toujours senti, au cours de ces trois années, ton regard à la fois discret et bienveillant sur mon travail. Merci de m'avoir toujours fait confiance et de m'avoir permis d'avancer à mon rythme avec optimisme. Merci à toi de l'énergie que tu as investie pour me permettre d'effectuer cette thèse sous contrat CIFRE. Je terminerai en remerciant ainsi également l'Association Nationale de Recherche et de Technologie (ANRT) pour avoir contribué à ma rémunération.

Ce travail de thèse n'aurait pas pu voir le jour sans l'ensemble des acteurs du Parc Amazonien de Guyane (PAG), initiateur du Programme Pêche et appuyé également par la DEAL et l'Office de l'Eau de la Guyane. Je remercie tout d'abord Raphaëlle Rinaldo, responsable scientifique du PAG, pour avoir mené le pilotage administratif et technique de ce projet de grande envergure. Un grand merci également à Guillaume Longin, qui non-content d'avoir brillamment mené de front son poste de chargé biodiversité au PAG et son stage de fin de Master (clôturant ainsi deux des trois grands volets du Programme Pêche), a également toujours su répondre avec efficacité à mes questions. Guillaume, ta connaissance extrêmement poussée des réalités sociales et géographiques de notre aire d'étude commune m'a été d'un très grand secours.

Je ne saurais oublier les nombreux autres employés du PAG qui ont joué un rôle déterminant dans la constitution du jeu de données utilisé dans cette thèse. Je remercie ainsi tous les référents PAG, habitants des villages Amérindiens et Boni du Haut-Maroni, qui ont assuré la mise en relation quotidienne entre les personnes mobilisées sur le terrain et les pêcheurs. A tous, un très grand merci d'avoir partagé votre quotidien durant toutes les campagnes de collecte de matériel biologique :

Kutaka Aïtalé, Palanaiwa Alikana, Sébastien Amaïpetit (pour ta bonne humeur et les virées tardives au Chinois), Moloko Atuwaike, Alfred Djaba, Kalou, Janakalé Makiloewala (pour ta pêche miraculeuse et ta superbe cuisine), Pitoma Mekouanali, Kuliwaike Menali, Marc Pinson, Mones Tokotoko, Hervé Tolinga et Etulano Yamo. Une pensée très émue pour Rastaman et Raymond Essimon ainsi que leurs proches. Merci à toutes les autres personnes employées au PAG, à Hydreco Guyane, à l'INRA ou au MNHG qui ont été mobilisées pour les campagnes de collecte. En outre, j'adresse toute ma reconnaissance aux pêcheurs des villages du Haut-Maroni et à leurs familles qui font preuve d'une bien grande patience vis-à-vis de tous ces visiteurs déambulant jour après jour.

Les échantillons biologiques récoltés sur le Haut-Maroni ont été traités (extraction, quantification et dilutions d'ADN) en laboratoire avec le concours précieux d'Adrien Oger, Anne-Laure Besnard et Thibaut Jousseume. Merci à tous les trois du temps passé à me former et m'aider, toujours dans la bonne humeur malgré les quelques frayeurs et casses-tête rencontrés (petit clin d'œil spécial Pherastar à Anne-Laure !). De même, je remercie la plateforme Génome-Transcriptome de Pierroton, qui a effectué toute la procédure de génotypage, de m'avoir accueillie dans ses locaux pour me montrer le fonctionnement de la spectrométrie de masse MassArray et me former à l'analyse des données brutes. Erwan Guichoux, Adline Delcamp, Emilie Chancerel et Marie Massot, je suis heureuse de vous avoir apporté un peu de poisson parmi votre multitude de plantes.

Ce travail de thèse n'aurait pas pu aboutir sans une première étape décisive : le développement de ressources moléculaires. A ce titre, je souhaite remercier vivement Eric Petit pour la conception et le pilotage du projet inter-équipes « RAD4SNPs », qui a abouti à la mise au point de marqueurs SNP pour plusieurs espèces étudiées au sein de l'UMR, y compris bien entendu nos chers poissons de Guyane. Merci Eric, pour ton pragmatisme et ta façon de détecter puis signaler, toujours avec justesse, pertinence et sans détours, les failles de raisonnement ou autres fausses routes qui ont pu être miennes. J'espère acquérir un jour le recul impressionnant que tu possèdes sur le paysage bigarré de la génétique des populations. Venons-en à tout le développement pratique de « RAD4SNPs » ! Un très, très grand merci à Gilles Lassalle, alias « Le Hibou », et bioinformaticien de son métier... oups, que dis-je, biologiste ET informaticien ! Que n'aurais-je fait sans ton aide, Gilles, bien des choses. Les connaissances que j'ai acquises grâce à toi quant aux logiciels de génomique, l'environnement Linux et la programmation Python ont déclenché chez moi un intérêt pour le sujet à un niveau que je n'aurais pas soupçonné, qui avais toujours été si effrayée par les ordinateurs. Merci pour ces longues heures passées à se brûler les neurones à deux derrière notre écran. Merci de m'avoir forcée... euh, encouragée à me réinscrire pour passer mon permis de conduire.

Je remercie également les personnes qui m'ont fait bénéficier de leur intérêt et de leur expertise tout au long de cette aventure quelle qu'en soit l'étape : Eric Petit, Olivier Lepais, Raphaël Covain, Ivan Paz-Vinas, Jérôme Prunier, Simon Blanchet, Quentin Rougemont, Guillaume Evanno, Damien Fourcy, Louis de Beaufort, Guy Fontenelle, Sébastien Brosse et tous ceux que je ne peux pas citer ici mais que je n'oublie pas. Je remercie chaleureusement Oscar Gaggiotti, du *Scottish Oceans Institute*, de m'avoir accueillie dans son équipe durant trois mois en Ecosse et de m'avoir apporté son expertise concernant l'utilisation des simulations en génétique des populations. Merci à Marie Louis et Emma Carroll pour leur gentillesse et leur soutien au cours de séjour. Par ailleurs, j'ai eu la chance d'effectuer ce séjour dans le cadre du programme EIR-A proposé par le contorsium Agreenium, qui m'a permis de belles rencontres avec de nombreux chercheurs et doctorants.

Merci à tous mes collègues des UMR ESE et U3E. Un grand merci à Gervaise Février, Marie-Thérèse Delaroché et Jessica Vincent de leur titanesque travail d'administration : les choses iraient bien de travers sans vous ! Jessica, grâce à toi je ramènerai même chez moi une nouvelle amie à mauvais caractère et longues moustaches. Merci à Gilles Maubert, administrateur réseau, d'avoir toléré mes infidélités à Windows et d'être constamment resté disponible pour faire face à mes déboires informatiques. Merci à tous ceux qui m'ont entourée au quotidien. Marc Collinet, tes crocs roses fushia, ta serviabilité et ton ouverture d'esprit resteront longtemps dans ma mémoire comme un véritable modèle de vie. Les cafés de la pause de 8:00 m'ont donné cœur à l'ouvrage jusqu'à la (presque) toute fin (eh oui, le lever pouvait alors devenir plus difficile ;-)!). Merci à Clarisse, Anne et Aurélie de leur gentillesse. Merci Clarisse, pour ton soutien sans faille y compris durant les phases les plus critiques. Tu as beaucoup contribué à me faire voir les choses de manière plus positive lorsqu'il le fallait et tu en auras toujours toute ma reconnaissance. Merci à toutes celles et ceux que je ne cite pas, mais que je n'oublie pas pour autant !

Je remercie tous les compagnons de galères qui ont partagé avec moi le « bureau des thésards »/« garde-manger officiel du bâtiment 33 » au cours de ces trois années : Lucie, Pierre-Loup, Fanny, Julien, Ahmed et Cécile. Certains d'entre vous sont passés de l'autre côté de la barrière il y a un moment déjà, Cécile je ne me fais aucun souci pour toi, tu en sortiras plus que victorieuse lorsque le temps sera venu ! Merci à tous pour votre bonté et votre humour, les batailles d'élastiques, les histoires d'aérobic ou de bootyshake et toutes nos conversations cultes (« J'aime bien l'ambiance dans le bureau, vous me faites rire. » « Oui c'est réciproque, sauf que nous on le fait exprès. » « Comment ça ? »). Merci pour ces belles années passées en votre compagnie. Merci aux stagiaires de Master d'avoir contribué à rafraîchir notre quotidien de doctorant encroûté !

Lorsque je suis arrivée à Rennes, je n'y connaissais quasiment personne. Au cours de ces trois années, j'ai pu y construire un quotidien enrichissant et réconfortant y compris en dehors du cadre professionnel. Merci à mes « colocatrices » de 1^{ère} année, Amanda, Manon, Dior et Divna grâce à qui je ne me suis jamais sentie seule le soir ; au département Urgence et Secourisme de l'Unité Locale de la Croix-Rouge de Rennes de m'avoir accueillie à bras grands ouverts dès mon arrivée et de tout de suite m'avoir fait me sentir à ma place. Merci à la Scène Breizh'Illienne (aujourd'hui Rennes Musical) pour les merveilleux moments de danse et de chansons partagés le mercredi soir. Je me souviendrai avec émotion de chacun et chacune d'entre vous.

Tout cela n'aurait pas été possible dans le soutien inconditionnel de ma famille. Merci à mes parents et à ma grand-mère, qui ont toujours eu confiance en mes choix professionnels et personnels. Merci à mon petit frère Emmanuel d'avoir supporté ma ronchonnerie d'étudiante plongée dans ses bouquins. Merci à tous mes amis de m'avoir adoptée et de rester malgré la distance géographique qui nous sépare, certains depuis un bon bout de temps déjà. Merci à Marylène, Manon & Manon, Sissou & Nassou d'être là pour moi depuis plus de dix ans. Merci à tous les copains et copines de fac de me laisser suivre la suite de vos aventures. Merci à celui qui ne m'a plus jamais quittée depuis notre première année de licence : Amiel, tu n'es pas encore débarrassé de moi, j'espère bien rester à tes côtés pour encore un bon moment.

Table des matières

Liste des figures	iii
Liste des tableaux	vi
Liste des annexes	vii

Introduction

I. Contexte sociologique de l'étude	2
I.1. La Guyane française, un cas particulier pour la gestion des ressources naturelles	6
I.2. La pêche vivrière et commerciale sur le Haut-Maroni, un bassin de vie	7
I.3. Le Parc Amazonien de Guyane et les acteurs de l'eau en territoire guyanais	9
II. Contexte scientifique et état de l'art	12
II.1. Biodiversité et conservation des écosystèmes aquatiques néotropicaux : le cas de la Guyane française	14
II.1.1. Les écosystèmes d'eau douce : rôles, menaces et conservation	14
II.1.2. Diversité et répartition des espèces néotropicales	15
II.1.3. Préservation de la diversité ichtyque des fleuves guyanais	18
II.1.3.1. Qualité d'habitat et richesse spécifique ou fonctionnelle	18
II.1.3.2. Les principales caractéristiques biologiques des espèces : identifier des espèces plus vulnérables	20
II.1.3.3. Traits d'histoire de vie et dynamique des populations	23
II.2. La variabilité génétique comme support de gestion des populations naturelles	25
II.2.1. Variabilité génétique, traits d'histoire de vie et environnement	25
II.2.1.1. Mécanismes de la variabilité génétique et marqueurs moléculaires	25
II.2.1.2. Forçages biotiques et abiotiques de la variabilité génétique	29
II.2.1.3. Comprendre et prédire la variabilité génétique : l'apport de la génétique comparative des populations	31
II.2.2. Variabilité génétique et pressions anthropiques	34
II.2.2.1. Diversité et structure génétique de populations exploitées	34
II.2.2.2. Des indices de variabilité génétique comme support de gestion	36
III. Enjeux méthodologiques de la thèse	38
III.1. Le développement de ressources moléculaires	38
III.2. Appliquer une approche comparative entre espèces	39
III.3. Révéler la mise en place et l'issue de processus complexes : Simuler des processus et explorer des hypothèses	41
IV. Objectifs de l'étude et stratégie	44

Chapitre 1 :

Développer les ressources moléculaires nécessaires pour l'étude de la variabilité génétique chez des espèces non-modèles	48
1. Résumé étendu en français	48
2. Article n°1 : <i>A cost-and-time effective procedure to develop SNP markers for multiple species: A support for community genetics</i>	52

Chapitre 2 :

Description des patrons de variabilité génétique chez 18 espèces de poissons d'intérêt halieutique à l'échelle du Haut-Maroni : Approche comparative entre sites et entre espèces	80
1. Résumé étendu en français	80
2. Article n°2 : <i>Riverscape patterns of within and between-population genetic diversity in eighteen harvested Neotropical freshwater fish species</i>	84

Chapitre 3 :

Analyse exploratoire de l'évolution de la variabilité génétique au sein de populations d'un réseau hydrographique dendritique	120
1. Introduction	121
2. Méthodes : Simulation d'un scénario simple avec et sans pression de pêche	123
3. Résultats et discussion préliminaires	125
4. Références	130

Discussion générale et perspectives 142

1. Réponse aux enjeux méthodologiques de l'étude	144
1.1. Le développement des ressources moléculaires	144
1.1.1. Des SNP disponibles pour les espèces de poissons du Haut-Maroni	144
1.1.2. Difficultés rencontrées et limites de l'approche	144
1.1.3. Recommandations pratiques	147
1.2. Appliquer une approche comparative entre espèces	148
1.3. Simuler des processus et explorer des hypothèses	149
2. Pratiques de pêche et conservation des populations de poissons du Maroni	150
2.1. Variabilité génétique et paysage aquatique	151
2.1.1. Influence de l'environnement naturel	151
2.1.2. Contrastes géographiques des rendements de captures	153
2.2. Variabilité génétique et traits d'histoire de vie	154
2.2.1. Relation entre les caractéristiques biologiques et la structure génétique des populations	154
2.2.2. Composition spécifique des captures sur le Haut-Maroni	155
2.3. Conclusion et recommandations	156

Bibliographie 162**Annexes** 186

Liste des figures

Introduction :

Figure I.1: Carte générale de la Guyane française, avec ses principaux cours d'eau et ses principales communes	5
Figure I.2: Carte des bassins versants de Guyane française	8
Figure I.3: Carte générale du bouclier guyanais	17
Figure I.4: Caractéristiques et comparatif des trois principaux types de marqueurs moléculaires : ADN mitochondrial, microsatellites et SNPs	28
Figure I.5: Positionnement relatif de différentes approches moléculaires en écologie aquatique, le long d'un gradient de recherche fondamentale à appliquée (abscisse) et de différents niveaux d'organisation de la diversité, des gènes aux communautés (ordonnée)	33

Chapitre 1 :

Figure C1.1: Summarized description of data treatment from reception of raw paired-end reads, to the obtaining of amplifiable single-nucleotide-polymorphism (SNP) markers. Previous demultiplexing, raw reads cleaning, and Stacks catalogue building need to be done manually before the automatized process throughout the pipeline begins	61
Figure C1.2: (A) Total number of paired-end reads after PCR duplicates removal, per species and per library (2 libraries/species) (grey bars) and mean per-RAD-tag depth of coverage, per species and per library (black dots). (B) Number of polymorphic RAD-tags per species and number of RAD-tags containing a single detected SNP	64
Figure C1.3: (A) Analysis of molecular variance showing within-site (framed values), between-site within river tributary (light), and between tributaries (dark) sources of variation (B) Evolution of the Pearson's correlation coefficient between within-site polymorphism rate (P) and the distance to the most downstream site across species.	67

Chapitre 2 :

Figure C2.1: Map of French Guiana (A) and of the Upper-Maroni (B) with its main tributaries. Sampling locations are indicated with villages (i.e., high fishing pressure, closed circles) and reference sites (i.e., low fishing pressure, open circles). Streams in black are main channels with Strahler index > 4 while streams in grey are streams with Strahler index ≤ 4.....	88
Figure C2.2: Values of (A) internal diversity ID and (B) standardized local isolation (F_{ST-p} , i.e., average pairwise F_{ST}) per sampling site across species	97
Figure C2.3: Graphical representation of the estimated marginal means (grey dots) with 95% confidence intervals (grey bars) for each species (rows), obtained from the three covariance analyses (columns) performed to test the effect of the upstream distance (i.e. the distance from the Abattis Cottica) and the longitude (i.e. the site position on tributaries from West to East) on local genetic diversity (P_{95} , H_S) and mean genetic isolation (F_{ST-p}) across species	100

Figure C2.3' : Values of expected heterozygosity H_S per sites grouped into three categories (reference sites from Litani and Marouini: LM.trib; sites with harvesting pressure: Villages, and reference sites from the Tampok and the Waki: TW.trib), for each species	101
Figure C2.3'' : Values of mean local isolation $F_{ST.p}$ per sites grouped into three categories (reference sites from Litani and Marouini: LM.trib; sites with harvesting pressure: Villages, and reference sites from the Tampok and the Waki: TW.trib), for each species	102
Figure C2.4 : Biplot of the principle-components analysis used to describe the relative position of all species regarding the strenght of isolation-by-distance (IBD_R) and of the evolution of expected heterozygosity (H_S), polymorphism rate (P_{95}) and local differentiation ($F_{ST.p}$) along the upstream gradient	104

Chapitre 3 :

Figure C3.1 : Représentation schématique du réseau dendritique servant de support à l'ensemble des simulations présentées dans ce chapitre. Les patchs 0, 7, 8, 10 et 14 représentent des zones peu connectées au reste du réseau tandis que les patchs 3, 4, 5 et 13 représentent une aire de plus forte connectivité	123
Figure C3.2 à C3.5 : Evolution, respectivement, du r de Mantel (IBD_r) (C3.2), de COR_{FST} (C3.3), de COR_p (C3.4) et du F_{ST} global (C3.5), au cours des générations n°140 à 200, pour chacune des huit combinaisons de paramètres explorées durant cette procédure de simulations d'une métapopulation de type dendritique (taille des patch $N=100$ ou 1000 ; probabilités de migration $m=0,05$ ou $0,20$; proportion d'individus éliminés dans patchs centraux $H=0$ ou $0,125$).	132
Figure C3.6 : Evolution de COR_{FST} , au cours des générations 140 à 350 pour plusieurs combinaisons de paramètres explorées durant cette procédure de simulations d'une métapopulation dendritique (taille des patchs $N=100$; probabilités de migration $m=0,05$; $0,10$; $0,20$ ou $0,30$; proportion d'individus éliminés dans patchs centraux $H=0$ à $0,50$)	136
Figure C3.7 : Evolution de COR_{FST} , au cours des générations 140 à 350 pour plusieurs combinaisons de paramètres explorées durant cette procédure de simulations d'une métapopulation dendritique (taille des patchs $N=1000$; probabilités de migration $m=0,05$; $0,10$; $0,20$ ou $0,30$; proportion d'individus éliminés dans patchs centraux $H=0$ à $0,50$).	137
Figure C3.8 : Evolution de COR_p , au cours des générations 140 à 350 pour plusieurs combinaisons de paramètres explorées durant cette procédure de simulations d'une métapopulation dendritique (taille des patchs $N=100$; probabilités de migration $m=0,05$; $0,10$; $0,20$ ou $0,30$; proportion d'individus éliminés dans les patchs centraux $H=0$ à $0,50$).	138
Figure C3.9 : Evolution de COR_p , au cours des générations 140 à 350 pour plusieurs combinaisons de paramètres explorées durant cette procédure de simulations d'une métapopulation dendritique (taille des patchs $N=1000$; probabilités de migration $m=0,05$; $0,10$; $0,20$ ou $0,30$; proportion d'individus éliminés dans patchs centraux $H=0$ à $0,50$).	139

Discussion :

Figure D.1: Schéma décrivant les étapes successives de sélection de marqueurs moléculaires SNP pour les espèces focales du Haut-Maroni, depuis la première étape de la constitution du panel de détection de SNP jusqu'à obtention du set de marqueurs final ayant servi de base au génotypage..... 145

Figure D.2: Principaux cours d'eau impactés par l'orpaillage sur le territoire du Parc Amazonien de Guyane..... 152

Liste des tableaux

Introduction :

Table I.1: Avantages et inconvénients des principales approches en génétique comparative des populations	40
---	----

Chapitre 1 :

Table C1.1: Phylogenetic status of sampled species, geographic origin and number N_s of samples included in the pools used for building the libraries.....	56
---	----

Table C1.2: Number of individuals and markers genotyped for each fish species from French Guiana using the MassARRAY facilities, and genotyping success per species	65
--	----

Table C1.3: Per-species evolution of the number of markers between RAD-tags assembly and pipeline output, with proportion of loss for each of the three filtering steps, when using identical, default parameter values of the pipeline for all species.....	66
---	----

Chapitre 2 :

Table C2.1: Information about the sites sampled across the Upper-Maroni for this study	88
---	----

Table C2.2: Information about the eighteen species chosen to conduct this study	91
--	----

Table C2.3: Values of IBD_R (and slope of the isolation by distance IBD_{Slope}), $Z_r.P_{95}$, $Z_r.H_s$ and $Z_r.F_{ST.p}$ per species)	94
--	----

Table C2.4: Estimates of local contemporary effective size N_e {with 95% parametric confidence intervals} for each species and sampling site. Estimates were obtained from the linkage-disequilibrium method implemented within the software NeEstimator with $P_{CRIT}=0.05$. Estimates with finite confidence intervals are grey-shaded	95
---	----

Table C2.5: Results of the covariance analyses (multiple linear regression models) performed to test the effect of the upstream distance and the longitude on local genetic diversity (P_{95} , H_s) and mean genetic isolation ($F_{ST.p}$) across species	99
--	----

Chapitre 3 :

Table C3.1: Liste des statistiques et indices de diversité génétique calculés sur chaque jeu de pseudo-données simulé.....	125
---	-----

Liste des annexes

Annexe I: Planches descriptives des principales espèces d'intérêt halieutique rencontrées dans le Haut-Maroni	187
Annexe II: Understanding and managing fish populations: keeping the toolbox fit for purpose, <i>Journal of Fish Biology</i> , 92, 727–751	191
Annexe III: Informations pratiques sur le protocole d'échantillonnage du matériel biologique effectué pour ce travail de thèse	217
Annexe IV: Génotypage : La spectrométrie de masse MALDI-TOF.....	221
Annexe V: Détection et gestion des phénomènes de contaminations croisées rencontrés lors de la procédure de séquençage RADSeq pour la mise au point des marqueurs moléculaires utilisés durant la thèse	225
Annexe VI: Matériel supplémentaire de l'article n°1 présenté en Chapitre 1	230
Annexe VII: Scripts Python constitutifs du pipeline informatique présenté en Chapitre 1	239
Annexe VIII: Exemple de fichier final issu du pipeline informatique présenté en Chapitre 1 pour l'espèce <i>Serrasalmus rhombeus</i>	264
Annexe IX: Matériel supplémentaire de l'article n°2 présenté en Chapitre 2	265
Annexe X: Scripts SimuPOP commentés ayant fourni les pseudo-données analysées dans le Chapitre 3	280
Annexe XI: Figures supplémentaires du Chapitre 3	283
Annexe XII: Informations supplémentaires relatives aux marqueurs sélectionnés au cours de la phase de mise au point des ressources génétiques pour les espèces étudiées	285
Annexe XIII: Liste des productions et communications scientifiques effectuées au cours de la thèse	290

Introduction:

I. Contexte sociologique de l'étude



Sébastien Amaïpetit et sa compagne, relevant un filet dormant en aval du saut Domofou Soula, près du village de Taluen, au crépuscule (Photographie : R. Covain, MNHG - Décembre 2015)



Figure I.1: Carte générale de la Guyane française, avec ses principaux cours d'eau et ses principales communes

I.1. La Guyane française, un cas particulier pour la gestion des ressources naturelles

La Guyane française (Fig. 1), « *Pays des milles eaux* » (Lézy 2000) situé en zone équatoriale humide d'Amérique du Sud, se caractérise par un milieu de type tropical amazonien connu pour sa très riche diversité faunistique et floristique (Haffer 1982). Longtemps préservé de l'occupation humaine moderne en dehors du développement relativement récent de ses zones côtières, le territoire guyanais ne fit objet que de peu de politiques suivies de conservation de ses ressources naturelles, et ce, malgré ses quatre siècles d'occupation par le gouvernement français. Bien que le département de Guyane soit théoriquement concerné par la réglementation en vigueur en France métropolitaine, il est aisé d'imaginer de quelle façon ses spécificités géographiques (zone moins densément peuplée, biodiversité riche mais restant partiellement documentée) et sociologiques (présence antérieure de peuples autochtones ayant vécu, et vivant toujours, en lien étroit avec les ressources naturelles) compliquent l'application de ces règlements. Cependant, les pressions anthropiques sur le milieu naturel (déforestation, activité d'orpaillage légal ou illégal, fragmentation des habitats ou encore exploitation des ressources) se sont accrues au cours des dernières décennies. Aujourd'hui, ces pressions rendent plus urgente la mise au point de mesures de gestion adaptées afin d'assurer la pérennité des ressources du territoire (Meunier et al. 1998).

La riche faune ichthyque, dulçaquicole et marine, de Guyane française fait l'objet d'une exploitation par la pêche à plusieurs niveaux. Tandis que pêches commerciale et de loisir se concentrent principalement sur les zones côtières et estuariennes, la pêche de subsistance, elle, persiste plus largement dans les fleuves de l'intérieur des terres, parfois renforcée localement par une pêche commerciale de moindre envergure (Balland & Roux 2005). Des tentatives d'élevage de certaines espèces (par exemple l'atipa *Hoplosternum littorale*, le paya *Astronotus ocellatus* et la carpe locale *Leporinus friderici*) ont été menées afin de renforcer la mise à disposition locale de poisson en zone côtière (Boujard et al. 1988). Son implantation est cependant plus complexe au niveau des fleuves (Giron et al. 2001) où, comme en milieu côtier, la pêche reste prédominante. La pêche maritime commerciale et de loisir font l'objet de quelques arrêtés préfectoraux qui définissent principalement les engins et lieux de pêche autorisés¹. En eau douce, il existe une réglementation succincte relative à l'usage du filet dormant (arrêté préfectoral n° 936 ID/2B de Mai 1978), portant sur la longueur totale par embarcation, la largeur de cours d'eau couverte et la durée maximale de pose autorisées. Il est cependant délicat d'imposer, et difficile de vérifier, l'application d'un tel règlement au quotidien surtout dans les zones où le pouvoir d'achat reste limité et où la pêche vivrière, fortement intégrée dans les mœurs, permet d'assurer un apport alimentaire régulier (Richard-Hansen 2002 ; Martin 2014). En outre et jusqu'à la fin des années 2000, l'absence de texte référent établissant la liste des poissons d'eau douce de Guyane limitait la portée d'une réglementation concernant leur exploitation. Balland et Roux (2005), dans leur rapport de l'Inspection Générale de l'Environnement, considèrent ainsi le poisson guyanais au sens large comme un « non-être juridique ».

¹ La pêche maritime de loisir est encadrée juridiquement par les articles R921-83 à R921-93 du code rural et de la pêche maritime. La pêche commerciale, elle, est réglementée par quelques arrêtés de 1984, 1999 et 2010 ainsi que par une décision du CRPMEM (Comité Régional des Pêches Maritimes et des Elevages Marins) de 2009 et 2016.

En comparaison, la gestion de la faune terrestre vis-à-vis de la chasse (vivrière, commerciale ou de loisir) a davantage bénéficié de réglementations successives portant sur des espèces particulières, protégées à différents niveaux de tolérance par des arrêtés préfectoraux dédiés. Ce processus a permis la validation, en 2004, d'un ensemble d'ORGFH² complémentées par des rencontres entre divers acteurs de la chasse en Guyane et a abouti à la signature, le 10 août 2006, d'un arrêté ministériel relatif à la capture d'animaux sauvages sur le territoire guyanais. Un socle juridique commun devenait alors disponible pour l'élaboration de réglementations de la chasse tenant compte des spécificités et des contraintes socio-écologiques locales, telles que la dimension culturelle de la chasse vivrière (Hansen 2009 ; Richard-Hansen 2002).

En l'an 2000, le premier SDAGE³ en application au niveau des bassins versants guyanais faisait état de la nécessité de mieux connaître la composition et le fonctionnement biologique des peuplements piscicoles exploités par la pêche, en vue de la mise en place de mesures plus adaptées à leur conservation. Entre temps, les travaux de description et inventaires de la diversité ichtyque guyanaise ont permis de quantifier son importance et sa répartition (Le Bail et al. 2012). D'autres travaux, plus spécifiquement dédiés à la place de la pêche dans le quotidien des peuples autochtones de Guyane, ont également permis de caractériser les espèces de forte importance halieutique et patrimoniale (e.g. Pagézy et Jégu 2002; Meunier 2004). Cependant, les pratiques et impacts de la pêche restent actuellement peu connus à l'échelle du territoire (voir cependant Hurault 1965; Richard-Hansen 2002; Longin 2016 en ce qui concerne les milieux dulçaquicoles; ou encore Cissé et al. 2014 pour un exemple en milieu marin), et il apparaît nécessaire de développer ces connaissances dans une perspective de préservation des espèces exploitées.

I.2. La pêche vivrière et commerciale sur le Haut-Maroni, un bassin de vie

La Guyane Française bénéficie d'un réseau hydrographique particulièrement dense comparativement au territoire français métropolitain, avec quatre grands bassins versants principaux (l'Oyapock et le Maroni étant les plus importants en termes de superficie, suivis de l'Approuague et de la Mana) et plusieurs autres fleuves côtiers (Fig.2). Bien que les larges tributaires représentent moins de 30% du réseau hydrographique guyanais (la majorité - jusqu'à 70% - étant représenté par les petites masses d'eau (PME) ou « criques », Dedieu et al. 2014), ils concentrent des liens socio-culturels importants entre les populations humaines riveraines, tout particulièrement dès lors que l'on s'éloigne de la zone côtière, les fleuves devenant alors les principales voies de communication et le support essentiel de la vie quotidienne. Ceci est particulièrement vrai pour l'Oyapock et le Maroni, qui délimitent le territoire guyanais à l'Est et à l'Ouest, respectivement, faisant frontière avec le Brésil et le Suriname. Ils abritent depuis la fin du XVIII^{ème} siècle plusieurs populations natives Amérindiennes Wayana, Teko et Wayãpi mais également des populations Bushinengués, essentiellement Aluku, fondées sur les rives du Maroni par d'anciens esclaves originaires des exploitations hollandaises du Suriname. Aujourd'hui, ces populations, parfois renforcées par

² « Orientations régionales de gestion de la faune sauvage et d'amélioration de la qualité de ses habitats », énoncées à l'échelle nationale en collaboration avec les Directions régionales de l'Environnement (DIREN) et l'Office national de la Chasse et de la Faune Sauvage (ONCFS) par la loi 2000-698 du 26 juillet 2000.

³ « Schéma directeur d'aménagement et de gestion des eaux »

d'autres communautés extérieures (Créoles, Haïtiennes, Brésiliennes, Surinamiennes ou encore métropolitaines) au niveau de certains bourgs, constituent l'essentielle mosaïque humaine présente sur ces bassins versants, en particulier le Maroni⁴.

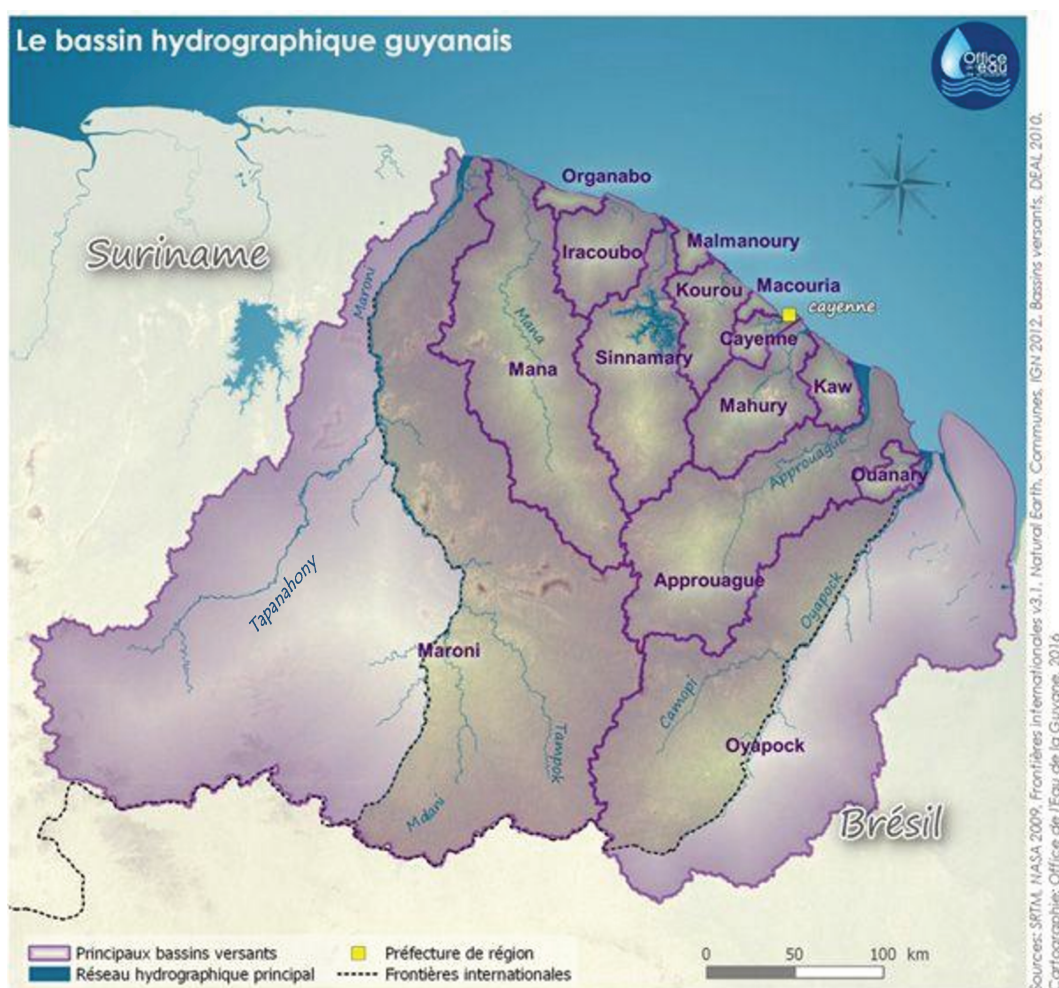


Figure I.2 : Carte des bassins versants de Guyane française (Site web de l'Office de l'eau de la Guyane)

Le bassin versant du Maroni est le plus important de Guyane française, fort de ses 611,70 km de linéaire et ses plus de 65 000 km² de superficie. La dénomination de « Haut-Maroni » fait référence à toute la partie du bassin versant s'écoulant jusqu'au sud de la ville de Grand-Santi, en amont de la confluence entre le lit principal du fleuve, la Lawa (qui marque la frontière avec le Suriname) et l'affluent de la Tapanahony, prenant sa source au Suriname (Fig.2). Trois divisions administratives recouvrent le territoire: Grand-Santi au Nord, Papaïchton en position intermédiaire, et Maripasoula, au Sud, la plus grande commune de tout le territoire français (métropolitain et outremer) en termes de superficie. Au 1^{er} janvier 2014, les référencements INSEE faisaient état de plus de 10 000 habitants à Maripasoula (bourg principal et sous-communes annexes confondues) et plus de 6 000 à Papaïchton et à Grand-Santi. Bien que ces chiffres soient nettement inférieurs à ceux des communes côtières de Guyane française, les communes du Haut-Maroni accusent le plus fort taux d'évolution

⁴ Une description plus exhaustive de l'ethnogenèse des peuplements du Haut-Maroni est disponible en introduction du rapport de stage de Master 2 de Guillaume Longin (Parc Amazonien de Guyane, 2016). Voir également Fleury, *Les populations du Haut-Maroni et le projet de parc national de la Guyane* (Revue d'ethnobiologie JATBA, 1998).

annuel depuis la fin des années 2000 et contribuent à la croissance démographique de l'ensemble du territoire guyanais. Plus précisément, un taux d'évolution annuel de +4.5% est enregistré pour les communes de l'Ouest guyanais, contre +2.4% pour l'ensemble du territoire et seulement +0.5% sur les communes de l'Est guyanais, sur l'Oyapock. Or, les zones habitées du Haut-Maroni restent très liées au fleuve dans le quotidien de leurs occupants. Déplacements, hygiène quotidienne, commodités, et bien entendu accès aux ressources naturelles, sont autant d'activités auxquelles le fleuve fournit un support indispensable. Plusieurs dizaines d'espèces de poissons sont régulièrement capturées (Annexe I), et consommées par les habitants de tous âges au quotidien (Pagézy et Jégu 2002; Meunier 2004; Martin 2014).

Dans le même temps, les phénomènes de sédentarisation des populations humaines et la modernisation de la vie quotidienne ont été accompagnés de significatifs bouleversements dans les pratiques d'exploitation des ressources. L'utilisation de certaines techniques traditionnelles (comme la pêche à la canne ou à l'arc pour les Amérindiens, le barrage de crique et la nasse *bakisi* pour les Bushinengués) a peu à peu diminué au profit du filet épervier, du fusil-harpon ou de la trappe⁵ mais, surtout, du filet dormant, plus efficace et moins sélectif. En outre, l'arrivée des moteurs hors-bord et des congélateurs alimentés par générateurs thermiques, de plus en plus accessibles à l'achat, ont permis de pêcher à plus longue distance et de mieux conserver les prises à court-terme avant de les saler ou de les boucaner pour la consommation différée, voire la vente. Le commerce local à petite échelle (à proximité du lieu d'habitation du pêcheur ou au sein du grand bourg de Maripasoula) a en effet eu tendance à se développer, bien que la pêche vivrière ait gardé une place généralement prépondérante surtout chez les Amérindiens (Longin 2016). Ainsi par exemple, la pêche à la nivrée, une autre méthode traditionnelle utilisant la sève de la liane *hali-hali* contenant un principe actif similaire à la roténone (Meunier 2004), s'est vue transitoirement mobilisée de façon plus intensive pour la pêche commerciale, avant qu'un moratoire imposé par les chefs coutumiers de certaines communautés locales n'en limite l'usage.

Ces modifications de pratiques illustrent l'évolution des modes de vie à la frontière entre préservation de la culture traditionnelle (comme la pêche vivrière), et les promesses apportées par la modernité. La préservation des ressources biologiques, naturelles et culturelles du Haut-Maroni (et du sud de la Guyane en général) reste essentielle, tant pour des raisons patrimoniales qu'écologiques. Ceci est d'autant plus vrai que les zones géographiques concernées sont également impactées par des activités anthropiques difficiles à suivre et à quantifier, notamment l'orpaillage illégal. C'est dans ce contexte que fut créé, en 2007, le Parc Amazonien de Guyane.

I.3. Le Parc Amazonien de Guyane et les acteurs de l'eau en territoire guyanais

Le Parc Amazonien de Guyane (PAG) couvre un espace compris entre une ligne recoupant les communes de Grand-Santi, Saül et Camopi au Nord, la frontière avec le Brésil à l'Est et au Sud, et la frontière avec le Suriname à l'Ouest (Fig.1). De par la superficie de sa zone de cœur (20 300 km²) et de son aire d'adhésion (13 600 km²), il représente la plus grande zone

⁵ Une trappe est un dispositif de piégeage avec appât, dédié à la capture de grandes espèces de poissons du fleuve.

protégée de l'Union Européenne. Il définit ses principaux objectifs autour d'enjeux relatifs à la préservation du patrimoine naturel et la pérennisation des activités traditionnelles des communautés humaines locales. Ainsi, le PAG tente de mener à bien ses actions en les co-construisant avec les communautés concernées et les différents acteurs de l'environnement. Le PAG est également investi d'une mission scientifique, en collaboration avec différents organismes de recherche, afin d'étudier l'évolution des ressources naturelles et la biodiversité présentes sur son territoire. Par exemple, le PAG et ses partenaires ont procédé à divers travaux de description des usages des ressources naturelles, comme la chasse, la cueillette de l'Arouman⁶, ou encore la dynamique des pratiques agricoles. Concernant les milieux aquatiques, le PAG a défini des orientations de recherche qui rejoignent le programme du SDAGE de 2000: établir un diagnostic de l'état actuel des ressources piscicoles en vue d'une gestion adaptée. L'orpaillage illégal et les enjeux de santé publique associés (Fréry et al. 2001; Durrieu et al. 2005) d'une part, et la surpêche d'autre part, sont identifiés comme les pressions ayant le plus fort impact potentiel sur la ressource naturelle dulçaquicole, constituant ainsi des thématiques scientifiques prioritaires.

En ce qui concerne l'exploitation des ressources halieutiques, plusieurs travaux menés entre le milieu des années 1960 (Hurault 1965) et le début des années 2000 (Richard-Hansen 2002) avaient permis une première description des pratiques de pêche. Ces travaux ont également permis d'approcher la part d'alimentation quotidienne occupée par le poisson, et ce chez plusieurs communautés autochtones. Cependant, les modifications plus récentes des pratiques de pêche au sein de ces communautés, conjuguées au sentiment d'une baisse générale de la qualité des captures exprimé par les pêcheurs (Fleury 1998, Martin 2014), rendaient nécessaire un nouvel état des lieux et une appréhension plus approfondie des caractéristiques socio-écologiques de la pêche.

A cette fin, un projet de description précise (1) des caractéristiques sociologiques et pratiques de la pêche artisanale, comparées dans l'espace et entre communautés et (2) de l'état écologique des populations des espèces de poissons d'intérêt halieutique a été engagé. Ce double-objectif a fait l'objet du « programme Pêche », initié en 2013 par le PAG et mené en collaboration avec le bureau d'études Hydreco Guyane et l'INRA autour des principaux villages autochtones du Haut-Maroni. Le premier objectif fit plus précisément l'objet d'un stage de Master 2, mené en formation continue par Guillaume Longin (PAG) dont la très grande connaissance de terrain lui a permis une immersion particulièrement informative dans le quotidien des pêcheurs et leur perception de l'activité de pêche artisanale.

Ce travail a apporté de nombreux éclairages sur les pratiques de pêche sur le HM. En résumé :

- Les engins de pêche utilisés ont bel et bien évolué, devenant plus efficaces et moins sélectifs, de manière généralisée pour l'ensemble des villages du HM. Le filet tramail monofilament devient l'engin majoritairement mobilisé, pouvant être posé au crépuscule et relevé à l'aube.
- Les sorties de pêches s'effectuent en grande majorité dans un périmètre restreint autour des lieux habités, de 5 à 10 km alentours. En revanche, elles sont plus étalées dans

⁶ L'arouman est une plante de la famille des Marantacées largement utilisée par les communautés de Guyane pour la confection de vanneries.

l'année qu'autrefois, devenant plus nombreuses en saisons des pluies notamment, la pose de filet étant possible à toute période de l'année contrairement, par exemple, à la pêche à la nivrée plus efficace en période d'étiage.

➤ Malgré un effort de pêche généralement corrélé au nombre de foyers présents dans la zone d'habitation considérée ou encore à sa finalité commerciale, le rendement par unité d'effort (en termes de biomasse ou de captures, toutes espèces confondues) ne suit pas systématiquement la même tendance et varie suivant les localités étudiées, certaines zones aval devant déployer un effort de pêche plus important pour un nombre égal de captures. Les données disponibles ne permettaient cependant pas, à ce stade, d'expliquer ce phénomène par une éventuelle atteinte biologique à la ressource halieutique.

➤ Malgré tout, la quantité de poisson disponible à la consommation par habitant est considérée en diminution au sein des communautés Amérindiennes et Bushinengués depuis une cinquantaine d'années.

➤ Des entretiens semi-directifs auprès des pêcheurs ont, entre autres, confirmé le ressenti d'une baisse de la ressource halieutique disponible.

Le rôle de l'approche biologique est de compléter les observations fournies par cette première partie du programme Pêche. Il s'agit de fournir une première visualisation de la vulnérabilité des populations de poissons du Haut-Maroni, notamment au niveau des zones géographiques associées à une forte pression de pêche, c'est-à-dire les zones d'habitation.

Le second objectif du programme Pêche vise donc à approfondir les connaissances disponibles quant à l'état écologique de ces populations, pour lesquelles des données d'abondance ne sont actuellement pas accessibles. Une étude sur la diversité génétique des populations de plusieurs espèces d'intérêt a donc été programmée et fait l'objet du présent travail de thèse. A partir d'échantillons biologiques prélevés sur la partie guyanaise du HM, elle a pour but de décrire la variabilité génétique intra- et interpopulations chez ces différentes espèces afin de mettre en évidence, le cas échéant, l'influence de certains facteurs anthropiques ou naturels (capacité de dispersion de l'espèce, fragmentation de l'habitat, exploitation halieutique) structurant les patrons de dispersion et pouvant avoir un effet sur le potentiel adaptatif des espèces.

Introduction:

II. Contexte scientifique et état de l'art

II.1. Biodiversité et conservation des écosystèmes aquatiques néotropicaux : le cas de la Guyane française

II.1.1. Les écosystèmes d'eau douce : rôles, menaces et conservation

Les milieux dulçaquicoles représentent 1% de la surface de la planète Terre, mais abritent à eux seuls 6% des espèces animales et un tiers des Vertébrés connus à ce jour (Strayer & Dudgeon 2010). Ils fournissent de nombreux services écosystémiques indispensables à l'humanité sur le plan alimentaire (approvisionnement en eau de consommation, ressources halieutiques), socio-économique et culturel. L'implication des activités humaines dans le déclin de certaines espèces commença à être reconnu dans le monde occidental au cours du 19^{ème} siècle (e.g. Lyell 1832) et de premiers constats d'atteinte à la biodiversité⁷ des milieux aquatiques ont été reportés dès le début du 20^{ème} siècle (e.g. Ortmann 1909; Hasler 1947). Cependant, la recherche dédiée à leur préservation ne s'est clairement établie comme sous-discipline scientifique distincte qu'avec l'avènement de la biologie de la conservation au cours des années 1980 (Wilson 1985). Aujourd'hui, l'impact des activités humaines sur l'évolution de la biodiversité aquatique est largement reconnu : changement climatique, destruction, dégradation ou fragmentation des habitats naturels, surexploitation des ressources, domestication, invasions biologiques suite à l'introduction d'espèces non-natives sont autant de facteurs susceptibles de fragiliser les écosystèmes et les populations naturelles, *modulo* leurs capacités de résistance⁸ ou de résilience⁹. Ainsi, la biodiversité aquatique continue de décroître de manière intense à ce jour (Collen et al. 2014) et les mesures de conservations efficaces sont plus que jamais nécessaires pour enrayer ce processus. L'observatoire mondial GEO et son réseau d'observation (GEO BON) ont défini un ensemble de variables essentielles pour décrire de la biodiversité (*essential biodiversity variables*, EBV, Pereira et al. 2013). Turak et al. (2016) ont souligné l'importance toute particulière de certaines classes d'EBV au regard des écosystèmes aquatiques. Pourtant, Jucker et al. (2018), en mesurant l'effort de recherche en biologie de la conservation via la littérature scientifique produite au cours de la décennie écoulée, soulignent un défaut d'investissement envers les écosystèmes aquatiques continentaux comparativement aux compartiments terrestres ou marins. Selon leurs calculs, les écosystèmes marins seraient crédités de 72% de publications de plus que leurs pendants dulçaquicoles.

Les milieux aquatiques néotropicaux - y compris guyanais, en dépit de leur biodiversité restée longtemps préservée, sont également soumis à des contraintes anthropiques croissantes (Barletta et al. 2010; Allard 2014; Reis et al. 2016). Ils ont néanmoins bénéficié d'une considération particulièrement tardive en ce qui concerne la préservation de leur biodiversité (Castello et al. 2013 ; Reis et al. 2016). Pourtant, Antunes et al. (2016) suggèrent que le bassin Amazonien aurait davantage souffert de déclin de richesse spécifique dans ses

⁷ Le terme de biodiversité recouvre la variabilité biologique à différents niveaux d'organisation, des écosystèmes aux gènes. Sa définition officielle, selon l'article 2 de la Convention sur la diversité biologique (1992) est la suivante : « La variabilité des êtres vivants de toute origine incluant entre autres, les écosystèmes terrestres et aquatiques et les complexes écologiques dont ils font partie : cela comprend la diversité au sein des espèces, ainsi que celle des écosystèmes. »

⁸ Résistance : Capacité d'un compartiment biologique à conserver son état initial en dépit d'une situation de stress consécutif à une contrainte ou une perturbation extérieure.

⁹ Résilience : Capacité d'un compartiment biologique à s'adapter, en modifiant son état d'équilibre, à une situation de stress consécutif à une contrainte ou une perturbation extérieure.

milieux aquatiques que dans les forêts adjacentes. Les travaux d'évaluation du statut des écosystèmes dulçaquicoles néotropicaux sont d'autant plus complexes à mettre en place que, par rapport aux milieux tempérés, la diversité des habitats et des espèces est très forte et le déploiement d'infrastructures scientifiques et techniques, nécessaires à leur étude, est limité par le contexte socio-économique des aires concernées (Barletta et al. 2010). Dans bien des cas, les pays concernés abritent des populations humaines, y compris des ethnies autochtones, qui dépendent fortement des ressources apportées par les milieux aquatiques. L'importance de tenir compte de telles spécificités pour l'élaboration de programmes de suivi et pour l'encadrement de l'usage des ressources représente donc un défi indispensable.

Sur le territoire français, l'application de mesures de conservation des milieux aquatiques dans les territoires d'outre-mer nécessite également de tenir compte des spécificités géographiques et socio-économiques de ces zones (c.f. Section I.1. pour l'exemple de la pêche en eau douce en Guyane). Rappelons en outre l'écart qui existe entre les connaissances biologiques disponibles sur l'ichtyodiversité en Guyane par rapport à celle de France métropolitaine, moins riche mais suivie depuis de nombreuses décennies. En Guyane, on dénombre actuellement 367 espèces de poissons strictement dulçaquicoles (Le Bail et al. 2012) contre 103 en France métropolitaine (Keith et al. 2011), avec une superficie du département guyanais presque sept fois inférieure. Or, les outils de diagnostic et de suivi des populations et des écosystèmes, parfois appliqués de façon quasi-routinière en métropole, sont difficiles à mobiliser dans les milieux tropicaux.

II.1.2. Diversité et répartition des espèces néotropicales

La biodiversité sur Terre ne se répartit pas de façon uniforme (Gaston 2000). L'intervention de différents « filtres », à des échelles spatiales et temporelles variables et d'origine biotique ou abiotique, influencent la distribution des habitats, des espèces ou des groupes fonctionnels (Devictor et al. 2010; Mouillot et al. 2013; Toussaint et al. 2016). Ils agissent aussi sur d'autres niveaux de diversité, comme la diversité des traits histoires de vie inter- (Pecuchet et al. 2018) ou intra-spécifiques (Bonfanti et al. 2018), ou encore la diversité génétique (Miraldo et al. 2016).

La région néotropicale est une région biogéographique correspondant à l'un des huit écozones terrestres tels que définis par Miklos Udvardy en 1975, et qui regroupe les aires géographiques de l'Amérique centrale et du Sud, des Antilles et des îles Galapagos. Elle s'est peu à peu distinguée des autres écozones au cours de la séparation des continents Sud-américain et Africain à la fin du Crétacé inférieur, il y a 100 millions d'années (Lundberg et al. 1998; Brito et al. 2007). Plusieurs événements géologiques successifs, survenus entre le Crétacé supérieur et le Miocène inférieur, se sont accompagnés d'alternances glaciaires et interglaciaires à l'origine de la formation de refuges (Gentili 1949; Haffer 1982; Haffer 1997) ayant permis les phénomènes de vicariance et de radiation évolutive qui expliquent la large richesse taxonomique (Brito et al. 2007) caractérisant aujourd'hui cette zone.

Notamment, la diversité ichtyque de la région néotropicale est la plus riche à l'échelle de tous les continents (Lévêque et al. 2008; Oberdorff et al. 2011). Reis et al. (2003) rapportaient un total de 4475 espèces dulçaquicoles décrites et estimaient que plusieurs centaines restaient encore à découvrir. En 2012, le nombre d'espèces décrites a été porté à 5200 (Reis

2013) et par ailleurs, Vari et Malabarba (1998) et Schaefer (1998) avançaient déjà une estimation supérieure à 8000 de la richesse spécifique des poissons d'eau douce Néotropicaux. Ce chiffre correspondrait à 25% de l'ichtyodiversité mondiale décrite, à la hauteur de ce qui, mieux connu du grand public, est observé chez les poissons marins au niveau des écosystèmes de récifs coralliens tropicaux (McAllister, 1991).

Lundberg et al. (1998) ont décrit les processus géologiques successifs ayant délimité les principales grandes régions et rivières Néotropicales les unes des autres : parmi elles, le bouclier guyanais constitue une unité géologique particulière (Fig.3), physiquement bien séparée du bouclier brésilien depuis la période médiane du Précambrien (108-109 Ma) par le lit dans lequel s'écoule aujourd'hui le fleuve Amazone. Le bouclier guyanais recouvre le Guyana, le Suriname et la Guyane française, ainsi qu'une partie du Venezuela, de la Colombie et du Brésil. Délimité à l'Ouest et au Sud par les deux importants bassins versants que sont l'Orénoque et l'Amazone, le plateau guyanais est considéré comme un ancien refuge de biodiversité, ce qui expliquerait par exemple les phénomènes de diversification chez certaines familles de poissons (Boujard & Tito de Morais 1990; Boujard et al. 1997). En outre, le plateau guyanais représente une zone de richesse spécifique et d'endémisme ichtyologique particuliers, y compris dans le contexte néotropical (Abell et al. 2008; Covain et al. 2012), 23% des espèces de poissons d'eau douce décrites sur l'écozone y étant représentées d'après l'inventaire le plus récent fourni par Vari et al. (2009). L'ichtyodiversité ne s'y répartit pas de façon homogène : plusieurs écorégions y sont délimitées (Abell et al. 2008) et les aires biogéographiques significatives y ont été redéfinies au cours de travaux récents (Lemopoulos & Covain 2018).

La faune ichtyque de Guyane, quant à elle, s'est vue documentée à partir de la fin des années 1940 (Puyo 1949) et de façon beaucoup plus approfondie à partir de la fin des années 1970, devenant au fil des années l'une des mieux décrites d'Amérique du Sud. Les nombreuses expéditions menées par l'ORSTOM (ex-IRD), l'INRA et le MNHN ont permis l'élaboration des inventaires et guides de détermination (Planquette et al. 1996; Keith et al. 2000; Le Bail et al. 2000 ; Le Bail et al. 2012) ayant servi de support à la plupart des travaux portant sur les relations entre la faune ichtyque et son milieu. Les assemblages d'espèces ichtyques sont différemment composés de part et d'autre du fleuve Kourou, qui scinde le territoire guyanais entre deux portions Est et Ouest (Le Bail et al. 2012). Lemopoulos et Covain (2018) définissent plus précisément trois blocs biogéographiquement distincts. L'un des blocs comprend les bassins du Maroni et de la Mana, qui couvrent une large superficie et abritent une richesse spécifique élevée (Le Bail et al. 2012; Mol et al. 2012). Des connexions anciennes avec d'autres bassins adjacents, notamment l'Amazone via le tributaire de la Jari, auraient favorisé des échanges d'espèces. Les autres bassins guyanais, plus petits, ont vraisemblablement été colonisés par les estuaires lors des reculs épisodiques du niveau des mers au cours des temps géologiques. En Guyane française, deux tiers des 367 espèces dulçaquicoles actuellement décrites peuvent être rencontrées dans le bassin versant du Maroni. Le Maroni abrite par ailleurs 16.9% d'espèces endémiques, soit 50% des espèces endémiques du département (Le Bail et al. 2012).

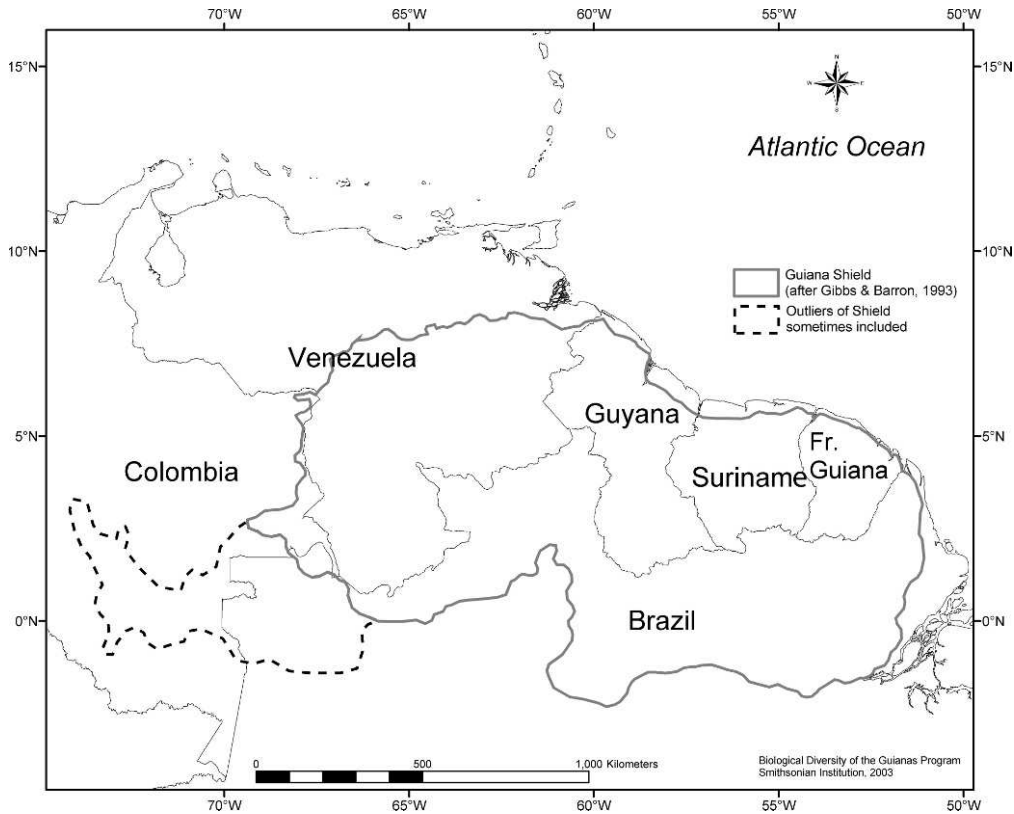


Figure I.3 : Carte générale du bouclier guyanais, issue de Vari et al. (2009).

L'influence des facteurs hydrogéographiques et environnementaux remplace peu à peu celle des filtres biogéographiques lorsque l'on passe de l'échelle régionale à celle du bassin versant. On distingue deux grandes catégories de milieux aquatiques présentant des fonctionnements différents en Guyane Française (Boujard 1992) : les eaux courantes (les fleuves et leurs affluents ainsi que les criques adjacentes) ; et les eaux stagnantes (comme les lacs, les plaines savanes inondées ou encore les marécages ou « pripri »). En ce qui concerne les eaux courantes, la zonation longitudinale varie légèrement par rapport à celle que l'on connaît au niveau des cours d'eau en zone tempérée, les niveaux d'élévation étant faibles en Guyane française. Sur le fleuve Sinnamary, Boujard (1992) distingue le cours principal, large et ouvert, souvent turbide et avec un fort étalement vers les forêts inondées lors de la saison des pluies. Il distingue également des cours secondaires (ou affluents principaux) de largeur intermédiaire ainsi que des zones de « sauts » - associées aux rapides, et les criques (petits cours d'eau étroits, peu profonds, translucides mais teintés d'humus forestier). Les criques constituent une proportion importante des petits cours d'eau guyanais qui alimentent les lits principaux. Ces différents types de milieux aquatiques se retrouvent sur l'ensemble des bassins versants du territoire et s'associent souvent à des communautés ichthyques plus ou moins spécialisées. Les assemblages de poissons de grands cours d'eau sont, par exemple, distincts de ceux des criques et des petites masses d'eau (Tejerina-Garro et al. 2006; De Mérona et al. 2012; Cilleros et al. 2017). Les têtes de bassins peuvent aussi présenter des communautés spécifiques, par exemple au niveau des altitudes plus élevées (Brosse et al. 2013). Enfin, l'influence de variables plus localisées dans l'espace (comme la largeur et profondeur des cours d'eau, la nature des sédiments, le pH, le couvert forestier adjacent ou la variabilité des micro-habitats benthiques) intervient conjointement (bien qu'à

moindre magnitude) à ces facteurs de plus grande échelle (Tejerina-Garro & De Mérona 2001; Cilleros et al. 2017).

II.1.3. Préservation de la diversité ichtyque des fleuves guyanais

II.1.3.1. Qualité d'habitat et richesse spécifique ou fonctionnelle

En Guyane, les impacts de la fragmentation des habitats (notamment suite à l'implantation de barrages hydroélectriques) et de la dégradation des milieux (notamment suite à l'exploitation forestière et les activités d'orpaillage) sur la richesse des communautés piscicoles ont été les plus largement étudiés.

Tito de Morais et al. (1995) soulignaient un impact à prévoir de la construction du barrage de Petit-Saut (bassin versant du Sinnamary) sur les communautés présentes, si celle-ci devait modifier la surface de l'écotone rivière-forêt servant de refuge aux stades juvéniles de nombreuses espèces. Des études ultérieures (Ponton & Copp 1997; De Mérona & Vigouroux 2012) ont confirmé la diminution d'abondance, voire la disparition totale de certaines espèces, au sein de la zone réservoir formée suite à l'implantation du barrage. Dedieu et al. (2014) ont mis en évidence des impacts physiques de l'orpaillage à l'échelle du territoire guyanais, visibles à travers la nature du sédiment. Laperche et al. (2014) décrivent les différentes concentrations en mercure total dans les sédiments et démontraient que celles-ci étaient plus élevées au niveau de sites impactés par l'orpaillage. La biomagnification du mercure le long des réseaux trophiques avait été déjà observée quelques années auparavant (Maury-Brachet et al. 2006) chez plusieurs espèces de poissons, avec notamment une plus forte concentration du mercure sous sa forme organique méthylée (et toxique) chez les espèces de haut niveau trophique, comme les omnivores et piscivores. Enfin, Allard et al. (2016) ont montré que l'activité d'orpaillage et, dans une moindre mesure, de déforestation à petite échelle pouvait entraîner une perte de richesse taxonomique et renforcer la prédominance d'espèces opportunistes. Les développements méthodologiques visant à optimiser et faciliter au maximum les collectes de données *in situ* persistent. Ils tentent par exemple d'optimiser des protocoles non-invasifs comme la pêche électrique en remplacement de l'échantillonnage par roténone (Pottier et al. *en préparation*) - bien que son efficacité se restreigne jusqu'à présent à certaines conditions particulières (Allard et al. 2014). L'ADN environnemental (Taberlet et al. 2012) a également été présenté comme une piste pertinente pour dégrossir les patrons de diversité piscicole, à échelle régionale, dans les cours d'eau guyanais (Cilleros et al. 2018).

L'étude de la composition des assemblages ichtyques a servi de support au développement préliminaire de bio-indicateurs de la qualité écologique des cours d'eau guyanais, inspirés des indices d'intégrité biotique (IBI, Karr 1981) appliqués en milieux tempérés. Au cours de son travail de thèse, Allard (2014) s'est attaché à développer un indice multi-métrique adapté à l'ichtyofaune des petits cours d'eau, permettant de discriminer des sites vierges de sites impactés par les activités d'orpaillage et/ou de déforestation de petite échelle. Tejerina-Garro et al. (2006) avaient déjà mené une réflexion similaire en considérant, eux, les cours d'eau principaux. Leur indice avait permis de distinguer les modifications des communautés ichtyques consécutives à l'implantation du barrage de Petit-Saut.

Evaluer la relation entre pression de pêche et composition des communautés semblerait particulièrement pertinent pour des pêcheries multispécifiques comme les pêcheries guyanaises et, plus généralement, les pêcheries tropicales y compris en eau douce (Welcomme et al. 1999). En outre, de telles approches vont dans le sens d'une vision écosystémique de l'impact de la pêche. Cependant, le suivi direct de la composition spécifique des assemblages a jusqu'ici difficilement permis de détecter, sans ambiguïté, l'influence de la pression de pêche (Rochet & Trenkel 2003; Kantoussan et al. 2018). Compte-tenu de la complexité des milieux aquatiques de Guyane, appréhender l'impact de la pression de pêche par le biais de l'étude des assemblages est quasiment impossible. En effet, bien que prometteuse, l'étude des assemblages spécifiques et la constitution d'indices biotiques en milieux néotropicaux représentent toujours aujourd'hui un véritable défi méthodologique. En milieux tempérés où la richesse spécifique est plus faible et les connaissances biologiques sur les espèces beaucoup plus poussées (bien souvent jusqu'à la variabilité intra-spécifique de leurs traits d'histoire de vie), il est possible de tenir compte de ces informations dans la construction et le calcul des indices d'intégrité biotique. Dans les écosystèmes néotropicaux, une telle résolution est inenvisageable du fait de nos connaissances biologiques encore limitées pour la plupart des nombreuses espèces de ces cours d'eau très diversifiés. Les métriques constituant les indices restent donc souvent de nature qualitative et nécessitent une comparaison directe entre des sites géographiques perturbés ou intacts (sites de référence)¹⁰. En outre, la capacité des indices biotiques à discriminer des sites subissant des pressions multifactorielles (par exemple, une action conjointe de la pression d'orpillage et de la fragmentation des habitats sur la bioaccumulation du méthylmercure, comme suggéré par Boudou et al. 2005) n'a pas été éprouvée, et leur capacité à détecter d'autres types d'anomalies, comme par exemple la réminiscence d'anciennes pêches à la nivree, reste également à implémenter (Allard 2014).

Des indicateurs indirects, basés sur les caractéristiques des captures débarquées (e.g. composition spécifique, tailles des individus, ratio proies/prédateurs) ont été proposés pour évaluer l'influence de la pression de pêche sur les communautés ciblées. Le suivi de tels indicateurs nécessite toutefois que les données de captures soient disponibles de façon robuste et sur long-terme, pour les pêcheries auxquelles l'on s'intéresse. Nous pouvons citer un exemple, quoique pas uniquement axé sur des considérations écologiques, de l'utilisation d'indicateurs pour évaluer la durabilité relative de certaines flottilles de pêche pour lesquelles l'IFREMER disposait de séries temporelles de captures (initiées en 2006) en zone côtière guyanaise. Une approche Rapfish¹¹ (Pitcher & Preikshot 2001) a en effet permis de comparer la durabilité écologique, sociale, économique et technologique de ces différentes flottilles (Cissé et al. 2014). Les auteurs avaient choisi d'intégrer à leur analyse, parmi les indicateurs de type écologique, la taille moyenne des spécimens capturés, leur niveau trophique moyen ainsi que son évolution au cours du temps, le coefficient de variation du

¹⁰ En effet, ceci permet une calibration spatiale directe entre une situation « attendue » et observée et donc, de s'affranchir de poser des hypothèses sur la réponse biologique ou physiologique attendue de différentes espèces face à une contrainte environnementale donnée (Pont et al. 2006).

¹¹ La méthode Rapfish s'appuie sur une analyse multidimensionnelle regroupant des flottilles de pêche distinctes en tant qu'individus statistiques, et divers indicateurs (économiques, écologiques, techniques, etc) en tant que variables dont les modalités sont associées à des scores (de 0 à 10 en fonction de leur « qualité », élevée à basse). Ex. : indicateur écologique de type évolution du niveau trophique au cours des cinq dernières années: 0-3 (faible évolution), 4-6 (évolution intermédiaire), 7-10 (évolution forte, ex. diminution de la proportion de top-prédateurs parmi les captures).

rendement en captures entre le début et la fin des séries temporelles, et enfin la quantité de captures accessoires. L'analyse n'a pas permis de montrer de différences significatives entre flottilles dans les scores de ces indicateurs, essentiellement parce que ces flottilles exploitaient des stocks relativement homogènes le long des côtes guyanaises. Elle fournit cependant un bon exemple de méthode de suivi de pêche multispécifique, applicable lorsque des données sont disponibles de manière suffisante et standardisée, ce qui n'est actuellement pas le cas sur le Haut-Maroni. Nous pouvons noter, à titre informatif, qu'une comparaison qualitative et préliminaire des données de captures obtenues par Longin (2016) à celles obtenues quelques années plus tôt (Richard-Hansen 2002), sur le village d'Antecume-Pata en saison sèche, ne suggérait aucune différence notable dans leur composition spécifique.

Retenons donc, ici, que l'étude de la composition des communautés dans les cours d'eau guyanais pour détecter leur réponse face à des contraintes anthropiques nécessite d'importants investissements, logistiques et analytiques, pour obtenir des données en quantité suffisante et alimenter des modèles écologiques pertinents. De telles approches ne sont, tout du moins à l'heure actuelle, que difficilement applicables en routine. Sur le Haut-Maroni, l'absence de descripteurs des paramètres physico-chimique locaux et de la diversité des habitats naturels, la présence de la zone urbanisée de Maripasoula, ou encore les connaissances limitées du cycle de déplacement des différentes espèces exigeraient une phase d'acquisition de données particulièrement laborieuse pour pouvoir distinguer l'influence de la pêche sur les communautés.

En outre, afin de comprendre les mécanismes biologiques qui sous-tendent l'évolution spatio-temporelle des communautés, il est important d'avoir connaissance des traits d'histoire de vie spécifiques, du moins dans leurs grandes catégories.

II.1.3.2. Les principales caractéristiques biologiques des espèces : identifier des espèces plus vulnérables

Sources d'énergie trophique et hydrologie des fleuves néotropicaux

Différentes stratégies permettent aux espèces vivantes d'assurer leurs principales fonctions de survie et de reproduction. Ces stratégies sont la résultante de compromis (*trade-offs*) énergétiques entre différents traits physiologiques dont l'issue peut fortement dépendre des caractéristiques ou de la stabilité spatiale et temporelle des habitats (Southwood 1977). Les écosystèmes aquatiques néotropicaux, d'une manière générale, présentent deux particularités.

Tout d'abord, ce sont souvent des milieux ayant une faible production primaire endogène, la faible charge en minéraux constituant un facteur limitant (Goulding 1980). Ainsi, la végétation rivulaire représente une importante part de la production primaire alimentant les fleuves guyanais via les feuilles, fruits ou graines qui peuvent soit tomber à la surface de l'eau et être alors consommés par les espèces aquatiques¹², soit être directement consommés par

¹² Dans certains systèmes comme le bassin Amazonien, on constate même un phénomène d'ichtyochorie, c'est-à-dire la dispersion passive, par les poissons granivores, d'espèces végétales rivulaires (Correa et al. 2015, *Biol. Reviews* 90(4), 1263-1278).

les poissons lors de la montée des eaux qui favorise le repli des individus auprès des rives (inondation de portions de forêt associées), leur donnant ainsi accès à cette source de nourriture (Goulding 1980; Méricoux et Ponton 1998; De Mérona et al. 2008). Le piranha herbivore *Myloplus rhomboidalis* possède des rangées de dents molariformes, adaptées au broyage de ces fruits et graines. Outre cette production primaire principalement exogène, la végétation rivulaire fournit également une source de nourriture de niveau trophique plus élevé, notamment les insectes ou petits mammifères tombant régulièrement dans l'eau. Cependant, la production primaire endogène existe, notamment au niveau des zones de faible profondeur comme les sauts, où peuvent pousser des végétaux aquatiques tels que les herbiers de Podostemacés, une source de nourriture importante pour certaines espèces de Myleinae (piranhas herbivores) (Meunier 2004). Le pakou *Tometes lebaili* s'est spécialisé sur cette source d'alimentation. Ses dents tranchantes lui permettent de brouter les herbiers de Podostemacées, dont les pêcheurs Amériendiens se servent souvent comme appâts. En outre, certaines espèces de poissons strictement rhéophiles peuvent se nourrir du biofilm algal qui se développe sur les roches. C'est le cas de nombreuses espèces de Loricariidae (Lujan et al. 2012), des poissons-chats cuirassés dont la variété d'espèces, de formes et de couleurs est très importante.

Par ailleurs, les écosystèmes Néotropicaux sont des milieux très fortement soumis au cycle d'alternance de saisons sèches (étiage) et de saisons des pluies (crués) (Boujard et al. 1992; Lima de Melo et al. 2009; De Mérona et al. 2012). Ces régimes hydrologiques contribuent à structurer les communautés (Boujard et al. 1992), influencent la disponibilité des différentes ressources alimentaires (e.g. l'accès à la végétation terrestre et aux ressources exogènes étant accru en saison de pluies) et affectent les régimes de reproduction chez la plupart des poissons Néotropicaux (Arrington et al. 2006).

Stratégies de reproduction

Lowe McConnell (1987) fournit une des premières documentations des diverses stratégies de reproduction chez les poissons Néotropicaux. Ces stratégies de reproduction combinent divers traits d'histoire de vie, comme le diamètre moyen des ovocytes, l'âge ou la taille à maturité ou encore la durée de période de reproduction. En Guyane, plusieurs auteurs (Ponton & Tito de Morais 1994; Ponton & De Mérona 1998; Méricoux et al. 2001) ont mis en relation ces diverses caractéristiques phénotypiques avec différentes variables d'habitat du fleuve Sinnamary, afin de visualiser leur évolution suite à la fermeture du barrage de Petit-Saut. Plus précisément, et de manière similaire aux travaux menés au Venezuela par Kirk Winemiller (1989), Méricoux et al. (2001) ont défini trois grandes classes d'espèces. Une première classe correspond à des espèces à fécondité modérée produisant peu de descendants mais pratiquant la garde parentale, espèces généralement associées à des milieux stables ou prévisibles dans le temps (stratégie « équilibrée », Winemiller 1989). Une seconde classe correspond à des espèces généralement de petite taille, à fécondité modérée et ne pratiquant pas la garde parentale, mais présentant par contre une propension à se reproduire tout au long de l'année et à recoloniser facilement des milieux après une perturbation (stratégie « opportuniste », Winemiller 1989). Une troisième et dernière classe correspond à des espèces généralement de plus grande taille produisant de nombreux

(fécondité élevée) ovocytes de faible diamètre au moment de l'année le plus favorable, bien souvent en cours de saison des pluies (stratégie « periodic-strategist », Winemiller 1989).

Comportements de dispersion ou de migration

On rencontre, dans la plupart des fleuves néotropicaux et dans le Haut-Maroni, différentes stratégies de dispersion active (comportement migrateur). Certaines espèces effectuent des migrations de « longue-distance ». Le plus souvent, il s'agit des espèces à fécondité élevée, qui se reproduisent généralement en groupe durant la saison des pluies où la disponibilité alimentaire est plus forte, et produisent de nombreux ovocytes de petite taille (Carolsfeld et al. 2004) - c'est le cas par exemple de la plupart des Pimelodidae ; qui sont parmi les plus grandes espèces que l'on peut rencontrer en Guyane française, de la plupart des Characidae de grands cours d'eau ou encore de certains Anostomidae. Certaines espèces effectuent des migrations de plus courte distance (généralement du cours principal au cours secondaire adjacent via des migrations qualifiées de « latérales », ou bien vers les petits cours d'eau amont). Ces migrations sont souvent dédiées à la reproduction de ces espèces dont la ponte, de taille intermédiaire, est généralement déposée sur un substrat ou au creux de la végétation rivulaire - c'est le cas de nombreux Serrasalminidae (piranhas et pacus) (Planquette et al. 1996). Enfin, les espèces à fécondité faible, souvent inféodées certains types d'habitats particuliers, restent généralement sédentaires et investissent leur énergie dans la production puis la maintenance d'œufs de grande taille et des larves, plutôt que dans la dispersion (Comte & Olden 2018a) - c'est le cas par exemple des Loricariidae et des Cichlidae (Keith et al. 2000; Le Bail et al. 2000).

Facteurs de vulnérabilité des espèces

Certaines espèces, de par leur distribution géographique restreinte (endémisme voire micro-endémisme), leur niche écologique étroite, leur faible fécondité ou encore leur capacité de dispersion restreinte, présentent une vulnérabilité particulière face aux contraintes environnementales naturelles ou anthropiques. Au niveau local, c'est ce qui peut expliquer qu'une perturbation environnementale s'accompagne de modifications de la structure fonctionnelle d'une communauté par l'élimination préférentielle de certaines espèces associées à des traits particuliers (comme observé en Guyane par Allard et al. (2016), consulter également Liu et al. (2016) pour une méta-analyse globale). En ce qui concerne l'exploitation des espèces, Winemiller (2005) souligne l'intérêt de tenir compte des différentes stratégies reproductives dans les tentatives de modélisation stock-recrutement pour le suivi démographique des populations impactées, en soulignant notamment la plus faible résilience démographique des taxons de stratégie « équilibrée » (e.g. investissement reproducteur important, ovocytes de grande taille, soins parentaux). Une maturité tardive chez des espèces de grande taille peut également représenter un facteur de vulnérabilité, notamment si les individus entrent dans la phase exploitable d'une pêcherie (i.e. section du cycle de vie pour laquelle les spécimens peuvent être exploités par cette pêcherie), alors qu'ils sont encore immatures et n'ont donc pas encore eu la possibilité de se reproduire.

Jusqu'ici, les travaux récents d'évaluation menés au Brésil, en Uruguay et Argentine ainsi qu'en Colombie, ont mené à classer sur les listes UICN une proportion de 4 à 10% des espèces de poissons d'eau douce évaluées comme étant vulnérables (VU), menacées (EN) ou en

danger critique d'extinction (CR) (Reis et al. 2016). Depuis peu, la Guyane française possède également sa propre liste rouge des espèces menacées (Allard et al. 2017) sur laquelle figurent 45 espèces de poissons d'eau douce, essentiellement de petits cours d'eau. Les critères principaux ayant mené à l'attribution d'un statut vulnérable, voire menacé, à certaines espèces sont très majoritairement liés à leur endémisme. C'est le cas de nombreuses petites espèces du genre *Hartiella* (Loricariidae) dont beaucoup ne sont rencontrées que dans un seul massif montagneux (comme par exemple le Mont Itoupé à l'extrême Sud de la Guyane). L'espèce *Tometes lebailli*, très emblématique et très recherchée par les pêcheurs du Haut-Maroni dont elle est également endémique, a été placée sous le statut « Quasi-menacé » de par sa spécialisation alimentaire sur les herbiers de Podostemaceae, eux-même fortement dépendants de la qualité du milieu.

II.1.3.3. Traits d'histoire de vie et dynamique des populations

Le suivi des populations sauvages et de leur vulnérabilité face aux pressions anthropiques au sens large peut s'effectuer à l'aide de différents outils analytiques et techniques (Paris et al. 2018, c.f. Annexe II; Gordon et al. 2018).

Ces outils peuvent permettre de suivre la distribution géographique locale ou globale de ces populations (Guisan & Zimmermann 2000; Elith et al. 2006; Grenouillet et al. 2011) et de la mettre en relation avec divers facteurs agissant à des échelles variables, comme ceux liés aux changements climatiques (Cheung et al. 2008; Comte et al. 2013; Petitgas et al. 2013). Ces outils peuvent également permettre d'étudier l'évolution temporelle des abondances (dynamique des populations), que ce soit globalement (Begon et al. 1996) ou bien en distinguant la contribution relative de différentes classes d'âge (Leslie 1945) ou de taille (Lefkovitch 1965). Les outils de modélisation de dynamique de populations peuvent intégrer des paramètres tels que la survie, le succès reproducteur, ou encore la vitesse de croissance des individus. Ces paramètres de dynamique de population peuvent être plus ou moins sensibles à différents facteurs locaux ou globaux, biotiques ou abiotiques, naturels ou anthropiques.

En France métropolitaine, il existe des suivis de long terme des abondance de certaines populations. Ces estimations sont permises grâce à des suivis par pêche électrique (Poulet et al. 2011; Montorio et al. 2018) ou encore via les données de capture collectées par les pêcheurs (par exemple, la déclaration des captures de Saumon atlantique auprès de l'ONEMA¹³ est obligatoire depuis 1987). Des modèles de dynamique des populations peuvent aider à la gestion des espèces exploitées tant en milieu marin (Fromentin & Powers 2005 pour le thon rouge *Thunnus thynnus*; Archambault et al. 2018 pour la sole commune *Solea solea*) qu'en eau douce (Boulenger et al. 2016 pour l'anguille européenne *Anguilla anguilla*). L'estimation des paramètres démographiques nécessaire à la construction de tels modèles peut s'effectuer via diverses méthodes comme les protocoles de capture-marquage-recapture (Lebreton et al. 1992; 2009), la télémétrie (Cooke et al. 2013), la reconstitution de généalogies (Jones & Wang 2010) et bien d'autres. Outre l'acquisition de données de dynamique de populations, d'autres types de données intraspécifiques, comme par exemple des données morphométriques illustrant l'évolution de la relation taille-poids des individus

¹³ Office National de l'Eau et des Milieux Aquatiques

au cours du temps, peuvent aussi permettre de visualiser l'effet de l'exploitation sur la condition physique des individus et des populations ciblées (Hutchings & Baum 2005).

En Guyane française, la croissance des individus a été étudiée pour certaines espèces dans une perspective d'élevage en aquaculture (Boujard et al. 1988) ou lorsque des données individuelles étaient disponibles en quantité suffisante. Ceci a notamment permis de définir les courbes de croissance de quelques grandes espèces de fleuves, notamment des Anostomidae et des Serrasalminidae (Boujard et al. 1991; Lecomte et al. 1993; Meunier et al. 1994; Lord et al. 2007), et de constater leur nature discontinue au gré des alternances saisonnières en lien avec des disponibilités alimentaires (saison des pluies : disponibilité alimentaire plus forte, associée également à la saison de reproduction ; saison sèche : période d'« hiver physiologique »). Des courbes de relation taille-poids ont également été tracées pour un nombre d'espèces plus conséquent fréquentant les petits cours d'eau de Guyane (Allard et al. 2015) ou les cours d'eau principaux du Haut-Maroni (Longin 2016). Des descriptions de paramètres liés à la reproduction (fécondité relative, nombre et taille des ovocytes, ovogenèse) sont également disponibles pour certaines espèces (Planquette et al. 1996; Keith et al. 2000; Le Bail et al. 2000). Par contre, les données d'abondance sont inexistantes ou fragmentaires (ex. issue d'observations ponctuelles lors d'inventaires) et il n'existe actuellement aucun suivi de dynamique de populations de poissons d'eau douce guyanais. Ceci s'explique aisément par les contraintes imposées par un tel milieu, et un important déploiement logistique indispensable à l'acquisition de données temporelles (voir Table 1 de Annexe II). Des suivis réguliers d'abondance ou des approches par capture-marquage-recapture par exemple, sont difficilement envisageables en Guyane (Meunier et al. 1994), la surface et la diversité des habitats étant telle qu'il paraît peu réaliste d'en couvrir une proportion suffisante pour assurer des observations ou des recaptures informatives. A titre d'exemple dans une autre zone néotropicale, sur le bassin du Paraná (Brésil), Antonio et al. (2007) ont marqué 6604 individus de différentes espèces afin d'observer leurs mouvements en amont et aval d'un barrage hydroélectrique. Seuls 2,8% ont pu être observés à nouveau après un seul épisode de recapture.

Effectuer des études de vulnérabilité des populations de poissons à travers des suivis d'abondance ou de paramètres démographiques apparaît donc compromis, tout du moins à l'heure actuelle, en ce qui concerne les espèces du territoire guyanais. De telles études sont, d'une manière générale, rares en zone néotropicale, à moins que des enregistrements de captures ne soient disponibles au niveau de pêcheries commerciales (e.g. pour *Arapaima gigas* dans le bassin Amazonien, Castello 2004; Castello et al. 2009). Dans le contexte relativement vierge de données intraspécifiques pour les populations de poissons exploitées du Haut-Maroni, la méthodologie choisie pour mener ce travail de thèse est orientée par une double condition : elle doit pouvoir être menée avec un investissement technique et logistique aussi peu exigeant que possible et fournir une visualisation la plus intégrée possible des facteurs passés et actuels influençant la variabilité intraspécifique des individus.

L'étude de la variabilité génétique, niveau le plus élémentaire de la diversité intraspécifique, peut permettre de répondre à ces objectifs.

II.2. La variabilité génétique comme support de gestion des populations naturelles

La diversité génétique constitue l'un des trois niveaux de la biodiversité. A ce titre, elle porte des enjeux spécifiques, à l'instar des autres niveaux de biodiversité que sont la diversité des espèces ou des écosystèmes, en ce qui concerne la conservation du monde vivant en général. En effet, elle contribue tout aussi bien la capacité de résilience à l'échelle individuelle (Dobzhansky 1965) qu'à l'échelle populationnelle (Frankham 2005) ou même écologique (Crutsinger et al. 2016).

II.2.1. Variabilité génétique, traits d'histoire de vie et environnement

II.2.1.1. Mécanismes de la variabilité génétique et marqueurs moléculaires

Grands principes de la génétique des populations

Les informations portées par la molécule d'ADN, à travers la séquence de ses différents nucléotides, varient entre individus et confèrent à chacun des caractéristiques génétiques, métaboliques et phénotypiques propres. Le polymorphisme génétique à l'échelle de la molécule d'ADN désigne alors les différentes versions (allèles) d'un même locus. Le terme locus désigne l'emplacement d'un marqueur moléculaire, comme un gène, le long d'un chromosome. Un locus est qualifié de polymorphe¹⁴ s'il présente au minimum deux allèles différents observables à une certaine fréquence à l'échelle du groupe d'individus considéré. La molécule d'ADN, dénominateur commun du monde vivant, peut donc présenter du polymorphisme intra-, mais aussi inter-spécifique. L'ensemble des polymorphismes, quelle que soit leur nature, leur fréquence d'observation au sein d'un groupe et la façon dont ils se répartissent entre individus, représente ce que l'on appelle la variabilité génétique. L'étude de la variabilité génétique permet d'approcher les mécanismes successifs à l'origine de la diversification des espèces (phylogénétique), d'analyser les processus démographiques, environnementaux ou comportementaux caractérisant le fonctionnement historique et contemporain des populations (phylogéographie, génétique des populations), ou encore d'anticiper les probabilités de persistance de cette variabilité en elle-même au regard de contraintes extérieures, naturelles ou anthropiques, susceptibles de porter atteinte à son intégrité (génétique de la conservation) (Allendorf et al. 2010).

A l'instar de la diversité spécifique, la diversité génétique peut se décomposer en trois échelles. La diversité intrapopulation (*alpha*) représente la diversité génétique portée par les individus constituant une entité unique et homogène. La diversité interpopulation (*beta*) correspond à la proportion de variation génétique observable (évolution de la composition allélique) lorsque l'on passe d'une population à l'autre. La diversité globale (*gamma*) correspond à la diversité observée sur l'ensemble des populations putatives considérées (e.g. la zone d'étude). La génétique des populations s'attache, entre autres, à décrire les facteurs et les mécanismes qui influencent ces différentes échelles de diversité. Ces mécanismes sont

¹⁴ Un locus est considéré comme polymorphe lorsqu'il présente au moins deux états alléliques dont la fréquence dans la population considérée est supérieure à un certain seuil, généralement 0,01 ou 0,05. La proportion de loci polymorphes est appelée « taux de polymorphisme » d'une population.

étroitement liés aux quatre forces évolutives que sont la mutation, la sélection, la migration et la dérive, qui peuvent augmenter ou au contraire appauvrir la diversité génétique, influençant ainsi le potentiel adaptatif des populations et donc leur capacité de résilience face à un environnement changeant ou contraignant (Frankel 1974; Frankham 2005). Les mutations sont des modifications aléatoires de la séquence d'ADN qui surviennent lors des événements de réplication et de division cellulaire. Elles sont à l'origine de la diversité génétique par l'apparition de nouveaux allèles. La sélection naturelle est le processus par lequel les gènes et allèles associés à certains phénotypes seront avantagés (et donc conservés à une fréquence supérieure dans la population) si ces phénotypes permettent une plus grande capacité de survie aux individus. Elle peut favoriser ou réduire la diversité génétique en fonction de la manière dont elle influence les phénotypes associés. La migration correspond au brassage génétique (flux de gènes) entre individus consécutif aux événements de reproduction. Elle tend à réduire les variations de composition génétique entre populations par une homogénéisation de leurs fréquences alléliques, et donc à maximiser la diversité génétique moyenne entre populations. Enfin, la dérive génétique correspond à l'évolution stochastique des fréquences alléliques d'une génération à la suivante. Elle s'explique par un effet d'échantillonnage aléatoire des gamètes au cours de la reproduction, dont la représentativité du pool génétique parental peut n'être que partielle. La dérive génétique est d'autant plus forte que l'effectif de la population est faible. Elle peut aboutir à la fixation de certains allèles et donc à une perte de diversité génétique (Frankham 1996).

Dans la nature, les processus démographiques tels que les fluctuations locales et globales de l'abondance réelle et de la taille efficace¹⁵ des individus (associée à une dérive génétique plus ou moins forte), ou encore l'échelle géographique des flux de gènes entre individus (dépendante de l'intensité de la migration), influencent les trois échelles de diversité génétique pour l'ensemble du génome. Les processus adaptatifs liés à la sélection, eux, influencent les parties du génome impliquées dans l'expression de phénotypes sélectionnés ou contre-sélectionnés. Afin d'éviter l'influence de cet élément confondant lorsque l'on s'intéresse uniquement à l'influence de processus démographiques, on s'appuie alors sur l'étude de la variabilité génétique dite « neutre », c'est-à-dire portée par les régions du génome qui ne sont *a priori* pas soumises à la sélection. En outre, lorsque l'on s'intéresse à la variabilité génétique sur des échelles de temps contemporaines, on considère généralement l'influence des mutations comme négligeable.

Les marqueurs moléculaires en génétique des populations

Quantifier la variabilité génétique entre individus ou entre espèces implique l'utilisation de marqueurs moléculaires. Un marqueur moléculaire est une séquence nucléique, facilement identifiable au sein du génome et homologue entre individus, que l'on utilise pour tracer un locus porteur de polymorphisme entre individus et/ou espèces à l'échelle géographique

¹⁵ La taille efficace d'une population correspond à la taille (= effectif) d'une population « idéale » (*population de taille constante, au sein de laquelle tous les individus se reproduisent de façon aléatoire les uns avec les autres et produisent un nombre identique de descendants*) dont la diversité génétique évolue au cours du temps de la même manière que celle de la population étudiée. Une population dans laquelle seule une petite proportion d'individus se reproduisent présentera une faible taille efficace. En outre, l'impossibilité pour les autres individus de transmettre leur patrimoine génétique à une descendance risque d'engendrer un fort effet de dérive génétique. Plus la taille efficace est faible et plus le risque de dérive, et donc d'érosion génétique est élevé.

et/ou taxonomique considérée. Ce locus sert alors de support à l'étude de la variabilité génétique de par l'analyse de la nature et de la fréquence de ses différents allèles. Trois types de marqueurs moléculaires prédominent actuellement dans les études de variabilité génétique chez les populations sauvages (Fig.4).

➤ L'ADN mitochondrial est une molécule d'ADN de petite taille qui constitue le support de l'héritabilité génétique de la mitochondrie, un organite indispensable au bon fonctionnement de la cellule eucaryote animale. Cette molécule haploïde, à transmission presque exclusivement maternelle, se comporte comme un locus unique mais comprend néanmoins plusieurs régions distinctes dont les taux moyens de mutation génétique varient, ce qui les rend plus ou moins adaptées en fonction de l'objectif visé. Par exemple, le gène mitochondrial codant pour la protéine cytochrome oxidase I (COI), est suffisamment variable pour distinguer entre elles différentes espèces même si elles sont phylogénétiquement très proches. Il est donc très utilisé, par exemple, pour des applications de métabarcoding¹⁶ (Hebert et al. 2003) ou pour résoudre des phylogénies à forte résolution (à l'échelle d'espèces très proches, e.g. Arce et al. 2013). Les gènes codant pour les sous-unités ribosomales (e.g. 12S ou 16S) présentent généralement des taux de mutation similaires et peuvent être utilisés pour résoudre des relations phylogénétiques interspécifiques ou intergénériques (e.g. Covain et al. 2016). La région de contrôle, elle, est généralement la plus variable du génome mitochondrial et peut donc permettre de distinguer entre elles différentes populations d'une même espèce, en observant la distribution spatiale des différents haplotypes¹⁷ portés par les individus (e.g. Santos et al. 2007). Chez les poissons néotropicaux, beaucoup d'études de génétique de populations s'appuient uniquement sur la région de contrôle de l'ADN mitochondrial faute d'autres ressources génétiques disponibles (e.g. Iervolino et al. 2010; Aguirre et al. 2013; Nogueira-Machado et al. 2016). Cependant, de par sa nature haploïde et sa transmission maternelle, les marqueurs génétiques de la mitochondrie répondent de manière plus sensible aux forces évolutives telles que la dérive génétique et en conservent donc la mémoire jusqu'à des échelles de temps plus anciennes, ce qui limite leur capacité à fournir des informations à des échelles de temps contemporaines. En outre, de tels marqueurs fournissent une vision partielle des patrons de variation génétique puisqu'ils ne permettent pas d'accéder à l'information génétique transmises par les lignées paternelles. Les patrons de variabilité génétique fournis par l'ADN mitochondrial peuvent être influencés par des facteurs non-démographiques, certains de ses gènes étant sensibles à la sélection naturelle car associés à des fonctions spécifiques (Galtier et al. 2009). C'est pourquoi, généralement, les marqueurs mitochondriaux en génétique des populations sont plutôt utilisés conjointement avec des marqueurs nucléaires neutres, et notamment les microsatellites, qui présentent des taux de mutation plus élevés ainsi qu'une transmission biparentale (e.g. Watson et al. 2016).

¹⁶ Le métabarcoding s'appuie sur les « barcodes » moléculaires naturels, comme la séquence espèce-spécifique du gène COI, pour résoudre l'origine d'un individu, d'un échantillon, d'un résidu, ou d'un fragment d'ADN libre (e.g. ADN environnemental) non-identifié. Ceci s'effectue par comparaison directe de la séquence du barcode avec des références disponibles en lignes, comme la base de données BOLD qui répertorie de nombreuses séquences appartenant à diverses espèces.

¹⁷ Un haplotype correspond à une séquence unique d'un gène (copie d'un gène d'une cellule haploïde ou de l'un de deux chromosomes d'une cellule diploïde), différant d'autres haplotypes par une ou plusieurs substitutions ou indels.

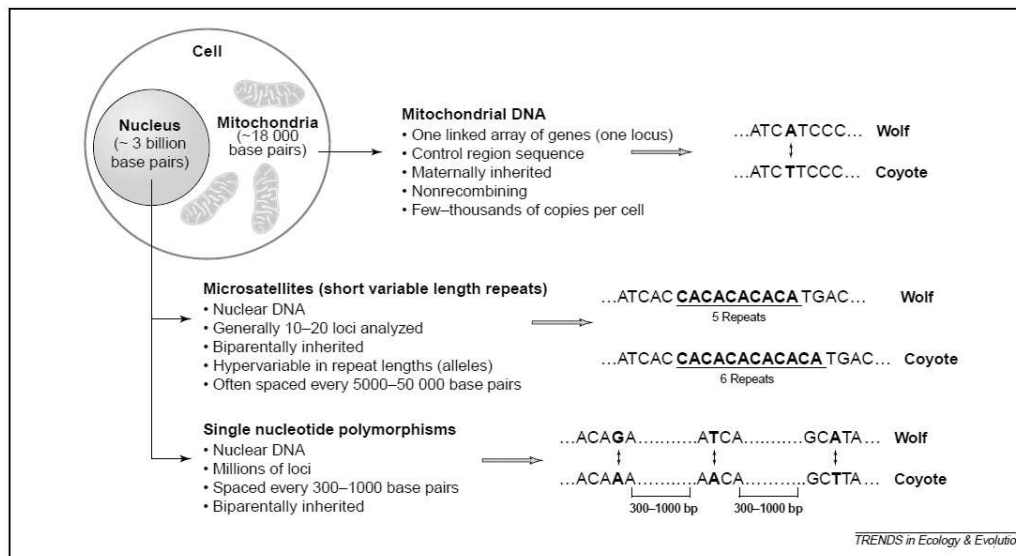


Figure I.4 : Caractéristiques et comparatif des trois principaux types de marqueurs moléculaires : ADN mitochondrial, microsatellites et SNPs. Issu de Morin et al. (2004).

- Les marqueurs STR (« short tandem repeat »), et tout particulièrement les microsatellites, sont longtemps restés les marqueurs de choix en génétique de la conservation (Schlötterer 2004). Ils se caractérisent par des répétitions, dans le génome, de petits « motifs » constitués de 2 à 6 nucléotides (e.g. « GCGCGCGCGCGC »). Le nombre de ces répétitions peut changer fortement d'un individu à l'autre, de par le taux de mutation moyen particulièrement élevé des marqueurs microsatellites par rapport au reste du génome (10^{-5} à 10^{-3} par locus et par génération, Estoup & Angers 1998). Ces marqueurs nucléaires sont donc fortement polymorphes et la fréquence de leurs nombreux allèles répond de manière relativement rapide aux différentes forces évolutives influençant la démographie des populations, comme les fluctuations d'effectif ou la migration. En outre, il s'agit de marqueurs *a priori* neutres. Les microsatellites permettent donc d'étudier les flux de gènes contemporains, de détecter des structures génétiques spatiales de faibles amplitudes, ou encore de détecter des fluctuations d'effectifs (réduction ou expansion de populations) survenues à des échelles de temps relativement récentes. Des marqueurs microsatellites sont développés pour un nombre croissant de poissons néotropicaux (e.g. pour le piranha rouge, Nascimento et al. 2011; pour *Hypostomus* Sp. – souvent désignés sous le nom de « pléco commun », Telles et al. 2010; Galindo et al. 2015; ou encore pour la torche - ou surubí, Revaldaves et al. 2005). Ils restent cependant très peu disponibles en ce qui concerne les espèces de Guyane française dont les aires de répartition géographique restent, pour certaines, très localisées. Or, les marqueurs microsatellites, en dépit de leur utilité et de leur faible coût de développement et de génotypage (Vignal et al. 2002), nécessitent des phases de mise au point pouvant se révéler laborieuses, qu'ils soient inédits (développement de marqueurs nouveaux) ou non (réutilisation de marqueurs issus d'un autre laboratoire ou d'une espèce proche).
- Les marqueurs SNP (« single-nucleotide polymorphism ») représentent la forme de polymorphisme la plus abondante dans le génome. Ils se caractérisent par une mutation

ponctuelle, généralement une substitution (e.g. remplacement d'un nucléotide C par un nucléotide T) dans la séquence d'ADN).

La grande majorité des SNP restent bialléliques du fait de leurs taux moyens de mutation généralement faibles, de l'ordre de 10^{-8} par locus et par génération (Brumfield et al. 2003). De fait, ils sont individuellement peu informatifs en comparaison aux marqueurs microsatellites et de leur plus fort taux de polymorphisme. Ceci est facilement compensé par l'étude d'un grand nombre de SNP. On considère qu'il est suffisant d'utiliser 3 (Brumfield et al. 2003) à 10 (Morin et al. 2004; 2009) fois plus de SNP que de marqueurs microsatellites pour obtenir une vision équivalente des patrons de diversité intra- et interpopulations. Les SNP sont largement répartis dans le génome et peuvent concerner des régions codantes ou non. Leur étude peut donc permettre de dissocier l'influence relative de facteurs purement démographiques (variabilité génétique neutre) ou de facteurs de sélection naturelle (variabilité génétique non-neutre), pour autant que l'accès à un nombre suffisant de ces marqueurs soit assuré (Luikart et al. 2003; Allendorf et al. 2010). Jusqu'au début des années 2000, ces marqueurs étaient largement utilisés pour l'étude de la génétique humaine (e.g. Akey et al. 2002; Pluzhnikov et al. 2002) ou chez des espèces modèles telles que *Drosophila sp.* (e.g. Machado et al. 2002). Cependant à la même période, Brumfield et al. (2003) présentaient les SNP comme futurs marqueurs de choix pour les études de génétique des populations naturelles. Ils sont neutres ou potentiellement soumis à sélection, plus représentatifs du génome que les microsatellites, suivent des modèles de mutations moins complexes (Vignal et al. 2002) et peuvent être utilisés comme autant de loci indépendants. Ils constituent donc une alternative utile à l'application des microsatellites ou des séquences d'ADN non-recombinantes (mono-locus). Ainsi, les SNP peuvent permettre d'étudier les trajectoires évolutives des populations selon des forces évolutives démographiques ou sélectives, et ce à des échelles temporelles d'envergure variable et intégratives. L'obtention d'une telle résolution nécessite cependant de pouvoir accéder simultanément à un grand nombre de ces marqueurs et de telles ressources génétiques ne sont pas disponibles pour les espèces non-modèles. Les connaissances de la nature du génome nucléaire et de la structuration naturelle des populations restent ainsi fragmentaires, voire inexistantes pour de telles espèces – c'est le cas de la grande majorité des poissons néotropicaux guyanais.

II.2.1.2. Forçages biotiques et abiotiques de la variabilité génétique

Les différents processus démographiques qui influencent la répartition spatio-temporelle de la diversité génétique intra et interpopulation ne dépendent pas uniquement de facteurs historiques ou contemporains, mais aussi de la biologie de la ou des espèces considérées (traits d'histoire de vie) et de l'environnement dans lequel elles évoluent ou ont évolué.

Variabilité génétique et traits d'histoire de vie

Les traits d'histoire de vie des espèces influencent naturellement leurs patrons spatiaux de diversité génétique, en agissant à la fois sur les flux de gènes et sur la taille efficace. Par exemple, une espèce de nature sédentaire est moins susceptible de présenter des flux de gènes à large échelle qu'une espèce qui se déplace sur de moyennes ou longues distances (Comte & Olden 2018b) et devrait présenter une diversité *beta* (différentiation génétique

spatiale) plus importante. On s'attend à un phénomène similaire chez les espèces de petite taille, qui présentent souvent une plus faible mobilité (Radinger & Wolter 2014). La taille ou le mode de transport des gamètes ou des propagules (e.g. œufs, larves, pollen, graines) influencent leur distance de dispersion et jouent également un rôle dans les patrons de structuration génétique (Meirmans et al. 2011; Selkoe et al. 2014), tout comme les stratégies de reproduction elles-mêmes. Par exemple, l'autofécondation, en réduisant la taille efficace des populations, peut diminuer la variation génétique intrapopulation et augmenter la différenciation (Charlesworth & Wright 2011). L'endémisme, potentiellement associé à de plus faibles capacités de dispersion ou des abondances plus faibles, peut être associé à une structuration génétique plus forte (Eble et al. 2009). Enfin, le degré de spécialisation écologique peut constituer un facteur supplémentaire de structuration génétique de par la fragmentation spatiale des habitats favorables, ou encore une plus forte vulnérabilité face aux variations environnementales entraînant des fluctuations d'effectifs (Hung et al. 2017; Pilger et al. 2017).

Variabilité génétique et hétérogénéité de l'environnement

Le paysage, c'est-à-dire l'environnement hétérogène au sein duquel évoluent espèces et populations, influence les patrons de distribution de biodiversité spécifique et génétique (Manel et al. 2003). A large échelle, des variables régionales peuvent déjà intervenir : par exemple, la structuration génétique serait globalement plus forte en altitudes et latitudes élevées, d'après une analyse comparative menée par Pelletier et al. (2018) sur plusieurs taxa et divers marqueurs haplotypiques géoréférencés. A échelle géographique plus restreinte, la distance géographique entre deux ou plusieurs populations peut constituer un obstacle à la dispersion des individus et ainsi limiter les flux de gènes entre elles : on parle alors d'isolement par la distance, IBD (Wright 1943; Rousset 1997). L'hétérogénéité des habitats peut agir comme une barrière aux flux de gènes entre milieux de nature différente (isolement par environnement, IBE, Sexton et al. 2014) ou comme une matrice s'opposant aux flux de gènes selon la capacité des individus à franchir des aires géographiques moins favorables (isolement par résistance, IBR, McRae 2006 ; Quéméré et al. 2016).

Cas particulier des réseaux dendritiques

Certains milieux physiques fortement structurés, comme le sont les bassins versants, contraignent la distribution de la diversité. Il peut en résulter une organisation « en réseau » d'habitats plus ou moins connectés entre eux par des « couloirs » permettant la dispersion des individus. Par exemple, les espèces strictement inféodées aux milieux hydrographiques d'eau douce verront nécessairement leur répartition délimitée par la surface, l'agencement et l'accessibilité des différents tributaires qu'elles occupent. Ainsi, les milieux hydrographiques présentent des patrons de diversité spécifique (Carrara et al. 2014) et génétique (Chaput-Bardy et al. 2009) particuliers du fait de leur structuration dite « dendritique ». Cette structuration, analogue à la configuration d'un arbre ou d'un neurone, se caractérise généralement par un axe principal se subdivisant hiérarchiquement en un nombre croissant de ramifications (branches ou dendrites). Les habitats situés au bord du réseau, au niveau des ramifications les plus extrêmes (comme les zones amont de bassins versants), présentent une connectivité plus faible que les autres et bénéficient donc des flux d'individus ou de gènes de façon modérée. A l'inverse, les habitats localisés au centre du

réseau présentent des connexions plus nombreuses avec les habitats environnants. C'est pourquoi la diversité spécifique ou génétique a davantage tendance à se concentrer au niveaux des nœuds centraux des réseaux dendritiques, là où la connectivité est maximale (Labonne et al. 2008), alors que cette même diversité a au contraire tendance à décroître au fur et à mesure que l'on se rapproche des bords (Paz-Vinas & Blanchet 2015), de manière d'autant plus importante que le réseau dendritique est fortement ramifié (Thomaz et al. 2016). L'écoulement de l'eau d'amont vers l'aval, plus favorable au transport physique des individus en direction des cours principaux (migration asymétrique), en particulier durant les stades de vie où leurs capacités de déplacement sont faibles, peut renforcer ce phénomène d'isolement des populations situées en têtes de bassin (Morrissey & Kerckhove 2009; Paz-Vinas et al. 2015). De même, une surface d'habitat plus restreinte dans les têtes de bassin (cours d'eau plus étroits, moins profonds, dénivelé plus marqué...), en diminuant l'abondance et l'effectif efficace des populations, peut également renforcer ce phénomène en favorisant l'intervention d'une dérive génétique plus forte dans ces zones (Meffe & Vrijenhoek 1988; Paz-Vinas et al. 2015). Les réseaux dendritiques constituent donc un bon exemple de la manière dont la structure physique du paysage *per se* peut influencer les patrons de biodiversité. En agissant de la même manière sur la diversité génétique et spécifique, elle peut même expliquer une covariation des patrons de diversité à ces deux niveaux (Vellend & Geber 2005) comme observé à plusieurs reprises chez les invertébrés aquatiques (Finn & Poff 2011) ou les poissons (Fournelle et al. 2016).

Cependant, outre cette configuration physique dendritique, d'autres facteurs peuvent intervenir et influencer les patrons de diversité génétique dans les réseaux hydrographiques. Les traits d'histoire de vie des espèces en font partie : puisque la connectivité des habitats est un élément-clé des patrons de diversité dans ces réseaux, le potentiel dispersif des espèces intervient donc également. Un fort potentiel dispersif aura tendance à contrebalancer l'influence des différents niveaux de connectivités des habitats explorés par les individus. Au contraire, des espèces à faible capacité de dispersion seront d'autant plus influencées par la structuration naturelle du réseau qu'elles seront moins aptes à assurer des flux de gènes à grande échelle (Alp et al. 2012; Paz-Vinas et al. 2015; Pilger et al. 2017). De même, des espèces présentant de plus faibles tailles efficaces locales seront davantage touchées par la dérive génétique et présenteront également une différenciation génétique plus forte entre populations d'un réseau (Paz-Vinas et al. 2015). Les historiques de colonisation de l'aire hydrographique considérée peuvent également agir en renforçant la structure génétique dendritique (colonisation d'aval vers amont, Paz-Vinas et al. 2015) ou en agissant, au contraire, de façon plus confondante (Salisbury et al. 2016; Pilger et al. 2017). Par ailleurs, l'influence de facteurs agissant plus localement (dégradation de l'habitat, nature du substrat, etc) peut complexifier les patrons de diversité qui seraient attendus sous seule influence de la configuration dendritique, ou du paysage d'une manière générale.

II.2.1.3. Comprendre et prédire la variabilité génétique : l'apport de la génétique comparative des populations

Le constat de la relation entre diversité génétique, traits d'histoire de vie et environnement pousse les généticiens de populations à s'interroger sur la prédictabilité de la diversité génétique. On cherche ainsi à savoir s'il existe des règles générales (valables pour plusieurs

espèces) pouvant décrire la façon dont la diversité génétique, et donc le potentiel évolutif des espèces, est susceptible d'évoluer en fonction de paramètres communs.

A large échelle temporelle et spatiale, la phylogéographie comparative par exemple permet une vision plus intégrée des processus évolutifs communs ayant abouti aux patrons de diversité génétique contemporains présentés par différentes espèces (Andrew et al. 2013; Jenkins et al. 2018).

Même en l'absence d'histoire de colonisation commune, la génétique comparative des populations peut permettre de détecter l'influence de processus biologiques ou géographiques contemporains et communs d'une espèce à l'autre, mais dont l'influence passerait inaperçue avec une approche monospécifique. Par exemple, Ward et al. (1994) et DeWoody & Avise (2000) ont respectivement regroupé des données génétiques issues de marqueurs allozymes¹⁸ et microsatellites pour comparer les niveaux globaux de diversité génétique entre plusieurs espèces de poissons marins et d'eau douce, concluant dans les deux cas à une diversité globale plus faible chez ces derniers. Ces approches ont également permis de détecter des « points chauds » de diversité génétique chez plusieurs espèces associées aux récifs coralliens (Selkoe et al. 2016a). Dans une méta-analyse rassemblant 100 publications de génétique du paysage marin, Selkoe et al (2016b) ont mis en évidence l'influence majeure de la température et de paramètres océanographiques sur la connectivité génétique entre populations marines. Il est aussi possible de mettre en évidence le rôle de traits d'histoire de vie via des approches comparatives. Cela a par exemple permis d'évaluer l'influence de la durée de phase larvaire sur les patrons de structure génétique chez des espèces marines, à large échelle géographique (Chust et al. 2016) ou encore dans le cas plus précis des écosystèmes de récifs coralliens (Selkoe et al. 2014, Liggins et al. 2016). Gomez-Uchida et al. (2009) et Pilger et al. (2017) ont mis en évidence l'interférence entre les traits d'histoire de vie de trois et neuf espèces de poissons d'eau douce, respectivement, et l'influence de variables hydrographiques (distance riparienne entre populations, présence de chutes d'eau, taille efficace locale des populations, altitude...) sur la diversité génétique intra et interpopulations. A l'inverse, Liggins et al. (2016) n'ont pas détecté d'influence des traits d'histoire de vie sur la structuration génétique de quatre espèces de poissons récifaux, mais ont souligné en revanche l'influence récurrente de variables océanologiques de large échelle géographique (Sud-Ouest du Pacifique) sur certaines de ces espèces malgré leur biologie différente. Stanley et al. (2018), à l'aide de marqueurs microsatellites, SNP et de technologies plus récentes de séquençage (c.f. section II.2.3.1) ont également identifié des facteurs physiques à l'échelle Nord-Atlantique influençant la structure génétique de cinq espèces marines de différents taxa, indépendamment de leurs traits d'histoire de vie. On voit, à la lumière de ces quelques exemples, que les influences relatives des traits d'histoire de vie des espèces et de leur environnement physique entretiennent des interactions complexes, probablement très dépendantes de l'échelle géographique considérée, du nombre d'espèces étudiées conjointement, ou encore du type de marqueurs génétiques utilisés.

¹⁸ Les allozymes sont des marqueurs protéiques (et plus précisément des enzymes) homologues, c'est-à-dire issus de l'expression d'un même gène, mais dont la taille et/ou la charge varient en fonction des allèles exprimés de ce gène. Les allozymes permettent ainsi un génotypage indirect par leur caractérisation au laboratoire, via électrophorèse par exemple.

La génétique comparative peut ainsi permettre de faire des prédictions sur des patrons spatiaux de diversité génétique attendus et de vérifier s'ils sont effectivement observés, et ce de façon répétable d'une espèce à l'autre (Paz-Vinas et al. 2015; Anthonysamy et al. 2017). De telles prédictions offrent de larges perspectives, notamment dans le domaine de la génétique de la conservation. Elles permettraient par exemple d'optimiser les mesures spatiales de protection des espèces, puisqu'il serait possible de préserver prioritairement les zones géographiques les plus susceptibles d'abriter les diversités génétiques maximales pour une classe d'espèces plus vulnérables que d'autres, voire pour l'ensemble des espèces évoluant de manière sympatrique dans ces zones. Elle permettrait plus généralement de mieux intégrer le niveau génétique de la biodiversité dans les mesures concrètes de conservation des milieux naturels (Nielsen et al. 2016; Paz-Vinas et al. 2018).

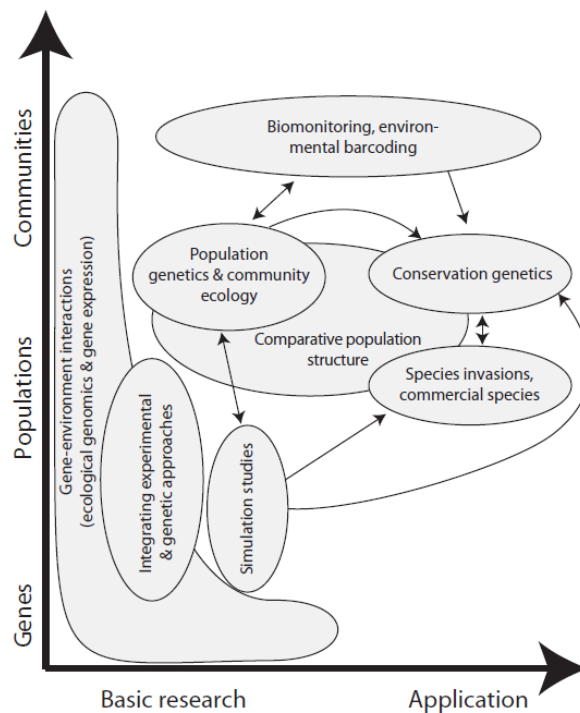


Figure I.5 : Positionnement relatif de différentes approches moléculaires en écologie aquatique, le long d'un gradient de recherche fondamentale à appliquée (abscisse) et de différents niveaux d'organisation de la diversité, des gènes aux communautés (ordonnée). Les superpositions entre cercles indiquent de possibles interconnexions entre approches. Les flèches indiquent des liens méthodologiques, réciproques ou non, entre approches. Issu de Pauls et al. (2014).

Dans cette optique, la génétique comparative des populations montre un intérêt majeur pour la compréhension de l'impact des activités humaines sur la diversité génétique des espèces. Comparer des patrons de diversité génétique entre espèces permet de voir s'ils répondent de façon similaire ou non à certaines perturbations. Khimoun et al. (2016) ont par exemple mis en évidence la plus forte vulnérabilité d'espèces d'oiseaux habitat-spécialistes face à la fragmentation de leurs habitats forestiers. Lino et al. (2018) ont choisi une approche méta-analytique pour mesurer l'effet de la fragmentation des habitats sur les niveaux globaux de diversité génétique chez 32 espèces de mammifères présentant des caractéristiques écologiques différentes, et analysées sur la base de marqueurs microsatellites et mitochondriaux. Ils ont ainsi mis en évidence une plus forte vulnérabilité, face à la fragmentation, des mammifères de grande taille, non-volants, herbivores ou encore inféodés

aux habitats forestiers. Ainsi pour les poissons, Blanchet et al. (2010) ont montré que plusieurs espèces présentant des capacités de dispersion contrastées de par leur taille adulte présentaient des patrons de structure génétique qui variaient différemment entre un habitat fragmenté ou non.

En apportant une vision multi-espèces (donc plus proche du fonctionnement des communautés) et par ses applications directes en génétique de la conservation, la génétique comparative des populations occupe un positionnement central en écologie (Fig 1.5; Pauls et al. 2014). Actuellement, la génétique comparative des populations reste une discipline émergente (Pauls et al. 2014; Horne 2014). Son développement dépendra notamment de la capacité des chercheurs à obtenir des données couvrant des aires géographiques communes -ou présentant des caractéristiques comparables, et suffisamment nombreuses pour capturer la variation génétique spatiale des différentes espèces d'intérêt dans ces zones. Cela implique aussi que les zones considérées soient représentées de façon homogène par un nombre suffisant d'espèces et d'individus afin d'optimiser la résolution des différents jeux de données à comparer (Horne 2014).

II.2.2. Variabilité génétique et pressions anthropiques

II.2.2.1. Diversité et structure génétique de populations exploitées

L'exploitation des populations naturelles, c'est-à-dire le prélèvement d'individus dans leur milieu par les outils de chasse ou de pêche, agit de manière directe sur leur abondance en générant un facteur de mortalité supplémentaire en plus de la mortalité naturelle. Ceci peut impacter la taille efficace locale des populations, ainsi que les flux de gènes à plus grande échelle (Allendorf et al. 2008). Ce phénomène peut entraîner d'importantes variations de la diversité génétique des populations ciblées via trois mécanismes principaux : une perte de variabilité génétique, une altération possible de la structure génétique, ou encore l'émergence d'une force de sélection modifiant les caractéristiques des traits d'histoire de vie intraspécifiques.

Exploitation et perte de variabilité génétique

Une diminution de la taille efficace des populations exploitées s'accompagne d'une augmentation de l'effet de la dérive génétique. Elle peut ainsi entraîner, à plus ou moins long terme, une perte de diversité génétique (érosion) par fixation de certains allèles (Allendorf et al. 2008). Ainsi par exemple, Hauser et al. (2002) ont mis en évidence une diminution de la taille efficace et du nombre moyen d'allèles par locus (richesse allélique¹⁹) chez le Sparidae *Pagrus auratus*, depuis le début de son exploitation en Nouvelle-Zélande. Cette perte de variabilité génétique peut être associée à un risque accru de baisse du potentiel adaptatif des populations impactées (Frankham 1995a; Spielman et al. 2004), et ce de manière irréversible si les allèles perdus via l'effet de dérive ne peuvent être recouverts par l'arrivée d'immigrants par exemple.

La réduction de taille efficace peut également modifier les régimes de reproduction au sein d'une population, en générant par exemple l'apparition de consanguinité par une

¹⁹ La richesse allélique peut correspondre au nombre total d'allèles observés dans une population sur l'ensemble des loci considérés, ou au nombre moyen d'allèles par locus.

augmentation de la probabilité des croisements entre individus apparentés (Frankham 1995b; Lowe et al. 2005). Des indices de fixation tels que le F_{IS} de Wright-Fisher (1931), qui mesure l'écart entre les taux d'hétérozygotie observé par rapport à l'attendu²⁰, peuvent permettre de détecter des croisements préférentiels entre individus apparentés, qui entraînent une diminution de la diversité génétique individuelle et donc du taux d'hétérozygotie observé.

Les exemples d'érosion génétique chez des espèces exploitées sont nombreux, même quand l'abondance observée des individus dans leur milieu reste très importante (c'est par exemple le cas pour *Pagrus auratus* susmentionné, Hauser et al. 2002). En 2014, Pinsky & Palumbi, en procédant à une méta-analyse portant sur 140 espèces marines et en comparant la richesse allélique et l'hétérozygotie moyennes de marqueurs microsatellites, ont montré que les espèces exploitées présentaient de plus faibles valeurs pour ces deux indicateurs. Ardura et al. (2011) ont mis en évidence une diversité nucléotidique et haplotypique relativement faible, pour le marqueur mitochondrial COI, chez des espèces de poissons d'eau douce exploitées du bassin Amazonien, en comparaison à des assemblages de milieux marins moins ciblés et en dépit de la diversité phylogénétique des espèces étudiées²¹. Chez le piranha à ventre rouge (*Pygocentrus nattereri*) d'un bassin versant du Rio Solimões (bassin Amazonien), une baisse de diversité génétique locale et une augmentation de l'indice de fixation F_{IS} ont conduit Santos et ses collaborateurs (2016) à soupçonner une surexploitation par la pêche dans certaines aires de leur zone d'étude.

Exploitation et modification des patrons de structure génétique

L'exploitation d'une ou plusieurs populations d'une espèce dans son milieu naturel peut également influencer et modifier les patrons spatiaux de structuration génétique. L'érosion locale de la diversité génétique peut par exemple augmenter le degré de structure, si les effets de dérive génétique deviennent plus importants que l'apport de diversité génétique par les flux de gènes depuis les populations adjacentes. Ce phénomène peut également intervenir si l'exploitation diminue l'intensité de la migration, par exemple en diminuant les probabilités locales de migration densité-dépendante (Allendorf et al. 2008). *A contrario*, l'exploitation peut aussi augmenter les flux de gènes entre certaines populations si une diminution de densité locale entraîne un effet « appel d'air » de migrants d'autres aires géographiques, effaçant l'existence de structures existantes et contrebalançant la mise en place d'adaptations locales (Allendorf et al. 2008).

Il est difficile d'anticiper les effets de l'exploitation sur le devenir des patrons de structure génétique, les caractéristiques précises d'une méta-population (flux de gènes, tailles efficaces locales et globale, etc) étant souvent difficiles à estimer. Cela peut néanmoins s'effectuer via des approches temporelles, c'est-à-dire en suivant l'évolution des indices de diversité et de structure génétique au cours du temps (*temporal genetic monitoring*,

²⁰ Le taux d'hétérozygotie correspond à la proportion moyenne d'individus hétérozygotes, c'est-à-dire porteurs de deux allèles différents pour un locus donné. Moyenné sur l'ensemble des loci considérés, il fournit une mesure de diversité génétique à l'échelle d'une population. Le taux d'hétérozygotie peut-être déduit des fréquences alléliques (valeur attendue) ou du comptage direct du nombre d'individus hétérozygotes dans la population considérée (valeur observée).

²¹ Il faut cependant rappeler, comme le soulignent également les auteurs, la possibilité qu'un tel résultat soit également influencé par la diversité génétique naturellement plus faible en milieu d'eau douce que marin (DeWoody & Avise, 2000).

Allendorf et al. 2008). Néanmoins, la connaissance des patrons de structuration originels est importante pour le suivi et la gestion des populations exploitées (Palsbøll et al. 2007). Ceci permet de mieux délimiter à quelle échelle peuvent agir les impacts significatifs de l'exploitation et de savoir si celle-ci touche une seule, ou plusieurs populations d'un ensemble plus vaste régis par des flux de gènes d'intensité variables ou des variations de taille efficaces locales. Toute pression d'exploitation plus ou moins localisée spatialement est alors susceptible d'impacter les patrons de diversité génétique à plus large échelle, et donc de modifier les patrons de structuration génétique globaux. En outre, une pression d'exploitation portant sur plusieurs populations différentes comporte le risque d'aboutir à l'extinction plus rapide d'une des sous-populations par rapport aux autres. Par exemple, il a été montré chez le saumon atlantique en Norvège (Hindar et al. 2004) que les sous-populations de petit effectif contribuaient davantage à la taille efficace globale (et donc à la diversité génétique totale à l'échelle de l'ensemble des sous-populations) que les sous-populations plus abondantes. Or la pression de pêche en mer, ciblant de manière égale et sans distinction d'origine l'ensemble des juvéniles et adultes, agit donc de façon plus délétère aux sous-populations de petite taille. Ce phénomène peut ainsi engendrer de profonds bouleversements dans la dynamique globale de la métapopulation de saumon atlantique de cette zone.

Exploitation et sélection orientée des traits d'histoire de vie intraspécifiques

Bien souvent, et en fonction des engins de capture utilisés, la pression d'exploitation porte préférentiellement sur certaines catégories d'individus, par exemple ceux ayant dépassé un certain poids ou une certaine taille. Ces individus, contre-sélectionnés via l'exploitation, verront leur succès reproducteur réduit, ce qui à moyen-terme peut entraîner une micro-évolution des stratégies d'histoire de vie induite de manière artificielle, par exemple en faveur d'individus à maturité sexuelle précoce. Ceci explique par exemple la réduction de la longueur moyenne des individus capturés, au cours du temps, dans certaines pêcheries (e.g. Quinn et al. 2006 chez le saumon Atlantique; Swain et al. 2007 chez le cabillaud). De telles modifications de traits d'histoire de vie peuvent freiner la résilience des populations même suite à un arrêt de la pression de pêche, et donc être défavorables à leur persistance (Walsh et al. 2006). Cette sélection orientée sur certains traits d'histoire de vie intraspécifiques peut donc impacter, indirectement, les fréquences alléliques au niveau des gènes impliqués dans l'expression de ces traits phénotypiques. L'étude des patrons de diversité génétique au niveau de tels marqueurs peut donc permettre de détecter des modifications des stratégies d'histoire de vie des espèces (Kuparinen & Merilä 2007). Les effets de sélection orientée et l'érosion de la variabilité génétique neutre induits par la pression de pêche et la dérive génétique associée, peuvent également agir de manière synergique ou antagoniste (Marty et al. 2015).

II.2.2.2. Des indices de variabilité génétique comme support de gestion

C'est à partir du début des années 1970 que les disciplines de conservation de la biodiversité ont commencé à intégrer la dimension génétique au cœur de leurs enjeux. Au cours des années 1980, la génétique de la conservation a alors effectué son apparition comme champ

de réflexion scientifique -et transdisciplinaire- à part entière (Allendorf et al. 2008). Aujourd'hui, les indices de diversité génétique basiques sont calculables de façon relativement aisée et constituent le cadre de travail standard du généticien de la conservation. La proportion de marqueurs polymorphes, la richesse allélique, les taux d'hétérozygotie attendu et/ou observé, la structuration globale et entre paires de sites géographiquement distincts ou encore l'analyse des relations d'apparentement entre individus sont autant d'indicateurs indirects des processus démographiques (dérive génétique, flux de gènes) qui gouvernent le fonctionnement des populations (Pereira et al. 2013). Les efforts de recherche en génétique de la conservation visent également une meilleure appropriation des outils de la génétique des populations par les gestionnaires eux-même. Par exemple, nous pouvons citer Hoban et al. (2013a) qui ont développé un ensemble de ressources, disponibles en ligne, permettant de construire un plan d'échantillonnage et d'effectuer des analyses génétiques préliminaires. Scribner et al. (2016) présentent une revue vulgarisée de l'intérêt des outils de génétique des populations et de génomique pour l'étude de la biodiversité en milieux hydrographiques dendritiques. Ils soulignent également l'intérêt d'intégrer, à l'avenir, les données génétiques aux données écologiques pour optimiser les mesures de conservations des populations de poissons d'eau douce.

L'écart entre l'effort alloué à l'étude des niveaux supérieurs de biodiversité (assemblages spécifiques, communautés et paysages) et de la variabilité génétique s'amenuise désormais peu à peu. Ceci s'explique en grande partie par le développement d'outils analytiques de plus en plus perfectionnés, mais aussi par la forte décroissance des investissements financiers nécessaires au développement et à l'analyse de marqueurs moléculaires. Cette dernière ouvre la voie vers une plus grande capacité d'échantillonnage et d'accumulation de données génétiques. Tout ceci s'inscrit également dans l'émergence de nouveaux questionnements théoriques concernant la façon dont la diversité génétique évolue en synergie entre les caractéristiques intrinsèques aux espèces et celles de leur environnement, empreinte anthropique comprise.

Cependant, et malgré les nombreux développements en cours, la disponibilité des données génétiques pertinentes reste encore très limitée pour de nombreuses espèces de poissons d'eau douce (ex. Paris et al. 2018, Annexe II). Turak et ses collaborateurs (2016), à la lumière des objectifs de l'observatoire GEO BON appliqués aux milieux aquatiques, ont défini plusieurs axes prioritaires concernant l'application des outils de génétique. Ils suggèrent notamment, d'ici 2030, de prioriser le développement de ressources moléculaires standard pour l'étude de la diversité intraspécifique, de définir des protocoles et des bases de données standardisés pour le suivi de la diversité génétique intrapopulation, et de répertorier les aires géographiques faisant l'objet de tels suivis.

Introduction:

III. Enjeux méthodologiques de la thèse

III.1. Le développement de ressources moléculaires

Le développement et la mise à disposition de marqueurs moléculaires constitue le socle de base de toute étude de la variabilité génétique chez une espèce non-modèle. Lorsque le programme Pêche a vu le jour en 2013, cette première étape a été identifiée comme une proportion significative à prévoir du présent travail de thèse. Les méthodes classiques de mise au point de marqueurs microsatellites, généralement basées sur des méthodes de séquençage classique type Sanger, sont peu coûteuses mais de longue haleine. Pour cette raison, il n'est pas rare que le développement de ces marqueurs s'effectue à partir de microsatellites préalablement développés pour une espèce phylogénétiquement proche. Il suffit alors de conserver les marqueurs qui se montrent également polymorphes pour la nouvelle espèce considérée. Cela permet de s'affranchir de certaines phases laborieuses de développement mais n'est pas sans inconvénients, notamment liés à ce que l'on appelle l'*ascertainment bias* (c.f section suivante). En outre, ce type de développement peut devenir particulièrement chronophage dès lors que l'on considère non plus une, mais plusieurs espèces pour lesquelles des ressources génétiques sont à développer. Ces considérations très pratiques, conjuguées aux avantages analytiques des marqueurs SNP soulevés en section II.2.1.1, ont orienté notre choix vers ces marqueurs-ci pour décrire la variabilité génétique chez les espèces focales du Haut-Maroni.

En effet, le développement des nouvelles générations de séquençage (NGS) a propulsé les marqueurs SNP au rang de marqueurs de choix en les rendant accessibles rapidement et en très grand nombre (plusieurs centaines à plusieurs milliers de marqueurs), y compris pour les espèces non-modèles (Ekblom & Galindo 2011; Seeb et al. 2011²²). On regroupe sous le terme de NGS l'ensemble des technologies de séquençage d'acides nucléiques (ADN ou ARN) à *haut-débit*, c'est-à-dire capables de séquencer des milliards de paires de bases et donc de générer de très nombreuses lectures ('reads') de fragments du génome. Différentes technologies NGS existent : leurs caractéristiques, avantages et inconvénients ne seront pas détaillés ici mais sont consultables dans de nombreuses revues (consulter par exemple Zhang et al. 2011 ou Da Fonseca et al. 2016). Il faut néanmoins distinguer deux manières de préparer les échantillons d'ADN (bibliothèques) à incorporer sur le séquenceur. Les approches type *whole-genome-sequencing* (WGS) visent à séquencer la plus grande portion possible d'un génome, quitte à obtenir un nombre de lectures chevauchantes limité (on accepte alors qu'une fraction donnée du génome ne soit lue qu'un petit nombre de fois, avec les sources d'incertitudes que cela implique). D'autres approches, au contraire, nécessitent que le nombre de lectures d'une même portion du génome, ce que l'on appelle *profondeur de couverture*, soit élevé : en effet cela permet, par comparaison croisée des différentes lectures chevauchantes, d'y détecter des sites variables, et plus particulièrement des SNP. Les NGS

²² Cette référence constitue l'introduction d'une issue spéciale de la revue *Molecular Ecology Resources*, publiée au mois de Mars 2011 et consacrée à l'application des marqueurs SNP chez les espèces non-modèles.

permettent ainsi de détecter des loci polymorphes entre différents individus séquencés, mais aussi pour un seul et même individu. De cette manière, on peut ainsi obtenir des informations sur son génotype. Ainsi le séquençage haut-débit, offrant une résolution sans précédent de la composition et de la variabilité du génome au sein et entre individus, a ouvert de nombreuses perspectives dans le domaine de la génétique des populations. Il permet par exemple de mettre en évidence des niveaux de structuration génétique subtils, qui seraient indétectables avec des marqueurs plus classiques (Angeloni et al. 2011). Les NGS ont également permis des avancées importantes dans le domaine de la génomique (cartographie et annotation de génome, détection d'îlots de différenciation ou de gènes impliqués dans la sélection naturelle, Chutimanitsakun et al. 2011; Hohenlohe et al. 2010), de la phylogénie (Pante et al. 2015) et de la phylogéographie (Emerson et al. 2010).

En outre, les NGS peuvent également se substituer aux méthodes de séquençage plus classiques pour le développement de ressources moléculaires à plus faible densité. Il n'est pas toujours indispensable de travailler sur des centaines de milliers de SNP : dans certains cas, quelques dizaines peuvent suffire, de manière par exemple à obtenir une information similaire à celle apportée par un lot de marqueurs microsatellites usuel. Ainsi, on peut s'appuyer sur les avantages des NGS non pas pour développer des milliers de marqueurs pour une espèce donnée, mais des dizaines de marqueurs pour un grand nombre d'espèces en simultané. Ces marqueurs, suffisamment peu nombreux, peuvent être génotypés via des technologies éprouvées dont beaucoup sont consultables dans la revue de Vignal et al. (2002), ce qui a pour avantage de diminuer drastiquement le coût financier de la procédure lorsqu'un grand nombre d'individus (typiquement supérieur à 100) doit être génotypé pour chaque espèce. C'est cette stratégie que nous avons choisi de suivre pour développer des ressources moléculaires pour les espèces focales du Haut-Maroni. Elle sera présentée de façon beaucoup plus détaillée au cours du Chapitre 1.

III.2. Appliquer une approche comparative entre espèces

Les intérêts et les perspectives de la génétique comparative des populations ont été abordés tantôt. Intéressons-nous de façon plus détaillée, à présent, à certaines de ses limites méthodologiques actuelles. Rappelons que le peu de données biologiques disponibles à forte résolution (c'est-à-dire, disponibilité d'un échantillonnage suffisant d'individus de plusieurs espèces et sur plusieurs sites d'une même zone géographique) et le manque de ressources moléculaires multiloci sont les principaux facteurs qui limitent encore aujourd'hui l'application en routine de la génétique comparative des populations. Nous pouvons citer, à titre informatif, des exemples d'efforts effectués pour remédier à cela. Yashima & Innan (2017) ont développé la base de données VARVER (Variations in Vertebrates) qui répertorie pour différentes espèces de vertébrés (2301 en date de publication de l'article) l'ensemble des données génétiques intra et interpopulations ayant pu être collectées au gré de différentes publications, sous forme téléchargeable et manipulable. Lawrence et al. (2019) proposent la base de données MacroPopGen, qui regroupe également des données de variabilité génétique intra et interpopulations pour de nombreuses espèces sur l'ensemble du continent américain. Deck et al. (2017) ont développé la base de métadonnées GeOMe, qui regroupe des informations précises et standardisées sur de nombreux échantillons biologiques (et marqueurs génomiques associés) collectés à l'échelle mondiale.

Les études existantes de génétique comparative utilisent globalement deux types d'approches : (1) les méta-analyses issues du regroupement de données déjà publiées antérieurement ou collectées via des collaborations, et portant sur une zone géographique commune ou non, ou bien (2) l'échantillonnage simultané de quelques espèces sympatriques dans une même zone géographique. Ces deux approches présentent chacune leurs avantages et inconvénients (Table I.1).

Table I.1 : Avantages et inconvénients des principales approches en génétique comparative des populations.

	Avantages	Inconvénients	Exemples
Méta-analyses	<ul style="list-style-type: none"> - Possibilité d'étudier un grand nombre d'espèces (parfois >10) - Faible coût financier et logistique 	<ul style="list-style-type: none"> - Difficulté d'accès aux données brutes ou manque d'informations sur leur origine spatio-temporelle (Pope et al. 2015) - Echantillonnage de différentes origines temporelles et/ou géographiques - <i>Ascertainment bias</i> 	<ul style="list-style-type: none"> Pelc et al. 2009 Dawson et al. 2014 Paz-Vinas et al. 2015 Selkoe et al. 2014 Selkoe et al. 2016a Chust et al. 2016 Costantini et al. 2018 Lino et al. 2018 (...)
Co-échantillonnage	<ul style="list-style-type: none"> - Standardisation de l'échantillonnage - Identification d'aires géographiques d'intérêt (forte diversité génétique, flux de gènes) 	<ul style="list-style-type: none"> - Echantillonnage coûteux et laborieux - Peu d'espèces (généralement <10) - <i>Ascertainment bias</i> 	<ul style="list-style-type: none"> Gomez-Uchida et al. 2009 Blanchet et al. 2010 Villamor et al. 2014 Khimoun et al. 2016 Liggins et al. 2016 Anthonyamy et al. 2017 Pilger et al. 2017 (...)

Il est également possible de combiner ces deux approches afin de comparer davantage d'espèces (Kelly & Palumbi 2010; Cahill et al. 2018). En outre, certaines approches méta-analytique regroupent des données d'origines géographiques très similaires (Selkoe et al. 2014; 2016a) ce qui permet d'en tirer des informations spatiales de même envergure que les approches de co-échantillonnage, à ceci près que les données peuvent avoir été collectées à des périodes différentes les unes des autres. Les approches de co-échantillonnage restent malgré tout, a priori, les plus à même de permettre des applications directes en biologie de la conservation, puisqu'elles portent une dimension spatiale beaucoup plus marquée et reposent sur des données collectées de façon standardisée. Cependant, les études de ce type restent actuellement limitées à un petit nombre d'espèces.

Parmi les sources possibles de biais dans une démarche de génétique comparative des populations, l'utilisation de différents lots de marqueurs moléculaires (développés à partir de matériel biologique d'origine différentes par exemple) pour comparer des espèces entre elles, est rarement évoquée. En effet, les études de génétique des populations font généralement l'hypothèse que les marqueurs moléculaires utilisés sont représentatifs du génome sur la zone d'étude, ou bien que les estimations effectuées à partir de ces marqueurs (qui dépendent de l'objectif de l'étude) est robuste à une représentativité partielle de ce génome (biais d'incertitude ou *ascertainment bias*, Fisher 1934). On appelle biais d'incertitude, plus précisément, l'écart entre le spectre des fréquences alléliques fourni par

un lot de marqueurs moléculaires par rapport au spectre global que l'on obtiendrait en observant l'ensemble du génome à l'échelle de la zone d'étude et/ou pour l'ensemble des individus. En d'autres termes, le biais d'incertitude correspond à un échantillonnage non représentatif des différents loci du génome d'une espèce, du moins pour la zone géographique considérée. Concrètement, ce biais peut résulter des particularités, du nombre ou de l'origine géographique des individus utilisés pour le développement originel d'un lot de marqueurs donnés, des modèles de mutations qui peuvent différer entre marqueurs (Ellegren & Galtier 2016) ou encore du transfert de marqueurs entre espèces phylogénétiquement proches, comme cela se fait beaucoup avec les marqueurs microsatellites (Pinsky & Palumbi 2014).

En outre, le biais d'incertitude peut se révéler problématique s'il affecte de façon imprévisible l'évaluation de la diversité intra et interpopulations pour les différentes espèces, qui risquent alors de s'en retrouver plus difficilement comparables (Li & Kimmel 2013). La plupart des études de génétique comparative s'attachent ainsi à décrire les patrons de diversité génétique de manière qualitative, visant par exemple des conclusions de type « Sur trois espèces co-échantillonnées sur dix localités, deux montrent un isolement par la distance significatif ». Les hypothèses qui sous-tendent cette conclusion sont moins fortes, a priori, que si l'on affirmait « Sur trois espèces co-échantillonnées sur dix localités, deux montrent un isolement par la distance significatif dont une présente une pente d'IBD deux fois supérieure à l'autre ». Avancer cette seconde conclusion sans risque de se tromper reste actuellement risqué, les outils analytiques corrigeant les biais d'incertitude restant difficiles à mobiliser en routine (c.f. Discussion). Cependant, il semble possible de s'en approcher davantage en comparant des espèces génotypées à l'aide de marqueurs moléculaires qui auraient été développés selon un protocole commun. Le Chapitre 2 de cette thèse s'attachera ainsi à comparer les patrons de variabilité génétique entre les différentes espèces co-échantillonnées sur le Haut-Maroni.

III.3. Révéler la mise en place et l'issue de processus complexes : Simuler des processus et explorer des hypothèses

Lorsque l'on s'intéresse à l'influence de multiples facteurs sur les niveaux de diversité génétique d'une population (matrice paysagère, structure d'un bassin versant, intensité et directionnalité des flux de gènes, effectif des populations...), il est difficile de mettre en exergue leurs impacts relatifs. En écologie, les approches de modélisation constituent un moyen de contourner la complexité des systèmes naturels en tenant compte uniquement de l'influence théorique d'un nombre limité de facteurs. En d'autres termes, les modèles tendent à une simplification du monde réel et leur intérêt réside précisément dans cette caractéristique.

En génétique, par exemple, le développement (Kimura & Ohta 1978) puis l'optimisation (Valdes et al. 1993; Ellegren 2004) du modèle de mutation pas-à-pas (*stepwise mutation model*, SMM) ont permis de mieux comprendre les subtilités des processus mutationnels à l'origine du fort polymorphisme des marqueurs microsatellites. Le modèle de génétique de populations le plus largement connu est le modèle d'équilibre reposant sur la loi de Hardy-Weinberg (1908), qui permet de déduire la fréquence des génotypes homo- et hétérozygotes d'individus d'une population idéale à partir des fréquences alléliques parentales. Un autre

exemple est le modèle en îles de Wright-Fisher (1931; 1943), qui permet d'estimer le nombre moyen de migrants par génération entre deux populations distinctes (« îles ») à partir de mesures de différenciation génétique. Bien que l'on sache aujourd'hui que ces modèles ne sont jamais parfaitement vérifiés dans les populations naturelles, leur développement a permis de nombreuses avancées dans la compréhension des mécanismes évolutifs qui régissent les populations.

Au cours des deux dernières décennies, la puissance de calcul croissante des ordinateurs s'est accompagnée d'une généralisation des outils de simulations de données génétiques (Hoban et al. 2012). Les simulations permettent de générer des données génétiques *in silico*, c'est-à-dire de manière informatique (on les qualifie alors de « pseudo-données »). Ces pseudo-données sont générées selon des scénarii évolutifs et/ou démographiques fixés par l'utilisateur, qui peut donc créer de toutes pièces un système biologique virtuel dont il contrôlera l'ensemble des paramètres (ex. flux migratoires, taille des populations, etc.). L'utilisateur est ainsi en mesure d'analyser comment ce système virtuel évolue sous l'influence des facteurs de son choix, en faisant varier leur intensité. Par ailleurs, en répétant l'opération un grand nombre de fois, il peut analyser le degré de stochasticité inhérent aux processus démographiques ou génétiques simulés.

Les simulations permettent ainsi d'analyser l'issue de scénarii évolutifs complexes où interagissent plusieurs facteurs contrôlés par l'utilisateur, dans un but prédictif ou exploratoire. Elles permettent également de reconstituer les processus évolutifs et les éventuels facteurs confondants qui expliquent les patrons de variabilité génétique observés à partir des données empiriques, dans un but inférentiel ou de validation de méthode analytique. On cherche alors, dans ce cas, à trouver le scénario qui maximise l'adéquation entre données empiriques et simulées (=inférence), par exemple via les procédures de calcul bayésien approché (ABC, Csilléry et al. 2010). On cherche également à vérifier qu'il n'existe pas d'autre scénario permettant d'aboutir aux mêmes observations empiriques et d'invalider les interprétations (=validation). Enfin, les simulations peuvent également permettre d'optimiser un plan d'échantillonnage en estimant le nombre d'individus, de sites géographiques ou de marqueurs génétiques nécessaires à la détection de l'effet recherché lorsqu'il existe. Morin et al. (2009) ont par exemple utilisé une approche par simulations pour analyser le nombre de marqueurs SNP nécessaires (et leur fréquence allélique moyenne idéale) pour la détection de différents niveaux de différenciation génétique.

Les outils informatiques de simulation sont nombreux et possèdent chacun leurs spécificités (Hoban et al. 2012; Peng et al. 2014; Yuan et al. 2014). Cependant, on distingue deux grandes catégories de simulateurs qui s'appuient sur deux approches distinctes. Les simulateurs *backward-in-time*, basés sur la théorie du coalescent (Kingman 1982), sont particulièrement adaptés pour l'analyse de processus agissant à grande échelle de temps évolutive et aux approches inférentielles en simulations (Hoban et al. 2012). Les simulateurs *forward-in-time*, ou individu-centrés, modélisent au contraire l'évolution chronologique de la variabilité génétique d'une ou plusieurs populations composées d'individus agissant de manière indépendante. Les « actions » de chaque individu (e.g. naissance, migration, reproduction, décès) sont définies à chaque pas de temps par des matrices de transition permettant de passer d'une génération à la suivante. Il est donc possible d'analyser l'évolution de la composition génétique des populations simulées au cours du temps. Les simulateurs

forward-in-time peuvent devenir computationnellement intensifs lorsqu'un grand nombre d'individus et/ou de générations sont ainsi impliqués. Cependant, ils offrent une large diversité de scénarii pouvant être testés, y compris sur des échelles temporelles très courtes de l'ordre de quelques générations (Hoban et al. 2012). Ils peuvent également intégrer des informations très précises tenant compte des traits d'histoire de vie des espèces d'intérêt. Par exemple, le logiciel individu-centré SimuPOP (Peng & Amos, 2008) permet de définir, au sein d'une population donnée, des classes d'individus pouvant varier par leur succès reproducteur ou leur probabilité de se déplacer vers une autre population. Ils sont donc particulièrement adaptés pour des approches prédictives ou exploratives en simulations.

Dans les milieux hydrographiques, l'influence de la structure dendritique d'un bassin versant et de facteurs supplémentaires liés à l'environnement ou aux traits de vie des espèces (surface d'habitat disponible, asymétrie des flux de migrations, etc), peuvent aboutir à des patrons de diversité génétique particuliers. Nous pouvons par exemple rappeler les travaux d'Ivan Paz-Vinas et de ses collaborateurs (2015) qui ont utilisé une approche par simulations pour montrer que la présence de migration asymétrique dans les bassins versants pouvait renforcer l'effet d'accumulation de diversité génétique au niveau des confluences et des zones aval, par rapport aux régions plus amont des tributaires. Les simulations pourraient ainsi, de manière similaire, permettre d'explorer l'influence éventuelle des impacts anthropiques tels que la pression de pêche sur les patrons de variabilité génétique dans les réseaux hydrographiques dendritiques. Des résultats préliminaires en lien avec cet objectif seront présentés en Chapitre 3 de cette thèse.

Introduction:

IV. Objectifs de l'étude et stratégie

L'objectif scientifique principal de cette thèse est d'appréhender l'impact de la pression de pêche sur la variabilité génétique neutre chez plusieurs espèces (Annexe I) de poissons d'intérêt halieutique du Haut-Maroni, en Guyane française. Dans une perspective de gestion et de conservation des ressources, nous cherchions plus précisément à savoir si l'exploitation avait entraîné une réduction de diversité génétique localisée, diminuant alors le potentiel adaptatif des populations. Nous nous sommes attachés à décrire, pour chaque espèce, les patrons de variabilité génétique intra- et interpopulations à l'échelle de la zone d'étude en relation avec la variabilité spatiale de la pression de pêche et la configuration spatiale naturelle du bassin versant. Nous avons également comparé les patrons globaux de variabilité génétique entre les espèces à la lumière de leurs traits d'histoire de vie et de leur intérêt halieutique.

Pour répondre à notre objectif, nous avons constitué un jeu de données regroupant les génotypes de plusieurs spécimens (1) collectés sur douze sites de prélèvement, répartis le long de différents cours d'eau (principal et secondaires) du réseau hydrographique du HM, dans ou au dehors des zones de pêche principales (2) appartenant à 18 espèces présentant des traits biologiques contrastés et d'intérêt halieutique variable, mais aussi suffisamment abondantes pour pouvoir être collectées de façon relativement uniforme sur l'ensemble de la zone d'étude et permettre ainsi une approche comparative.

Dans un premier temps, il s'agissait de développer, de façon rapide et à coût aussi réduit que possible, les ressources moléculaires nécessaires pour caractériser la variabilité génétique spatiale pour chacune des 18 espèces. Dans le Chapitre 1 de ce manuscrit, je présente le développement méthodologique et pratique du protocole m'ayant permis la mise au point de marqueurs de type SNP (*single-nucleotide polymorphism*) qui ont ensuite été validés par génotypage pour ces espèces non-modèles. Cette méthode s'appuie sur la technologie de séquençage à haut-débit RADSeq et sur l'élaboration d'un pipeline informatique cherchant à maximiser la facilité d'un génotypage de marqueurs par amplification PCR classique. Ce travail a permis de générer l'ensemble des données génétiques sur lesquelles s'appuie la suite de cette thèse. Il a fait l'objet d'un article publié dans la revue *Methods in Ecology and Evolution* au mois de Septembre 2018.

Dans un second temps, j'ai cherché à évaluer si les sites de collecte concernés par une forte pression de pêche, localisée au niveau des villages proches du cours principal du Haut-Maroni (Annexe III), présentaient un appauvrissement de diversité génétique par rapport aux sites de collecte localisés hors des zones de pêche. Je me suis appuyée sur la théorie selon laquelle la diversité génétique, naturellement et en l'absence de facteur antagoniste, ne se répartit pas toujours de façon uniforme au sein des réseaux hydrographiques dendritiques (*dendritic-ecological-network*, DEN) mais augmente dans les zones de plus forte connectivité telles que les confluences ou le cours d'eau principal, là où la pression de pêche se trouve être la plus

forte au sein de la zone d'étude. En combinant des modèles de partitionnement de variance à des approches classiques de génétique des populations, j'ai recherché si les espèces présentaient des tendances spatiales communes en termes de variabilité génétique intrapopulation et plus généralement si elles se conformaient ou non à la théorie DEN. Cette approche a permis de mettre en évidence à la fois des patrons communs et espèce-spécifiques, dont je propose une interprétation à la lumière conjointe des connaissances disponibles sur les traits d'histoire de vie des espèces (notamment le comportement de migration des adultes et les stratégies de reproduction) et leur intérêt pour les pêcheurs qui les recherchent de façon plus ou moins importante. Les résultats de cette étape de travail sont présentés dans le Chapitre 2 de cette thèse, centré sur un manuscrit d'article en préparation pour la revue *Molecular Ecology*.

Dans un troisième temps et considérant qu'une pression de pêche, même géographiquement concentrée, pouvait n'être pas immédiatement détectable localement, je me suis interrogée quant à l'évolution temporelle de la variabilité génétique dans une métapopulation exploitée de type DEN. A l'aide d'un outil de simulation individu-centrée (*forward-in-time*), j'ai commencé à analyser l'évolution théorique d'un patron de variabilité génétique dendritique dans le cas où une pression de pêche cible les zones de plus forte connectivité. Dans le Chapitre 3 de ce manuscrit, je présente les résultats préliminaires de cette approche ainsi que ses perspectives d'approfondissement.

Je clôture ce manuscrit par une discussion générale synthétique revenant sur les résultats-clé de l'ensemble de ce travail et leurs implications pour la gestion des ressources halieutiques du Haut-Maroni. J'intègre également ces éléments à la lumière des connaissances plus largement acquises au cours du programme Pêche développé par le Parc Amazonien de Guyane et dont ce projet de thèse fait partie intégrante.

Chapitre 1:

Développer les ressources moléculaires nécessaires pour l'étude de la variabilité génétique chez des espèces non-modèles

1. Résumé étendu en français

La génétique des populations comparée entre plusieurs espèces constitue une approche émergente pour appréhender les questions relatives à la biologie de la conservation et à l'écologie des communautés. Elle peut permettre la mise en évidence de mécanismes communs régissant les niveaux de diversité génétique et spécifiques à l'échelle d'une communauté, l'analyse des patterns de variabilité génétique intra- et inter-populations des espèces en lien avec leurs caractéristiques écologiques ou biologiques, ou encore d'entrevoir de quelle façon divers facteurs environnementaux peuvent prédire les patrons de diversité génétique globaux. Cependant, de telles approches restent à ce jour restreintes aux espèces pour lesquelles des ressources génétiques sont disponibles. L'utilisation de plusieurs milliers de marqueurs de type SNP (*single-nucleotide polymorphism*) développés via les technologies de génotypage-par-séquençage présente un avantage certain pour l'étude d'espèces non-modèles, mais garde un coût prohibitif lorsque plusieurs espèces nouvelles sont impliquées. En outre, ces approches à très haute densité de marqueurs ne sont pas nécessairement indispensables à l'étude des patrons de variabilité génétique neutre chez les populations naturelles, telles qu'elles peuvent être effectuées avec des marqueurs moléculaires plus classiques et bon marché comme les microsatellites lorsqu'ils sont disponibles. L'établissement d'un compromis entre la densité d'informations apportées par les nouvelles technologies de séquençage à haut débit, et la possibilité d'un développement simultané, rapide et bon marché de ressources génétiques pour plusieurs espèces, apparaît donc comme un objectif pertinent.

Dans cette étude, nous proposons une méthode permettant de développer plusieurs panels de marqueurs SNP, en nombre réduit mais suffisant pour des études classiques de génétique des populations (quelques dizaines à quelques centaines), pour plusieurs espèces et avec un temps et un budget réduits. Nous présentons et appliquons cette méthode en vue du développement de marqueurs SNP pour 40 espèces de divers taxa incluant plantes, invertébrés, poissons et mammifères. Les espèces choisies représentent les principaux modèles biologiques d'intérêt au sein de notre laboratoire et ont été choisies pour cette raison, dans une démarche transversale inter-équipes de recherche. Dans un premier temps, un mélange d'ADN (pool) est préparé pour chacune des espèces à partir d'extraits d'ADN génomique de 10 à 20 spécimens, couvrant la zone géographique d'intérêt, incorporés en quantités équimolaires dans le pool final. Pour la plupart des espèces, chacun des pools est ensuite divisé en deux afin d'assurer la présence d'un réplicat (dans certains cas particuliers, un second pool d'ADN, différent, a été préparé indépendamment du précédent). Ces pools d'ADN sont ensuite utilisés pour la préparation d'autant de bibliothèques de séquençage RAD-Seq

(*restriction-site associated DNA sequencing*), chacune identifiée par un barcode moléculaire unique permettant de distinguer espèces et réplicats les uns des autres. L'ensemble des librairies est ainsi séquencé en « paired-end » sur un seul et même run Illumina HiSeq3000, et les données brutes obtenues (plusieurs millions de lectures individuelles de fragments d'ADN : *reads*) sont ensuite réassignées à leur librairie d'origine. Pour chaque librairie associée à son réplicat, un catalogue de RAD-tags (c'est-à-dire, de sous-ensembles de reads homologues dont le regroupement permet la reconstitution de différents loci présents dans le génome de l'espèce), est construit en utilisant le logiciel Stacks 1.37. Nous présentons alors un pipeline (bio-)informatique (Annexe VII) ayant pour but d'isoler, pour chaque espèce et à partir du catalogue de RAD-tags préalablement construit, 100 à 500 marqueurs SNPs de haute qualité: c'est-à-dire, pouvant ensuite fournir sans ambiguïté un génotype lisible de façon répétable entre individus, via diverses techniques impliquant une amplification par PCR. Ce pipeline se caractérise par l'intervention successive de trois filtres. Chaque filtre est associé à un petit nombre de paramètres modifiables par l'utilisateur, de façon à pouvoir augmenter ou diminuer la stringence de l'ensemble du processus et augmenter ou réduire le nombre final de marqueurs SNP isolés.

Un premier filtre s'appuie sur la base de données générée par Stacks. Il sélectionne, via une requête SQL et à partir du catalogue précédemment créé, des SNPs « candidats » vérifiant un certain nombre de critères. Les SNP candidats sélectionnés sont tous bialléliques (afin de favoriser l'étude de loci soumis à des modèles de mutation homogènes) et communs à chacune des deux librairies homologues. Chacun de leurs allèles a une profondeur de couverture minimale de 5X dans chacune des deux librairies. Ces deux dernières contraintes permettent d'augmenter la probabilité qu'un marqueur candidat corresponde réellement à un site variable existant dans le génome, et d'éviter ainsi les faux-positifs consécutifs à d'éventuelles erreurs de séquençage. Les séquences flanquant la position d'un SNP candidat doivent être d'une taille suffisante pour permettre le dessin de couples d'amorces PCR très spécifiques, et ne pas comporter elles-même de polymorphisme. Afin d'éviter un déséquilibre résultant d'une liaison physique entre marqueurs, deux SNPs candidats ne peuvent pas se trouver sur le même RAD-tag. Les séquences consensus de chaque RAD-tag associé à un SNP sélectionné sont ensuite exportées dans un fichier *fasta*, qui servira de support de référence au filtre suivant. Le second filtre procède alors au réalignement des reads single-end (*reads 1*) issus du séquençage contre les séquences consensus exportées précédemment. Cette étape a pour objectif de cibler et d'éliminer, autant que possible, tout SNP candidat dont la séquence associée présenterait des homologies partielles avec d'autres portions de reads ne provenant pas du même locus. De cette façon, nous souhaitons favoriser au maximum une amplification très spécifique de chaque SNP identifié, et éviter les amplifications croisées. Le second filtre permet également de valider le statut biallélique de chaque SNP candidat, et de vérifier de façon plus fine l'absence de polymorphismes flanquants qui n'auraient pas été détectés par le logiciel Stacks. Suite à une nouvelle exportation *fasta* des marqueurs ayant passé ce second filtre, un troisième et dernier filtre utilise cette fois-ci l'information, plus résiduelle, des reads paired-end (*reads 2*) issus du séquençage afin d'achever la détection des polymorphismes flanquant de part et d'autres de chacun des SNPs candidats. En effet, dans certains cas, la séquence des reads 2 peut chevaucher celle des reads 1, fournissant ainsi une profondeur de couverture

supplémentaire à l'analyse des séquences flanquantes d'un SNP candidat. Chaque SNP candidat ayant passé avec succès les trois étapes du pipeline bioinformatique sera référencé dans un fichier *fasta* final contenant la séquence consensus du RAD-tag correspondant et qui servira de référence pour la mise au point de couples d'amorces de génotypage (Un exemple est disponible Annexe VIII).

Le séquençage RAD a fourni des données brutes en quantité variable d'une espèce à l'autre. Néanmoins, l'application du pipeline a permis d'isoler entre 130 et 3492 SNPs pour 30 des 40 espèces focales. Les 10 espèces à moindre succès sont, pour la plupart, des végétaux ou des invertébrés pour lesquels des caractéristiques particulières du génome (e.g., méthylation, polyploidie, hypervariabilité) peuvent être soupçonnées. Des paramètres moins strigents permettent cependant d'augmenter le nombre de SNPs isolés pour certaines de ces espèces.

La validité et la qualité des marqueurs SNP ainsi isolés a été éprouvée via le génotypage de quelques dizaines d'entre eux pour plusieurs espèces focales de cette étude. Plus précisément, 124 à 160 SNPs fournis par le pipeline bioinformatique ont été génotypés pour 190 à 284 individus chez 18 espèces de poissons néotropicaux, échantillonnées sur le bassin versant du Maroni en Guyane française (Amérique du Sud). Le génotypage est effectué après avoir constitué, pour chaque espèce, quatre multiplexes de 35 à 40 SNPs chacun, et via la technologie de spectrométrie de masse MassARRAY iPLEX (Agena Bioscience) dont une présentation plus détaillée est disponible en Annexe IV. La procédure de génotypage appliquée aux 18 espèces du bassin du Maroni a montré un taux de succès élevé des marqueurs SNPs retenus à l'issue du pipeline. Les cas d'échec les plus importants s'expliquaient généralement par la survenue d'événements ponctuels de contaminations croisées, qu'il a été possible de contourner (Annexe V) afin d'obtenir un nombre de SNP validés à la hauteur de nos attentes. En moyenne, 85% des marqueurs étaient amplifiés avec succès chez au moins 90% des individus génotypés pour chacune des espèces, ce score pouvant monter jusqu'à 96% chez certaines espèces. La grande majorité (>98%) de ces marqueurs était à l'équilibre de Hardy-Weinberg dans chacun des sites d'échantillonnage du bassin du Maroni. Une analyse de variance moléculaire (AMOVA) a permis de détecter, sur la base des marqueurs validés, une structuration génétique significative entre tributaires pour les espèces dont la biologie permettait de s'attendre à l'existence d'une variabilité génétique spatiale.

Grâce à la méthode proposée dans ce travail, le coût financier du développement de marqueurs SNPs, tenant compte de la préparation des librairies et de la procédure de séquençage Illumina (sans tenir compte des coûts d'extraction d'ADN), avoisine 100 euros pour chacune des 40 espèces. La procédure de génotypage consécutive avoisine 6200 euros pour chacune des 18 espèces de Guyane française, avec en moyenne 200 individus génotypés par espèce, pour un rendement élevé entre le nombre de marqueurs génotypés et le nombre de marqueurs se révélant informatifs pour des études classiques de génétique de populations. Suite à l'obtention des données brutes issues du séquençage Illumina, quelques heures à quelques demi-journées suffisent désormais à en isoler des marqueurs SNP de qualité pour plusieurs espèces, en fonction des caractéristiques inhérentes à chacune d'elles (succès de séquençage, particularités du génome, etc, susceptibles d'allonger notamment la phase de traitement par le logiciel Stacks), le pipeline informatique lui-même ne demandant

que quelques minutes de calcul pour chacune des espèces. La stratégie proposée s'avère donc efficace pour développer jusqu'à plusieurs centaines de marqueurs SNPs pour une large gamme d'espèces non-modèles, contribuant ainsi à l'élargissement des perspectives d'étude de questions écologiques ou évolutives, comparée ou non entre espèces, et qui ne nécessiteraient toutefois pas d'approche génomique à très haute densité de marqueurs.

2. Article n°1 :

A cost-and-time effective procedure to develop SNP markers for multiple species: A support for community genetics

Delord C.^{1,2}, Lassalle G.¹, Oger A.¹, Barloy D.¹, Coutellec M.A.¹, Delcamp A.³, Evanno G.¹, Genthon C.⁴, Guichoux E.³, Le Bail P.Y.⁵, Le Quilliec P.¹, Longin G.⁶, Lorvelec O.¹, Massot M.³, Reveillac E.¹, Rinaldo R.⁶, Roussel J.-M.¹, Vigouroux R.², Launey S.¹, Petit E.J.¹.

¹ESE, Ecology and Ecosystem Health, Agrocampus Ouest, INRA, 35042, Rennes, France

²HYDRECO Guyane SARL, Laboratoire-Environnement de Petit Saut, 97310, Kourou, France

³BIOGECO, INRA, Université de Bordeaux, 33610, Cestas, France

⁴Plateforme génomique INRA GenoToul, 31320, Castanet-Tolosan, France

⁵LPGP, INRA, 35042, Rennes, France

⁶PARC AMAZONIEN DE GUYANE, 97354, Rémire-Montjoly, France

Publié dans la revue **Methods in Ecology and Evolution**, Issue 9 (9), pp 1959-1974.

(<https://doi.org/10.1111/2041-210X.13034>)

Le matériel supplémentaire associé à cet article est consultable Annexe VI.

La première version des scripts rédigés en vue du développement du pipeline informatique présenté ici est consultable en ligne à l'adresse suivante : <https://doi.org/10.5281/zenodo.1252193>, ainsi qu'en Annexe VII.

2.1. Abstract

Multi-species population genetics is an emerging field that provides insight relevant to conservation biology and community ecology. However, to date, this approach is limited to species with available genetic resources. The use of thousands of single nucleotide polymorphism (SNP) markers developed from recent genotyping-by-sequencing (GBS) technologies is a roadmap for the study of non-model species, but remains cost prohibitive when several, distantly related species are involved. We aimed to overcome this issue by using a single HiSeq3000 run of restriction-site associated DNA sequencing (RAD-Seq) to retrieve SNP markers for 40 diverse species including plants, invertebrates, fish and mammals. We developed a Python-based pipeline to isolate ~100-500 high-quality SNP markers for each species that could be genotyped through classical PCR amplification methods. To assess the quality of these markers, we validated our approach on ~160 of the characterized SNPs for each of 18 Neotropical fish species from the river Maroni (French Guiana, South America), using the MassARRAY iPLEX platform from Agena Bioscience (San Diego, CA, USA). A run of the pipeline applying stringent filtering parameters enabled the successful design of between 130 and 3492 SNP markers for 30 of the 40 study species. Relaxing pipeline parameters allows for an increase in the number of detected SNPs. Across the 18 species from French Guiana, an average of 85% of markers were successfully amplified, polymorphic, and scored in $\geq 90\%$ of individuals (~200 individuals per species). The great majority (>98%) of these markers were at Hardy-Weinberg equilibrium in each sampling site from the river Maroni. This SNP discovery was performed at the cost of ~\$US110 for each of the 40 species. Genotyping was performed at the cost of ~\$US6000 for each of the 18 fish species with an average of 200 individuals per species. This strategy was found cost-and-time efficient to develop hundreds of SNP markers for a large range of non-model species, which can be used to investigate ecological and evolutionary questions that do not require whole-genome coverage.

Keywords: community genetics, comparative genetic studies, Guiana shield fish, MassARRAY, *non-model* organisms, pipeline, RAD-Seq, SNP development.

2.2. Introduction

Community genetics usually focuses on multi-species approaches to decipher the influence of intraspecific genetic variation and local adaptation on community structure (Neuhauser et al., 2003; Pauls et al., 2014). However, another critical field of inquiry in community genetics aims to unravel the drivers of neutral genetic diversity across taxa (Crutsinger, 2016). For instance, positive relationships between intraspecific genetic diversity and local species richness could be observed within an ecosystem and is referred to as species-genetic diversity correlation (SGDC, Vellend and Geber, 2005; Lamy et al., 2017). Although the underlying mechanisms of SGDC have not been fully elucidated yet, examples from freshwater fish (Fourtune et al., 2016), macroinvertebrates (Seymour et al., 2016), or a coral reef community (Selkoe et al., 2016) have highlighted the interest of studying SGDC for conservation perspectives, particularly in systems facing anthropogenic or natural disturbances (Evanno et al., 2009; Frey et al., 2016). Likewise, life history traits are worthy of interest to explain neutral patterns of genetic diversity (Ellegren & Galtier, 2016; Romiguier et al., 2014). Variation in fecundity, mean body size or ecological specialization might partly explain genetic variability across populations (Blanchet et al., 2010; Khimoun et al., 2016) and studies have tested whether neutral patterns of genetic variation were consistent across species displaying similar life history traits (Hung et al., 2017; Selkoe et al., 2014). However, comparative population genetics (or ‘macrogenetics’, Blanchet et al., 2017), requires the availability of molecular markers. This is not trivial when working with non-model organisms lacking genomic information. Mitochondrial DNA or other single-locus genes have been successfully used for phylogeographic analyses (Hung et al., 2016), but provide low levels of resolution for questions regarding intra-population diversity (but see Papadopoulou et al., 2011). Furthermore, and despite recent applications (e.g., De Barba et al., 2017), the development of classical markers such as microsatellites remains time-prohibitive to complete a multi-species study within the short time-frame relevant to management.

The emergence of modern molecular biology techniques for population genomics, especially high-throughput sequencing (HTS), allows accessing genomic information for species without available molecular resources. Restriction-site associated DNA sequencing (RAD-Seq, Baird et al., 2008) is a widely employed genotyping-by-sequencing (GBS, Narum et al., 2013) method for single-nucleotide polymorphism (SNP) discovery and direct genotyping. However, although such techniques have become increasingly affordable through sequencing costs reduction, expanding their use to several species at a time would tremendously increase the number of individuals and libraries to be processed, and exceed a typical research grant. Furthermore, a genome-wide dataset is not necessarily essential for comparative population genetics application, where – as least currently – most concerns rely on approaches typically performed with low-density and low-cost markers, such as microsatellites. Pool-Seq, i.e., sequencing a DNA mixture of individuals from the same geographical site and pooled in the same RAD library (Futschik & Schlötterer, 2010), was partly developed to reduce the costs of the original RAD-Seq technique whilst also providing a reliable coverage of genome variability. However, Pool-Seq requires a large number of individuals to be mixed together in an equimolar way, along with a high depth of coverage, otherwise allele frequencies inferred from the number of reads might be strongly biased due to ascertainment bias (Schlötterer et al., 2014). Moreover, as genotypic information on an individual-level is not provided, Pool-seq cannot yield population genetic indices such as inbreeding levels or gametic disequilibrium.

Thus, whilst classical molecular resources (e.g. microsatellites) are efficient but require fastidious development, the opportunities opened by GBS are promising though somehow outsized or unfitted for multi-species population genetics. Alternatively, HTS approaches could help to develop a few tens to hundreds of SNPs markers, possibly more efficiently than Sanger sequencing did in the past (Garvin et al., 2010). Thrasher et al. (In press) used ddRAD to perform relatedness analyses based on 150-600 optimized SNP markers. Other authors characterized small-sets of SNP markers to be genotyped through a classical, affordable technology compared to GBS for large numbers of samples (e.g., Ackerman & Campbell, 2012; Westgaard et al., 2017). Mixing and sequencing DNA of individuals from a SNP discovery-set is technically close to Pool-Seq, and represents a valuable way to access molecular resources. However, to our knowledge, such approaches usually focus on a single species. Adapting this strategy for several species, without increasing the financial and technical cost of SNP-design, could be widely useful.

We aimed to produce a set of high-quality SNPs for 40 species from different taxa in parallel. We used a small-scale RAD-Seq protocol to sequence DNA-pools of individuals from all 40 species at once. We then used a custom pipeline to isolate for each species independently, ideally 100-500 SNPs to provide species-specific molecular resources. We validated a subset of markers through MassARRAY mass-spectrometry genotyping, and basic population genetics analyses, for 18 fish species all from the same geographical area (river Maroni, French Guiana, South America). We expect this strategy to be a cost-effective way to simultaneously develop molecular resources for several species.

2.3. Material and Methods

DNA samples

We collected DNA from 40 species currently of interest for our lab's current research projects. These species cover a wide taxonomic range (Table C1.1) and embed a subset of 18 fish species all collected from French Guiana that were used for validating our approach through MassARRAY genotyping.

For each species, samples used for SNP design consisted of three to ~50 individuals collected at different geographical locations (details in Table C1.1), in accordance with recommendations from Morin et al. (2009). Samples consisted of white muscle, skin or fin fragment for all animals; a mixture of leaves for mosses (each comprising ~10 to ~20 individuals – approximation due to the small size of individuals hampering precise counts); leaf fragments for the two *Ludwigia* Sp. (each sample containing tissue from a full clonal population); or aqueous suspensions for cyanobacteria. Samples were stored in 99% ethanol and/or frozen at -20°C.

DNA from epithelial samples was isolated using the extraction kit NucleoSpin® 96 Tissue (Macherey-Nagel) or NucleoSpin® 96 Food (for high fat-level tissue), following the manufacturer's protocol. We used the NucleoSpin® Plant Kit for cyanobacteria and plant samples. We assessed DNA yield and quality by fluorescence (Qbit 2.0, Thermo Fisher Scientific) and spectrophotometry (DS-11, DeNovix) in order to check for the conformity of DNA with Illumina platform requirements.

Table C1.1: Phylogenetic status of sampled species, geographic origin and number N_s of samples included in the pools used for building the libraries. †: 5 and 3 samples each from a distinct, full-clonal population. See text for details.






Class, Order	Family	Genus and species (Vernacular name)	Sampling geographic coverage	N_s
Cyanophyceae,				
Oscillatoriales	Phormidiaceae	<i>Phormidium</i> Sp. (Cyanobacteria)	Normandie (Selune watershed)	~20
Equisetopsidae,				
Hypnales	Brachyteciaceae	<i>Rhynchostegium riparioides</i> (Long-beaked water feather-moss)	Normandie (Selune watershed)	~40
Isobryales	Fontinalaceae	<i>Fontinalis antipyretica</i> (Greater water-moss)	Normandie (Selune watershed)	~50
Lamiales	Plantaginaceae	<i>Callitriche hamulata</i> (Intermediate water-starwort)	Normandie (Selune watershed)	~70
Magnoliopsida,				
Myrtales	Onagraceae	<i>Ludwigia grandiflora</i> (Perennial water-primrose)	North-West France, Auvergne, Aquitaine	5†
		<i>Ludwigia peploides</i> (Floating primrose-widow)	Pays de la Loire, Occitanie	3†
Cephalopoda,				
Sepiida	Sepiidae	<i>Sepia officinalis</i> (Common cuttlefish)	Bay of Biscay (coastline)	20
				
Gastropoda,				
Stylommatophora	Helicidae	<i>Helix aspersa</i> (syn. <i>Cornu aspersum</i> - Garden snail)	Bretagne, Pays de la Loire	24
Hygrophila	Lymnaeidae	<i>Lymnaea stagnalis</i> (Great pond snail)	Renilys, Amsterdam & Sønderborg strains (cultures from INRA U3E, Rennes, France)	18
				
Malacostraca,				
Decapoda				
	Astacidae	<i>Pacifastacus leniusculus</i> (Signal crayfish)	Normandie	17
	Cambaridae	<i>Procambarus clarkii</i> (Red swamp crayfish)	Pays de la Loire	15
Petromyzonti,				
Petromyzontiformes	Petromyzontidae	<i>Petromyzon marinus</i> (Sea lamprey)	Bretagne, Rhone, Camargue, Centre (France), USA	40
				

Table C1.1 (Continued):





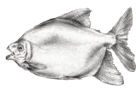





Actinopterygii,					
	Characiformes	Characidae	<i>Brycon falcatus</i> (Piabanha)	French Guiana (Maroni watershed)	18
			<i>Cynodon meionactis</i> (Minguilista)	French Guiana (Maroni watershed)	21
		Triportheidae	<i>Triportheus brachypomus</i> (Sardinha)	French Guiana (Maroni watershed)	15
		Erythrinidae	<i>Hoplias aimara</i> (Giant wolf fish)	French Guiana (Maroni watershed)	19
		Anostomidae	<i>Leporinus friderici</i> (Aracú-comum)	French Guiana (Maroni watershed)	19
			<i>Leporinus lebaili</i> (Leporinus de Le Bail)	French Guiana (Maroni watershed)	19
		Serrasalmidae	<i>Acnodon oligacanthus</i> (Pakousinn)	French Guiana (Maroni watershed)	20
			<i>Myloplus rhomboidalis</i> (Pacu-açu, Koumarou)	French Guiana (Maroni watershed)	16
			<i>Myloplus rubripimmis</i> (Pakousinn, Pakou-branco)	French Guiana (Maroni watershed)	18
			<i>Tometes lebaili</i> (Watau, Koumarou)	French Guiana (Maroni watershed)	15
			<i>Serrasalmus eigenmanni</i> (Pinch piranha)	French Guiana (Maroni watershed)	17
			<i>Serrasalmus rhombeus</i> (Black piranha)	French Guiana (Maroni watershed)	17
	Siluriformes	Auchenipteridae	<i>Ageneiosus inermis</i> (Slopehead or driftwood catfish)	French Guiana (Maroni watershed)	23
		Doradidae	<i>Doras micropoeus</i> (Botinho)	French Guiana (Maroni watershed)	17
		Loricariidae	<i>Harttia guyanensis</i> (Stick catfish)	French Guiana (Maroni watershed)	20
			<i>Hypostomus gymnorhynchus</i> (Pleco)	French Guiana (Maroni watershed)	17
			<i>Pseudancistrus barbatus</i> (Bearded catfish)	French Guiana (Maroni watershed)	21
	Clupeiformes	Clupeidae	<i>Alosa alosa</i> (Hallis shad)	Atlantic and Channel coast (France)	20
			<i>Alosa fallax</i> (Twate shad)	Atlantic and Mediterranean coast (France)	19
	Pleuronectiformes	Soleidae	<i>Solea solea</i> (Common sole)	English channel (East)	20
	Perciformes	Cichlidae	<i>Geophagus harreri</i> (Maroni eartheater)	French Guiana (Maroni watershed)	19
	Salmoniformes	Salmonidae	<i>Salmo salar</i> (Atlantic salmon)	Bretagne centre, Pyrénées (France)	20

Table C1.1 (Concluded):

Mammalia,					
	Carnivora	Herpestidae	<i>Urva auro-punctata</i> (Small indian mongoose)	French caribeans	9
	Chiroptera	Rhinolophidae	<i>Rhinolophus hipposideros</i> (Lesser horseshoe bat)	Bretagne and Lorraine (France), Germany, Ireland	11
	Soricomorpha	Soricidae	<i>Crocidura russula</i> (Greater white-toothed shrew)	Bretagne (France)	13
			<i>Crocidura suaveolens</i> (Lesser white-toothed shrew)	Bretagne (France)	13
	Rodentia	Muridae	<i>Rattus exulans</i> (Pacific rat)	Vanuatu and French Polynesia	20
			<i>Rattus rattus</i> (Ship rat)	Bretagne (France), French caribeans, Tunisia	19

We prepared two DNA pools (technical replicates) for all species. For 34 species, both pools were identical, for six species each pool contained different individuals (for future additional analyses that will not be detailed here). We manually prepared each pool by mixing DNA from the N_5 aforementioned samples in approximately equimolar amounts.

DNA was extracted from additional fish samples used for MassARRAY genotyping as described above. These datasets comprised 190 to 284 (N_T) samples, including those used for the SNP-discovery set, from each of 18 fish species (Table C1.2).

RAD-sequencing

Considering the wide taxonomic range of species and the absence of knowledge about their genomes, we assumed a mean genome size of 1 Gb and mean GC content of ~50% and chose the commonly used, high-fidelity SbfI restriction enzyme. A DNA library was prepared at the Get-Plage platform (Toulouse, France) from each pool with the usual precautions to avoid cross-contamination. All 80 libraries were sequenced together on an Illumina HiSeq3000 lane (150 bp paired-end reads). We expected our sequencing design to provide a mean per-library depth coverage of 20-60X.

Bioinformatics

Preliminary treatment of raw data and RAD-tag catalog building

Raw-reads were trimmed at 145 bp and demultiplexed using the `process_radtags` command from Stacks 1.37 software (Catchen et al., 2011). Low-quality reads (i.e., showing a mean Phred score $Q < 10$ upon a sliding window of 6 bp), or reads with an uncalled base were discarded. We eliminated PCR duplicates using the `clone_filter` command. Finally, we built a catalog of RAD-tag loci for each species separately (thus combining information from its two libraries) through single-end (SE) reads assembly and global alignment using the `denovo_map` command.

The following steps describe actions performed successively on each of the 40 species using a set of custom Bash and Python scripts as summarized in Fig.C1.1. Overall, the pipeline aims to provide informative and readily amplifiable markers for a given species. It uses each catalog of RAD-tag loci as a baseline to (1) pre-select candidate SNP markers for each species, and (2) only retrieve candidates that are the most likely to be unambiguously amplified by PCR. We ran the pipeline using default parameters for each of the 40 species successively. Indeed, a common configuration can initially be used to compare outputs across species and to speed up computation time when analyzing a large number of species. The pipeline can be run with modified parameter values for each species separately if needed. For instance, constraints could be relaxed if the final number of SNPs is too low (<100) or strengthened if the final number of SNPs is unnecessarily high.

First filtering: Candidate SNP selection

The objective of this first step was to select SNPs based on coverage, position and polymorphism. By default, selected SNPs had a minimum depth coverage of 5X for each allele within each library (pipeline parameter *depthAlleleMin*), and a maximum coverage of 100X (*depthAlleleMax*). SNPs had to be positioned at least at 30 bp from both sequence ends (-Pos5prim=30 and -Pos3prim=110) and had no flanking variant, to facilitate primer design. Selected SNPs were bi-allelic and found in both libraries of the focal species (*numSamplesHomologous=2* and *bothSamplePolymorphic=True*). Candidate SNPs were automatically exported in a fasta file along with their catalog ID and their complete 145 bp sequence. Generation of the fasta file marks the end of the first step of the pipeline.

Second filtering: Checking the specificity of SNP flanking sequences for good amplification

Stacks catalog assembly being based on full-length alignment, there may be homologies between shorter portions of sequences from two different RAD-tags. In addition, flanking sequences of selected SNPs may harbor polymorphisms that were initially not detected by Stacks because of the bounded maximum-likelihood SNP-calling model implemented in this software. This model is not able to detect variants with minor allele count (MAC) levels less than 10% (Cammen et al., 2015; Catchen et al., 2013). Both features could be detrimental for subsequent genotyping because they could lead to the design of unspecific or unreliable primers.

This second step in our pipeline searches for homologies between flanking sequences from all candidate markers and all available SE reads. The pipeline here uses the BWA-MEM algorithm (Li, 2013) and SAMtools (Li et al., 2009) to re-align SE reads against the previously exported candidates (seed length set to default value of 19 bp), and filters out any candidate sequence that displays homologies with reads that were not used to define the focal SNP while building Stacks catalogue. Homologies could be either imperfect matches in BWA full-length alignments, partial alignments, or a mix of both. The user could specify the maximum tolerated number of these “outlier” reads (*nbReadsSam*). We set the default to three, as we considered that one to three outlier reads would likely come from sequencing artifacts. Then, the pipeline launches the samtools pileup command and uses the output file to check for the absence of flanking variants. Again, the user could specify the maximum tolerated depth of coverage for such variants (*depthFlankMax*). Following this second step, the pipeline generates a new “SE-

validated" fasta file for the species, containing only sequences that successfully passed the filters.

Third filtering: Checking the absence of SNP flanking polymorphisms with PE reads

The new fasta file was used as a template to align paired-end (PE) reads using BWA-MEM and SAMtools, because PE reads may overlap with sequences issued from their corresponding SE reads, thus adding information about potential flanking polymorphisms. Again, the pipeline checked for the absence of flanking variants in the same way as in the previous filter on SE reads (*depthFlankmax*). Following this filter step, the output is a final fasta file of in silico validated SNP markers for the species, containing only candidates that successfully passed all three steps.

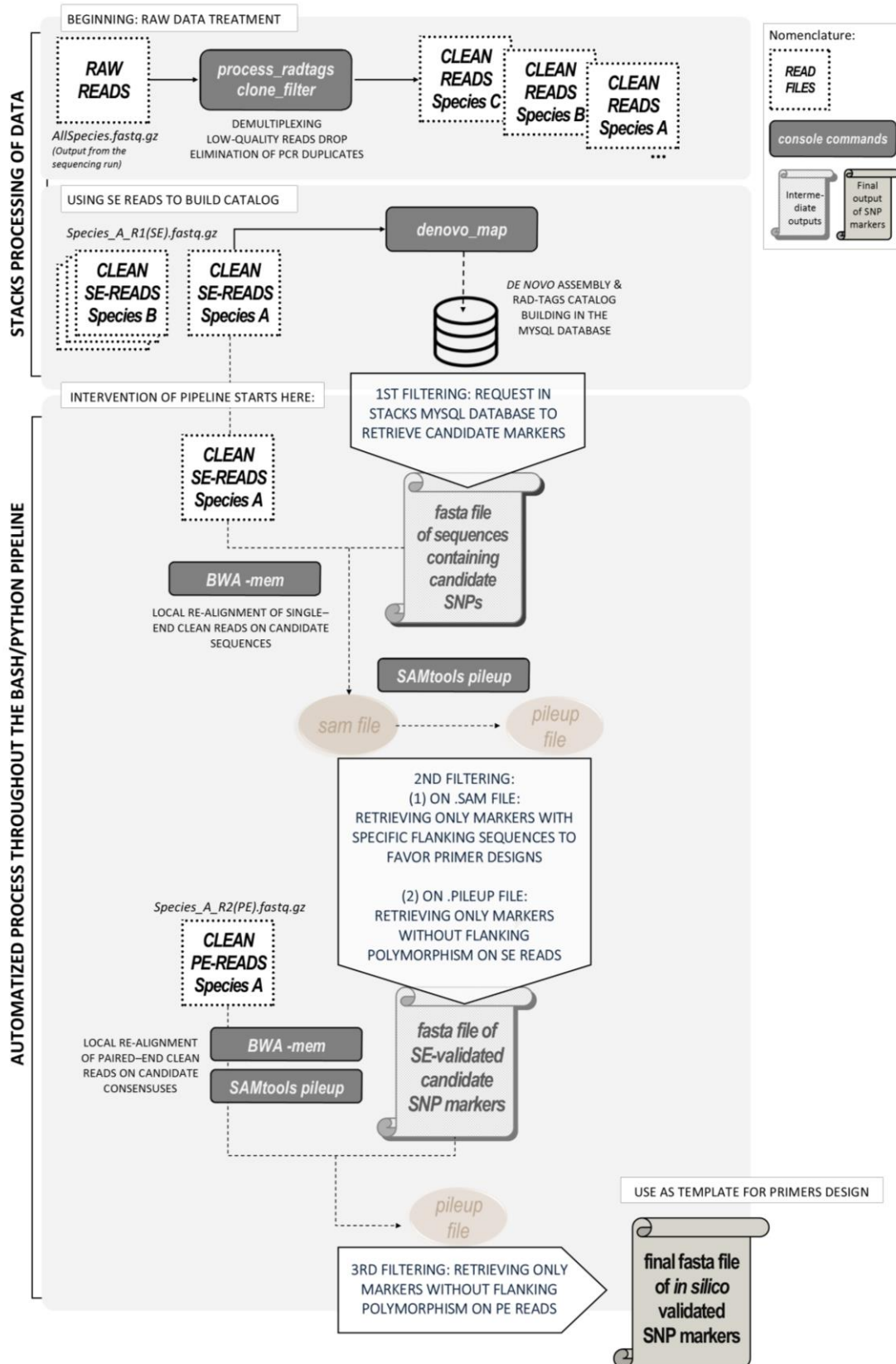


Figure C1.1: Summarized description of data treatment from reception of raw paired-end reads, to the obtaining of amplifiable single-nucleotide-polymorphism (SNP) markers. Previous demultiplexing, raw reads cleaning, and Stacks catalogue building need to be done manually before the automatized process throughout the pipeline begins.

Description of the final set of SNP markers

After we obtained a final list of markers for each species, we calculated the transition to transversion ratio (TS/TV ratio) as a rough indication of each final SNP set's quality, based on the assumption that TS/TV is $\sim 2:1$ for vertebrates (Smitz et al., 2016) and $\sim 1.5:1$ for invertebrates (Keller et al., 2007). According to Cooper & Krawczak (1989), a bias due to a higher rate of cytosine-to-thymine mutations increases the proportion of transitions upon transversions, even though at random we could expect the opposite since there are twice as many possible transversions than transitions. Thus, a significant amount of false-positive variants due to sequencing error would lead to random permutation between all four different nucleotides, so we would expect the TS/TV ratio to be closer to 1:2.

SNP genotyping through mass-spectrometry for fish from the Maroni

We validated our approach by genotyping 18 fish species from French Guiana at a subset of the markers that were previously developed. SNP markers were retrieved from the pipeline output with default parameters, or adapted when necessary, in order to reach 100-500 SNPs for each species to fit our MassARRAY assay design. In this latter case, we modified either *depthAlleleMin* (switching from 5 to 4) or *depthFlankMax* (switching from 2 to 3) parameters. Four multiplexes of 37 to 40 SNP primer sets were designed by the Genome Transcriptome Facility of Bordeaux (France) using Array Designer 4.0 (Agena Biosciences) for 157 to 160 SNPs, except for *Myloplus rubripinnis* (124 SNPs). Primer design typically targeted a mean amplicon length of 100-bp for elongation PCR, and 15 to 30 bp for extension PCR. We chose a constant mass difference of 25 daltons to discriminate elongation products from each other through mass spectrometry. Genotyping was performed on a MassARRAY system combined with iPLEX chemistry (Agena Biosciences, San Diego, CA, USA). Seventy-nine samples from various species were replicated and genotyped twice to check for reproducibility. We manually checked each genotype using MassARRAY TyperViewer 4.0 software (Agena Bioscience) to detect potential ambiguous genotype assignment or unusable SNPs.

Final evaluation of genotyping success and marker relevance

For each marker, we counted the percentage of individuals successfully genotyped. Next, we calculated the global genotyping success (GS) for a species across all markers as the proportion of SNPs that were (1) amplified (2) polymorphic with unambiguous genotype calls and (3) genotyped for at least 90% of individuals. All uninformative SNPs (either unamplified or displaying unclear genotype assignment), and monomorphic SNPs (minor allele frequency (MAF) being globally < 0.01 or null) were removed for subsequent population genetic analyses. SNPs genotyped in less than 60% of the whole set of NT individuals were also discarded from the analyses. We estimated local deviance from Hardy-Weinberg equilibrium (HWE) for each species and several (9 to 12) sampling sites from the river Maroni (Appendix S2) with a classical goodness-of-fit chi-square (χ^2) test using the packages *adegenet* (Jombart, 2008) and *pegas* 0.9 (Paradis, 2010) under the R software (version 3.3.2). For each sampling site, we computed 95% polymorphism rate P and expected heterozygosity H_S averaged across loci. We computed the Pearson's coefficient of correlation between P and the distance to the most downstream site, hereafter referred as COR.P. In river networks, we might expect a downstream increase of diversity for species with limited dispersal abilities (Paz-Vinas & Blanchet, 2015) so COR.P

should be negative. We performed an AMOVA analysis using the R package poppr (Kamvar et al., 2014) for each species across the study area to decompose the overall genetic variance into within-site, between-sites within a river tributary, and between-tributaries relative contributions.

2.4. Results

Raw data information and preparation

We obtained 582,194,084 (SE and PE) reads from the HiSeq3000 run. After demultiplexing, 494,361,178 reads were retained (the remaining were dropped due to ambiguous barcodes or Q-score <10). The number of reads varied from 0.38 to 4.81 million after PCR duplicate removal (Fig.2A). After RAD-tag assembly and catalog building, we obtained ~12,000 to ~133,000 RAD-tags per species (Fig.2B and Table C1.3). Mean RAD-tag depth coverage in a library ranged from 8.23X to 57.61X (Fig.2A).

SNP characterization for the 40 species set

The pipeline provided a highly variable number of markers across species (Table C1.3). We obtained ≥ 250 markers for 18 species and ≥ 100 for 12 other species. Five species yielded between 34 and 94 markers, and five yielded 11 markers or less. The first filter removed 97% of RAD-tags, on average, and led to a drastic loss of loci for the less successful species (e.g., for *Pacifastacus leniusculus* and most plant species). The loss of candidates was mainly due to a RAD-tag or a SNP being found in only one of the two libraries, usually linked to low depth of coverage. Also, the four lowest scores (lightest boxes of Fig. 2C) correspond to species with particularly low numbers of RAD-tags, and more importantly, low proportions of RAD-tags with only one SNP (Fig.2B). The second filter was also critical to the final number of SNPs and resulted in the elimination of 68% of the remaining markers, on average. The third filter led to the elimination of a much lower proportion (4%) of remaining markers, on average.

Mean values of the TS/TV ratio (\pm standard deviation) on final SNP markers were 2.08 ± 0.63 for vertebrate species (fish and mammals), and 1.52 ± 0.14 for invertebrate species for which the final SNP number was above 30 (*Helix aspersa*, *Sepia officinalis* and *Procambarus clarkii*). With the 200 to 514 SNPs obtained after a species-focused run of the pipeline on fish species from the Maroni, TS/TV ratio values were 1.98 ± 0.19 across all 18 species.

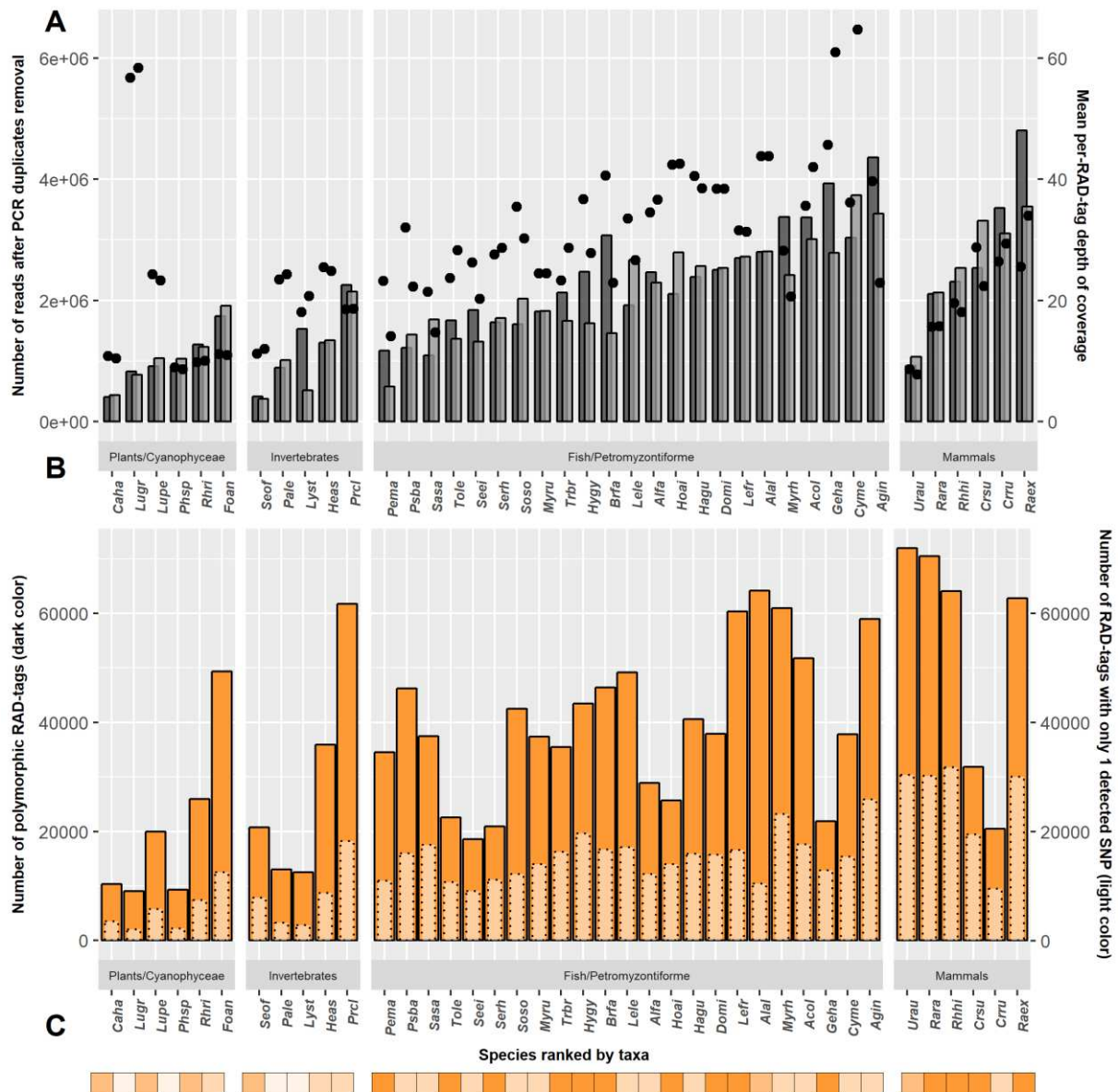


Figure C1.2: (A) Total number of paired-end reads after PCR duplicates removal, per species and per library (2 libraries/species) (grey bars) and mean per-RAD-tag depth of coverage, per species and per library (black dots). (B) Number of polymorphic RAD-tags per species (dark) and number of RAD-tags containing a single detected single nucleotide polymorphism (SNP) (light). (C) Colour scale indicating the success of markers design after pipeline processing. The darker the colour, the higher the number of in silico validated SNP markers. The colour scale has four shades corresponding to [0;10], [10;100], [100;500] and >500 respectively.

Table C1.2: Number of individuals and markers genotyped for each fish species from French Guiana using the MassARRAY facilities, and genotyping success (Gs, defined in the text) per species.

Species	Number of genotyped markers	Number of genotyped individuals (N _T)	Uninformative markers	Gs values (with % of genotyped markers)
<i>Brycon falcatus</i>	160	190	5	154 (96.3%)
<i>Cynodon meionactis</i>	160	208	3	150 (94.0%)
<i>Triportheus brachypomus</i>	160	186	10	146 (91.2%)
<i>Hoplias aimara</i>	160	236	6	151 (94.4%)
<i>Leporinus friderici</i>	160	210	15	143 (89.4%)
<i>Leporinus lebaili</i>	160	164	6	80 (50.0%)
<i>Acnodon oligacanthus</i>	159	235	11	136 (85.5%)
<i>Myloplus rhomdoidalis</i>	157	252	9	141 (89.8%)
<i>Myloplus rubripinnis</i>	124	236	6	104 (83.9%)
<i>Tometes lebaili</i>	160	190	4	146 (91.2%)
<i>Serrasalmus eigenmanni</i>	160	190	6	151 (94.3%)
<i>Serrasalmus rhombeus</i>	160	271	6	56 (35.0%)
<i>Ageneiosus inermis</i>	158	249	47	80 (50.6%)
<i>Doras micropoeus</i>	160	190	10	146 (91.2%)
<i>Harttia guianensis</i>	158	220	4	152 (96.2%)
<i>Hypostomus gymnorhynchus</i>	160	236	7	148 (92.5%)
<i>Pseudancistrus barbatus</i>	160	284	9	150 (93.8%)
<i>Geophagus harreri</i>	159	190	9	146 (91.8%)

Genotyping success and marker reliability for the 18 fish species from the Maroni

Between 124 and 160 markers were individually genotyped for each of the 18 fish species (Table C1.2). Fifty-six to 154 were polymorphic and successfully scored in more than 90% of individuals in a given species, corresponding to a global genotyping success (GS) of 35.0% to 96.3%. Lowest GS scores (<80%) were observed for *Ageneiosus inermis*, *Leporinus lebaili* and *Serrasalmus rhombeus*. For these species, all informative markers were still genotyped in more than 75% of individuals (except for 2 markers in *Serrasalmus rhombeus* and 7 markers in *Leporinus lebaili*, genotyped for 60–65% of individuals). We detected less than 4% divergence, on average, between two outputs from the same sample when checking for genotyping reproducibility, and most incongruences resulted from sporadic missing genotype calls. Four markers at maximum (i.e., for *Acnodon oligacanthus* and *Leporinus friderici*) were removed from species-specific datasets due to deviance from HWE in more than a third of the sampling sites. These corresponded to markers flagged as slightly ambiguous through MassARRAY TyperViewer analysis, but initially kept in the dataset.

Levels of diversity varied between species and sites, with a mean expected heterozygosity HS ranging from 0.257 to 0.427, and a within-site polymorphism rate P ranging from 0.69 to unity. Although within-site genetic variability explained the highest proportion of molecular variance, AMOVA also detected significant genetic differentiation between sites and/or between tributaries for several species (Fig.C1.3A). COR.P values were also variable across species, but tend to decrease with levels of population structure (Fig.C1.3B).

Table C1.3: Per-species evolution of the number of markers between RAD-tags assembly and pipeline output, with proportion of loss for each of the three filtering steps, when using identical, default parameter values of the pipeline for all species.

Species	Total number of RAD-tags	Proportion of marker loss at step 1	Proportion of marker loss at step 2	Proportion of marker loss at step 3	Final number of SNP markers (pipeline output):
<i>Pacifascatus leniusculus</i>	21976	1,000	-	-	0
<i>Phormidium</i> Sp.	15363	0,999	1,000	-	0
<i>Ludwigia grandiflora</i>	14233	0,995	0,987	0,000	1
<i>Lymnaea stagnalis</i>	18273	0,988	0,968	0,000	7
<i>Callitriche hamulata</i>	17672	0,999	0,522	0,000	11
<i>Ludwigia peploides</i>	34287	0,997	0,650	0,029	34
<i>Rhynchosstegium riparioides</i>	43325	0,998	0,478	0,028	35
<i>Sepia officinalis</i>	32625	0,997	0,518	0,100	36
<i>Urva auropunctata</i>	111080	0,999	0,240	0,026	37
<i>Alosa alosa</i>	53731	0,955	0,960	0,031	94
<i>Myloplus rubripinnis</i>	69971	0,981	0,897	0,071	130
<i>Pseudancistrus barbatus</i>	68258	0,987	0,810	0,049	155
<i>Acnodon oligacanthus</i>	76873	0,953	0,956	0,019	156
<i>Ageneiosus inermis</i>	152681	0,994	0,812	0,049	156
<i>Alosa falax</i>	53624	0,948	0,941	0,043	157
<i>Helix aspersa</i>	49489	0,982	0,788	0,064	176
<i>Myloplus rhomboidalis</i>	117084	0,984	0,900	0,043	178
<i>Fontinalis antipyretica</i>	73766	0,996	0,354	0,036	185
<i>Procambarus clarkii</i>	88249	0,991	0,760	0,031	186
<i>Cynodon meionactis</i>	93065	0,990	0,776	0,043	198
<i>Solea solea</i>	55667	0,964	0,896	0,029	203
<i>Salmo salar</i>	72520	0,992	0,610	0,095	210
<i>Harttia guianensis</i>	64799	0,951	0,918	0,027	253
<i>Serrasalmus eigenmanni</i>	58666	0,984	0,598	0,051	351
<i>Crossidura russula</i>	102427	0,993	0,467	0,045	359
<i>Leporinus lebaili</i>	74728	0,960	0,845	0,045	443
<i>Tometes lebaili</i>	55029	0,970	0,654	0,066	542
<i>Leporinus friderici</i>	83621	0,966	0,802	0,028	554
<i>Petromyzon marinus</i>	48083	0,979	0,412	0,045	570
<i>Doras micropoeus</i>	69898	0,959	0,739	0,047	706
<i>Brycon falcatus</i>	70049	0,958	0,716	0,030	810
<i>Serrasalmus rhombeus</i>	54679	0,962	0,553	0,030	905
<i>Triporthus brachypomus</i>	67920	0,957	0,664	0,034	937
<i>Hoplias aimara</i>	66167	0,942	0,678	0,050	1176
<i>Hypostomus gymnorhynchus</i>	79512	0,949	0,653	0,055	1319
<i>Rattus rattus</i>	127647	0,985	0,250	0,037	1369
<i>Geophagus harreri</i>	61088	0,912	0,632	0,033	1902
<i>Rhinolophus hipposideros</i>	122091	0,978	0,235	0,030	1994
<i>Crossidura suaveolens</i>	106948	0,951	0,426	0,038	2869
<i>Rattus exulans</i>	132013	0,949	0,463	0,027	3492

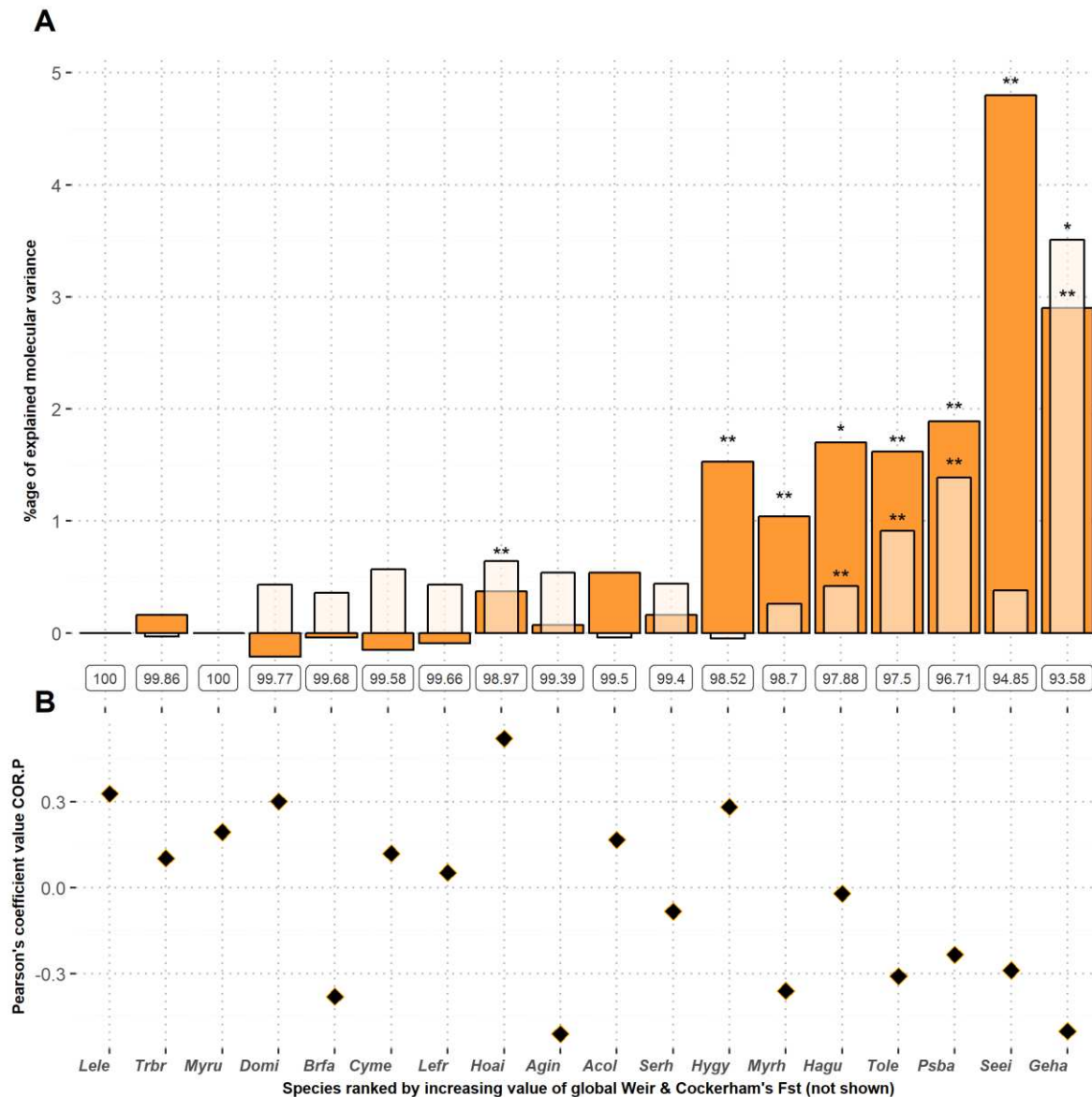


Figure C1.3 : (A) Analysis of molecular variance showing within-site (framed values), between-site within river tributary (light), and between tributaries (dark) sources of variation (* $p < 0.05$, ** $p < 0.01$). (B) Evolution of the Pearson's correlation coefficient between within-site polymorphism rate (P) and the distance to the most downstream site across species.

2.5. Discussion

Sequencing DNA from 40 species on a single HiSeq3000 lane enabled efficient SNP discovery for 30 of them (75%). Successful genotyping of the majority of the 124-160 SNPs for each of 18 fish species from the Maroni (on average 85% of markers being successfully amplified, polymorphic and scored in $\geq 90\%$ of individuals), illustrated the reliability of SNP selection. Most markers were at HWE and provided consistent information when applied to several fish populations within a river network.

Multi-species DNA sequencing

HTS or other approaches (e.g. Sanger sequencing, also see Bayerl et al., 2017 for a recent application using RAPD) have been widely used to develop SNP markers through cost-effective, semi-deep sequencing. Here, we validated the possibility of using a small-scale RAD-Seq

protocol to extend these applications to dozens of species at a time. Mean per-library depth of coverage reflected our expectations (20-60X) in most species, though this, along with number of raw reads and number of RAD-tags within the species catalog, could vary.

The first pipeline filter had the most drastic effect on SNP selection: across the 10 species with the lowest number of final SNP markers (<100), most RAD-tags were discarded at this step. Indeed, the number of sequencing reads after PCR duplicate removal was low for some of these species leading to a high number of RAD-tags displaying low depth of coverage. This phenomenon could result from the choice of the restriction enzyme. It should be noted that independent of any filtering step, the raw number of RAD-tags obtained is sensitive to intrinsic properties of both the species genome combined with those the restriction enzyme used, such as the GC content. Thus, using Sbf1 (the recognition site of which is 75% GC rich), fewer cutting sites were expected from genomes with lower GC content (e.g., *L. stagnalis*, 35%). Another explanation for lower success was a high variability or repetitiveness within the genome. This might have affected the efficiency of the enzyme digestion or amplification steps before sequencing (e.g., if there were too many cutting sites or restriction-site mutations generating allele dropout). This might also have led to over-splitting during catalog building (e.g., splitting reads over two or several different RAD-tags even though these were biologically coming from the same locus). Genome variability might also reduce the number of polymorphic RAD-tags containing a single SNP, which was also observed in some species with lower success, as perceptible in Fig.2B. Modifying pipeline parameters might increase success rate in some cases. Tolerating candidate loci that appear in only one of two replicated libraries (*numSamplesHomologous*=1), instead of considering only loci and SNPs present in both libraries, lead to >100 final SNP markers for eight on the 10 lower-success species. For the least successful species (no marker detected with the most stringent parameter set, see first row of Table C1.3), *P. leniusculus*, ongoing genotyping experiments of a set of 66 SNPs selected with this approach have already provided encouraging results (work in progress).

Finally (and preferably, when a basic knowledge about the species genome is available), prior adjustments in library preparation could avoid sequencing biases across taxa by providing a better-balanced sequencing coverage across species. Users could use different restriction enzymes and adaptors for library preparation before the common sequencing run, based on the available knowledge concerning the genome of the species of interest (e.g., see Herrera et al., 2015). For instance, a library preparation for plant species, separately from animals, could use a methylation-sensitive restriction enzyme to avoid cutting repetitive regions of the genome (Davey et al., 2011).

Sequencing success was acceptable for the other 30 species, yielding the targeted number of SNP markers. The TS/TV ratio for the final set of SNPs fitted our expectations for vertebrate and invertebrate species (test not performed for plants considering the low SNP number), corroborating that markers were not the result of sequencing errors. In the end, the number and identity of species and libraries to be sequenced together should only depend on (1) the desired final mean depth coverage for loci along with their respective alleles (e.g., expected MAC within a RAD-tag), and (2) the available knowledge about the characteristics of their genome (potentially interesting to optimize library preparation).

SNP-discovery for PCR genotyping: *in silico* validation

Whether developed from HTS to build high-density arrays (e.g., Faivre-Rampant et al., 2016), or for more classical amplification-based genotyping (e.g., Westgaard et al., 2017), potential SNP markers are first detected from the assembly of sequencing reads, which is performed by means of increasingly-argued and documented decision steps (Mastretta-Yanes et al., 2015; Shafer et al., 2016). However, to date, precise strategies that favor the successful amplification by PCR are case-dependent and without an existing, standardized framework. Westgaard et al. (2017) used RAD-Seq on DNA pools and Stacks software to isolate 200 SNP markers, but did not run further *in silico* filtering. Ninety-six SNPs were actually polymorphic and at HWE after MassARRAY genotyping, corresponding to a 50% genotyping success. When designing their SNP-set, Smitz et al. (2016) isolated 2,534 SNPs for non-model *Syncerus caffer* using the closely related *Bos taurus* reference genome, which also served for primer design. Fifteen of these markers were validated through Sanger sequencing with an 87% success rate (13 SNPs being successfully amplified and polymorphic), emphasizing the effectiveness of a reference genome when available. Without a reference genome, efficient SNP discovery for subsequent amplification-based genotyping must be performed with stringency, and *in silico* filtering is of high importance. Lepoittevin et al. (2015) illustrated this while building an Illumina high-density array for European white oaks. They used several thousand SNPs from both previous classical Sanger development (*in vitro*) and Roche 454 HTS technology (*in silico*). They observed a higher genotyping success for markers developed with the *in vitro* method, and pointed out several possible causes that might overcome the reliability of *in silico* isolated SNPs, including SNP flanking sequences of insufficient quality for primer design. As underlined, we designed our pipeline to maximize this quality as much as possible. Notably, the second step of our pipeline acts as a critical filter of candidate SNPs due to partial homology of their flanking sequences with other RAD-tags. Such cases pose a potential risk of false amplification if subsequent primer design is based upon these homologous sequences. Conversely, keeping only candidate SNPs with the most dissimilar flanking sequences avoids false amplification, and strongly facilitates the further building of reliable SNP multiplexes.

We avoided the use of probability-based filters for our *in silico* validation of candidate SNPs, giving priority to a small number of high-quality SNPs rather than an exhaustive selection of all potentially amplifiable SNPs amongst candidates. Consequently, drastic and non-probabilistic filtering might lead to the elimination of a number of potentially relevant SNP markers, which does not matter if the final number of markers remains around 100-500, as targeted. Slightly relaxing the pipeline parameters can increase the final number of SNPs if needed. The pipeline could easily be complemented with other analyses, e.g., when a reference genome is available.

SNP laboratory validation through mass-spectrometry genotyping for fish species from the Maroni

Fifteen of the 18 species displayed a GS score exceeding 80%, and three displayed GS between 35% and 50%. Low success in these species could a posteriori be explained by low quality samples (overall success increased from 35% to 80% when excluding 23 poor-quality samples, showing consistent failure, in *Serrasalmus rhombus*), technical issues on the MassARRAY platform (two of the four SNP multiplexes displayed lower success in *Leporinus lebaili*, where approximately 20% of individuals, at random for each locus, failed to amplify), or cross-contamination between two DNA libraries (the two DNA pools of the *Ageneiosus inermis* were

later found to be contaminated by DNA from the plant species *Ludwigia grandiflora*, leading to the design of primers for 47 SNPs that were not in the fish genome). Pairwise BLAST comparisons between RAD-tag loci for each possible pair of libraries across species (data not shown) detected no other similar issue in our dataset, making *Ageneiosus inermis* an isolated case whose low-success genotyping shows no relationship with marker design itself. We advise those using our pipeline to perform pairwise BLAST comparisons systematically in order to detect such an issue. In conclusion, technical or human artifacts explained most cases of unsuccessful genotyping.

All species with significant between-site or between-tributary molecular variance components are known to display low dispersal abilities linked to parental care behavior or a narrow habitat range. These species also displayed small, but negative COR.P values; thus, polymorphism tends to decline when moving upstream, as theoretically expected in dendritic ecological networks for species with limited gene flow. Thus, the developed markers provided sufficient information to highlight expected patterns with regard to biological knowledge about the focal species.

Cost/time-effectiveness and tractability as key advantages of the approach

Sequencing DNA from 40 species on a single HiSeq3000 lane enabled SNP discovery at the cost of ~\$US110 per species (DNA extraction not included). Individual genotyping of ~160 SNPs and ~200 individuals from each of the 18 fish species from the Maroni cost ~\$US107,000 in total. This corresponds to an average cost less than \$US6,000 per species and \$30 per sample. Although our approach ultimately yields less genomic information compared to high-throughput sequencing such as classical RAD-Seq, it is nevertheless a highly cost effective means of providing enough markers to detect patterns of spatial genetic variability even at small geographical scale for a number of non-model species. Once markers are available, genotyping per-sample costs then decrease with the number of individuals within a species: a study of 160 SNPs for 1000 samples would cost ~\$US10 per sample assuming easy access to MassARRAY facilities. In our design, since we had a limited number of samples per species, we assessed MassARRAY to be more affordable than, for instance, technologies based on targeted sequencing such as Tru-Seq or GT-seq (Campbell et al., 2015) which require the use of sequencing adaptors and barcodes. Similar methods are available for multiplexed genotyping, such as SNPstream (Beckman Fullerton, CA) or SNaPshot (Applied Biosystems, Foster City, CA). These also provide high accuracy and reproducibility, while GBS is sometimes associated with biases, e.g., locus or allele dropout (Cariou et al., 2016), generating data loss that might be difficult to handle (Huang & Knowles, 2016). In our case, the proportion of missing data across informative loci was relatively low. Furthermore, GBS analyses usually require accessibility to calculation clusters and bioinformatics skills regardless of the number of markers, which might increase time-investment in data analyses, whereas MassARRAY outputs are easily tractable on a classical spreadsheet format. In our design, after treatment of raw data, catalog building and automatic pipeline implementation on all species could be performed in a few hours on an 8-CPU, 32 GB cluster node. Genotyping of 200 individuals and 160 SNPs takes around three days.

Conclusion: opening opportunities for multi-species population genetics

Facilitating access to molecular resources for several non-model species is a first step towards a better understanding of patterns and mechanisms occurring in any species assemblage. In an era of high-density data acquisition, the availability of datasets for comparative population genetics has increasingly arisen as a key issue (e.g., Horne, 2014). Reconciling datasets obtained from different strategies and different periods for an integrated analysis is fastidious, particularly with HTS data where user-bias on bioinformatics processing could seriously affect subsequent population genetics analyses (Shafer et al., 2016) and comparisons, if no adequate precautions are taken (e.g., Campbell et al., 2017). Therefore, any multi-species data obtained from a common and standardized procedure could help resolve this issue. Here, SNP markers have been isolated using a standardized protocol and are easily transferable to any other laboratory, in an easier way than microsatellites since genotyping does not rely on an estimation of allele size (but see De Barba et al., 2017). Due to relatively small amplicon sizes, markers are also usable for degraded DNA, e.g., with samples from non-invasive protocols or from museum collections (Mendisco et al., 2011; Morin & McCarthy, 2007).

Our study showed the possibility of extracting a reliable number of SNPs from a straightforward RAD-Seq protocol with several species for population genetics applications, making the strategy usable in any assemblage or system where a multi-species approach is required. As a concrete outcome of our experiment, we provide 110 to 157 validated SNP markers for 18 non-model Neotropical fish species, usable for population genetics studies. Such molecular resources provide a basis for many applications: SNP markers have proven useful to assess gene diversity, population structure and effective size assessment, as well as relatedness and parentage analyses (Morin et al., 2004; Seeb et al., 2011; Weinman et al., 2015).

Acknowledgements

We are grateful to researchers, assistants, technicians, and students who helped with samples collection or laboratory tasks. We thank for the warm welcome from Wayana and Aluku fishermen who enabled the most efficient sample collection on the river Maroni. Samples from fish species of the Maroni were also collected in collaboration and with technical support from the National Amazonian Park of French Guiana, Hydreco Guyane SARL and the Natural History Museum of Geneva (MNHG), in accordance with ethical considerations defined in the convention APA-973-7. This study was funded by the National Amazonian Park of French Guiana under contract R&D-2003-06 to Pierre-Yves Le Bail and Jean-Marc Roussel with financial support from the DREAL and the Office de l'Eau de Guyane. This study was also funded by the Agence de l'Eau Seine-Normandie and the Association Nationale de la Recherche et de la Technologie. We thank Dr. Olivier Lepais (INRA Biogeco, Bordeaux, France), Dr. Emma L. Carroll (University of St Andrews, SOI, Scotland), Dr. Scott McCairns (INRA ESE, Rennes, France) and three anonymous associate editor and reviewers for providing valuable comments on previous versions of this manuscript.

Data accessibility

Custom Bash and Python scripts constituting the pipeline are fully available online on GitHub through the repository <https://doi.org/10.5281/zenodo.1252193>. Pipeline outputs (final fasta files of *in silico* validated SNPs) for each of the 40 species included in this study, along with technical information about SNP MassARRAY genotyping for fish species from French Guiana (including SNP multiplexes composition, amplification and extension primers sequences), are deposited in the Dryad repository <http://datadryad.org/resource/doi:10.5061/dryad.2b6b43k>.

2.6. References

- Ackerman, M., & Campbell, M. (2012). Chinook and steelhead genotyping for genetic stock identification at lower granite dam. Annual Progress Report July 1, 2010 - June 30, 2011. Submitted to Bonneville Power Administration. Contract #48347; Project # 2010-026-00
- Baird, N. A., Etter, P. D., Atwood, T. S., Currey, M. C., Shiver, A. L., Lewis, Z. A., ... Johnson, E. A. (2008). Rapid SNP Discovery and Genetic Mapping Using Sequenced RAD Markers. *PLoS ONE*, 3(10), e3376. <https://doi.org/10.1371/journal.pone.0003376>
- Bayerl, H., Kraus, R.H.S., Nowak, C., Foerster, D.W., Fickel, J. & Kuehn, R. (2017). Fast and cost-effective single nucleotide polymorphism (SNP) detection in the absence of a reference genome using semi-deep next generation random amplicon sequencing (RAMseq). *Molecular Ecology Resources*, In press. doi:10.1111/1755-0998.12717.
- Blanchet, S., Prunier, J. G., & De Kort, H. (2017). Time to Go Bigger: Emerging Patterns in Macrogenetics. *Trends in Genetics*, 33(9), 579-580. <https://doi.org/10.1016/j.tig.2017.06.007>
- Blanchet, S., Rey, O., Etienne, R., Lek, S., & Loot, G. (2010). Species-specific responses to landscape fragmentation: implications for management strategies. *Evolutionary Applications*, 3(3), 291–304. <https://doi.org/10.1111/j.1752-4571.2009.00110.x>
- Cammen, K.M., Schultz, T.F., Rosel, P.E., Wells, R.S., Read, A.J. (2015) Genomewide investigation of adaptation to harmful algal blooms in common bottlenose dolphins (*Tursiops truncatus*). *Molecular Ecology*, 24, 4697–4710.
- Campbell, E.O., Davis, C.S., Dupuis, J.R., Muirhead, K., & Sperling, F.A.H. (2017) Cross-platform compatibility of de novo-aligned SNPs in a non-model butterfly genus. *Molecular Ecology Resources* 17, e84-e93. <https://doi.org/10.1111/1755-0998.12695>.
- Campbell, N. R., Harmon, S. A., & Narum, S. R. (2015). Genotyping-in-Thousands by sequencing (GT-seq): A cost effective SNP genotyping method based on custom amplicon sequencing. *Molecular Ecology Resources*, 15(4), 855–867. <https://doi.org/10.1111/1755-0998.12357>
- Cariou, M., Duret, L., & Charlat, S. (2016). How and how much does RAD-seq bias genetic diversity estimates? *BMC Evolutionary Biology*, 16(1). <https://doi.org/10.1186/s12862-016-0791-0>
- Catchen, J. M., Amores, A., Hohenlohe, P., Cresko, W., & Postlethwait, J. H. (2011). Stacks : Building and Genotyping loci De Novo from Short-Read Sequences. *G3: Genes|Genomes|Genetics*, 1(3), 171–182. <https://doi.org/10.1534/g3.111.000240>
- Catchen, J., Hohenlohe, P.A., Bassham, S., Amores, A., Cresko, W.A. (2013) Stacks: an analysis tool set for population genomics. *Molecular Ecology*, 22, 3124–3140.
- Cooper, D. N., & Krawczak, M. (1989). Cytosine methylation and the fate of CpG dinucleotides in vertebrate genomes. *Human Genetics*, 83(2), 181–188.
- Crutsinger, G. M. (2016). A community genetics perspective: opportunities for the coming decade. *New Phytologist*, 210(1), 65–70. <https://doi.org/10.1111/nph.13537>
- Davey, J. W., Hohenlohe, P. A., Etter, P. D., Boone, J. Q., Catchen, J. M., & Blaxter, M. L. (2011). Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics*, 12(7), 499–510. <https://doi.org/10.1038/nrg3012>
- De Barba, M., Miquel, C., Lobréaux, S., Quenette, P. Y., Swenson, J. E., & Taberlet, P. (2017). High-throughput microsatellite genotyping in ecology: improved accuracy, efficiency, standardization and success with low-quantity and degraded DNA. *Molecular Ecology Resources*, 17(3), 492–507. <https://doi.org/10.1111/1755-0998.12594>

- Delord, C., Lassalle, G., Oger, A., Barloy, D., Coutellec, M., Delcamp A., Evanno, G., Genthon, C., Guichoux, E., Le Bail, P., Le Quilliec, P., Longin, G., Lorvelec, O., Massot, M., Reveillac, E., Rinaldo, R., Roussel, J., Vigouroux, R., Launey, S., Petit, E. Data from: A cost-and-time effective procedure to develop SNP markers for multiple species: a support for community genetics. Dryad Digital Repository. <https://doi.org/10.5061/dryad.2b6b43k>
- Ellegren, H., & Galtier, N. (2016). Determinants of genetic diversity. *Nature Reviews Genetics*, 17(7), 422–433. <https://doi.org/10.1038/nrg.2016.58>
- Evanno, G., Castella, E., Antoine, C., Paillat, G., & Goudet, J. (2009). Parallel changes in genetic diversity and species diversity following a natural disturbance. *Molecular Ecology*, 18(6), 1137–1144. <https://doi.org/10.1111/j.1365-294X.2009.04102.x>
- Faivre-Rampant, P., Zaina, G., Jorge, V., Giacomello, S., Segura, V., Scalabrin, S., ... Morgante, M. (2016). New resources for genetic studies in *Populus nigra* : genome-wide SNP discovery and development of a 12k Infinium array. *Molecular Ecology Resources*, 16(4), 1023–1036. <https://doi.org/10.1111/1755-0998.12513>
- Fourtune, L., Paz-Vinas, I., Loot, G., Prunier, J. G., & Blanchet, S. (2016). Lessons from the fish: a multi-species analysis reveals common processes underlying similar species-genetic diversity correlations. *Freshwater Biology*, 61(11), 1830–1845. <https://doi.org/10.1111/fwb.12826>
- Frey, D., Arrigo, N., Granereau, G., Sarr, A., Felber, F., & Kozłowski, G. (2016). Parallel declines in species and genetic diversity driven by anthropogenic disturbance: a multispecies approach in a French Atlantic dune system. *Evolutionary Applications*, 9(3), 479–488. <https://doi.org/10.1111/eva.12351>
- Futschik, A. & Schlötterer, C. (2010). The Next Generation of Molecular Markers from Massively Parallel Sequencing of Pooled DNA Samples. *Genetics*, 186(1), 207–218. <https://doi.org/10.1534/genetics.110.114397>
- Garvin, M.R., Saitoh, K., & Gharrett, A.J. (2010). Application of single nucleotide polymorphisms to non-model species: a technical review. *Molecular Ecology Resources*, 10, 915–934.
- Herrera, S., Reyes-Herrera, P.H., & Shank, T.M. (2015). Predicting RAD-seq marker numbers across the eukaryotic Tree of Life. *Genome Biology and Evolution*, 7(12), 3207–3225.
- Horne, J. B. (2014) Emerging patterns and emerging challenges of comparative phylogeography. *Frontiers of Biogeography*, 6(4), 166–168.
- Huang, H., & Knowles, L. L. (2016). Unforeseen Consequences of Excluding Missing Data from Next-Generation Sequences: Simulation Study of RAD Sequences. *Systematic Biology* 65(3):357–365. <https://doi.org/10.1093/sysbio/syu046>.
- Hung, C.-M., Drovetski, S. V., & Zink, R. M. (2016). Matching loci surveyed to questions asked in phylogeography. *Proceedings of the Royal Society B: Biological Sciences*, 283(1826), 20152340. <https://doi.org/10.1098/rspb.2015.2340>
- Hung, C.-M., Drovetski, S. V., & Zink, R. M. (2017). The roles of ecology, behaviour and effective population size in the evolution of a community. *Molecular Ecology*, 26(14), 3775–3784. <https://doi.org/10.1111/mec.14152>
- Jombart, T. (2008). adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, 24(11), 1403–1405. <https://doi.org/10.1093/bioinformatics/btn129>
- Kamvar, Z. N., Tabima, J. F., & Grünwald, N. J. (2014). poppr : An R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ*, 2, e281. <https://doi.org/10.7717/peerj.281>

- Keller, I., Bensasson, D., & Nichols, R. A. (2007). Transition-Transversion Bias Is Not Universal: A Counter Example from Grasshopper Pseudogenes. *PLoS Genetics*, 3(2), e22. <https://doi.org/10.1371/journal.pgen.0030022>
- Khimoun, A., Eraud, C., Ollivier, A., Arnoux, E., Rocheteau, V., Bely, M., ... Garnier, S. (2016). Habitat specialization predicts genetic response to fragmentation in tropical birds. *Molecular Ecology*, 25(16), 3831–3844. <https://doi.org/10.1111/mec.13733>
- Lamy, T., Laroche, F., David, P., Massol, F., & Jarne, P. (2017). The contribution of species–genetic diversity correlations to the understanding of community assembly rules. *Oikos*, 126(6), 759–771. <https://doi.org/10.1111/oik.03997>
- Lepoittevin, C., Bodénès, C., Chancerel, E., Villate, L., Lang, T., Lesur, I., ... Kremer, A. (2015). Single-nucleotide polymorphism discovery and validation in high-density SNP array for genetic analysis in European white oaks. *Molecular Ecology Resources*, 15(6), 1446–1459. <https://doi.org/10.1111/1755-0998.12407>
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., ... 1000 Genome Project Data Processing Subgroup. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25(16), 2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>
- Li, Heng. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv Preprint arXiv:1303.3997*. Retrieved from <https://arxiv.org/abs/1303.3997>
- Mastretta-Yanes, A., Arrigo, N., Alvarez, N., Jorgensen, T. H., Piñero, D., & Emerson, B. C. (2015). Restriction site-associated DNA sequencing, genotyping error estimation and de novo assembly optimization for population genetic inference. *Molecular Ecology Resources*, 15(1), 28–41. <https://doi.org/10.1111/1755-0998.12291>
- Mendisco, F., Keyser, C., Hollard, C., Seldes, V., Nielsen, A. E., Crubézy, E., & Ludes, B. (2011). Application of the iPLEXTM Gold SNP genotyping method for the analysis of Amerindian ancient DNA samples: Benefits for ancient population studies. *ELECTROPHORESIS*, 32(3–4), 386–393. <https://doi.org/10.1002/elps.201000483>
- Morin, P.A., Luikart, G., Wayne, R.K., SNP_Workshop_Group (2004). SNPs in ecology, evolution and conservation. *Trends in Ecology & Evolution*, 19, 208–216.
- Morin, P.A., Martien, K.K., Taylor, B.L. (2009). Assessing statistical power of SNPs for population structure and conservation studies. *Molecular Ecology Resources*, 9, 66–73.
- Morin, P. A., & McCarthy, M. (2007). Highly accurate SNP genotyping from historical and low-quality samples. *Molecular Ecology Notes*, 7(6), 937–946. <https://doi.org/10.1111/j.1471-8286.2007.01804.x>.
- Narum, S. R., Buerkle, C. A., Davey, J. W., Miller, M. R., & Hohenlohe, P. A. (2013). Genotyping-by-sequencing in ecological and conservation genomics. *Molecular Ecology*, 22(11), 2841–2847. <https://doi.org/10.1111/mec.12350>
- Neuhauser, C., Andow, D. A., Heimpel, G. E., May, G., Shaw, R. G., & Wagenius, S. (2003). Community Genetics: Expanding the Synthesis of Ecology and Genetics. *Ecology*, 84(3), 545–558. [https://doi.org/10.1890/0012-9658\(2003\)084\[0545:CGETSO\]2.0.CO;2](https://doi.org/10.1890/0012-9658(2003)084[0545:CGETSO]2.0.CO;2)
- Papadopoulou, A., Anastasiou, I., Spagopoulou, F., Stalimerou, M., Terzopoulou, S., Legakis, A., & Vogler, A. P. (2011). Testing the Species–Genetic Diversity Correlation in the Aegean Archipelago: Toward a Haplotype-Based Macroecology? *The American Naturalist*, 178(2), 241–255. <https://doi.org/10.1086/660828>
- Paradis, E. (2010). pegas: an R package for population genetics with an integrated-modular approach. *Bioinformatics*, 26(3), 419–420. <https://doi.org/10.1093/bioinformatics/btp696>

- Pauls, S. U., Alp, M., Bálint, M., Bernabò, P., Čiampor, F., Čiamporová-Zaťovičová, Z., ... Monaghan, M. T. (2014). Integrating molecular tools into freshwater ecology: developments and opportunities. *Freshwater Biology*, 59(8), 1559–1576. <https://doi.org/10.1111/fwb.12381>
- Paz-Vinas, I., & Blanchet, S. (2015) Dendritic connectivity shapes spatial patterns of genetic diversity: a simulation-based study. *Journal of Evolutionary Biology*, 28, 986–994. doi: 10.1111/jeb.12626
- Romiguier, J., Gayral, P., Ballenghien, M., Bernard, A., Cahais, V., Chenuil, A., ... Galtier, N. (2014). Comparative population genomics in animals uncovers the determinants of genetic diversity. *Nature*, 515(7526), 261–263. <https://doi.org/10.1038/nature13685>
- Schlötterer, C., Tobler, R., Kofler, R., & Nolte, V. (2014). Sequencing pools of individuals — mining genome-wide polymorphism data without big funding. *Nature Reviews Genetics*, 15(11), 749–763. <https://doi.org/10.1038/nrg3803>
- Seeb, J.E., Carvalho, G., Hauser, L et al. (2011) Single-nucleotide polymorphism (SNP) discovery and applications of SNP genotyping in nonmodel organisms: INTRODUCTION. *Molecular Ecology Resources*, 11, 1–8.
- Selkoe, K. A., Gaggiotti, O. E., ToBo Laboratory, Bowen, B. W., & Toonen, R. J. (2014). Emergent patterns of population genetic structure for a coral reef community. *Molecular Ecology*, 23(12), 3064–3079. <https://doi.org/10.1111/mec.12804>
- Selkoe, K. A., Gaggiotti, O. E., Treml, E. A., Wren, J. L. K., Donovan, M. K., Hawai'i Reef Connectivity Consortium, & Toonen, R. J. (2016). The DNA of coral reef biodiversity: predicting and protecting genetic diversity of reef assemblages. *Proceedings of the Royal Society B: Biological Sciences*, 283(1829), 20160354. <https://doi.org/10.1098/rspb.2016.0354>
- Seymour, M., Seppälä, K., Mächler, E., & Altermatt, F. (2016). Lessons from the macroinvertebrates: species-genetic diversity correlations highlight important dissimilar relationships. *Freshwater Biology*, 61(11), 1819–1829. <https://doi.org/10.1111/fwb.12816>
- Shafer, A.B.A., Peart, C.L., Tusso, S., Maayan, I., Brelsford, A., Wheat, C.W., & Wolf, J.B.W. (2017) Bioinformatic processing of RAD-seq data dramatically impacts downstream population genetic inference. *Methods in Ecology and Evolution* 8(8): 907-917. <https://doi.org/10.1111/2041-210x.12700>.
- Smitz, N., Van Hooft, P., Heller, R., Cornélis, D., Chardonnet, P., Kraus, R., ... Michaux, J. (2016). Genome-wide single nucleotide polymorphism (SNP) identification and characterization in a non-model organism, the African buffalo (*Syncerus caffer*), using next generation sequencing. *Mammalian Biology - Zeitschrift Für Säugetierkunde*, 81(6), 595–603. <https://doi.org/10.1016/j.mambio.2016.07.047>
- Thrasher, D. J., Butcher, B. G., Campagna, L., Webster, M. S. and Lovette, I. J. (In press), Double-digest RAD sequencing outperforms microsatellite loci at assigning paternity and estimating relatedness: a proof of concept in a highly promiscuous bird. *Molecular Ecology Resources*. doi:10.1111/1755-0998.12771
- Vellend, M., & Geber, M. A. (2005). Connections between species diversity and genetic diversity: Species diversity and genetic diversity. *Ecology Letters*, 8(7), 767–781. <https://doi.org/10.1111/j.1461-0248.2005.00775.x>
- Weinman, L.R., Solomon, J.W., & Rubenstein, D.R. (2015). A comparison of single nucleotide polymorphism and microsatellite markers for analysis of parentage and kinship in a cooperatively breeding bird. *Molecular Ecology Resources*, 15, 502-511.
- Westgaard, J.-I., Saha, A., Kent, M., Hansen, H. H., Knutsen, H., Hauser, L., ... Johansen, T. (2017). Genetic population structure in Greenland halibut (*Reinhardtius hippoglossoides*) and its relevance

to fishery management. *Canadian Journal of Fisheries and Aquatic Sciences*, 74(4), 475–485.
<https://doi.org/10.1139/cjfas-2015-043>

Chapitre 2:

Description des patrons de variabilité génétique chez 18 espèces de poissons d'intérêt halieutique à l'échelle du Haut-Maroni :

Approche comparative entre sites et entre espèces

1. Résumé étendu en français

Les milieux aquatiques néotropicaux abritent une large diversité ichtyque de plus en plus menacée par les activités anthropiques. En Guyane française, le bassin versant du Maroni constitue un support de vie quotidienne indispensable pour les communautés humaines qui l'habitent, notamment via l'activité de pêche de subsistance qui cible quelques dizaines d'espèces. Or, ce territoire se caractérise par un accroissement démographique élevé et une modification des pratiques traditionnelles de pêche. Une diminution des rendements est perçue depuis quelques années par les communautés natives, et la question de la vulnérabilité des populations de poissons face à l'exploitation par la pêche se pose. Cependant dans ces milieux, les outils de diagnostic visant à évaluer la dynamique temporelle des populations naturelles sont difficiles à mobiliser logistiquement. L'étude de la variabilité génétique neutre chez les populations permet, via un unique échantillonnage, d'appréhender l'influence des facteurs naturels ou anthropiques sur l'évolution de la variabilité génétique et les phénomènes d'érosion génétique dans une perspective de conservation.

La variabilité génétique et sa répartition spatiale peuvent dépendre des traits d'histoire de vie des espèces mais aussi de contraintes liées aux caractéristiques du paysage, qui peuvent limiter les flux de gènes ou les tailles efficaces locales. Ceci est particulièrement vrai pour les espèces aquatiques au sein des bassins versants, comme suggéré par la théorie des réseaux écologiques dendritiques (*dendritic-ecological-networks*, DEN). De par leur configuration spatiale, ces milieux ont souvent tendance à présenter une diversité génétique plus forte au niveau des zones de forte connectivité que sont les confluences ou le cours principal. Cependant, d'autres facteurs environnementaux locaux peuvent intervenir, comme la disponibilité des habitats favorables. Une surpêche localisée, en influençant potentiellement les tailles efficaces locales, peut entraîner des phénomènes d'érosion et de dérive génétique susceptibles d'influencer la variabilité génétique et donc le potentiel adaptatif des populations.

Nous proposons ici une description des patrons de variabilité génétique intra- et interpopulations chez plusieurs espèces de poissons du Haut-Maroni, afin de fournir des éléments de décision pour leur gestion et leur conservation. Nous comparons 18 espèces relativement abondantes sur l'ensemble de la zone d'étude, mais plus ou moins recherchées

par les pêcheurs et présentant des traits d'histoire de vie contrastés - notamment au niveau de leur stratégie de reproduction et de leur comportement migrateur. Ces espèces ont été co-échantillonnées sur 12 sites répartis sur le bassin versant du Haut-Maroni. Cinq des sites correspondent aux zones habitées (villages) proches du cours d'eau principal, où l'activité de pêche se concentre en grande majorité. Sept autres sites sont localisés plus en amont sur les quatre principaux tributaires du Haut-Maroni (Litani et Marouini à l'Ouest, Tampok et Waki à l'Est), où l'activité de pêche diminue, voire disparaît, graduellement vers l'amont (sites de référence). L'objectif général de cette étude est de savoir si, au niveau des villages, la pression de pêche a généré une perte localisée de diversité génétique, cela malgré leur positionnement central dans le réseau hydrographique. Parmi les espèces étudiées, celles qui sont fortement ciblées par la pêche et présentant des capacités de dispersion limitées sont a priori les plus susceptibles de subir une perte locale de variabilité génétique. A l'inverse, des flux de gènes plus importants chez les espèces plus dispersives pourraient effacer les effets de dérive locale, les rendant plus difficiles, voire impossibles, à détecter par simple comparaison spatiale des niveaux de variabilité génétique.

Nous nous appuyons sur les données génétiques générées au cours du travail présenté en Chapitre 1 et plus précisément sur les génotypes de 145 à 283 individus par espèce, échantillonnés sur 9 à 12 des sites présentés ci-dessus et caractérisés pour 111 à 188 marqueurs SNP bialléliques. Pour chaque espèce, nous calculons la diversité génétique ainsi qu'une estimation de la taille efficace sur chaque site. Nous analysons les niveaux de différenciation génétique globale et le degré d'isolement par la distance. Nous calculons également, pour chaque espèce, un indice de diversité génétique correspondant à la contribution locale, de chacun des sites échantillonnés, à la diversité génétique globale sur la zone d'étude. A l'aide de modèles linéaires mixtes, nous comparons alors les niveaux de diversité locale et d'isolement génétique moyens entre sites, toutes espèces confondues, afin de détecter d'éventuelles tendances spatiales générales, en lien avec la pression de pêche ou la nature dendritique du réseau. Des analyses de covariance nous permettent ensuite de tester, globalement et par espèce, l'évolution graduelle du taux de polymorphisme, du taux d'hétérozygotie attendu et de l'isolement génétique moyen en fonction de descripteurs du réseau hydrographique (gradient de distance à l'aval et tributaire d'origine). Nous comparons également les niveaux moyens d'hétérozygotie attendue par regroupements de sites (villages, sites de référence des tributaires de l'ouest du bassin versant, ou sites de référence des tributaires de l'est du bassin versant). Enfin, nous visualisons comment les différentes espèces focales se regroupent sur des patterns génétiques communs en fonction de quelques caractéristiques biologiques connues, à l'aide d'une analyse en composantes principales, intégrant divers indices intraspécifiques de variation génétique spatiale.

Nous avons observé des patrons de différenciation génétique variables d'une espèce à l'autre, allant d'une différenciation nulle à modérée à l'échelle de la zone d'étude. En moyenne, les villages présentent une contribution à la diversité génétique globale plus importante, ainsi qu'un isolement génétique moyen plus faible que les sites de référence, en cohérence avec la théorie DEN. Cependant, la diversité génétique plus faible dans les sites de référence est vérifiée uniquement au niveau des tributaires du Tampok et de la Waki. La contribution moyenne (et le degré d'isolement génétique, respectivement) est significativement plus faible (et plus fort) au niveau de ces deux tributaires, alors que peu de différences significatives sont

détectées entre les autres sites, qu'ils se situent ou non dans les zones de pêche. Les analyses espèce-centrées confirment que la différenciation génétique spatiale, lorsqu'elle existe, consiste généralement en un isolement moyen plus important des sites de référence en zones amont alors que la diversité génétique plus faible, attendue dans ces mêmes zones amont, ne concerne en réalité que les tributaires situées à l'Est de la zone d'étude. L'absence de données environnementales précises disponibles limite la portée des interprétations possibles pour ce phénomène. Cependant, nous soupçonnons l'effet d'un habitat moins accessible et/ou plus dégradé au niveau de ces tributaires. En effet, ils sont caractérisés par un étalement moins important du cours principal en saison des pluies, par des variations annuelles du niveau d'eau plus marquées qu'ailleurs et par la présence de plusieurs sites d'orpillage actifs ces dernières années, susceptibles d'augmenter le colmatage du substrat par dépôt de sédiments et d'augmenter la turbidité.

L'analyse en composantes principales suggère que les traits d'histoire de vie considérés dans cette étude expliquent assez bien le degré de structuration génétique spatiale observé chez les différentes espèces. En effet, les espèces dont les dispersions larvaires et adultes sont supposées faibles présentent une structuration spatiale plus marquée. Cependant, l'interprétation plus précise des contrastes observés entre certaines espèces, présentant pourtant des capacités théoriques de dispersion similaires, se heurte à l'absence de connaissances biologiques plus précises. Certaines espèces montrent ainsi des patrons de variabilité génétique inattendus au regard des connaissances actuelles sur leur stratégie de dispersion, indiquant que d'autres paramètres écologiques, actuellement peu ou mal connus, peuvent également entrer en jeu. Cela souligne la nécessité d'augmenter les connaissances sur la biologie et l'écologie des espèces, afin de pouvoir affiner les prédictions sur la sensibilité de chaque espèce à la pression de pêche et savoir si une différenciation génétique spatiale plus importante peut être directement imputable à l'exploitation des populations.

Globalement, nous détectons peu de signes d'un impact local important de la pression de pêche sur la diversité génétique. Cependant, nous constatons une diversité génétique plus réduite au niveau de certains sites pour certaines espèces. Ainsi, Papaïchton, un village plus densément peuplé et situé en aval de la zone plus urbanisée de Maripasoula, montre un isolement génétique moyen plus marqué pour l'espèce emblématique *Hoplias aimara* notamment, et les estimations de taille efficace y sont généralement plus faibles pour de nombreuses d'espèces. Le village de Kayodé, situé sur les tributaires du Tampok et de la Waki, pourrait également pâtir des niveaux de diversité génétique plus faibles rencontrés sur ces tributaires. Nous recommanderions ainsi que ces sites fassent l'objet d'un suivi régulier, voire de mesures de gestion adaptées. Nous recommanderions également de maintenir une vigilance pour certaines espèces très recherchées par les pêcheurs et montrant des patrons de variabilité génétique atypiques (*Hoplias aimara*) ou relativement structurés par rapport à d'autres et donc plus susceptibles de souffrir d'une érosion génétique localisée (*Tometes lebaili*).

Nous terminons en soulignant que l'absence de différence détectable de diversité génétique entre zones impactées ou non par la pêche, n'est pas synonyme d'absence d'effet de l'exploitation des ressources. Pour les espèces montrant une structuration génétique conforme à l'attendu dendritique (diversité génétique plus forte en aval), une baisse de

diversité due à la pêche peut logiquement ne pas être détectée si elle est récente et/ou suffisamment faible pour être compensée par les flux de gènes depuis l'amont. Pour les espèces ne montrant pas de structure génétique spatiale, les flux de gènes importants brouillent la comparaison entre des zones à pression de pêche variable. Ainsi, l'absence de différence de diversité génétique entre zones aval (pêchées) et amont pourrait être due aux effets contraires de la pêche (qui diminue la diversité génétique en aval) et des flux de gènes (qui l'augmente). Ces phénomènes, s'ils existent, ne sont donc pas discernables sur la base de nos seules données génétiques. Pour les détecter, il faudrait pouvoir quantifier la déviation de la structure génétique de la métapopulation par rapport à celle attendue en l'absence de pression de pêche. Cela pourrait être abordé en modélisant l'impact d'un prélèvement localisé et répété d'individus (par la pêche) au sein d'un réseau dendritique, via des outils de simulation et sous réserve de pouvoir quantifier précisément les tailles efficaces locales et les taux de migration entre dèmes. Un suivi temporel permettrait également de mieux comprendre la dynamique de la diversité génétique sur le long terme.

2. Article n°2 :

Riverscape patterns of within and between-population genetic diversity in eighteen harvested Neotropical freshwater fish species

Delord C.^{1,2}, Le Bail P.-Y.³, Fourcy D.¹, Longin G.⁴, Rinaldo R.⁴, Vigouroux R.², Roussel J.-M.¹, Launey S.¹

¹ESE, Ecology and Ecosystem Health, Agrocampus Ouest, INRA, 35042, Rennes, France

²HYDRECO Guyane SARL, Laboratoire-Environnement de Petit Saut, 97310, Kourou, France

³LPGP, INRA, 35042, Rennes, France

⁴PARC AMAZONIEN DE GUYANE, 97354, Rémire-Montjoly, France

En préparation pour la revue *Molecular Ecology*

Le matériel supplémentaire associé à cet article en préparation est consultable Annexe IX.

2.1. Abstract

The River Maroni (French Guiana, South America) hosts a high fish diversity, including numerous endemic species, and provides subsistence resources for a growing community of native people. The risk of overfishing is the subject of ongoing concern, including from native communities, but investigations by conventional tools for stock assessments cannot apply for such multiple-species fishery in a vast, Neotropical river context. We thus described patterns of genetic variability, across 111 to 188 SNP loci, for eighteen species co-sampled in five villages (high harvesting pressure) and seven reference sites (low to nil harvesting pressure) across the Upper Maroni. Using a comparative riverscape genetics approach, we highlighted species-specific patterns of genetic diversity, from nil to moderate genetic structure. In the latter case, we often observed higher mean local differentiation in upstream, less connected sites, in accordance with the theoretical expectation in a dendritic ecological network (DEN). This pattern was particularly accented in eastern tributaries, which offer reduced habitat quality and/or availability, and for species with reduced dispersal capacity at both early and late stage. Finally, there was no clear evidence of a genetic drift on populations located under fishing areas, except for one densely populated village, and for one large, highly emblematic species for which the effect of overfishing is suspected. In conclusion, riverscape, environmental constraints, and species ecology, seem the main drivers of genetic structuring in this area, rather than fishing pressure. However, our investigation on a multiple-species fishery in a large river system allows identifying species and locations of particular concern for future conservation and fisheries management plans.

Keywords: ichtyodiversity, anthropic impact, comparative population genetics, dispersion, Maroni

2.2. Introduction

The Neotropical freshwater fish fauna encompasses a high specific and functional richness (Abell et al. 2008), with more than 5,600 species described to date. This unique biodiversity heritage provides cultural and commercial resources for human societies (Camargo et al. 2015). However, ichthyodiversity is currently threatened by water pollution, habitat degradation, water damming for hydroelectricity and irrigation, and overexploitation (Reis 2013; Smith 1985), leading to high risk of species extinction (Barletta et al. 2010; Reis et al. 2016; Rodríguez-Olarte et al. 2011). Such conservation concern has drawn attention but to date, most studies have focused on the community level and described the diversity of species (Bayley & Petrere 1989; Tejerina-Garro et al. 2006; Agostinho et al. 2016; Cilleros et al. 2018), or more recently the diversity of species roles, i.e. functional diversity (Carvalho & Tejerina-Garro 2015; Toussaint et al. 2016). However, intraspecific investigations to assess populations vulnerability are lacking, notably with regards to metapopulation structures and dynamics (Abell 2002; Thieme et al. 2007). Conventional tools for population or stock monitoring, such as capture-mark-recapture or telemetry are simply not applicable or fatally flawed in Neotropical rivers due to harsh environmental conditions and logistical constraints that strongly limit recapture probabilities (Meunier et al. 1994; Antonio et al. 2007). Consequently, demographic parameters are virtually unknown for most species. Intraspecific genetic variation, however, can help decoding the past and current drivers influencing micro-evolutionary patterns. Genetic diversity supports species adaptability to changing environment, and its loss might have irreversible impacts on populations resilience (Frankham 1995; Allendorf et al. 2008; Walsh et al. 2006). Nevertheless, population genetics studies of Neotropical freshwater fish are scarce, and mostly focused on species of high commercial interest (Piorsky et al. 2008; Oliveira et al. 2009; see also Hilsdorf & Hallerman 2017 for a detailed overview).

Genetic variation is strongly shaped by inherent biological characteristics of species such as mating strategies (Charlesworth & Wright 2001) or dispersal abilities influencing the extent of gene flow between populations (Comte & Olden 2018a; 2018b). However, the environment at the landscape level also influences spatial genetic variation (Manel et al. 2003). This is particularly true for aquatic species living in hydrographic networks (or *dendritic-ecological-networks*, DEN, Campbell-Grant et al. 2007; Chaput-Bardy et al. 2009) as demonstrated through theoretical (Paz-Vinas & Blanchet 2015; Thomaz et al. 2016) and empirical (e.g., Paz-Vinas et al. 2015; Pilger et al. 2017) approaches. According to the DEN theory, genetic diversity increases around the main confluences usually in the median or downstream parts of the watershed, where higher connectivity leads to higher genetic diversity (Labonne et al. 2008). By the same logic, genetic diversity decreases upstream where gene flow (i.e., the effective exchange of genetic material between distant individuals) or effective size (i.e., the number of breeding individuals effectively contributing to the next generation) are theoretically too low to counterbalance local genetic drift. Thus, a DEN pattern is mostly expected for species with low to moderate dispersal abilities (Paz-Vinas & Blanchet 2015) or displaying reproductive strategies such as long-term parental care (nest building and guarding) with spatially restrained spawning sites. In addition, if local effective size decreases upstream due to lower habitat availability or quality (Carrara et al. 2014), diversity loss could be even stronger in headwaters (Paz-Vinas et al. 2015), reinforcing the expected DEN pattern. Conversely though, historical or

environmental constraints independent from network properties might also induce local genetic drift and lead to a mismatch between the expected DEN pattern and the observed genetic variation (e.g. Salisbury et al. 2016; Pilger et al. 2017). Harvesting pressure, for instance, is a factor susceptible of impacting local effective sizes (Allendorf et al. 2008) by acting as an extended bottleneck.

French Guiana is a remarkable hotspot of Neotropical freshwater fish diversity (Abell et al. 2008). Its largest river catchment, the Maroni (> 65,800 km²), hosts 242 strictly freshwater fish species among which 16.9% are endemic according to the most recent report (Le Bail et al. 2012). The rainforest in the upstream part of the watershed (the Upper-Maroni: UM) includes inhabited areas, located in the close vicinity to the main river, where native human communities have settled for centuries. Today, towns and their close surroundings grow quickly (Fleury et al. 2016), but native societies still rely on food resources provided by the river. Subsistence fishing is a major contribution to daily nutrition (Meunier 2004) and around sixty species are regularly harvested (Longin 2016). During the past three decades, the increasing use of motor boats and large, weakly selective trammel nets has increased harvesting pressure on fish populations around inhabited areas. More recently, the trade of species of substantial economic value such as large herbivorous Piranhas of the Myleinae subfamily (Characidae) has been observed. In parallel, a decline in the quantity of fish proteins from local fisheries per inhabitant is reported, and native people started reporting a decrease in number and size of captures (Fleury 1998; Longin 2016). All these elements taken together ask the question of anthropic impacts, and particularly overfishing, on fish populations. Looking beyond conservation issues, an efficient management of this unique natural resource is essential for preserving traditional livelihood and subsistence of native communities. To date, there is no genetic information about fish populations from the Upper Maroni.

This study aims at analyzing the patterns of within and between population genetic variation for eighteen fish species of the UM, with the applied objective to provide a baseline for fisheries management. We focused on species from various families, commonly found in the UM but differentially targeted by fishermen depending on their cultural or economical values. To detect species or locations subjected to overfishing risk, we compared overall genetic structure and levels of genetic diversity between areas under contrasted harvesting pressure along the watershed for all species. We used a riverscape approach to see (1) whether species with contrasted dispersal abilities showed different patterns of genetic structure with respect to the DEN theory, and (2) whether harvested, downstream-located fish populations showed a loss of genetic diversity in spite of higher theoretical downstream connectivity, inducing in a deviance from the expected DEN pattern.

2.3. Material and Methods

Study area

The Upper-Maroni (UM) corresponds to the main Maroni River and its tributaries upstream from the village of Grand-Santi (Fig.C2.1A). Twelve sites were sampled (Fig.C2.1B, Table C2.1) across this study area in large (Strahler index > 4) and open channels. Five sites (V1 to V5) were located near villages (66 to over 3000 inhabitants) where harvesting pressure is the highest since people mostly fish every day within a ~5 km radius area around their home (Longin 2016).

Fishing pressure quickly decreases upstream from villages (Longin 2016), and seven sampling sites located on four major tributaries (R1 to R7) were considered as reference. A minimum of 3 days transit is required to reach the most upstream sites (R1, R3, R5 and R7), and these sites are more sporadically visited by fishermen (Longin 2016). Distances between sites (km) were measured along the waterways, and were obtained from the national GIS database *IGN BD Topo® Hydrographical Network* version 2.2. We used the distance to the most downstream location in the study zone (Abattis Cottica) to characterize the position of each site (Table C2.1). Given the geographical configuration of the studied network, with tributaries aligned from West to East (Waki, Tampok, Marouini, Litani, see Fig.C2.1B), the longitude was retained as a continuous variable to characterize each site position, allowing to distinguish between tributaries.

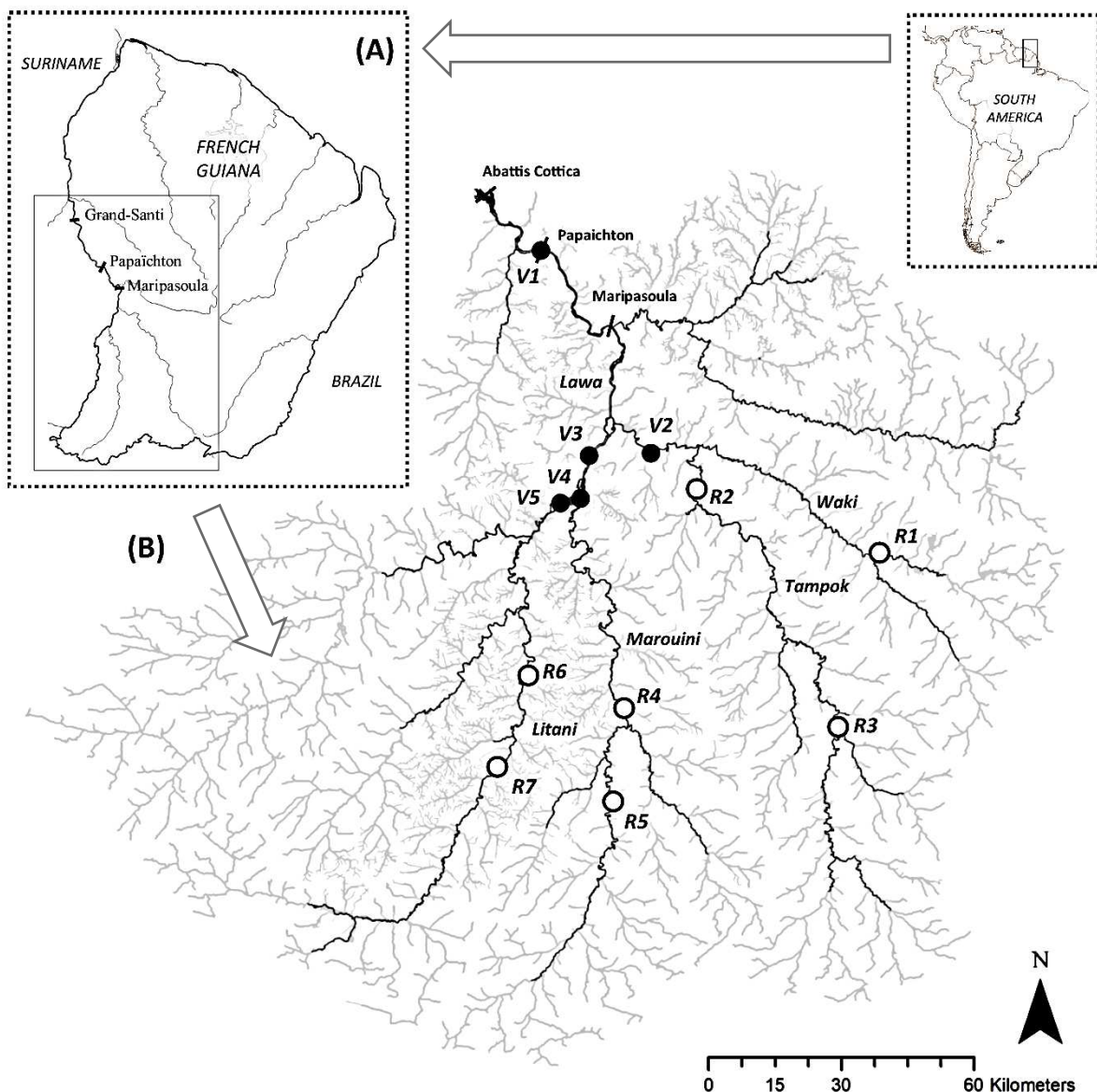


Figure C2.1: Map of French Guiana (A) and of the Upper-Maroni (B) with its main tributaries. Sampling locations are indicated with villages (i.e., high fishing pressure, closed circles) and reference sites (i.e., low fishing pressure, open circles). Streams in black are main channels with Strahler index > 4 while streams in grey are streams with Strahler index ≤ 4.

Table C2.1: Information about the sites sampled across the Upper-Maroni for this study.

Sampling location (local name)	Geog. coordinates		River	Human population ¹	Dist. to Cottica ²	Nb of Species ³	CPUE ⁴
	Lat.	Long.					
V1 (Papaïchton)	3.8083	-54.1491	Lawa	>3000	72	13	0.0437
V2 (Kayodé/Wempi)	3.3890	-53.9246	Tampok	228	144	14	0.0473
V3 (Opoya/Taluen)	3.3813	-54.0504	Lawa	263	141	15	0.0790
V4 (Antecume-Pata)	3.2980	-54.0713	Lawa	170	151	15	0.1169
V5 (Pilima-Pata)	3.2875	-54.1085	Litani	66	148	12	0.1070
R1 (Saut l'Inspecteur)	3.1879	-53.4592	Waki	-	218	13	0.0873
R2 (Saut Tampok)	3.3169	-53.8340	Tampok	-	177	18	0.3521
R3 (Saut Pierkourou)	2.8323	-53.5455	Tampok	-	274	17	0.2028
R4 (Langa Soula)	2.8704	-53.9770	Marouini	-	241	18	0.1175
R5 (Wayo Gaan Soula)	2.6771	-53.9985	Marouini	-	277	17	0.1391
R6 (Apsik Icholi)	2.9358	-54.1748	Litani	-	212	16	0.1964
R7 (Eléüéletpe icholi)	2.7496	-54.2359	Litani	-	244	17	0.0944

¹According to the most recent available figures from the French national institute for statistical and economic studies (INSEE)

²Waterway distance (in km) from each sampling site to the center of the Abattis Cottica = upstream position. ³Number of focal species that were successively sampled (i.e., with at least 10 specimens) within each site. ⁴Preliminary estimation of multi-species CPUE, in captures par square meter (m²) of gillnet, obtained from Longin (2016) and L.Bonneau de Beaufort (pers.comm.).

Focal species and sampling strategy

Eighteen species were chosen amongst the 60 that are regularly harvested in the UM (Longin 2016). We chose the species with regards to their sampling easiness in order to cover as many sampling sites as possible for all of them. We also selected species displaying various biological characteristics, including reproductive strategies (reflecting early dispersal ability) and migratory behaviors (late dispersal ability), according to current knowledge about their ecology (Table C2.2). All species display overlapping generations and a generation time empirically considered superior to one year. The actual extent of fishing-induced mortality rate for the focal species is currently unknown, however, all of them display contrasted interest for fishermen. Eight species, usually large body-sized or easy to catch, are actively targeted and/or caught on a daily basis. Six species are of lesser interest but willingly consumed when captured. Finally, four species are rarely captured because of small size or low occurrence. Six species have high fecundity and can move long distance to spawn during the rainy season. Conversely, seven species, including brood-hiders Serrasalminidae, are known to perform shorter migration before spawning, preferentially in the early rainy season. Finally, five species are non-migrant fish with lower fecundity but providing parental care to their progeny (Table C2.2), and five species are endemic to the Maroni River and its two adjacent drainages according to Le Bail et al. (2012).

Individuals were sampled between the first semester of 2014 and early 2016, using various types of fishing nets but mostly trammel and cast nets. In villages, samples were collected on fish captured by local fishermen, and additional campaigns were performed to increase sample size when needed. At reference sites, we performed specific surveys to collect fish sample. Fin

or muscle tissues were collected on each specimen, then immediately stored in 95% ethanol. Ideal sample size was set at 20 specimens per species and per site, but this objective was not always reached due to field constraints. When less than 10 specimens of a species were caught on a site, they were not considered for further analyses. Therefore, 145 to 283 individuals per species were finally retained, each species being sampled in at least 9 sites, and each site being represented by at least 12 species (Table C2.1, C2.2 and C2.S1). DNA was extracted and each sample was genotyped as described in Delord et al. (2018).

Genetic data and statistical analyses

We used the genotype data available from Delord et al. (2018) for all 18 species. These included 111 to 157 neutral biallelic SNP markers scored per species (full data available from Dryad through the repository <https://doi.org/10.5061/dryad.2b6b43k>). Following the method of Delord et al. (2018), an extra 31 SNP loci were obtained for the Characidae *Cynodon meionactis*. Markers were at Hardy-Weinberg equilibrium and we assumed no physical linkage between loci.

General description of intra and interpopulation genetic diversity

We described intra and inter-population genetic variability for each species, using several genetic metrics. We first calculated the proportion of loci that were polymorphic (i.e., frequency of the less common allele is ≥ 0.05 , P_{95}) within each site using a rarefaction procedure estimation for allelic richness (analogous to P_{95} when using biallelic markers) with a modified *rarefactor* function from the R package DIVERSITY (Keenan et al. 2013), with R version 3.4.2 (Core Team 2017). We computed the expected (H_S) and observed (H_O) heterozygosity, along with the fixation index F_{IS} , for each species and site using the R package HIERFSTAT (Goudet 2005; Goudet & Jombart 2017). We evaluated genetic structure using global and pairwise Weir & Cockerham (1984) F_{ST} with HIERFSTAT. We assessed the significance of global F_{ST} through the Goudet G-test (R function *gstat.randtest*) and we computed the 95% confidence intervals (CI95) for each pairwise F_{ST} using 999 bootstraps. Then, for each species, we calculated a population-specific F_{ST} , $F_{ST.p}$ (Pflüger & Balkenhol 2014) for each sampling site through an averaging of its pairwise F_{ST} values.

To obtain point estimates of effective size (N_e) for each species and sampling site, we used the linkage-disequilibrium (LD) method implemented in the NEESTIMATOR software (Do et al. 2014). Gilbert & Whitlock (2015) urged caution for interpretation of local LD estimates, especially when continuous genetic differentiation occurs (see also Neel et al. 2013), or when the actual N_e is large (e.g. ≥ 500 or significantly higher than sample size). Thus, we did not include N_e estimates in any further analysis. However, we used them to spot whether some sampling sites displayed particularly low N_e values with finite CIs with a relationship to low genetic diversity or strong isolation. We set the minimum allele frequency (P_{CRIT}) for NEESTIMATOR to 0.05, so monomorphic loci were discarded from the analysis.

Table C2.2: Information about the eighteen species chosen to conduct this study.

Species ¹	Family	Reproductive guild ²	Migratory behavior ²	Fishing interest	Other information ²	LSmax (cm)	Nb of loci	Nb of samples
1. <i>Brycon falcatus</i>	Characidae	Broadcast spawner	Long-distance	Moderate		32	155	174
2. <i>Cynodon meionactis</i>		Broadcast spawner	Long-distance	Low	Endemic	36	188	184
3. <i>Triportheus brachypomus</i>		Broadcast spawner	Long-distance	Low		30	148	174
4. <i>Hoplias aimara</i>	Erythrinidae	Guarder	Non-migrant	High		100	154	229
5. <i>Leporinus friderici</i>	Anostomidae	Broadcast spawner	Long-distance	High		35	141	208
6. <i>Leporinus lebaili</i>		Broadcast spawner	Long-distance	Moderate	Endemic	27	150	145
7. <i>Acnodon oligacanthus</i>	Serrasalmodae	Brood hider	Short distance	Moderate	Endemic	20	138	218
8. <i>Myloplus rhomboidalis</i>		Brood hider	Short distance	High		40	137	242
9. <i>Myloplus rubripinnis</i>		Brood hider	Short distance	Moderate		22	118	218
10. <i>Tometes lebaili</i>		Brood hider	Short distance	High	Endemic	53	156	171
11. <i>Serrasalmus eigenmanni</i>		Brood hider	Short distance	Moderate		20	150	178
12. <i>Serrasalmus rhombeus</i>		Brood hider	Short distance	High		45	154	253
13. <i>Ageneiosus inermis</i>	Auchenipteridae	Broadcast spawner	Short distance	High		45	111	232
14. <i>Doras micropoeus</i>	Doradidae	Broadcast spawner	Long-distance	Moderate		38	150	185
15. <i>Harttia guianensis</i>	Loricariidae	Guarder	Non-migrant	Low		19	154	212
16. <i>Hypostomus gymnorhynchus</i>		Guarder	Non-migrant	High		20	153	217
17. <i>Pseudancistrus barbatus</i>		Guarder	Non-migrant	High		22	151	283
18. <i>Geophagus harreri</i>	Cichlidae	Guarder	Non-migrant	Low	Endemic	23	150	164

¹Full given binomen are based on the most recent checklist of freshwater and estuary fish of French Guiana from Le Bail et al. (2012). *Serrasalmus eigenmanni*, formerly known as *Pristobrycon eigenmanni*, was reclassified (Géry 1972). ²References: Azevedo 2010; Carolsfeld et al. 2004; Cohen et al. 1999; Keith et al. 2000; Kullander 2003; Le Bail et al. 2000; Le Bail et al. 2012; Lord et al. 2007; Mérioux et al. 2001; Meunier 2004; Pagezy et Jégu 2002; Planquette et al. 1996; Ponton et De Mérona, 1998.

Multispecies patterns of genetic variability among harvested and unharvested sites

First, for each species, we computed local genetic diversity as the contribution of each sampling site to the full genetic dataset, using an estimation of the local *internal diversity* (I_D , Caballero & Toro 2002), as implemented in the software `MOLKIN` (Gutiérrez et al. 2005). I_D provides the magnitude of overall diversity loss when a putative population is removed from the full dataset. A negative (respectively positive) value indicates a high (respectively low) local genetic diversity. Expressed as a percentage of the sampled metapopulation diversity, I_D allows the contribution of each site to be compared between species. In addition, we computed for each species a normalized version of $F_{ST.p}$ ($StF_{ST.p}$) by rescaling all $F_{ST.p}$ values between 0 (lowest $F_{ST.p}$) and 1 (highest $F_{ST.p}$), in order to detect sites where some species would display their highest mean genetic differentiation.

Using all combined datasets (i.e., all species and all sampling sites, $N_{obs} = 185$ observations, Appendix C2.S1), we investigated if downstream sites with harvesting pressure (villages) had, on average, higher (in accordance with DEN theory) or lower (in accordance with a strong anthropic-induced genetic drift counterbalancing connectivity) genetic diversity across species than upstream reference sites. We used a mixed effect linear model (R function `lme`), with I_D as the response variable, and the type of sampling site (“village” versus “reference”) as a fixed effect predictor. Species acted as replicates and were included within the full model as a random effect. Using the same modelling approach, we then tested if some sampling sites were consistently associated with particular genetic diversity or mean local isolation patterns across species. To this end, I_D and $StF_{ST.p}$ were used as dependent variables and sampling sites as the fixed effect predictor. When a significant site influence was detected, we performed Tukey post-hoc tests to identify which sites displayed strong differences in I_D or $StF_{ST.p}$.

Riverscape patterns of genetic variability

We used a riverscape approach to investigate general and species-specific patterns of spatial genetic variation, and evaluate whether species with such variation associated to or deviated from the variability pattern expected under DEN architecture.

Genetic diversity across hydrographic position

We performed three covariance analyses (R function `lm`) to correlate within-site P_{95} , H_s and $F_{ST.p}$ (response variables) with the upstream distance (i.e. the distance from the Abattis Cottica). We integrated the geographical longitude as a second continuous variable, and the species as a categorical covariate, in order to test if the effect of upstream position would persist when discriminating between upstream areas and how spatial patterns of variability would change among species. This allowed detecting any upstream decrease of P_{95} and H_s , and any upstream increase of $F_{ST.p}$, expected under DEN theory. For each analysis, we used the R package `EMMEANS` (Lenth, 2018) to report the estimated marginal means (EMM, with CI95) which provides the magnitude (and associated uncertainty) of the response variable evolution along increasing upstream position or longitude for every species. To complement the information provided by EMM, for each species we also tested whether allelic richness

(analogous to P_{95}) and H_S differed between sites grouped into the two categories “villages” and “reference sites”, using F_{STATS} version 2.9.3.2 (Goudet 1995, 999 random permutations).

Isolation by stream distance

For each species, we assessed the effect of stream distance (km) on pairwise genetic differentiation (Linearized pairwise $F_{ST} = F_{ST}/(1 - F_{ST})$) using a simple Mantel test for isolation by distance (IBD) with the R package *VEGAN* (Oksanen et al. 2018) and 999 permutations. IBD is expected to be stronger for species with low dispersal abilities, and thus is also an expected feature of the DEN pattern. However, isolation by distance could be associated with additional structuring mechanisms such as natural fragmentation (e.g. rapids) or isolation by drift, especially within DEN (Brauer et al. 2016; Salisbury et al. 2016; Prunier et al. 2017). The underlying influence of isolation-by-drift has been tested in other freshwater systems, using for instance pairwise values of effective sizes (e.g., Pilger et al. 2017) or an environmental proxy of theoretical genetic drift, such as habitat surface (e.g., Prunier et al. 2017). However, the UM channels display widely and regularly distributed rapids, and none of the aforementioned descriptors are available in our case. Therefore, results of IBD analysis will be taken with caution.

Relationship between genetic structure patterns and species biological characteristics

Finally, we used a multivariate approach to help visualizing if species with similar biological characteristics (i.e., reproductive strategies, dispersal abilities, standard adult size or endemism) shared similar genetic variability patterns. For each species, we summarized the degree of overall spatial genetic variability by four variables: a Mantel test's r for isolation-by-distance (IBD_R), Pearson's coefficient of correlation between the distance to the Abattis Cottica and P_{95} , H_S and $F_{ST.p}$. These three coefficients were standardized (as $Zr.P_{95}$, $Zr.H_S$ and $Zr.F_{ST.p}$) using a r to Fisher's Z transformation to take varying sample sizes ($N_{obs} = 9$ to 12) into account (function *res* from the R package *COMPUTE.ES*; Del Re 2013), and describe the evolution of local genetic diversity from downstream to upstream. Then, we used those variables within a principle components analysis (PCA) including all 18 species, using the R package *FACTOMINER* (Lê et al. 2008).

2.4. Results

General description of intra and interpopulation genetic diversity

Detailed values of genetic diversity indices for each species and sampling site are available on Supporting Information (Appendix C2.S1). For all species, local distributions of minor allele frequencies showed no shift toward low or high frequency values from a sampling sites to another. However, polymorphism rate P_{95} and expected heterozygosity H_S both varied among sampling sites at various magnitude depending on species, ranging overall from 0.687 to 0.999 and from 0.257 to 0.436, respectively. Per species, coefficients of variation for P_{95} and H_S varied from 0.729 to 4.197 and 0.755 to 3.696. The Characidae *T. brachypomus* and *B. falcatus* had the lowest magnitude of variation while *H. gymnorhynchus* and *S. eigenmanni*, a Loricariidae and a Serrasalminidae, had the highest.

We found significant heterozygote deficiency (F_{IS}) in 26 on 185 cases (case = one species in one site) in total (14%). However, significant F_{IS} values never exceeded 0.089 (0.013 on average \pm 0.023 s.d.) and were not restricted to some sites or species in particular, suggesting no biologically -or spatially- explicit heterozygote deficiency. Overall levels of genetic differentiation were low as global F_{ST} ranged from zero to 0.038 and were significant in only seven species (Table C2.3). Pairwise F_{ST} values ranged from zero to 0.099 across species, and significant values mainly included upstream sites R1 and R3, as well as R2 and V2 (Appendix C2.S2). Bootstrapped Neighbor-Joining trees on Nei’s genetic distances among sampling sites, are available for each species in Appendix C2.S3. Point estimates of effective size N_e for each species and sampling site are available in Table C2.4. As expected, most N_e values were non-finite (i.e., with either infinite estimate or infinite lower/upper bound of the CI). However, two sites hosted finite N_e values, associated to lower polymorphism or expected heterozygosity, for several species. In the village of Papaïchton (V1), eight out of 13 sampled species had finite N_e values, including six with an estimate under 100. In the reference site Saut l’Inspecteur (R1), six of 13 species had finite N_e values.

Table C2.3: Values of IBD_R (and slope of the isolation by distance IBD_{Slope}), $Zr.P_{95}$, $Zr.H_s$ and $Zr.F_{ST.p}$ per species (i.e., effect sizes of the evolution of polymorphism rate, expected heterozygosity and average pairwise F_{ST} with the upstream distance - variables used for the multivariate PCA analysis, see also Fig.C2.4). Bold, underlined values are significantly different from zero.

Species	IBD_R	IBD_{Slope}	$Zr.P_{95}$	$Zr.H_s$	$Zr.F_{ST.p}$	H_T	F_{ST}
1. <i>Brycon falcatus</i>	0.032	0.14E-05	-0.286	-0.524	-0.229	0.398	0.001
2. <i>Cynodon meionactis</i>	0.329	1.45E-05	0.120	0.228	0.335	0.348	0.002
3. <i>Triportheus brachypomus</i>	0.006	0.02E-05	0.102	0.276	-0.229	0.430	0.001
4. <i>Hoplias aimara</i>	<u>0.446</u>	4.00E-05	0.684	-0.145	-0.147	0.388	0.005
5. <i>Leporinus friderici</i>	-0.055	-0.30E-05	0.071	0.089	-0.351	0.362	0.002
6. <i>Leporinus lebaili</i>	-0.184	-1.39E-05	0.332	0.636	-0.152	0.382	0.000
7. <i>Acnodon oligacanthus</i>	0.293	1.15E-05	0.200	-0.080	-0.050	0.311	0.000
8. <i>Myloplus rhomboidalis</i>	<u>0.745</u>	8.55E-05	-0.397	0.024	0.508	0.350	0.008
9. <i>Myloplus rubripinnis</i>	<u>0.425</u>	2.72E-05	0.196	-0.264	0.315	0.354	0.001
10. <i>Tometes lebaili</i>	<u>0.611</u>	9.28E-05	-0.436	-0.333	0.879	0.386	<u>0.012</u>
11. <i>Serrasalmus eigenmanni</i>	<u>0.520</u>	17.00E-05	-0.297	-0.530	0.212	0.349	<u>0.029</u>
12. <i>Serrasalmus rhombeus</i>	<u>0.289</u>	2.00E-05	-0.084	0.079	0.152	0.386	0.006
13. <i>Ageneiosus inermis</i>	<u>0.481</u>	2.79E-05	-0.360	0.290	<u>0.811</u>	0.373	<u>0.004</u>
14. <i>Doras micropoeus</i>	-0.204	-1.36E-05	0.244	0.151	0.105	0.372	0.001
15. <i>Harttia guianensis</i>	0.315	4.19E-05	-0.021	0.164	0.196	0.356	<u>0.013</u>
16. <i>Hypostomus gymnorhynchus</i>	<u>0.849</u>	9.91E-05	0.282	-0.103	<u>0.975</u>	0.268	<u>0.007</u>
17. <i>Pseudancistrus barbatus</i>	<u>0.716</u>	20.00E-05	-0.241	-0.139	0.693	0.319	<u>0.020</u>
18. <i>Geophagus harreri</i>	<u>0.515</u>	23.80E-05	-0.475	-0.520	0.695	0.389	<u>0.038</u>

Table C2.4: Estimates of local contemporary effective size N_e {with 95% parametric confidence intervals} for each species and sampling site. Estimates were obtained from the linkage-disequilibrium method implemented within the software NeEstimator with PCRI=0.05. Estimates with finite confidence intervals are grey-shaded. NA: absence of data.

Species	V1	V2	V3	V4	V5	R1	R2	R3	R4	R5	R6	R7
1. <i>Brycon falcatus</i>	25 {20;33}	NA	NA	>500 {171;∞}	NA	298 {82;∞}	470 {152;∞}	∞ {212;∞}	102 {67;200}	198 {96;∞}	412 {129;∞}	>500 {119;∞}
2. <i>Cynodon meionactis</i>	NA	363 {159;∞}	>500 {153;∞}	NA	NA	142 {86;380}	∞ {263;∞}	394 {143;∞}	∞ {293;∞}	∞ {416;∞}	403 {147;∞}	278 {124;∞}
3. <i>Triportheus brachypomus</i>	NA	90 {62;157}	249 {70;∞}	NA	NA	457 {169;∞}	344 {111;∞}	>500 {170;∞}	∞ {192;∞}	∞ {181;∞}	213 {80;∞}	∞ {218;∞}
4. <i>Hoplias aimara</i>	7 {7;9}	∞ {140;∞}	∞ {143;∞}	351 {105;∞}	∞ {121;∞}	259 {107;∞}	>500 {163;∞}	∞ {450;∞}	>500 {181;∞}	∞ {379;∞}	78 {57;119}	∞ {271;∞}
5. <i>Leporinus friderici</i>	17 {14;19}	80 {50;179}	49 {37;71}	251 {100;∞}	252 {67;∞}	73 {36;1168}	>500 {157;∞}	∞ {119;∞}	285 {121;∞}	∞ {1856;∞}	75 {35;∞}	320 {116;∞}
6. <i>Leporinus lebaili</i>	∞ {244;∞}	NA	NA	>500 {135;∞}	NA	121 {76;496}	∞ {84;∞}	∞ {180;∞}	∞ {289;∞}	∞ {377;∞}	58 {30;326}	183 {80;∞}
7. <i>Acnodon oligacanthus</i>	195 {86;∞}	NA	250 {98;∞}	∞ {264;∞}	>500 {158;∞}	NA	>500 {145;∞}	376 {106;∞}	>500 {147;∞}	>500 {148;∞}	>500 {119;∞}	∞ {334;∞}
8. <i>Myloplus rhomboidalis</i>	172 {88;1546}	149 {56;∞}	>500 {154;∞}	>500 {160;∞}	>500 {352;∞}	NA	454 {105;∞}	∞ {461;∞}	428 {114;∞}	>500 {121;∞}	>500 {169;∞}	210 {80;∞}
9. <i>Myloplus rubripinnis</i>	NA	∞ {382;∞}	∞ {318;∞}	464 {155;∞}	∞ {166;∞}	NA	∞ {141;∞}	∞ {198;∞}	>500 {96;∞}	20 {15;27}	∞ {257;∞}	∞ {160;∞}
10. <i>Tometes lebaili</i>	∞ {385;∞}	NA	∞ {212;∞}	>500 {246;∞}	332 {182;1549}	NA	>500 {131;∞}	109 {52;∞}	∞ {∞;∞}	102 {53;714}	NA	>500 {124;∞}
11. <i>Serrasalmus eigenmanni</i>	NA	∞ {165;∞}	∞ {83;∞}	NA	NA	∞ {332;∞}	266 {101;∞}	>500 {129;∞}	∞ {322;∞}	>500 {132;∞}	∞ {189;∞}	>500 {144;∞}
12. <i>Serrasalmus rhombeus</i>	24 {20;28}	>500 {175;∞}	12 {11;14}	∞ {151;∞}	248 {100;∞}	47 {33;75}	>500 {122;∞}	∞ {168;∞}	74 {47;156}	101 {64;226}	307 {322;∞}	92 {59;192}
13. <i>Ageneiosus inermis</i>	182 {69;∞}	72 {44;183}	NA	221 {89;∞}	>500 {182;∞}	103 {53;712}	166 {70;∞}	320 {93;∞}	143 {65;∞}	∞ {163;∞}	140 {67;∞}	457 {106;∞}
14. <i>Doras micropoeus</i>	>500 {146;∞}	∞ {217;∞}	>500 {239;∞}	39 {28;65}	64 {46;106}	NA	>500 {91;∞}	NA	>500 {111;∞}	∞ {209;∞}	229 {96;∞}	278 {89;∞}
15. <i>Harttia guianensis</i>	NA	70 {49;114}	411 {125;∞}	>500 {236;∞}	NA	>500 {174;∞}	299 {119;∞}	177 {87;9382}	∞ {250;∞}	>500 {181;∞}	296 {118;∞}	692 {149;∞}
16. <i>Hypostomus gymnorhynchus</i>	105 {59;374}	>500 {165;∞}	∞ {275;∞}	∞ {159;∞}	85 {36;∞}	>500 {140;∞}	499 {120;∞}	∞ {226;∞}	∞ {131;∞}	101 {58;324}	NA	318 {96;∞}
17. <i>Pseudancistrus barbatus</i>	60 {47;80}	>500 {218;∞}	472 {128;∞}	146 {83;496}	>500 {198;∞}	∞ {199;∞}	31 {25;39}	∞ {166;∞}	∞ {226;∞}	∞ {200;∞}	>500 {159;∞}	484 {129;∞}
18. <i>Geophagus harreri</i>	54 {40;80}	24 {17;39}	>500 {191;∞}	∞ {159;∞}	>500 {109;∞}	34 {26;47}	>500 {86;∞}	107 {61;351}	115 {58;1515}	NA	∞ {193;∞}	NA

Multispecies patterns of genetic variability across harvested and unharvested sites

Internal diversity I_D was significantly, although slightly (6.5%, CI95 = [0.1-12.8]), lower in villages than in reference sites, according to the mixed effects linear modelling (Fisher's statistic $F(1,166) = 4.073$, conditional $R^2 = 0.022$, marginal $R^2 = 0.022$, $P = 0.045$). This indicated a higher contribution of villages to the overall genetic diversity, on average across species. However, the local contribution of each sampling site, even within a same category, was not homogeneous, as suggested by the significant variation of I_D among sites ($F(11,156) = 2.587$, cond. $R^2 = 0.134$, marg. $R^2 = 0.134$, $P = 0.005$). Standardized local mean isolation $S_{TF_{ST.p}}$ was significantly higher within reference sites than downstream villages (Fisher's statistic $F(1,166) = 13.902$, conditional $R^2 = 0.089$, marginal $R^2 = 0.069$, $P < 0.0001$) but also varied among sampling sites ($F(11,156) = 5.436$, cond. $R^2 = 0.269$, marg. $R^2 = 0.240$, $P < 0.0001$).

Tukey post-hoc tests showed that R1 (and R1 with R3, respectively) were significantly showing among the highest I_D (and $S_{TF_{ST.p}}$) values across species (Fig.C2.2), meaning a low genetic diversity and a high local genetic isolation for these most upstream sites from the Tampok and the Waki tributaries. Other pairwise comparisons were non-significant, implying the lower genetic diversity of reference sites compared to villages to be mainly driven by these two tributaries. When considering the other reference sites, there was a trend for higher average I_D in R2 (Fig.C2.2A), and for higher average $S_{TF_{ST.p}}$ in R5 (Fig.C2.2B). Conversely, R7 showed the lowest $S_{TF_{ST.p}}$ values among reference sites and across species. Thus, globally, upstream sites did not have a similar genetic diversity and mean isolation signatures between the different catchments of the study area, with two tributaries, located on the Eastern portion of the study zone, showing particular signals of poor diversity and strong isolation.

Regarding villages, there was a visual trend for higher contribution to genetic diversity (i.e. lower I_D values) from V3 and V4, while other villages displayed more variable, but slightly higher I_D values on average (Fig.C2.2A). V3 and V4 had also lower-ranked mean isolation ($S_{TF_{ST.p}}$) (Fig.C2.2B), while higher values were observed in the three other villages, and especially in V1 (Fig.C2.2B). A mixed linear model on $S_{TF_{ST.p}}$ including only villages showed significant differences among villages ($F(4,47) = 3.631$, cond. $R^2 = 0.299$, marg. $R^2 = 0.152$, $P = 0.012$) with V1 displaying significantly higher $S_{TF_{ST.p}}$ values. This particular case was mostly driven by four species (*H. aimara*, *B. falcatius*, *S. rhombeus* and *L. friderici*) in which this sampling site harbored higher-ranked $F_{ST.p}$ values (Appendix S1), and also small N_e estimates as seen in Table C2.4.

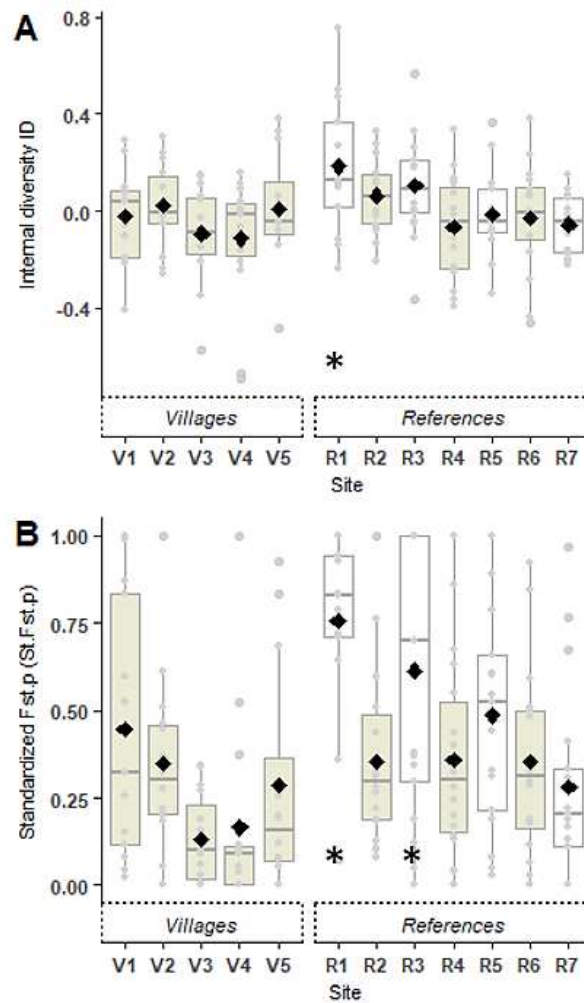


Figure C2.2: Values of (A) internal diversity I_D and (B) standardized local isolation ($F_{ST.p}$, i.e., average pairwise F_{ST}) (B) per sampling site across species. For each site, grey dots represent species, boxplot gives median and quartiles, and black diamond gives the average value. Asterisk (*) indicate sampling sites associated with significantly higher I_D or $F_{ST.p}$ values according to a Tukey post-hoc test following a linear mixed effect model to test for the effect of site on both variables. In the right panel of each plot, white boxplots represent headwaters.

Species-specific riverscape patterns of genetic variability

Effect of the hydrographic network on genetic diversity

Covariance analyses results are given in Table C2.5. The main source of variance for local genetic diversity metrics (P_{95} , H_5) and mean local genetic isolation ($F_{ST.p}$) was the species. However, here was a significant positive effect of the upstream position on mean isolation ($F_{ST.p}$), but not on genetic diversity (Table C2.5). Nevertheless, $F_{ST.p}$ (respectively P_{95} and H_5) significantly increased (decreased) with longitude, i.e. when going from Western to Eastern tributaries and particularly sites R1 and R3, as already suggested by the previous multispecies analysis. These general trends were differently supported among species, as seen from the significant interactions between predictors (Table C2.5).

EMM and their associated CIs (Fig.C2.3) showed that the decrease of P_{95} and/or H_5 with longitude was significant for six species: two Serrasalmidae (*S. eigenmanni* and *M.*

rhomboidalis), two Loricariidae (*P. barbatus* and *H. guianensis*), the Cichlidae *G. harreri* and the Characidae *C. meionactis*. This pattern also existed as a trend for a third Serrasalminidae *T. lebaili* and the two Characidae *Leporinus friderici* and *Leporinus lebaili*. Taking into account the previous results, we grouped sampling sites in three categories: Villages, Western reference sites (Litani/Marouini) and Eastern reference sites (Tampok/Waki). Testing for differences of expected heterozygosity H_s between these three categories, using permutation tests in FSTATS, showed that for seven of these species, villages and Western reference sites usually displayed similar H_s , whereas Eastern reference sites were significantly less variable (Appendix S4). FSTATS analyses also showed that *H. gymnorhynchus*, the third Loricariidae of our focal species, had an opposite pattern, being the only species where Western tributaries had a lower expected heterozygosity than Eastern tributaries, with Villages occupying an intermediate position. *S. eigenmanni* was the only species in which genetic diversity (H_s) also decreased with the upstream position, according to EMM (Fig.C2.3). No variation of local genetic diversity with longitude nor upstream position was observed for the other species.

The magnitude of the increase of mean isolation with longitude, as provided by EMM of the covariance analysis on $F_{ST.p}$, also varied among species (Table C2.5). The five previously mentioned species with global $F_{ST} \geq 0.008$, which showed a significant eastern decrease of genetic diversity, also had higher local mean genetic isolation with increasing longitude. Mean isolation also significantly increased with the upstream position for three species: *T. lebaili*, *P. barbatus* and *H. gymnorhynchus*, according to EMM. Values of $F_{ST.p}$ between the three groups of sampling sites, for each species, are available in Appendix C2.S4 and usually confirmed mean isolation to be higher within reference sites than villages for the aforementioned species. No variation of mean isolation with longitude nor upstream position was observed for the other species.

Isolation by stream distance

Stream distance between two sites was a significant predictor of pairwise genetic differentiation for all seven species with a significant global F_{ST} (Table C2.3). IBD was also weakly significant for three additional species (*H. aimara* and two Serrasalminidae *M. rubripinnis* and *S. rhombeus*). There was no detectable IBD for the eight remaining species, but Mantel's r went from 0.315 to significant 0.635 for the Loricariidae *H. guianensis* when removing R1 from the dataset.

The steepness of the linear relationship between stream distance and pairwise differentiation (e.g., IBD_{Slope} , e.g., Table C2.3) was the highest for *P. barbatus*, *G. harreri* and *S. eigenmanni*. However, the goodness of fit of this IBD (given by Mantel's r) was relatively low for the latter two because of the blurring effect of strong structure patterns for these species. Finally, IBD_{Slope} was higher for species that showed significant genetic diversity variation toward the east (i.e. increasing longitude) as obtained from EMM in the previous section.

Table C2.5: Results of the covariance analyses (multiple linear regression models) performed to test the effect of the upstream distance (i.e. the distance from the Abattis Cottica) and the longitude (i.e. the site position on tributaries from West to East) on local genetic diversity (P_{95} , H_s) and mean genetic isolation ($F_{ST.p}$) across species.

P_{95} $R^2 = 0.893$				
	df	F-value	p-value	%expl deviance
Upstream distance	1	1.573	0.212	0.091
Longitude	1	10.090	0.002	0.589
Species	17	89.195	< 0.0001	88.449
Upstream distance:Species	17	0.940	0.529	0.933
Longitude:Species	17	2.315	0.004	2.295
H_s $R^2 = 0.973$				
	df	F-value	p-value	%expl deviance
Upstream distance	1	3.203	0.076	0.047
Longitude	1	16.470	< 0.0001	0.243
Species	17	383.531	< 0.0001	96.379
Upstream distance:Species	17	1.201	0.272	0.302
Longitude:Species	17	2.646	0.001	0.665
$F_{ST.p}$ $R^2 = 0.888$				
	df	F-value	p-value	%expl deviance
Upstream distance	1	16.850	< 0.0001	1.025
Longitude	1	93.602	< 0.0001	5.695
Species	17	71.907	< 0.0001	74.377
Upstream distance:Species	17	4.366	< 0.0001	4.516
Longitude:Species	17	6.203	< 0.0001	6.416

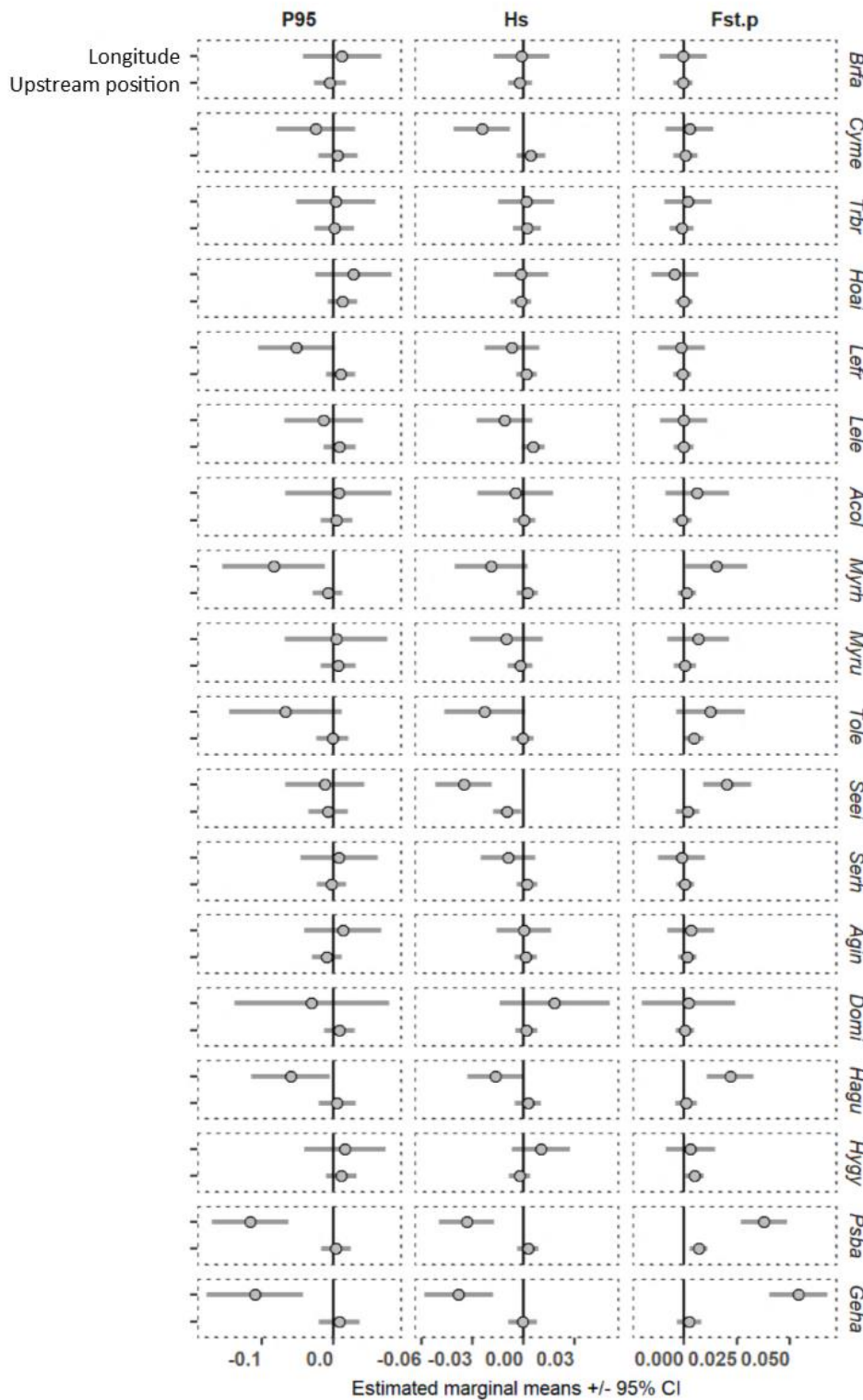


Figure C2.3: Graphical representation of the estimated marginal means (*grey dots*) with 95% confidence intervals (*grey bars*) for each species (rows), obtained from the three covariance analyses (columns) performed to test the effect of the upstream distance (i.e. the distance from the Abattis Cottica) and the longitude (i.e. the site position on tributaries from West to East) on local genetic diversity (P₉₅, H_s) and mean genetic isolation (F_{ST,p}) across species (See also Table 5). EMM and CI values for the upstream distance predictor were multiplied by 100 for better readability.

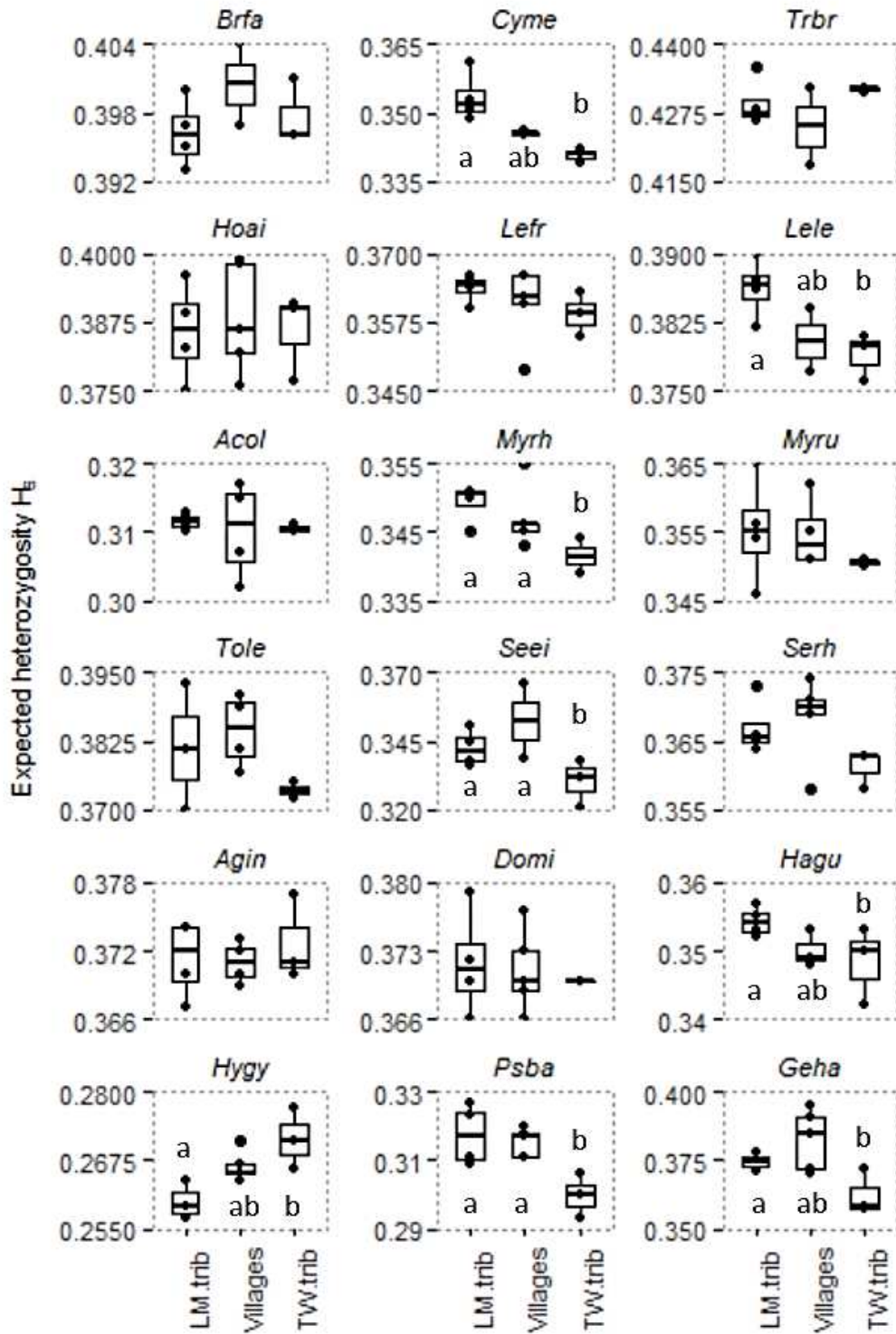


Figure C2.3': Values of expected heterozygosity H_e per sites grouped into three categories (reference sites from Litani and Marouini: LM.trib; sites with harvesting pressure: Villages, and reference sites from the Tampok and the Waki: TW.trib), for each species. Letters indicate significant differences of expected heterozygosity H_e detected between the three groups of sites, as provided by permutations test with the software FSTATS.

NB: Cette figure sera normalement intégrée au Matériel Supplémentaire de cet article en préparation.

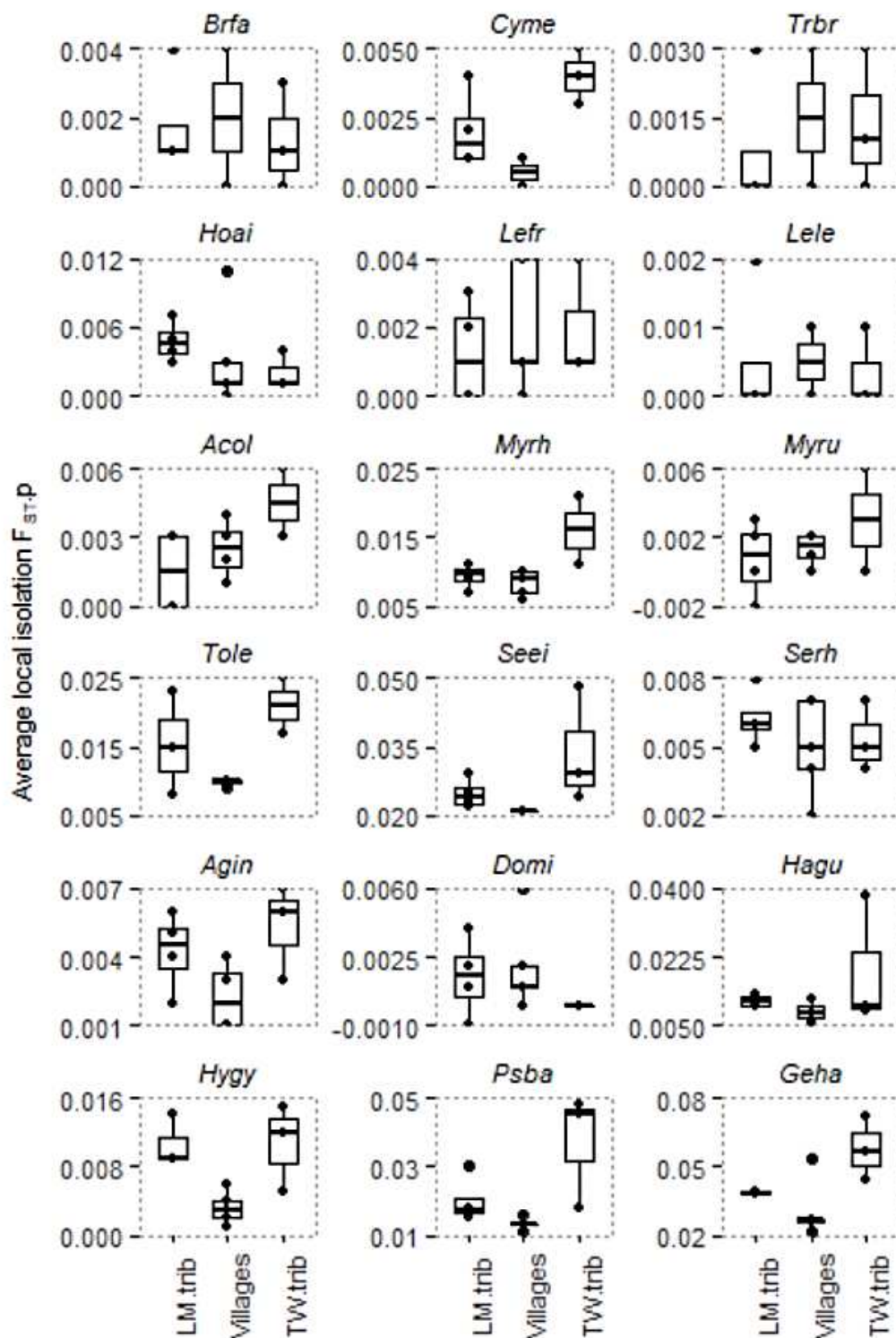


Figure C2.3'': Values of mean local isolation $F_{ST,p}$ per sites grouped into three categories (reference sites from Litani and Marouini: LM.trib; sites with harvesting pressure: Villages, and reference sites from the Tampok and the Waki: TW.trib), for each species.

NB: Cette figure sera normalement intégrée au Matériel Supplémentaire de cet article en préparation.

Relationship between genetic structure patterns and species biological characteristics

The first two components of the PCA explained more than 80% of the between-species variability, but only PC1 had an eigenvalue >1 (Fig.C2.4). PC1 highlighted the increase of spatial genetic variability for species from the left to the right of the PCA biplot, with support from increasing IBD_R and $Zr.F_{ST.p}$, which were strongly correlated ($N_{obs} = 18$, $adj. R^2 = 0.620$, $P < 0.001$), illustrating the co-occurrence of strong isolation by distance with steep upstream increase of mean isolation. This overall gradient of spatial genetic structure grossly reflected the known species dispersal abilities, both at the larval stage (through reproductive strategy) and the adult stage (migratory behavior). Broadcast spawners/long-distance migrants had no detectable genetic structure nor IBD, and were located on the left side of the biplot, whereas broadcast spawner/short-distance migrants showed more marked patterns of spatial genetic variation. Brood hidiers/short-distance migrants and guarders/sedentary species interleaved between the center and right part of the biplot, showing variability in relation to their very low (e.g., *H. aimara*, *A. oligacanthus* or *S. rhombeus*) to much stronger IBD patterns associated with increasing upstream isolation (e.g., *P. barbatus*, *T. lebaili* or *M. rhomboidalis*). For *H. guianensis*, the left position on the PCA biplot is linked to the fact we kept Mantel's r of 0.315 that included the strongly differentiated site R1 (see previous section). Adult size, fishing interest or endemism did not relate to similar genetic variability patterns, at least when considering all 18 species together.

Thus, the PCA also reflected between-species variability within each class of dispersal ability, as illustrated for instance by the dispersion of observations around the second principle component of the biplot. Indeed, PC2, in spite of a lower contribution to the total between-species variability (21.01%) and an eigenvalue smaller than unity, underlined the various trends of $Zr.P_{95}$ and $Zr.H_5$ across species. Besides, it appeared more clearly here that the presence of genetic differentiation (i.e. strong IBD_R and $Zr.F_{ST.p}$) was not systematically in relationship with a steep decrease of upstream genetic diversity, even for species with strong IBD_R (and IBD_{Slope} , Table C2.3), which could display a decrease of diversity in only two on four upstream areas (Eastern tributaries) as seen previously. *S.eigenmanni* arised once again as the only species with upstream decrease of genetic diversity (so did *G.harreri*, however sites R5 and R7 could not be sampled for this species so information about upstream areas from the Litani and Marouini were actually lacking).

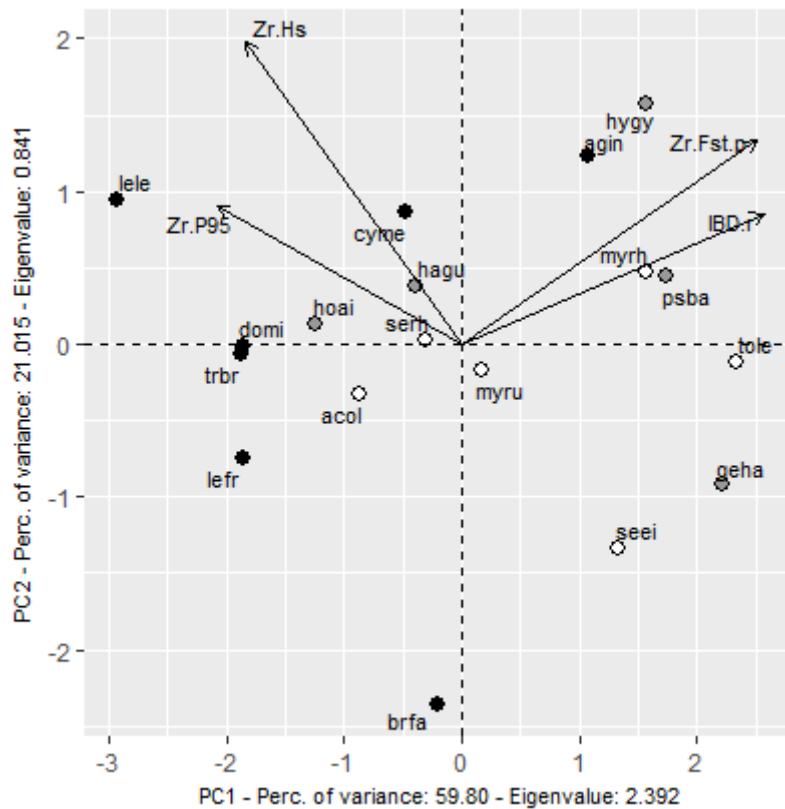


Figure C2.4: Biplot of the principle-components analysis used to describe the relative position of all species regarding the strenght of isolation-by-distance (IBD_R) and of the evolution of expected heterozygosity (H_s), polymorphism rate (P_{95}) and local differentiation ($F_{st.p}$) along the upstream gradient (respectively, Zr.Hs, Zr.P95 and Zr. Fst.p). Species are coloured according to their reproductive guild with black: *Broadcast spawners*, white: *Brood hidere* and grey: *Guarders*.

2.5. Discussion

So far, population genetic studies performed in Neotropics aimed to describe genetic variation for species of strong commercial value, like large Pimelodidae (*Pseudoplatystoma sp.* or *Brachyplatystoma sp.*, e.g. Pereira et al. 2009; Batista & Alvez-Gomez 2006), *Arapaima gigas* (e.g. Hrbek et al. 2007; Watson et al. 2016), or a few Serrasalminidae like *Colossoma macropomum* (e.g. Santos et al. 2007; Fazzi-Gomes et al. 2017), *Piaractus mesopotamicus* (e.g. Iervolino et al. 2010) or *Pygocentrus nattereri* (Santos et al. 2016). To our knowledge, comparative genetic studies with several species of subsistence fishing interest, co-sampled across a common area in the Neotropics, were never performed. In this study, we used a comparative riverscape approach to describe patterns of genetic variation for 18 sympatric fish species from the Upper-Maroni (French Guiana). We aimed at evaluating whether fishing pressure had led to a spatially detectable, downstream decrease of genetic diversity for some species, in spite of the river network architecture theoretically expected to concentrate genetic diversity downstream. The observed patterns of genetic variation are discussed hereafter with regards to river network architecture, species biology and fishing pressure in order to formulate a general synthesis regarding fish populations state and conservation within the Upper-Maroni.

Across species, we highlighted various patterns of genetic structuring, from nil (i.e. global panmixia) to migration-drift equilibrium (i.e. isolation by stream distance) associated to the expected DEN pattern of higher mean isolation in upstream sites. As expected, overall differentiation and spatial genetic variation were more pronounced for species with low dispersal abilities (i.e. sedentary or short-distance migrant with reproductive strategies involving medium to low fecundity), although the additional influence of less-documented biological attributes might be supported by the fact that some species, with similar theoretical dispersal abilities, showed slightly contrasted patterns of genetic variation. Genetic structuring was in accordance with DEN theory in most cases. However, network architecture, was not the only driver of spatial genetic variation. First, an upstream decrease of genetic diversity restrained to eastern tributaries was found in several species while one species, *H. gymnorhynchus*, showed an opposite pattern. Second, unidentified environmental features blurred isolation by stream distance for the two species with strongest genetic structure, *G. harreri* and *S. eigenmanni*. Last, the most downstream site of the study area, Papaïchton, was significantly isolated and showed lower polymorphism along with very small effective size for *H. aimara* which is highly targeted in this area.

Dendritic architecture shapes patterns of genetic structure but not those of local genetic diversity

The spatial configuration of stream networks is considered as a factor of genetic structure *per se*, all population-specific parameters (effective size and migration rates) being equal elsewhere (Paz-Vinas & Blanchet 2015). Areas occupying a central position within a network, which usually correspond to the mainstem, tend to concentrate genetic diversity due to higher connectivity (Labonne et al. 2008). Genetic diversity is then splitted along the successive stream ramifications toward river sources. Under this simplified representation, and assuming a regular and symmetric network, we expected genetic variation to be equally redistributed from mainstem to sources, hence genetic diversity (and mean isolation, respectively) being similarly reduced (increased) within all tributaries.

The UM upstream areas did show stronger genetic differentiation compared to the rest of the study area, according to pairwise F_{ST} values for species that displayed genetic structure. However, its magnitude varied between tributaries. The multispecies analysis suggested mean genetic isolation ranking to be the highest for the Tampok and the Waki upstream areas (R3 and R1), Marouini upstream areas (R5) occupying an intermediate situation, while Litani upstream areas (R7) was associated with the lowest-ranked mean differentiation. The significant genetic diversity loss within R3 and R1 in comparison to the other sites explained this pattern. Such a genetic diversity reduction was not observed within sites from the Litani and Marouini according to covariance analyses, except for *S.eigenmanni*. We detected no significant differences of genetic diversity between Litani and Marouini compared to villages of the downstream area.

Covariance analyses confirmed the “eastern-directed” decrease of genetic diversity for several species, including two (*C. meionactis* and *L. lebaili*), which did not displayed significant spatial structure (with a similar trend for *L.friderici*). This, taken together, suggests that diversity loss within headwaters is, in our case, influenced by additional factors than predicted in the DEN configuration *per se*; otherwise, it should be equal within all tributaries and be found only in species with low theoretical dispersal and substantial genetic structure.

In a meta-analysis of freshwater genetics studies, Paz-Vinas et al. (2015) described three evolutionary processes that successfully explained a steep diversity loss within headwaters, in addition to DEN configuration: downstream-directed (asymmetrical) migration, upstream-directed (re)colonization, and upstream decrease in habitat availability. All three processes involve heterogeneous patterns of migration rates, colonization dispersal or local effective sizes across the study zone. However, asymmetrical migration could explain an unequal decrease of genetic diversity between tributaries only if some local, habitat features (e.g. differences in elevation or water flowing) led to contrasted directionalities of gene flow. Therefore, this process alone cannot explain our observations and would necessarily act jointly with other, local factors. There is currently no evidence nor suspicion of ongoing, upstream-directed colonization for species of the UM. Consequently, variations in habitat availability (or quality) appear as the best candidate for a upstream decrease in genetic diversity occurring within the Tampok and the Waki.

The Waki has a smaller area than the other sampled catchments, implying a reduced habitat availability. Within the Tampok, water level is empirically known (G. Longin, pers. comm.) to have higher amplitude of variation between wet and dry seasons. This could possibly influence passive mobility patterns of fish compared to other tributaries (leading to downstream-directed transfert of individuals for instance) or limitate the accessibility to some habitats (e.g. shallow rocky areas of high water velocity and high importance for rheophilic species; or adjacent forest which provide exogen food supply for fish when flooded during rainy season). The Tampok and the Waki were also more impacted by localized small-scale goldmining activities during the past decade: several hundreds of small-scale goldmining spots were identified between 2009 and 2018, whereas these figures were much lower for the Marouini and the Litani (G. Longin, pers. comm.). Goldmining may generate water pollution, increase turbidity or deteriorate riverbanks, possibly reducing habitat quality and promoting diversity loss. The Marouini and the Litani did not display upstream decrease of genetic diversity, although there were, in contrast, differences in the magnitude of mean genetic isolation in upstream sites between these tributaries. GIS data suggest higher complexity of adjacent small streams for the Litani (streams with Strahler index < 4, Fig.C2.1B). A stronger density of small streams might enhance food availability or sheltering areas for juveniles (Ponton & Copp 1997), increasing local effective sizes and lowering genetic drift and mean isolation. This would be consistent with the fact R7 was the only sampling site where all N_e estimates but one, were large and non-finite. Thus, large upstream effective sizes and the absence of environmental constraints on both Litani and Marouini headwaters might explain that moderate genetic isolation have happened without a following decrease of genetic diversity, or decrease which is not strong enough to be detected using biallelic markers with relatively high frequencies. In conclusion, it is likely that the effect of DEN configuration on genetic variation in the UM is tightly linked to habitat characteristics along an eastern gradient of decreasing habitat quality and/or availability.

Early and late dispersal abilities partly explain global genetic structure patterns between species

We expected species with limited early or late dispersal abilities to show stronger patterns of spatial genetic variation, and to be closer to the DEN pattern if fishing pressure did not lead to strong downstream genetic drift. The more marked isolation by distance and eastern decrease

of genetic diversity were indeed shown by species which were short-migrants or sedentary and brood-hiders or guarders. The multivariate analysis of genetic patterns between the 18 species also revealed species grossly disseminating along increasing spatial differentiation according to their reproductive strategies and mobility, confirming that dispersal abilities, affecting the efficiency of gene flow between distant populations, shape spatial genetic structure in freshwater species (Gomez-Uchida & al. 2009; Comte & Olden 2018b). However, at a finer level, all species theoretically sharing common migratory or reproductive strategies did not systematically displayed the same patterns of genetic variation, which could be explained by more specific, less documented ecological attributes.

As expected, broadcast spawners, long distance migrants *B. falcatus*, *T. brachypomus*, *C. meionactis*, *L. friderici*, *L. lebaili* and *D. micropoeus* showed no patterns of spatial genetic structure, in spite of *C. meionactis* and the two *Leporinus sp.* supporting the eastern decrease of genetic diversity. *A. inermis* although a broadcast spawner, but considered short-distance migrant (Carolsfeld et al. 2004) displayed weak but significant patterns of isolation by stream distance, without any spatially explicit evolution of genetic diversity. Similar patterns were observed for three brood hiders short-distance migrants Serrasalminidae: *S. rhombeus*, *A. oligacanthus* and *M. rubripinnis* (although IBD was very weak - just above and under significance threshold, respectively - for *S. rhombeus* and *A. oligacanthus*). Finally, the expected DEN influence on genetic structure, associated to the eastern decrease of genetic diversity, was observed for three sedentary, guarder species (*P. barbatus*, *H. guianensis* and *G. harreri*) in accordance with these species' low fecundity, high habitat specialization and energetic investment in progeny. The Cichlidae *G. harreri* was the species with strongest spatial genetic variation, nearly all pairwise F_{ST} values being significant, and with steep IBD blurred by strong genetic structure, which was not unexpected given its extreme combination of dispersal-limiting traits: small body size, low fecundity, sophisticated parental care and strong habitat specialization (Kullander et al. 2003; Waters & Burridge 2016).

The five remaining species (*S. eigenmanni*, *T. lebaili*, *M. rhomboidalis*, *H. aimara* and *H. gymnorhynchus*), showed patterns of genetic variation less expected on the basis of their migratory and reproductive strategies. However, hypotheses can be made about the role of additional ecological traits. For instance, steeper isolation by distance, upstream increase of genetic differentiation and eastern decrease of genetic diversity were observed for *S. eigenmanni*, *M. rhomboidalis* and *T. lebaili* in comparison to the three other aforementioned short-migrants brood-hiders Serrasalminidae. The two small-bodied Myleinae *A. oligacanthus* and *M. rubripinnis* displayed less between-site genetic variation than their larger-bodied relatives *T. lebaili* and *M. rhomboidalis*, although body size is usually considered to associate with higher dispersal abilities (Radinger & Wolter 2014), especially for large species with high or moderate fecundity and delayed sexual maturity (Comte & Olden 2018a). A more restrained ecological niche might explain these observations. *A. oligacanthus*, although considered rheophilic, was documented to breed within small creeks apart from the main channel (Planquette et al. 1996). Oppositely, the larger rheophilic *T. lebaili* and *M. rhomboidalis* are more strictly infedated to flowing waters of large streams, including during breeding period (Planquette et al. 2016; Jégu & Keith 2005) which might restrain gene flow. In addition, *T. lebaili* is a rapid-dwelling species with an important foraging specialization (Jégu et al. 2002), exclusively feeding on Podostemaceous plants blooming within shallow rapids. The two

Serrasalminae *S. eigenmanni* and *S. rhombeus* displayed contrasted genetic structure patterns as well. *S. rhombeus* is a large-bodied, habitat-opportunistic and piscivorous species (Planquette et al. 1996) while its relative *S. eigenmanni* is small-bodied and lives along vegetated riverbanks (J. Makiloewala, pers. comm.). In addition, this species was difficult to sample within villages, and could be, perhaps, essentially inhabiting secondary streams.

In conclusion, contrasted patterns of genetic variation between Serrasalminidae species might be essentially linked to various degree of habitat specialization. Especially, additional studies about these species biology would be needed to confirm the potentially strongly restrained dispersal abilities, in spite of large body size, for *T. lebaili* and *M. rhomboidalis* which actually displayed patterns of genetic variation similar to sedentary guarders.

Among the sedentary guarder species, *H. gymnorhynchus* was the only one with significantly higher eastern than western genetic diversity, that is, an opposite pattern from *P. barbatus* and *H. guianensis*. These three Loricariidae species have apparent similar ecology (Le Bail et al. 2000), while their abundance seems to slightly vary as observed in field campaigns: *H. gymnorhynchus* showing lower abundance (R. Covain, pers. comm.). In addition to this lesser abundance, *H. gymnorhynchus* also harbored particularly low genetic diversity, all species considered, which has already been observed for this species in Brazil and across French Guiana (Telles et al. 2010; Weber et al. 2012) using microsatellites and allozyme markers. However, explaining the particular features of this species compared to the two other focal Loricariidae seems difficult at present and further studies might be needed to better understand these patterns. Another sedentary guarder, *H. aimara* displayed weaker spatial genetic variation. However, this species has a much larger fecundity (Planquette et al. 2016) which might compensate for restrained movements providing sufficient larvae dispersal. Besides, Tito de Morais & Raffray (1999) spotted a few mobile individuals when conducting a radio-tracking experiment within a population from the Sinnamary (central French Guiana). Punctual, but medium to long-distance dispersion might be a sufficient driver of effective gene flow. In spite of this overall weak differentiation, the significant genetic isolation between the most downstream site of the study zone (V1) and all other sites for *H. aimara* reflected the lower polymorphism and the very low estimate of effective size within this site, suggesting the occurrence of strong genetic drift.

The absence of relationship between patterns of genetic variability and endemism or fishing interest across all eighteen species, does not mean that these traits have no actual influence on between-site genetic diversity. Once reproductive or mobility pattern are taken into account, we may lack statistical power (i.e., too few species by traits combination) to enable to discriminate between them. Given our incomplete knowledge about finer specific species characteristics, it is also possible that unexpected patterns of genetic variation originate from undocumented, or wrongly documented, biological characteristics. For instance, *A. inermis* was considered a broadcast spawner because females release free larvae within the water column after fecundation. However, the fecundation is internal (Loir et al. 1989) and the actual larvae dispersal after spawning needs to be better documented. A restrained larvae dispersal could explain a pattern of spatial genetic variation that is similar between this species and some of the brood-hiders.

Impact of harvesting in fish from the Upper-Maroni

Few evidence of over-harvesting, yet particular areas and species deserve attention

A downstream decrease of genetic diversity, expected under an extended bottleneck resulting from high fishing pressure, should be qualitatively detectable if connectivity does no longer counterbalance genetic drift. This should indeed lead to the emergence of downstream, demographically independent units showing genetic differentiation. In our study though, none of the species showed genetic diversity constantly lower in villages than in reference sites. No significant inbreeding, which might suggest harvesting impact in mating strategies (Lowe et al. 2005) was detected. Overall, therefore, there was no evidence of fishing-induced local genetic drift – at least without a current compensation by genetic rescue through gene flow, in villages of the UM.

According to pairwise F_{ST} values, there was significant genetic differentiation between downstream sites (villages) only for *S. eigenmanni* and *H. guianensis* with isolated V2, for *G. harreri* with both V1 and V2 showing lower heterozygosity than the three others villages, and for *H. aimara* with isolated V1 (although V1 and V2 both showed lower polymorphism). V1 (Papaïchton) showed reduced levels of genetic diversity and increased mean isolation for *G. harreri* and *H. aimara*, along with the highest number of finite N_e values across species (8 finite values out of 13 species sampled), particularly for species with high fishing interest (6 out of 8 species). Although N_e data should be interpreted with caution since some samples were collected in two separate periods of time in 2014 and 2016 (but not for *H. aimara*), these results suggest a particular sensitivity of Papaïchton's fish populations to fishing pressure. Papaïchton is associated with the highest human population. Its surroundings have among the smallest multispecies catch per unit effort (CPUE) (Longin 2016, Table C2.1). Longin underlined fishing practices to be more dedicated to small-scale commercial trade around Papaïchton, where fishermen target bigger specimens including the emblematic species *H. aimara*, captured using dedicated traps. *H. aimara* is among the most captured species in Papaïchton and accounted for 56% of the harvested biomass in this site according to Longin (2016). Our results, in line with these considerations, might reflect local overharvesting for this species. There were also substantial practices of poison-fishing during past decades in Papaïchton surroundings, hence a past and current fishing pressure stronger than in other villages. Decrease of water quality as a result of goldmining downstream of Maripasoula might be an additional constraining factor. These results should encourage vigilance with regards to long-term monitoring of fish communities in Papaïchton. Another village, V2 (Kayodé) also showed higher differentiation and lower genetic diversity than other villages for *S. eigenmanni*, *H. guianensis* and *G. harreri*. In contrast with most other villages, V2 is not located on the main river, but is on the Waki/Tampok branch. Thus, this reduction in genetic variability and increase in genetic isolation could be explained by similar environmental feature as put forward above. Indeed, Longin (2016) invoked the effect of seasonal variations in water levels limiting species capturability through a "dilution effect" to explain smaller CPUE within V2 (Table C2.1).

To conclude, our study showed no overall evidence for harvesting-induced local genetic erosion subsequent to fishing pressure in the UM. However, V1 (Papaïchton), located in a more urbanized area and V2 (Kayodé), located on a tributary with habitat constraints potentially promoting diversity loss, should be of particular interest regarding conservation practices.

Conversely, reference sites from the Litani and the Marouini might act, under certain assumptions, as a genetic diversity backup for the UM since these areas are less impacted by fishing while keeping high levels of genetic diversity, in spite of their upstream position. In addition, some species might be more vulnerable to fishing and require specific consideration for monitoring. This is the case of *H. aimara*, but also *T. lebaili* and *M. rhomboidalis* for instance, these species being of very high cultural and economical interest for native societies and showing potentially lower gene flow at regional scale than other Serrasalminae species in our dataset.

Toward the understanding of harvesting pressure influence on metapopulations genetics in the UM

The possibility of taking the exact contribution of migration rates and local effective sizes into account when assessing genetic structure, might help fully resolve the influence of harvesting on fish populations in the UM. For instance, equal levels of genetic diversity between areas of various fishing pressure could be due to both forces (downstream-directed gene flow and fishing in downstream locations) counterbalancing each other. Similarly, actually observing an expected DEN pattern with higher downstream genetic variation should not lead to a conclusion of no influence of fishing pressure. The latter could exist but fail to be detected if genetic drift is too recent or not strong enough to counteract higher downstream connectivity. In such cases, downstream genetic diversity loss subsequent to fishing could just lower the expected DEN pattern without suppressing it, making any qualitative conclusion uncertain. We could even expect a persistence or reinforcement of the DEN pattern if downstream reduction of effective size diminishes emigration rates from these areas, turning them into sinks and upstream areas as sources. These so-called 'genetic mass effects' (Kunin 1998), induced by the resulting downstream-biased gene flow (e.g., Campbell-Grant et al. 2007) was observed for instance in the case of central stonerollers sampled across a catchment with urbanized downstream areas in North America (Waits et al. 2008).

In some of the aforementioned cases, detecting ongoing harvesting pressure would imply to quantitatively assess the actual deviation of a metapopulation structuring from what should be expected without fishing. That is, for instance, the gap between levels of downstream genetic diversity that are observed, and these that should actually be without fishing pressure and given various deme effective sizes and direction and intensity of migration rates. The use of more sophisticated modelling approach, e.g. the simulations of genetic data, would enable to test for a large range of these parameters and confront the output to empirical datasets. Ultimately, a temporal approach could be of the highest interest to monitor the longer-term dynamics of genetic diversity in the UM, as performed in other river systems (e.g. Mathieu-Bégné et al. 2018). Such considerations are important because localized constraints could actually have an extended impact onto the whole metapopulation dynamics (Palstra & Ruzzante 2011), influencing interdemographic variance of gene flow or local effective size (Hindar et al. 2004).

General conclusion

Overall, genetic diversity across eighteen species from the UM was higher in the downstream sites and western part of the watershed, advocating these reference areas to be preferentially

preserved for management applications. However, there was a suggestion, to be taken cautiously, of higher vulnerability for populations located around the town of Papaïchton. This phenomenon might be associated to overharvesting in some species, particularly *H.aimara*, which is actively targeted in this area, but this site deserves general attention with regards to the stronger influence of urbanization or water pollution through goldmining activities, for all species. Out of harvested areas, there was a decrease of genetic diversity - particularly for sedentary, habitat-specialist species - within eastern tributaries of the UM, likely to be associated to environmental and habitat constraints. We advocate long-term monitoring of fish metapopulations from the UM in order to preserve the current patterns of genetic diversity that were observed across the study area.

Acknowledgements

This work was funded by the Amazonian Park of French Guiana, the DEAL and the Office de l'Eau de Guyane. We thank all persons involved in fieldwork, including local mediators and fishermen who helped with sample collection in villages and reference sites of the study area: Raymond Essimon, Pitoma Mekouanali, Janakale Makiloewala, Kutaka Aïtalé, Kuliwaïkë Menali, Moloko Atuwaïke, Palanaïwa Alikana, Sébastien Amaïpetit, Etulano Yamo, Mones Tokotoko, Marc Pinson, Hervé Tolinga, Kalou and Alfred Djaba. We thank all inhabitants and fishermen of villages from the Upper-Maroni for their hospitality and support.

2.6. References

- Abell, R. (2002). Conservation Biology for the Biodiversity Crisis: a Freshwater Follow-up. *Conservation Biology*, 16(5), 1435-1437. <https://doi.org/10.1046/j.1523-1739.2002.01532.x>
- Abell, R., Thieme, M. L., Revenga, C., Bryer, M., Kottelat, M., Bogutskaya, N., ... Petry, P. (2008). Freshwater Ecoregions of the World: A New Map of Biogeographic Units for Freshwater Biodiversity Conservation. *BioScience*, 58(5), 403-414. <https://doi.org/10.1641/B580507>
- Agostinho, A. A., Gomes, L. C., Santos, N. C. L., Ortega, J. C. G., & Pelicice, F. M. (2016). Fish assemblages in Neotropical reservoirs: Colonization patterns, impacts and management. *Fisheries Research*, 173, 26-36. <https://doi.org/10.1016/j.fishres.2015.04.006>
- Allendorf, F., England, P., Luikart, G., Ritchie, P., & Ryman, N. (2008). Genetic effects of harvest on wild animal populations. *Trends in Ecology & Evolution*, 23(6), 327-337. <https://doi.org/10.1016/j.tree.2008.02.008>
- Antonio, R. R., Agostinho, A. A., Pelicice, F. M., Bailly, D., Okada, E. K., & Dias, J. H. P. (2007). Blockage of migration routes by dam construction: can migratory fish find alternative routes? *Neotropical Ichthyology*, 5(2), 177-184. <https://doi.org/10.1590/S1679-62252007000200012>
- Azevedo, M. A. (2010). Reproductive characteristics of characid fish species (Teleostei, Characiformes) and their relationship with body size and phylogeny. *Iheringia. Série Zoologia*, 100(4), 469-482. <https://doi.org/10.1590/S0073-47212010000400020>
- Barletta, M., Jaureguizar, A. J., Baigun, C., Fontoura, N. F., Agostinho, A. A., Almeida-Val, V. M. F., ... Corrêa, M. F. M. (2010). Fish and aquatic habitat conservation in South America: a continental overview with emphasis on neotropical systems. *Journal of Fish Biology*, 76(9), 2118-2176. <https://doi.org/10.1111/j.1095-8649.2010.02684.x>
- Batista, J. S., & Alves-Gomes, J. A. (2006). Phylogeography of *Brachyplatystoma rousseauxii* (Siluriformes - Pimelodidae) in the Amazon Basin offers preliminary evidence for the first case of "homing" for an Amazonian migratory catfish. *Genetics and Molecular Research*, 18.
- Bayley, P., & Petrere, M. (s. d.). Amazon fisheries: assessment methods, current status and management options.
- Brauer, C. J., Hammer, M. P., & Beheregaray, L. B. (2016). Riverscape genomics of a threatened fish across a hydroclimatically heterogeneous river basin. *Molecular Ecology*, 25(20), 5093-5113. <https://doi.org/10.1111/mec.13830>
- Caballero, A., & Toro, M. A. (2002). Analysis of genetic diversity for the management of conserved subdivided populations. *Conservation Genetics*, 3, 289-299.
- Camargo, M., Giarrizzo, T., & Isaac, V. (2015). Population and biological parameters of selected fish species from the middle Xingu River, Amazon Basin. *Brazilian Journal of Biology*, 75(3 suppl 1), 112-124. <https://doi.org/10.1590/1519-6984.01914BM>
- Campbell Grant, E. H., Lowe, W. H., & Fagan, W. F. (2007). Living in the branches: population dynamics and ecological processes in dendritic networks. *Ecology Letters*, 10(2), 165-175. <https://doi.org/10.1111/j.1461-0248.2006.01007.x>
- Carolsfeld, J., Harvey, B., Ross, C., & Baer, A. (2004). *Migratory Fishes of South America: Biology, Fisheries and Conservation Status*. (World Fisheries Trust, The World Bank, International Development Research Centre, Éd.). The World Bank. <https://doi.org/10.1596/1-5525-0114-0>
- Carrara, F., Rinaldo, A., Giometto, A., & Altermatt, F. (2014). Complex Interaction of Dendritic Connectivity and Hierarchical Patch Size on Biodiversity in River-Like Landscapes. *The American Naturalist*, 183(1), 13-25. <https://doi.org/10.1086/674009>

- Carvalho, R. A., & Tejerina-Garro, F. L. (2015). The influence of environmental variables on the functional structure of headwater stream fish assemblages: a study of two tropical basins in Central Brazil. *Neotropical Ichthyology*, *13*(2), 349-360. <https://doi.org/10.1590/1982-0224-20130148>
- Chaput-Bardy, A., Fleurant, C., Lemaire, C., & Secondi, J. (2009). Modelling the effect of in-stream and overland dispersal on gene flow in river networks. *Ecological Modelling*, *220*(24), 3589-3598. <https://doi.org/10.1016/j.ecolmodel.2009.06.027>
- Charlesworth, D., & Wright, S. I. (2001). Breeding systems and genome evolution. *Current Opinion in Genetics & Development*, *11*(6), 685-690. [https://doi.org/10.1016/S0959-437X\(00\)00254-9](https://doi.org/10.1016/S0959-437X(00)00254-9)
- Cilleros, K., Valentini, A., Allard, L., Dejean, T., Etienne, R., Grenouillet, G., ... Brosse, S. (2018). Unlocking biodiversity and conservation studies in high-diversity environments using environmental DNA (eDNA): A test with Guianese freshwater fishes. *Molecular Ecology Resources*. <https://doi.org/10.1111/1755-0998.12900>
- Cohen, M. A., Beaumont, R. C. B., & Thorp, N. C. (1999). Movement and activity patterns of the black piranha. *Environmental Biology of Fishes*, *54*, 45-52.
- Comte, L., & Olden, J. D. (2018a). Evidence for dispersal syndromes in freshwater fishes. *Proceedings of the Royal Society B: Biological Sciences*, *285*(1871), 20172214. <https://doi.org/10.1098/rspb.2017.2214>
- Comte, L., & Olden, J. D. (2018b). Fish dispersal in flowing waters: A synthesis of movement- and genetic-based studies. *Fish and Fisheries*. <https://doi.org/10.1111/faf.12312>
- Delord, C., Lassalle, G., Oger, A., Barloy, D., Coutellec, M.-A., Delcamp, A., ... Petit, E. J. (2018). A cost- and-time effective procedure to develop SNP markers for multiple species: A support for community genetics. *Methods in Ecology and Evolution*. <https://doi.org/10.1111/2041-210X.13034>
- Del Re, A.C. (2013). compute.es: Compute Effect Sizes. R package version 0.2-2. <http://cran.r-project.org/web/packages/compute.es>
- Do, C., Waples, R. S., Peel, D., Macbeth, G. M., Tillett, B. J., & Ovenden, J. R. (2014). NeEstimator v2: re-implementation of software for the estimation of contemporary effective population size (N_e) from genetic data. *Molecular Ecology Resources*, *14*(1), 209-214. <https://doi.org/10.1111/1755-0998.12157>
- Fazzi-Gomes, P., Guerreiro, S., Palheta, G. D. A., Melo, N. F. A. C. de, Santos, S., & Hamoy, I. (2017). High genetic diversity and connectivity in *Colossoma macropomum* in the Amazon basin revealed by microsatellite markers. *Genetics and Molecular Biology*, *40*(1), 142-146. <https://doi.org/10.1590/1678-4685-gmb-2015-0222>
- Fleury, M. (1998). Les populations du Haut-Maroni et le projet de Parc national de la Guyane. *Journal d'agriculture traditionnelle et de botanique appliquée*, *40*(1-2), 577-610. <https://doi.org/10.3406/jatba.1998.3693>
- Fleury, M., Alupki, T., Opoya, A., & Aloïké, W. (2016). Les Wayana de Guyane française sur les traces de leur histoire: Cartographie participative sur le Litani (Aletani) et mémoire orale. *Revue d'ethnoécologie*, (9). <https://doi.org/10.4000/ethnoecologie.2711>
- Frankham, R. (1995). Conservation Genetics. *Annual Review of Genetics*, (29), 305-327.
- Géry, J. 1972. Poissons characoïdes des Guyanes. I. Généralités. II. Famille des Serrasalmidæ. Zoologische Verhandelingen (Leiden) No. 122: 1-250, Pls. 1-16.
- Gilbert, K. J., & Whitlock, M. C. (2015). Evaluating methods for estimating local effective population size with and without migration: ESTIMATING N_e IN THE PRESENCE OF MIGRATION. *Evolution*, *69*(8), 2154-2166. <https://doi.org/10.1111/evo.12713>

- Gomez-Uchida, D., Knight, T. W., & Ruzzante, D. E. (2009). Interaction of landscape and life history attributes on genetic diversity, neutral divergence and gene flow in a pristine community of salmonids. *Molecular Ecology*, 18(23), 4854-4869. <https://doi.org/10.1111/j.1365-294X.2009.04409.x>
- Goudet, J. (1995). FSTAT (Version 1.2): A computer program to calculate F-Statistics. *Journal of Heredity*, 86(6), 485-486.
- Goudet, J. (2005). hierfstat, a package for r to compute and test hierarchical F-statistics. *Molecular Ecology Notes*, 5(1), 184-186. <https://doi.org/10.1111/j.1471-8286.2004.00828.x>
- Goudet, J. & Jombart, T. (2017). hierfstat: Estimation and Tests of Hierarchical F-Statistics. <http://www.r-project.org>, <http://github.com/jgx65/hierfstat>.
- Gutiérrez, J. P., Royo, L. J., Álvarez, I., & Goyache, F. (2005). MolKin v2.0: A Computer Program for Genetic Analysis of Populations Using Molecular Coancestry Information. *Journal of Heredity*, 96(6), 718-721. <https://doi.org/10.1093/jhered/esi118>
- Hiltsdorf, A. W. S., & Hallerman, E. M. (2017). *Genetic Resources of Neotropical Fishes*. Cham: Springer International Publishing. <https://doi.org/10.1007/978-3-319-55838-7>
- Hindar, K., Tufto, J., Sattem, L., & Balstad, T. (2004). Conservation of genetic variation in harvested salmon populations. *ICES Journal of Marine Science*, 61(8), 1389-1397. <https://doi.org/10.1016/j.icesjms.2004.08.011>
- Hrbek, T., Crossa, M., & Farias, I. (2007). Conservation strategies for *Arapaima gigas* (Schinz, 1822) and the Amazonian várzea ecosystem. *Brazilian Journal of Biology*, 67(4 suppl), 909-917. <https://doi.org/10.1590/S1519-69842007000500015>
- Iervolino, F., de Resende, E. K., & Hiltsdorf, A. W. S. (2010). The lack of genetic differentiation of pacu (*Piaractus mesopotamicus*) populations in the Upper-Paraguay Basin revealed by the mitochondrial DNA D-loop region: Implications for fishery management. *Fisheries Research*, 101(1-2), 27-31. <https://doi.org/10.1016/j.fishres.2009.09.003>
- Jari Oksanen, F. Guillaume Blanchet, Michael Friendly, Roeland Kindt, Pierre Legendre, Dan McGlenn, Peter R. Minchin, R. B. O'Hara, Gavin L. Simpson, Peter Solymos, M. Henry H. Stevens, Eduard Szoecs and Helene Wagner (2018). vegan: Community Ecology Package. R package version 2.4-6. <https://CRAN.R-project.org/package=vegan>
- Jégu, M., Keith, P., & Belmont-JéGu, E. (2002). UNE NOUVELLE ESPÈCE DE TOMETES (TELEOSTEI : CHARACIDAE : SERRASALMINAE) DU BOUCLIER GUYANAIS, TOMETES LEBAILI N. SP. *Bulletin Français de la Pêche et de la Pisciculture*, (364), 23-48. <https://doi.org/10.1051/kmae:2002002>
- Jégu, Michel, & Keith, P. (2005). Threatened fishes of the world: Tometes lebaili (Jégu, Keith & Belmont-Jégu 2002) (Characidae: Serrasalminae). *Environmental Biology of Fishes*, 72(4), 378-378. <https://doi.org/10.1007/s10641-004-4126-4>
- Keenan, K., McGinnity, P., Cross, T. F., Crozier, W. W., & Prodöhl, P. A. (2013). diveRsity : An R package for the estimation and exploration of population genetics parameters and their associated errors. *Methods in Ecology and Evolution*, 4(8), 782-788. <https://doi.org/10.1111/2041-210X.12067>
- Keith, P., Le Bail, P.-Y., & Planquette, P. (2000). *Atlas des poissons d'eau douce de Guyane - Tome 2, Fascicule I*. Paris.
- Kullander, S. O. (2003). Family Cichlidae. In *Checklist of the Freshwater fishes of South and Central America* (Edipucrs). Consulté à l'adresse <http://www.bioone.org/doi/abs/10.1643/OT-04-142>
- Kunin, W. E. (1998). Biodiversity at the edge: A test of the importance of spatial « mass effects » in the Rothamsted Park Grass experiments. *Proceedings of the National Academy of Sciences*, 95(1), 207-212. <https://doi.org/10.1073/pnas.95.1.207>

- Labonne, J., Ravigné, V., Parisi, B., & Gaucherel, C. (2008). Linking dendritic network structures to population demogenetics: The downside of connectivity. *Oikos*, *117*(10), 1479-1490. <https://doi.org/10.1111/j.0030-1299.2008.16976.x>
- Lê, S., Josse, J. & Husson, F. (2008). FactoMineR: An R Package for Multivariate Analysis. *Journal of Statistical Software*. *25*(1). pp. 1-18.
- Le Bail, P.-Y., Covain, R., Jégu, M., Fisch-Muller, S., Vigouroux, R., & Keith, P. (2012). Updated checklist of the freshwater and estuarine fishes of French Guiana, 28.
- Le Bail, P.-Y., Keith, P., & Planquette, P. (2000). *Atlas des poissons d'eau douce de Guyane - Tome 2, Fascicule II*. Paris.
- Lenth, R. (2018). emmeans: Estimated Marginal Means, aka Least-Squares Means. R package version 1.2.2. <https://CRAN.R-project.org/package=emmeans>
- Longin, G. (2016). *La pêche chez les Wayana, Teko et Aluku en 2014 sur le Haut-Maroni (Guyane française), complémentarité des enquêtes halieutiques et des cartes cognitives*. Parc Amazonien de Guyane.
- Lord, C., Fermon, Y., Meunier, F. J., Jégu, M., & Keith, P. (2007). Croissance et longévité du Watau yaïke, *Tometes lebailli* (Serrasalminae), dans le bassin du Haut Maroni (Guyane française). Résultats préliminaires, 9.
- Lowe, A. J., Boshier, D., Ward, M., Bacles, C. F. E., & Navarro, C. (2005). Genetic resource impacts of habitat loss and degradation; reconciling empirical evidence and predicted theory for neotropical trees. *Heredity*, *95*(4), 255-273. <https://doi.org/10.1038/sj.hdy.6800725>
- Manel, S., Schwartz, M. K., Luikart, G., & Taberlet, P. (2003). Landscape genetics: combining landscape ecology and population genetics. *Trends in Ecology & Evolution*, *18*(4), 189-197. [https://doi.org/10.1016/S0169-5347\(03\)00008-9](https://doi.org/10.1016/S0169-5347(03)00008-9)
- Mathieu-Bégné, E., Loot, G., Chevalier, M., Paz-Vinas, I., & Blanchet, S. (2018). Demographic and genetic collapses in spatially structured populations: insights from a long-term survey in wild fish metapopulations. *Oikos*. <https://doi.org/10.1111/oik.05511>
- Mérigoux, S., Dolédec, S., & Statzner, B. (2001). Species traits in relation to habitat variability and state: neotropical juvenile fish in floodplain creeks. *Freshwater Biology*, *46*(9), 1251-1267. <https://doi.org/10.1046/j.1365-2427.2001.00744.x>
- Meunier, F. J. (coord. . (2004). *Piranhas enivrés, des poissons et des hommes en Guyane*. SFI/RMN éditeurs.
- Meunier, F. J., Rojas-Beltran, R., Boujard, T., & Lecomte, F. (1994). Rythmes saisonniers de la croissance chez quelques Téléostéens de Guyane française. *Revue d'hydrobiologie tropicale*, *27*(4), 423-440.
- Neel, M. C., McKelvey, K., Ryman, N., Lloyd, M. W., Short Bull, R., Allendorf, F. W., ... Waples, R. S. (2013). Estimation of effective population size in continuously distributed populations: there goes the neighborhood. *Heredity*, *111*(3), 189-199. <https://doi.org/10.1038/hdy.2013.37>
- Oliveira, C., Foresti, F., & Hilsdorf, A. W. S. (2009). Genetics of neotropical fish: from chromosomes to populations. *Fish Physiology and Biochemistry*, *35*(1), 81-100. <https://doi.org/10.1007/s10695-008-9250-1>
- Pagezy, H., & Jégu, M. (2002). Valeur patrimoniale de Serrasalminae herbivores du Haut-Maroni (Guyane française): approches biologiques et socioculturelle en pays Wayana. *Bulletin Français de la Pêche et de la Pisciculture*, (364), 49-69. <https://doi.org/10.1051/kmae:2002003>
- Palstra, F. P., & Ruzzante, D. E. (2011). Demographic and genetic factors shaping contemporary metapopulation effective size and its empirical estimation in salmonid fish. *Heredity*, *107*(5), 444-455. <https://doi.org/10.1038/hdy.2011.31>

- Paz-Vinas, I., & Blanchet, S. (2015). Dendritic connectivity shapes spatial patterns of genetic diversity: a simulation-based study. *Journal of Evolutionary Biology*, 28(4), 986-994. <https://doi.org/10.1111/jeb.12626>
- Paz-Vinas, I., Loot, G., Stevens, V. M., & Blanchet, S. (2015). Evolutionary processes driving spatial patterns of intraspecific genetic diversity in river ecosystems. *Molecular Ecology*, 24(18), 4586-4604. <https://doi.org/10.1111/mec.13345>
- Pereira, L. H. G., Foresti, F., & Oliveira, C. (2009). Genetic structure of the migratory catfish *Pseudoplatystoma corruscans* (Siluriformes: Pimelodidae) suggests homing behaviour. *Ecology of Freshwater Fish*, 18(2), 215-225. <https://doi.org/10.1111/j.1600-0633.2008.00338.x>
- Pflüger, F. J., & Balkenhol, N. (2014). A plea for simultaneously considering matrix quality and local environmental conditions when analysing landscape impacts on effective dispersal. *Molecular Ecology*, 23(9), 2146-2156. <https://doi.org/10.1111/mec.12712>
- Pilger, T. J., Gido, K. B., Propst, D. L., Whitney, J. E., & Turner, T. F. (2017). River network architecture, genetic effective size and distributional patterns predict differences in genetic structure across species in a dryland stream fish community. *Molecular Ecology*, 26(10), 2687-2697. <https://doi.org/10.1111/mec.14079>
- Piorski, N., Sanches, A., Carvalho-Costa, L., Hatanaka, T., Carrillo-Avila, M., Freitas, P., & Galetti Jr., P. (2008). Contribution of conservation genetics in assessing neotropical freshwater fish biodiversity. *Brazilian Journal of Biology*, 68(4 suppl), 1039-1050. <https://doi.org/10.1590/S1519-69842008000500011>
- Planquette, P., Le Bail, P.-Y., & Keith, P. (1996). *Atlas des poissons d'eau douce de Guyane*. Paris.
- Ponton, D., & Copp, G. H. (1997). Early dry-season community structure and habitat use of young fish in tributaries of the River Sinnamary (French Guiana, South America) before and after hydrodam operation. *Environmental Biology of Fishes*, 50, 235-256.
- Ponton, D., & De Mérona, B. (1998). Fish life-history tactics in a neotropical river with a highly stochastic hydrological regime : the Sinnamary river, French Guiana, South America, 24.
- Prunier, J., Dubut, V., Chikhi, L., & Blanchet, S. (2017). Contribution of spatial heterogeneity in effective population sizes to the variance in pairwise measures of genetic differentiation. <https://doi.org/10.1101/031633>
- Radinger, J., & Wolter, C. (2014). Patterns and predictors of fish dispersal in rivers. *Fish and Fisheries*, 15(3), 456-473. <https://doi.org/10.1111/faf.12028>
- Reis, R. E. (2013). Conserving the freshwater fishes of South America: Conserving South American Freshwater Fishes. *International Zoo Yearbook*, 47(1), 65-70. <https://doi.org/10.1111/izy.12000>
- Reis, R. E., Albert, J. S., Di Dario, F., Mincarone, M. M., Petry, P., & Rocha, L. A. (2016). Fish biodiversity and conservation in South America: FISH BIODIVERSITY AND CONSERVATION. *Journal of Fish Biology*, 89(1), 12-47. <https://doi.org/10.1111/jfb.13016>
- Rodríguez-Olarte, D., Taphorn, D. C., & Lobón-Cerviá, J. (2011). Do protected areas conserve neotropical freshwater fishes? A case study of a biogeographic province in Venezuela. *Animal Biodiversity and Conservation*, 13.
- Salisbury, S. J., McCracken, G. R., Keefe, D., Perry, R., & Ruzzante, D. E. (2016). A portrait of a sucker using landscape genetics: how colonization and life history undermine the idealized dendritic metapopulation. *Molecular Ecology*, 25(17), 4126-4145. <https://doi.org/10.1111/mec.13757>
- Santos, C. H. dos A. dos, de Sá Leitão, C. S., Paula-Silva, M. de N., & Almeida-Val, V. M. F. (2016). Genetic differentiation in red-bellied piranha populations (*Pygocentrus nattereri* , Kner, 1858) from the

- Solimões-Amazonas River. *Ecology and Evolution*, 6(12), 4203-4213. <https://doi.org/10.1002/ece3.2195>
- Santos, M. C. F., Ruffino, M. L., & Farias, I. P. (2007). High levels of genetic variability and panmixia of the tambaqui *Colossoma macropomum* (Cuvier, 1816) in the main channel of the Amazon River. *Journal of Fish Biology*, 71(sa), 33-44. <https://doi.org/10.1111/j.1095-8649.2007.01514.x>
- Smith, N. J. H. (1985). The impact of cultural and ecological change on Amazonian fisheries. *Biological Conservation*, 32(4), 355-373. [https://doi.org/10.1016/0006-3207\(85\)90023-0](https://doi.org/10.1016/0006-3207(85)90023-0)
- Tejerina-Garro, F. L., de Mérona, B., Oberdorff, T., & Hugueny, B. (2006). A fish-based index of large river quality for French Guiana (South America): method and preliminary results. *Aquatic Living Resources*, 19(1), 31-46. <https://doi.org/10.1051/alr:2006003>
- Telles, M. P. C., Resende, L. V., Brondani, R. P. V., Collevatti, R. G., Costa, M. C., & Silva Junior, N. J. (2010). Short Communication Isolation and characterization of microsatellite markers in the armored catfish *Hypostomus gymnorhynchus* (Loricariidae). *Genetics and Molecular Research*, 9(3), 1770-1774. <https://doi.org/10.4238/vol9-3gmr868>
- Thieme, M., Lehner, B., Abell, R., Hamilton, S. K., Kellendorfer, J., Powell, G., & Riveros, J. C. (2007). Freshwater conservation planning in data-poor areas: An example from a remote Amazonian basin (Madre de Dios River, Peru and Bolivia). *Biological Conservation*, 135(4), 484-501. <https://doi.org/10.1016/j.biocon.2006.10.054>
- Thomaz, A. T., Christie, M. R., & Knowles, L. L. (2016). The architecture of river networks can drive the evolutionary dynamics of aquatic populations: BRIEF COMMUNICATION. *Evolution*, 70(3), 731-739. <https://doi.org/10.1111/evo.12883>
- Tito de Moraes, L., & Raffray, J. (1999). Movements of *Hoplias aimara* during the filling phase of the Petit-Saut dam, French Guyana. *Journal of Fish Biology*, 54(3), 627-635. <https://doi.org/10.1111/j.1095-8649.1999.tb00641.x>
- Toussaint, A., Charpin, N., Brosse, S., & Villéger, S. (2016). Global functional diversity of freshwater fish is concentrated in the Neotropics while functional vulnerability is widespread. *Scientific Reports*, 6(1). <https://doi.org/10.1038/srep22125>
- Waits, E. R., Bagley, M. J., Blum, M. J., McCormick, F. H., & Lazorchak, J. M. (2008). Source-sink dynamics sustain central stonerollers (*Camptostoma anomalum*) in a heavily urbanized catchment. *Freshwater Biology*, 53(10), 2061-2075. <https://doi.org/10.1111/j.1365-2427.2008.02030.x>
- Walsh, M. R., Munch, S. B., Chiba, S., & Conover, D. O. (2006). Maladaptive changes in multiple traits caused by fishing: impediments to population recovery: Evolutionary consequences of harvesting. *Ecology Letters*, 9(2), 142-148. <https://doi.org/10.1111/j.1461-0248.2005.00858.x>
- Waters, J. M., & Burrige, C. P. (2016). Fine-scale habitat preferences influence within-river population connectivity: a case-study using two sympatric New Zealand *Galaxias* fish species. *Freshwater Biology*, 61(1), 51-56. <https://doi.org/10.1111/fwb.12675>
- Watson, L. C., Stewart, D. J., & Kretzer, A. M. (2016). Genetic Diversity and Population Structure of the Threatened Giant *Arapaima* in Southwestern Guyana: Implications for Their Conservation. *Copeia*, 104(4), 864-872. <https://doi.org/10.1643/CG-15-293>
- Weber, C., Covain, R., & Fisch-Muller, S. (2012). Identity of *Hypostomus plecostomus* (Linnaeus, 1758), with an overview of *Hypostomus* species from the Guianas (Teleostei: Siluriformes: Loricariidae), 34.
- Weir, B. S., & Cockerham, C. C. (1984). Estimating F-Statistics for the Analysis of Population Structure. *Evolution*, 38(6), 1358. <https://doi.org/10.2307/2408641>

Chapitre 3:

Analyse exploratoire de l'évolution de la variabilité génétique au sein de populations d'un réseau hydrographique dendritique : influence d'une pression de pêche localisée

Au cours de ce travail de thèse, j'ai pu découvrir et expérimenter l'intérêt des simulations en génétique des populations, en complément de mon travail de description des patrons empiriques de variabilité génétique chez les espèces étudiées sur le Haut-Maroni. Dans un premier temps, j'ai utilisé ces outils pour confirmer l'influence attendue de la configuration spatiale dendritique sur la variabilité génétique de populations génotypées sur des loci bialléliques. J'ai ensuite initié de premières analyses visant à comprendre comment une pression de pêche localisée sur les zones de plus forte connectivité spatiale (comme les zones proches des confluences d'un bassin versant) pourrait influencer la structuration génétique globale des populations, ainsi que les niveaux de diversité locaux dans les réseaux hydrographiques dendritiques malgré l'influence de cette configuration spatiale *per se*. Il s'agit donc, à ce stade, d'une approche exploratoire visant principalement à soutenir qualitativement l'interprétation des données empiriques présentées dans le chapitre précédent. Cette approche permettra par ailleurs d'affiner ou d'énoncer des hypothèses complémentaires dont nous pourrions poursuivre l'étude par la suite à l'aide des outils de simulations.

Afin de mieux percevoir les avantages et inconvénients des différents outils de simulation existants, j'ai eu l'occasion d'échanger avec des chercheurs de différents laboratoires extérieurs. Afin de me former, de manière plus concrète, à l'utilisation d'un de ces outils de simulations, j'ai choisi d'y consacrer mon séjour de 3 mois à l'étranger, obligatoire dans le cadre du programme doctoral EIR-Agreenium. J'ai eu la chance de me rendre à l'Université de Saint-Andrews au sein du CREEM (*Centre for Research into Ecological and Environmental Modelling*), où j'ai eu la possibilité de travailler auprès du Professeur Oscar Gaggiotti. Ces travaux, encore préliminaires, seront présentés sous forme d'un chapitre court présentant d'abord la démarche et les principaux axes de réflexion explorés, puis les premiers résultats obtenus ainsi que les perspectives à poursuivre. Ce chapitre est rédigé en français.

1. Introduction

Actuellement, de nombreuses pressions anthropiques peuvent bouleverser la dynamique naturelle des métapopulations, par exemple en altérant les patrons de connectivité (Schick & Lindley 2007) ou en influençant les probabilités locales d'extinction (Carranza et al. 2012). L'influence de tels facteurs sur le maintien des écosystèmes peut être théorisée par des outils de modélisation de métapopulations (*patch occupancy models*, ex. Fagan 2002) ou d'abondance locale des populations (Labonne et al. 2008; Webb & Padgham 2013). La théorie des graphes (Urban et al. 2009) selon laquelle, par exemple, une métapopulation peut être représentée par un ensemble de « nœuds » (patchs d'habitats ou populations) et de « liens » (connexions entre patchs, représentatifs de la connectivité) constitue également un cadre de travail pertinent (Saunders et al. 2015). Dans les milieux aquatiques, ces différentes approches ont été très informatives pour évaluer l'impact de la fragmentation des habitats, tout particulièrement en eau douce, sur la persistance des métapopulations.

Ainsi dans les réseaux hydrographiques, la distribution spatiale naturelle de diversité spécifique et génétique est actuellement bien connue à la fois par les outils théoriques et les observations empiriques (Hughes et al. 2009; Carrara et al. 2012; Cañedo-Argüelles et al. 2015; Paz-Vinas & Blanchet 2015). Les réseaux hydrographiques présentent en effet souvent une structure dite dendritique (*dendritic-ecological-networks*, ou DEN), se caractérisant par une succession hiérarchique et ascendante d'embranchements depuis le cours principal jusqu'aux petits cours d'eau amont (Campbell-Grant et al. 2007). Cette structure spatiale contraint fortement la distribution de la diversité, tout du moins pour ce qui concerne les espèces exclusivement aquatiques (Hughes et al. 2013). Cette distribution dépend d'une part de facteurs régionaux (arrangement spatial des cours d'eau et complexité du réseau hydrographique), qui influencent principalement la connectivité entre patchs en fonction des capacités de dispersion des espèces ou des individus. Elle dépend aussi de facteurs locaux comme les conditions environnementales ou la capacité de charge des patchs (Hughes et al. 2013; Cañedo-Argüelles et al. 2015), qui influencent leur abondance locale ou leur taille efficace. Les patchs qui se trouvent en position centrale du réseau, comme les confluences et les cours principaux, présentent généralement une capacité de charge et un degré de connectivité supérieurs aux cours d'eau amont, plus petits et plus périphériques. Ils concentrent donc généralement une plus forte diversité spécifique et génétique (Labonne et al. 2008; Paz-Vinas & Blanchet 2015) et peuvent occuper un rôle de réservoir ou bien de puit de diversité.

L'influence de la fragmentation des habitats sur le maintien des métapopulations de type dendritique a été approchée théoriquement (Fagan 2002) et empiriquement. Paz-Vinas et al. (2015) ont répertorié plusieurs études menées sur des populations aquatiques afin de vérifier lesquelles rapportaient un accroissement de la diversité génétique vers les zones aval des bassins versants étudiés. La majorité des métapopulations ne vérifiant pas cette caractéristique avaient été étudiées dans des zones géographiques concernées par des phénomènes de fragmentation d'origine naturelle ou anthropique. Richmond et al. (2018) ont observé des patrons de structure génétique quasi-exclusivement dépendants de la dérive génétique locale chez l'espèce endémique *Catostomus santaanae* en Californie. Les auteurs ont ainsi souligné une perte d'influence de la connectivité théorique dendritique, résultant des effets de l'urbanisation et du cycle de sécheresse annuelle qui engendrent une fragmentation importante du milieu. Lean et al. (2017) ont également constaté des populations

génétiqnement très isolées et appauvries chez l'espèce *Gadopsis marmoratus* au cœur d'un bassin versant très fragmenté au Sud-Est de l'Australie.

Si les exemples d'effets de la fragmentation sur les patrons de variabilité génétique ne manquent pas, il n'en va pas de même pour l'effet de facteurs agissant davantage sur les abondances ou les tailles efficaces locales, comme une pression d'exploitation localisée sur certains patches. Par exemple, l'impact de l'exploitation par la pêche sur la variabilité génétique neutre reste fortement sous-évalué en dehors des réseaux d'aires marines protégées (Planque et al. 2010; Saunders et al. 2015). L'exploitation par la pêche impacte pourtant fortement les écosystèmes d'eau douce (Allan et al. 2005) et peut théoriquement agir sur différentes échelles (locale et régionale) d'une métapopulation de rivière (Allendorf et al. 2008; Schindler et al. 2010). Cette influence, cependant, peut n'être pas détectable spatialement si la connectivité est suffisamment forte pour contrebalancer des impacts localisés (*genetic rescue*). Cela pourrait s'observer par exemple dans le cas des réseaux hydrographiques dendritiques non fragmentés si la pression d'exploitation intervient au niveau des zones présentant la plus forte connectivité, comme le cours d'eau principal. Waits et al. (2008) soupçonnent ce phénomène chez l'espèce *Campostoma anomalum*, malgré de fortes contraintes environnementales liées à l'urbanisation dans les zones aval du bassin étudié.

Dans le bassin versant du Haut-Maroni, en Guyane française, les communautés humaines natives se sont principalement sédentarisées au niveau du cours principal, tirant en partie leur alimentation quotidienne de l'activité de pêche menée dans les zones environnantes. Ce travail de thèse (Chapitre 2) a permis de mettre en évidence divers patrons de variabilité génétique, essentiellement en lien avec la configuration hydrographique dendritique des cours d'eau considérés. En parallèle, un effet local de la pression de pêche sur la diversité génétique n'a été observé que pour un nombre réduit de sites et/ou d'espèces. Cette absence d'effet identifiable dans la plupart des cas s'explique soit par une réelle absence d'impact de la pêche, soit par des flux de gènes suffisamment importants pour contrebalancer, du moins à l'heure actuelle, les effets de la dérive locale induits par l'exploitation de la ressource. Ainsi, il m'a semblé important de visualiser quels scénarii théoriques pouvaient expliquer nos observations empiriques, et si ces dernières peuvent être compatibles avec un impact de la pêche qui ne serait pas détectable par comparaison spatiale directe entre localités situées dans, ou hors des zones de pêches. Ce type d'information pourrait être crucial pour éclairer les résultats observés Chapitre 2, et orienter les décisions de gestion des espèces exploitées du Haut Maroni.

Dans un premier temps, j'ai simplement cherché à reconstituer le patron de variabilité génétique attendu dans un réseau hydrographique dendritique et à observer dans quelle mesure il était possible de le détecter avec des marqueurs de type bialléliques - les travaux théoriques et empiriques actuellement publiés s'appuyant essentiellement sur des marqueurs microsatellites (ex. Paz-Vinas & Blanchet 2015; Pilger et al. 2017). Dans un second temps, j'ai commencé à explorer comment ces patrons de variabilité génétique évoluaient selon un scénario mimant une pression de pêche par la réduction régulière, à chaque génération, de la taille de certaines populations localisées au niveau de patches centraux du réseau. Ici, l'objectif principal était de mieux percevoir dans quelle mesure plusieurs combinaisons différentes de paramètres (tailles de populations et taux de migration combinés à une proportion variable d'individus éliminés par pression de pêche) pouvaient aboutir à des observations empiriques similaires. Pour répondre à ces deux objectifs, j'ai procédé à la simulation de données de génétique de populations sous différents scénarii démographiques.

2. Méthodes : Simulation d'un scénario simple, avec et sans pression de pêche

L'ensemble des pseudo-données dont nous analysons les patrons de variabilité génétique ont été simulées à l'aide de la librairie SimuPOP 1.1.7 (Peng & Amos 2008), qui s'appuie sur l'environnement Python 2.8. SimuPOP est un programme de simulation individu-centré qui permet la modélisation de processus démographiques et génétiques. Je présenterai ici les résultats d'un scénario très simple que j'ai expérimenté avec ce programme. Ce scénario correspond à une métapopulation de 16 patches, numérotés de 0 à 15 et organisés selon un schéma spatial de type dendritique simplifié, comme indiqué Fig.C3.1 ci-dessous.

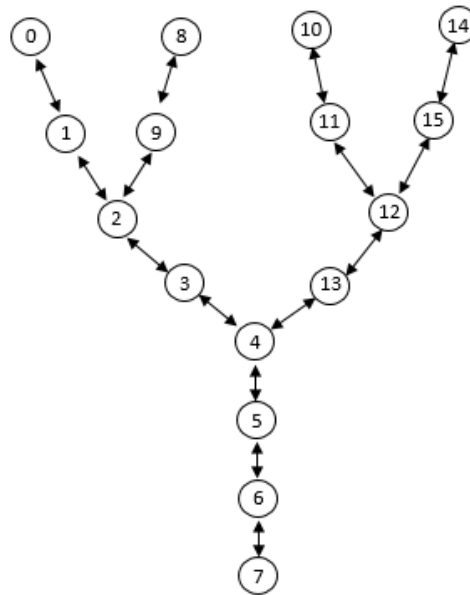


Figure C3.1: Représentation schématique du réseau dendritique servant de support à l'ensemble des simulations présentées dans ce chapitre. Les patches 0, 7, 8, 10 et 14 représentent des zones peu connectées au reste du réseau tandis que les patches 3, 4, 5 et 13 représentent une aire de plus forte connectivité.

Ces patches comportent tous un nombre identique d'individus 'N'. A chaque génération, chacun des patches échange, avec le ou les patches directement adjacents, des flux migratoires bidirectionnels (migration symétrique) selon le schéma type stepping-stone (migration de proche en proche), et selon une probabilité de migration 'm' identique pour tous les patches. Je simule des générations chevauchantes : chaque nouvelle génération se constitue arbitrairement de deux tiers d'individus déjà présents de la génération précédente, et d'un tiers de nouveaux individus correspondant à des descendants. Chaque descendant est généré par le croisement d'un couple de la génération précédente via un processus aléatoire (*random-mating*). Le sex-ratio est artificiellement maintenu équilibré au cours des générations afin d'éviter un arrêt de la simulation par impossibilité de former un couple mâle et femelle. Chaque croisement entre un mâle et une femelle génère un descendant.

➤ Dans un premier temps, je cherche à analyser si l'on retrouve le patron de variabilité génétique attendu selon la théorie DEN, et seuls les paramètres N et m interviennent. Parmi plusieurs valeurs de N testées (50; 100; 500 et 1000), je présenterai ici les résultats pour **N=100** et **N=1000**, correspondant à une taille de population modérée ou grande. Parmi plusieurs valeurs de m testées (0,001; 0,005; 0,01; 0,05; 0,10; 0,20 et 0,30), je présenterai ici les résultats pour **m=0,05** et **m=0,20**, correspondant respectivement à de faibles ou fortes probabilités de

migration entre patchs adjacents (faible ou forte connectivité génétique entre populations). Quatre combinaisons $N \times m$ seront ainsi présentées.

➤ Dans un second temps, un troisième paramètre 'H' entre en jeu dans chacune des quatre combinaisons $N \times m$. H représente une proportion de réduction de la taille des patchs n°3, 4, 5 et 13 de la métapopulation simulée. Ces quatre patchs se trouvent autour de la zone de confluence du réseau (Fig.C3.1) et constituent une représentation schématique de zones de pêches. A chaque génération, une proportion H d'individus de ces patchs est éliminée aléatoirement avant l'événement de reproduction. Au cours du temps, la taille des populations évolue donc au gré des prélèvements successifs et des flux de migrants qui surviennent entre patchs. Parmi plusieurs valeurs de H testées, allant de 0 à 0.50, je présenterai d'abord les résultats d'une valeur unique de **H=0,125**. Dans le cas précédent, H est donc considéré comme nul : **H=0**.

Un total de 30 réplicats est effectué pour chaque combinaison de N, m et H (30 procédures de simulations sont donc effectuées avec un état initial et des combinaisons de paramètres identiques). Pour chaque réplicat, 'N' génotypes sont générés pour 130 loci bialléliques sur la base de fréquences alléliques échantillonnées parmi celles observées chez les populations de *Serrasalmus rhombeus* (sélectionnée arbitrairement) du Haut-Maroni. De façon à permettre la redistribution « spatiale » des allèles au gré des différents croisements et flux migratoires successifs entre patchs, plusieurs générations sont alors simulées « à blanc » afin de permettre au système simulé d'atteindre un état d'équilibre. Les 130 premières générations sont ainsi éliminées des analyses consécutives. Par la suite, les génotypes de 20 individus, sélectionnés au hasard dans chacun des patchs, sont exportés sous forme d'un fichier de format Genepop. Un fichier Genepop est ainsi exporté toutes les 5 générations et de manière synchrone entre les réplicats. La pression de pêche simulée ('H') intervient à partir de la génération 150. Le script SimuPOP créé pour cette analyse est consultable en Annexe X.

Un ensemble de statistiques décrivant la variabilité génétique est ensuite calculé, à l'aide du logiciel R, pour chaque fichier Genepop exporté. L'évolution de ces statistiques peut alors être visualisée au cours des générations sous forme graphique. Les statistiques calculées ont été choisies pour leur simplicité et rapidité de calcul compte tenu du très grand nombre de fichiers à traiter par le logiciel R (plusieurs combinaisons de paramètres, plusieurs générations et plusieurs réplicats). Ces statistiques correspondent également à des indices très couramment calculés dans la littérature portant sur la diversité génétique dans les réseaux dendritiques (consulter par exemple Paz-Vinas et al. 2015). La liste des statistiques calculées pour chaque jeu de données est fournie ci-après (Table C3.1).

Table C3.1 : Liste des statistiques et indices de diversité génétique calculés sur chaque jeu de pseudo-données simulé :

<i>Statistique</i>	<i>Description</i>
COR_P	Coefficient de corrélation de Pearson entre le taux de polymorphisme P95 et la distance (en nombre de patchs) aux zones centrales du réseau (patch n°3 ou 13).
SP	Ecart-type de P95 à l'échelle des 16 patchs.
α_P	Pente de la relation entre le taux de polymorphisme P95 et la distance (en nombre de patchs) aux zones centrales du réseau (patch n°3 ou 13).
COR_{FST}	Coefficient de corrélation de Pearson entre l'isolement génétique moyen ($F_{ST.p}$) et la distance (en nombre de patchs) aux zones centrales du réseau (patch n°3 ou 13).
IBDr	R de Mantel évaluant la présence d'isolement génétique par la distance.
F_{ST}	F_{ST} global de Weir & Cockerham (1984)
COR_{HO}	Coefficient de corrélation de Pearson entre le taux d'hétérozygotie observé et la distance (en nombre de patchs) aux zones centrales du réseau (patch n°3 ou 13).
COR_{FIS}	Coefficient de corrélation de Pearson entre l'indice de fixation F_{IS} et la distance (en nombre de patchs) aux zones centrales du réseau (patch n°3 ou 13).
COR_{dg}	Coefficient de corrélation de Pearson entre le nombre moyen des différences génotypiques entre deux individus et la distance (en nombre de patchs) aux zones centrales du réseau (patch n°3 ou 13).
COR_{dG}	Coefficient de corrélation de Pearson entre le nombre moyen des différences alléliques entre deux individus et la distance (en nombre de patchs) aux zones centrales du réseau (patch n°3 ou 13).

Par souci de simplicité ou parce que ces statistiques ont montré des tendances plus informatives, nous ne montrerons ici que certains résultats obtenus pour les statistiques IBDr, COR_{FST} , COR_P et F_{ST} . Les figures mentionnées ci-après sont consultables en fin de Chapitre.

3. Résultats et discussion préliminaires

3.1. Cas où $H=0$ (aucune pression de pêche)

La figure C3.2.(a) et (c) nous montre que, quand $H=0$, l'isolement par la distance (IBDr) est marqué lorsque N est petit (100) et que la connectivité est faible ($m=0,05$). A l'inverse, l'isolement par la distance est peu marqué lorsque N et m sont tous deux élevés, et intermédiaire dans les deux autres cas. Nous retrouvons également (Fig.C3.3.(a) et (c)) des valeurs COR_{FST} positives, qui montrent que les patchs sont plus génétiquement isolés des autres lorsque l'on s'éloigne de la zone centrale du réseau, et ce de façon d'autant plus marquée que N et/ou m sont petits. Enfin, nous retrouvons des valeurs COR_P négatives (Fig.C3.4.(a) et (c)), qui montrent que les patchs présentent un taux de polymorphisme d'autant plus faible que l'on s'éloigne de la zone centrale du réseau (principalement lorsque $N=100$), ainsi qu'un F_{ST} qui augmente, naturellement, lorsque N et de m sont petits (Fig.C3.5(a) et (c)). Tout cela correspond à ce qui est attendu lorsque la configuration dendritique agit comme principal facteur de la structuration génétique (Paz-Vinas & Blanchet 2015; Paz-Vinas et al. 2015), avec un isolement par la distance plus marqué et une réduction du nombre d'allèles vers les patchs amonts lorsque la taille de patchs (N) et/ou les flux de migration (m)

sont modérés. Nous retrouvons donc bien ici le modèle théorique de la répartition de diversité génétique dans les réseaux hydrographiques avec des marqueurs bialléliques. Par rapport aux travaux publiés avec des marqueurs types microsatellites, COR_{FST} semble simplement plus marqué, en valeur absolue, que COR_P , même lorsque N ou m sont grands : ainsi, l'influence de la structure dendritique est plus facilement observée sur l'isolement génétique moyen des patchs, que sur leur taux de polymorphisme, ce qui est cohérent avec la nature biallélique des loci simulés ici.

3.2. Cas où $H=0.125$ (pression de pêche continue)

Nous pouvons ici détailler ce qu'il se passe lorsque la pression de pêche apparaît ($H=0,125$) dans chacune des quatre combinaisons de N et m possibles (Fig.C3.2 à C3.5) :

N=100 et m=0.05	<p>Nous observons ici, par rapport au cas l'analogue avec $H=0$, une diminution d'$IBDr$ (Fig.C3.2.(b)) et de COR_{FST} (Fig.C3.3.(b)) dès la génération n°160, donc 10 générations après l'intervention de H. Toutefois, l'évolution de ces deux statistiques est lente et graduelle et semble atteindre un « plateau » à la génération n°170, au-delà de laquelle elles n'évoluent plus. De façon similaire, COR_P (Fig.C3.4.(b)) augmente pour passer d'une valeur négative à une valeur plus proche de zéro, ce qui signifie que l'on égalise peu à peu les niveaux de polymorphisme entre patchs centraux et périphériques du réseau. Cela traduit une augmentation rapide de la structure génétique, observable via le F_{ST} global (Fig.C3.5.(b)), du fait d'une forte dérive dans les populations aval sous l'effet de la pêche. Ceci empêche l'observation du patron théorique de diversité génétique attendu dans les réseaux dendritiques, de façon analogue à ce qui fut observé avec des marqueurs microsatellites par Paz-Vinas & Blanchet (2015) pour des valeurs de N et m très petites. Dans dans un tel cas, chaque patch devient démographiquement indépendant des autres, les faibles flux de gènes et la forte dérive génétique supprimant l'effet d'accumulation de diversité même dans les patchs situés en position centrale du réseau.</p>
N=100 et m=0.20	<p>Dans cette situation, aucune évolution consécutive à l'apparition de H n'est observable quel que soit l'indice considéré (Fig.C3.2(b), C3.3.(b) et C3.4(b)) excepté le F_{ST} global qui augmente au cours des générations (Fig.C3.5.(b)). On constate que le patron général reste identique à celui que l'on observe avec $H=0$.</p>
N=1000 et m=0.05	<p>L'évolution des paramètres est similaire à ce que l'on observe avec $N=100$ et $m=0,05$. Elle est cependant moins marquée, avec une réduction d'$IBDr$ et COR_{FST} moins abrupte (Fig.C3.2.(d) et C3.3.(d)) et une augmentation de COR_P très faible (Fig.C3.4.(d)). La plus grande taille de population au niveau de chaque patch, contrebalance vraisemblablement l'effet du retrait aléatoire H d'individus en réduisant l'intensité de la dérive génétique. Comme dans les cas précédents, on observe toutefois toujours l'augmentation de la structure globale via le F_{ST} (Fig.C3.5.(d)).</p>

N=1000 et m=0.20

Initialement, dans cette situation, la structure génétique est quasi homogène à l'échelle des 16 patchs (Fig.C3.5.(d)), ce qui correspond à ce que l'on attend dans un réseau dendritique lorsque N et m sont élevés. L'apparition et persistance de la pêche ($H=0,125$) entraînent la hausse des indices $IBDr$ et COR_{FST} au cours des générations (Fig.C3.2.(d) et C3.3.(d)) ainsi qu'une diminution (peu marquée) de COR_P (Fig.C3.4.(d)). Cela correspond à l'apparition progressive du patron de variation génétique spatiale que l'on attendrait au regard de la configuration dendritique du réseau dans le cas de N et m plus faibles, mais qui est ici uniquement créé par l'effet du retrait d'individus au niveau des patchs centraux.

3.3. Discussion complémentaire et perspectives

Lorsque $m=20$ (taux de migration plus élevés), la diminution de l'effectif au niveau des patchs où un retrait d'individus s'exerce diminue du même coup le nombre de migrants qui quittent ces patchs, même si les probabilités de migration, elles, sont maintenues constantes entre patchs et au cours du temps. Au cours du temps, la quantité d'individus arrivant, depuis des patchs périphériques, au sein de patchs impactés par la pression de pêche simulée devient ainsi plus importante que la quantité d'individus qui en repart. Au gré de cette procédure de simulations, nous avons donc illustré un phénomène de rescousse génétique, associé à l'apport de diversité fourni par les patchs périphériques, se comportant comme des sources, vers les patchs centraux se comportant comme des puits (*genetic mass effect*) comme observé par exemple au cours de l'étude de Waits et al. (2008).

Pour approfondir ce résultat, j'ai effectué une autre série de simulations où, cette fois-ci, j'ai artificiellement forcé les effectifs de chaque patch à rester constants au cours des générations. Ainsi, même après le retrait aléatoire d'individus exercé par H au niveau des patchs centraux, la procédure de reproduction simulée fait en sorte que l'effectif de la génération suivante redevienne toujours égal à ce qu'il était avant le retrait d'individus. De cette façon, seule la diversité génétique originelle de chaque patch est impactée, du fait d'un déséquilibre de succès reproducteur mais le nombre de migrants entre patchs, lui, restera constant et symétrique. Une population de taille constante malgré une pression d'exploitation peut être plausible dans le cas, par exemple, d'espèces à forte fécondité comme certaines des espèces étudiées au cours de cette thèse. J'ai également fait varier le paramètre H sur une gamme de valeurs plus fournie, allant de 0 à 0,5. En observant les graphiques obtenus, on constate cette fois-ci que le patron de variabilité génétique attendu sous l'influence de la structuration dendritique est peu à peu effacé par l'intervention de H. En effet, COR_P et COR_{FST} se rapprochent de zéro (Fig.C3.6 à C3.9), comme observé précédemment (Point 3.2) lorsque la probabilité de migration était faible ($m=0,05$) -et ne pouvait donc que faiblement contrebalancer l'effet de dérive locale. Ceci est alors vrai quelle que soit la combinaison de N et m considérée, contrairement à ce que l'on observait lorsque la taille de patchs n'était pas maintenue constante. On constate par ailleurs que ces effets de H sur les patrons de variabilité génétique nécessitent ici, d'une part, un beaucoup plus grand nombre de générations pour apparaître, et d'autre part qu'ils apparaissent pour des valeurs de H plus fortes (Fig.C3.6 à C3.9) que précédemment, où la valeur $H=0.125$ suffisait généralement à observer une évolution de la variabilité génétique globale au cours des générations (Fig.C3.2 à C3.5) (voir également § suivant). Cela peut suggérer deux conclusions importantes : tout d'abord, l'influence de la pression de pêche sur les patrons de variabilité génétique observés est fortement liée à la

réponse démographique des populations exploitées comme l'on pouvait s'y attendre, pouvant même renforcer le patron de variabilité génétique théoriquement attendu au sein d'un réseau hydrographique dendritique. Cependant, la pression de pêche peut aussi influencer les patrons de variabilité génétique sans qu'une évolution d'abondance des patches (nombre 'N' d'individus) n'intervienne, même si son effet sera alors plus ténu et tardif.

Notons qu'il est actuellement impossible d'avoir une estimation de l'intensité de pêche réelle sur le Haut-Maroni. La valeur de $H=0.125$ a été choisie ici dans un premier temps car, a priori, biologiquement cohérente, mais également suite à des premiers essais de simulations. Ces essais se basaient sur des combinaisons de paramètres identiques à celles qui ont été simulées pour le présent Chapitre, et n'en différaient que par la procédure d'initialisation des génotypes au départ ainsi que par les numéros de générations considérées. En outre, lors de ces essais, j'ai fait varier la valeur H de 0 à 0,3 (Annexe XI) sans forcer l'abondance des patches à rester constante au cours du temps. Les effets de H sur COR_{FST} étaient alors observables pour des valeurs de H supérieures ou égales à 0,1, en particulier pour les taux de migration faibles (Fig.XI.1 et XI.2). Nous pouvons détailler un petit peu plus ce qui l'on a observé pour de telles valeurs de H :

- Lorsque $m=0,05$, on observe la même évolution de COR_{FST} , quel que soit N (100 ou 1000), que dans la situation présentée en section 3.2, évolution de plus en plus marquée cependant au fur et à mesure que H augmente.
- Le cas $m=0,20$ confirme également les observations fournies dans la section 3.2 pour $N=1000$, en ce qui concerne COR_{FST} (Fig.XI.2). Pour $N=100$ en revanche, on observe (Fig.XI.1), dans un premier temps, une diminution de COR_{FST} , qui augmente ensuite pour se stabiliser au bout de la génération n°110 environ. Tout se passe ici comme si, après la rupture d'un équilibre pendant quelques générations, un nouvel équilibre se mettait alors en place, le patron attendu au vu de la configuration dendritique du réseau étant alors toujours observé bien que moins marqué qu'avant l'intervention de H .

Ces premières analyses exploratoires permettent de montrer à quel point la taille des populations dans un réseau, l'intensité et la direction des flux de migration ainsi que l'impact réel de la pêche (en termes de mortalité et donc d'individus éliminés de la population naturelle à chaque génération) peuvent aboutir à des patrons de variabilité génétique parfois imprévisibles et, surtout, qui peuvent être communs à des combinaisons très différentes de ces paramètres. A ce stade, aucun des indices statistiques testés (Table C3.1) ne permet de discriminer entre elles les différentes combinaisons de paramètres, il n'est donc pas possible de les utiliser pour évaluer quel serait le scénario le plus probable chez les espèces étudiées au cours de cette thèse. Nous voyons cependant que l'absence d'une plus faible diversité génétique au niveau de zones contraintes par une réduction d'effectif, via prélèvement d'individus, ne signifie pas qu'une telle réduction d'effectif n'a pas d'effet. Nous avons visualisé, ici, le cas où une contrainte agissant en aval est contrebalancée par des flux d'individus et de gènes venant d'amont, cependant nous pouvons également imaginer d'autres situations dans lesquelles une contrainte agissant localement pourrait ne pas être non plus être détectée en recherchant un phénomène d'érosion génétique associé. Par exemple, dans une métapopulation de rivière dont le patch le plus important se situerait autour des confluences ou du cours principal et agirait comme source alimentant des populations secondaires localisées sur différents tributaires (migration asymétrique vers l'amont), on peut

supposer qu'une pression de pêche localisée sur cette source de diversité influencerait donc, du même coup, la diversité génétique des patchs secondaires qui en dépendent, rendant sa détection difficile par comparaison directe de ces différents types de patchs.

En perspective, pour poursuivre cette analyse exploratoire, je souhaiterais détailler davantage la dynamique temporelle du scénario testé ici, en intégrant dans une seule et même analyse une gamme plus détaillée de chaque paramètre N , m et H , et en analysant l'évolution globale sur un plus grand nombre de générations. Cela permettrait par exemple de connaître la durée au bout de laquelle un système simulé sous une certaine combinaison de paramètres atteint un équilibre, ou au contraire finit par s'éteindre. Afin d'améliorer la pertinence biologique d'une telle approche, il sera possible de tirer davantage parti des fonctionnalités de SimuPOP qui peut permettre, par exemple, de simuler un certain taux de croissance d'une population ou encore de subdiviser chaque patch en « sous-populations virtuelles » qui peuvent présenter une vulnérabilité différente face à une pression de pêche simulée. Cette fonctionnalité permet, par exemple, de simuler différentes classes d'âges.

Il faudrait, dans le même temps, suivre davantage l'évolution temporelle de caractéristiques globales ou « purement démographiques » de chaque métapopulation simulée, comme la diversité génétique totale, le nombre exact d'individus par patch, ou encore l'estimation des tailles efficaces locales – qui peuvent différer du nombre d'individus total par patchs si, par exemple, certains couples d'individus ont, de part la stochasticité de la phase de reproduction simulée, davantage contribué à la production de descendants que d'autres.

4. Références

- Allan, J. D., Abell, R., Hogan, Z., Revenga, C., Taylor, B. W., Welcomme, R. L., & Winemiller, K. (2005). Overfishing of Inland Waters. *BioScience*, 55(12), 1041. [https://doi.org/10.1641/0006-3568\(2005\)055\[1041:OOIW\]2.0.CO;2](https://doi.org/10.1641/0006-3568(2005)055[1041:OOIW]2.0.CO;2)
- Allendorf, F., England, P., Luikart, G., Ritchie, P., & Ryman, N. (2008). Genetic effects of harvest on wild animal populations. *Trends in Ecology & Evolution*, 23(6), 327-337. <https://doi.org/10.1016/j.tree.2008.02.008>
- Campbell Grant, E. H., Lowe, W. H., & Fagan, W. F. (2007). Living in the branches: population dynamics and ecological processes in dendritic networks. *Ecology Letters*, 10(2), 165-175. <https://doi.org/10.1111/j.1461-0248.2006.01007.x>
- Cañedo-Argüelles, M., Boersma, K. S., Bogan, M. T., Olden, J. D., Phillipsen, I., Schriever, T. A., & Lytle, D. A. (2015). Dispersal strength determines meta-community structure in a dendritic riverine network. *Journal of Biogeography*, 42(4), 778-790. <https://doi.org/10.1111/jbi.12457>
- Carranza, M. L., D'Alessandro, E., Saura, S., & Loy, A. (2012). Connectivity providers for semi-aquatic vertebrates: the case of the endangered otter in Italy. *Landscape Ecology*, 27(2), 281-290. <https://doi.org/10.1007/s10980-011-9682-3>
- Carrara, F., Altermatt, F., Rodriguez-Iturbe, I., & Rinaldo, A. (2012). Dendritic connectivity controls biodiversity patterns in experimental metacommunities. *Proceedings of the National Academy of Sciences*, 109(15), 5761-5766. <https://doi.org/10.1073/pnas.1119651109>
- Fagan, W. F. (2002). Connectivity, Fragmentation, and Extinction Risk in Dendritic Metapopulations. *Ecology*, 83(12), 3243-3249. [https://doi.org/10.1890/0012-9658\(2002\)083\[3243:CFAERI\]2.0.CO;2](https://doi.org/10.1890/0012-9658(2002)083[3243:CFAERI]2.0.CO;2)
- Hughes, J. M., Huey, J. A., & Schmidt, D. J. (2013). Is realised connectivity among populations of aquatic fauna predictable from potential connectivity?: *Connectivity among populations of aquatic fauna. Freshwater Biology*, 58(5), 951-966. <https://doi.org/10.1111/fwb.12099>
- Hughes, J. M., Schmidt, D. J., & Finn, D. S. (2009). Genes in Streams: Using DNA to Understand the Movement of Freshwater Fauna and Their Riverine Habitat. *BioScience*, 59(7), 573-583. <https://doi.org/10.1525/bio.2009.59.7.8>
- Labonne, J., Ravigné, V., Parisi, B., & Gaucherel, C. (2008). Linking dendritic network structures to population demogenetics: The downside of connectivity. *Oikos*, 117(10), 1479-1490. <https://doi.org/10.1111/j.0030-1299.2008.16976.x>
- Lean, J., Hammer, M. P., Unmack, P. J., Adams, M., & Beheregaray, L. B. (2017). Landscape genetics informs mesohabitat preference and conservation priorities for a surrogate indicator species in a highly fragmented river system. *Heredity*, 118(4), 374-384. <https://doi.org/10.1038/hdy.2016.111>
- Paz-Vinas, I., & Blanchet, S. (2015). Dendritic connectivity shapes spatial patterns of genetic diversity: a simulation-based study. *Journal of Evolutionary Biology*, 28(4), 986-994. <https://doi.org/10.1111/jeb.12626>
- Paz-Vinas, I., Loot, G., Stevens, V. M., & Blanchet, S. (2015). Evolutionary processes driving spatial patterns of intraspecific genetic diversity in river ecosystems. *Molecular Ecology*, 24(18), 4586-4604. <https://doi.org/10.1111/mec.13345>
- Peng, B., & Amos, C. I. (2008). Forward-time simulations of non-random mating populations using simuPOP. *Bioinformatics*, 24(11), 1408-1409. <https://doi.org/10.1093/bioinformatics/btn179>
- Pilger, T. J., Gido, K. B., Propst, D. L., Whitney, J. E., & Turner, T. F. (2017). River network architecture, genetic effective size and distributional patterns predict differences in genetic structure across species in a dryland stream fish community. *Molecular Ecology*, 26(10), 2687-2697. <https://doi.org/10.1111/mec.14079>

- Planque, B., Fromentin, J.-M., Cury, P., Drinkwater, K. F., Jennings, S., Perry, R. I., & Kifani, S. (2010). How does fishing alter marine populations and ecosystems sensitivity to climate? *Journal of Marine Systems*, 79(3-4), 403-417. <https://doi.org/10.1016/j.jmarsys.2008.12.018>
- Prunier, J. G., Dubut, V., Loot, G., Tudesque, L., & Blanchet, S. (2017). The relative contribution of river network structure and anthropogenic stressors to spatial patterns of genetic diversity in two freshwater fishes: A multiple-stressors approach. *Freshwater Biology*, 63(1), 6-21. <https://doi.org/10.1111/fwb.13034>
- Richmond, J. Q., Backlin, A. R., Galst-Cavalcante, C., O'Brien, J. W., & Fisher, R. N. (2018). Loss of dendritic connectivity in southern California's urban riverscape facilitates decline of an endemic freshwater fish. *Molecular Ecology*, 27(2), 369-386. <https://doi.org/10.1111/mec.14445>
- Saunders, M. I., Brown, C. J., Foley, M. M., Febria, C. M., Albright, R., Mehling, M. G., ... Burfeind, D. D. (2015). Human impacts on connectivity in marine and freshwater ecosystems assessed using graph theory: a review. *Marine and Freshwater Research*, 67(3), 277-290. <https://doi.org/10.1071/MF14358>
- Schick, R. S., & Lindley, S. T. (2007). Directed connectivity among fish populations in a riverine network: Directed connectivity. *Journal of Applied Ecology*, 44(6), 1116-1126. <https://doi.org/10.1111/j.1365-2664.2007.01383.x>
- Schindler, D. E., Hilborn, R., Chasco, B., Boatright, C. P., Quinn, T. P., Rogers, L. A., & Webster, M. S. (2010). Population diversity and the portfolio effect in an exploited species. *Nature*, 465(7298), 609-612. <https://doi.org/10.1038/nature09060>
- Urban, D. L., Minor, E. S., Treml, E. A., & Schick, R. S. (2009). Graph models of habitat mosaics. *Ecology Letters*, 12(3), 260-273. <https://doi.org/10.1111/j.1461-0248.2008.01271.x>
- Waits, E. R., Bagley, M. J., Blum, M. J., McCormick, F. H., & Lazorchak, J. M. (2008). Source-sink dynamics sustain central stonerollers (*Camptostoma anomalum*) in a heavily urbanized catchment. *Freshwater Biology*, 53(10), 2061-2075. <https://doi.org/10.1111/j.1365-2427.2008.02030.x>
- Webb, J. A., & Padgham, M. (2013). How does network structure and complexity in river systems affect population abundance and persistence? *Limnologica - Ecology and Management of Inland Waters*, 43(5), 399-403. <https://doi.org/10.1016/j.limno.2013.04.006>

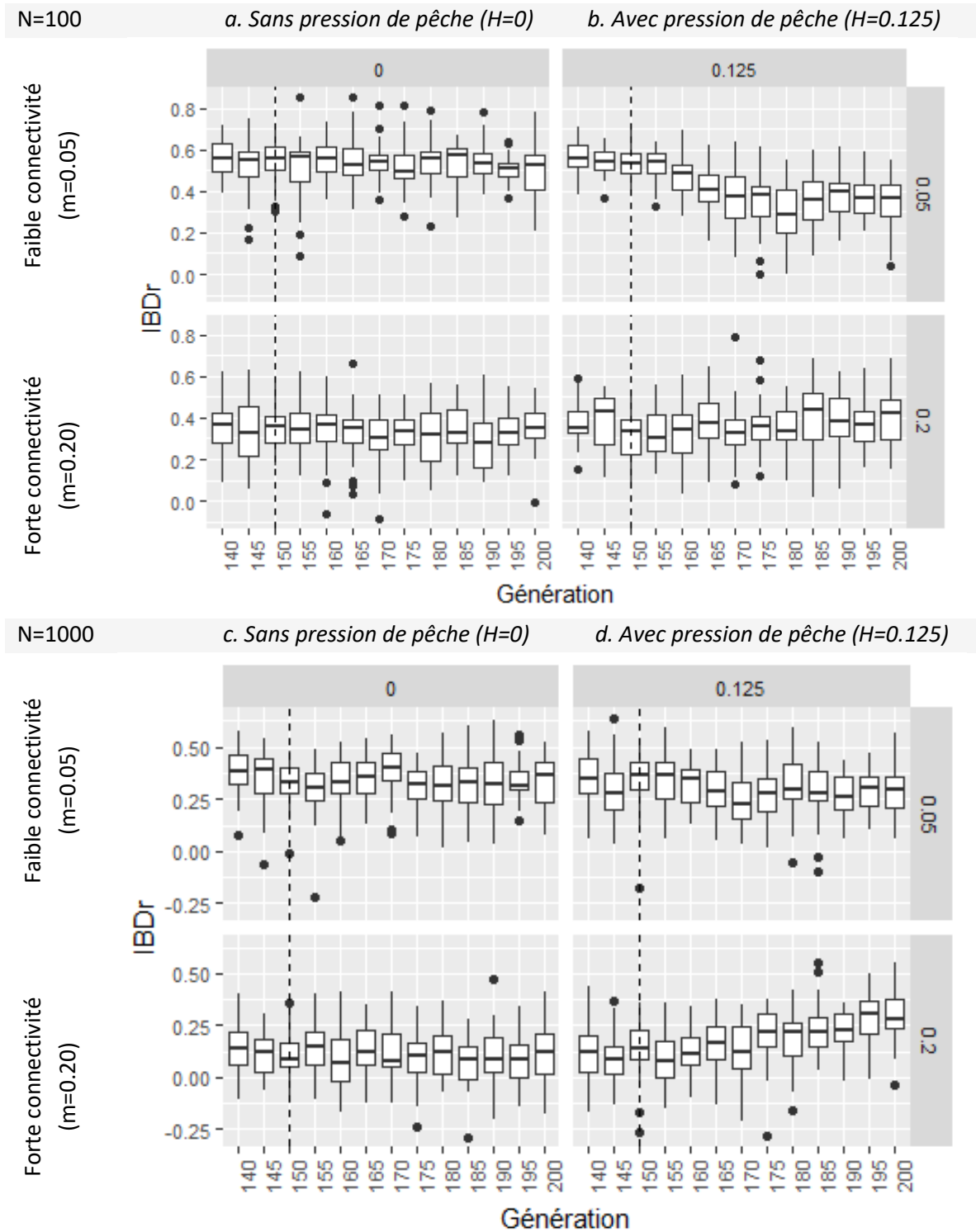


Figure C3.2: Evolution du r de Mantel (IBDr), au cours des générations n°140 à 200, pour chacune des huit combinaisons de paramètres explorées durant cette procédure de simulations d'une métapopulation de type dendritique (taille des patch N=100 ou 1000 ; probabilités de migration m=0,05 ou 0,20 ; proportion d'individus éliminés dans patchs centraux H=0 ou 0,125). Chaque boxplot représente l'ensemble des 30 répliques simulés. La ligne verticale en pointillés représente la génération à laquelle la valeur H intervient (H est donc toujours nul avant la génération 150). La taille des patchs varie au cours du temps au gré de la contrainte imposée par H et des flux de migrants.

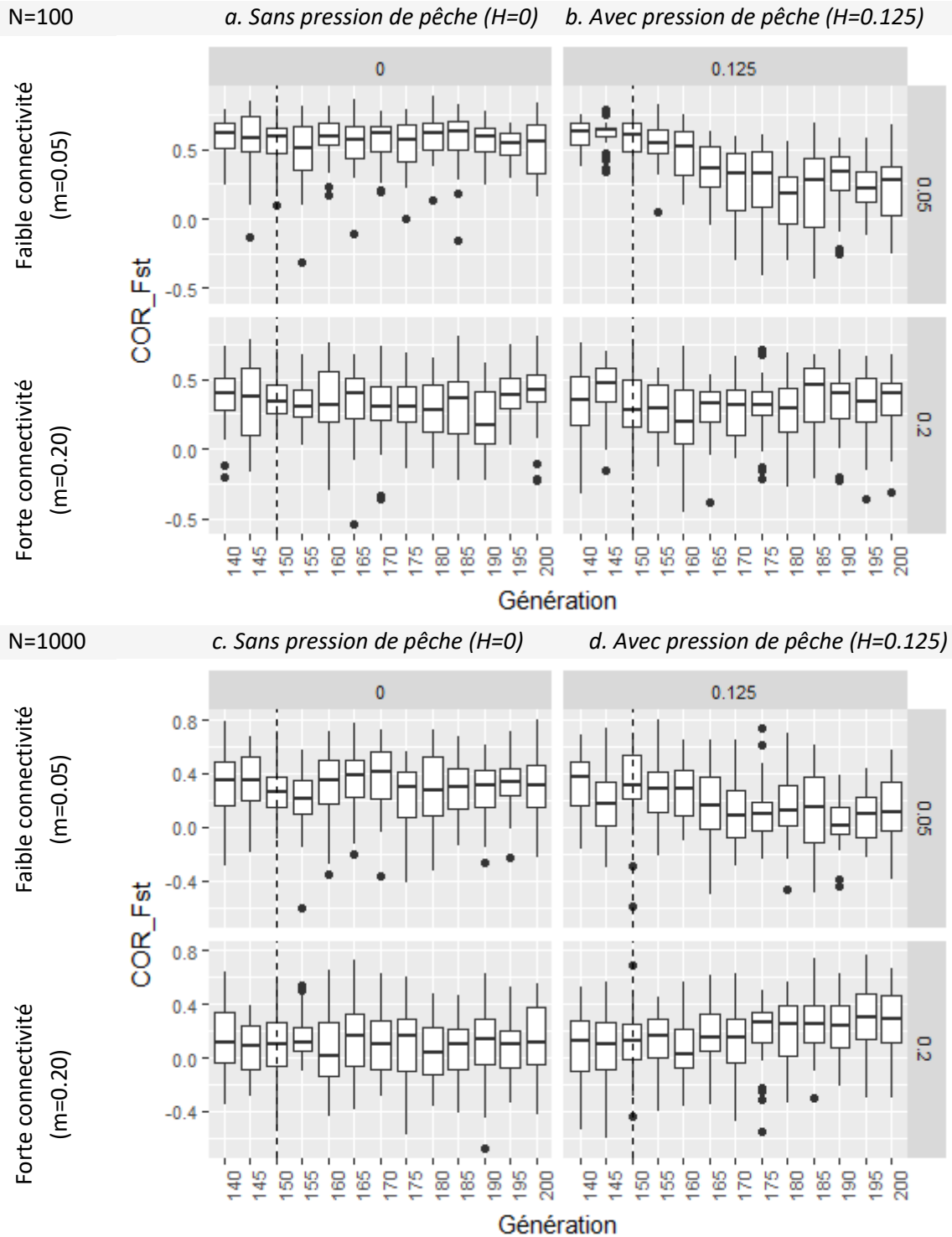


Figure C3.3: Evolution de COR_{Fst} au cours des générations n°140 à 200, pour chacune des huit combinaisons de paramètres explorées durant cette procédure de simulations d'une métapopulation de type dendritique (taille des patch $N=100$ ou 1000 ; probabilités de migration $m=0,05$ ou $0,20$; proportion d'individus éliminés dans patchs centraux $H=0$ ou $0,125$). Chaque boxplot représente l'ensemble des 30 réplicas simulés. La ligne verticale en pointillés représente la génération à laquelle la valeur H intervient (H est donc toujours nul avant la génération 150). La taille des patchs varie au cours du temps au gré de la contrainte imposée par H et des flux de migrants.

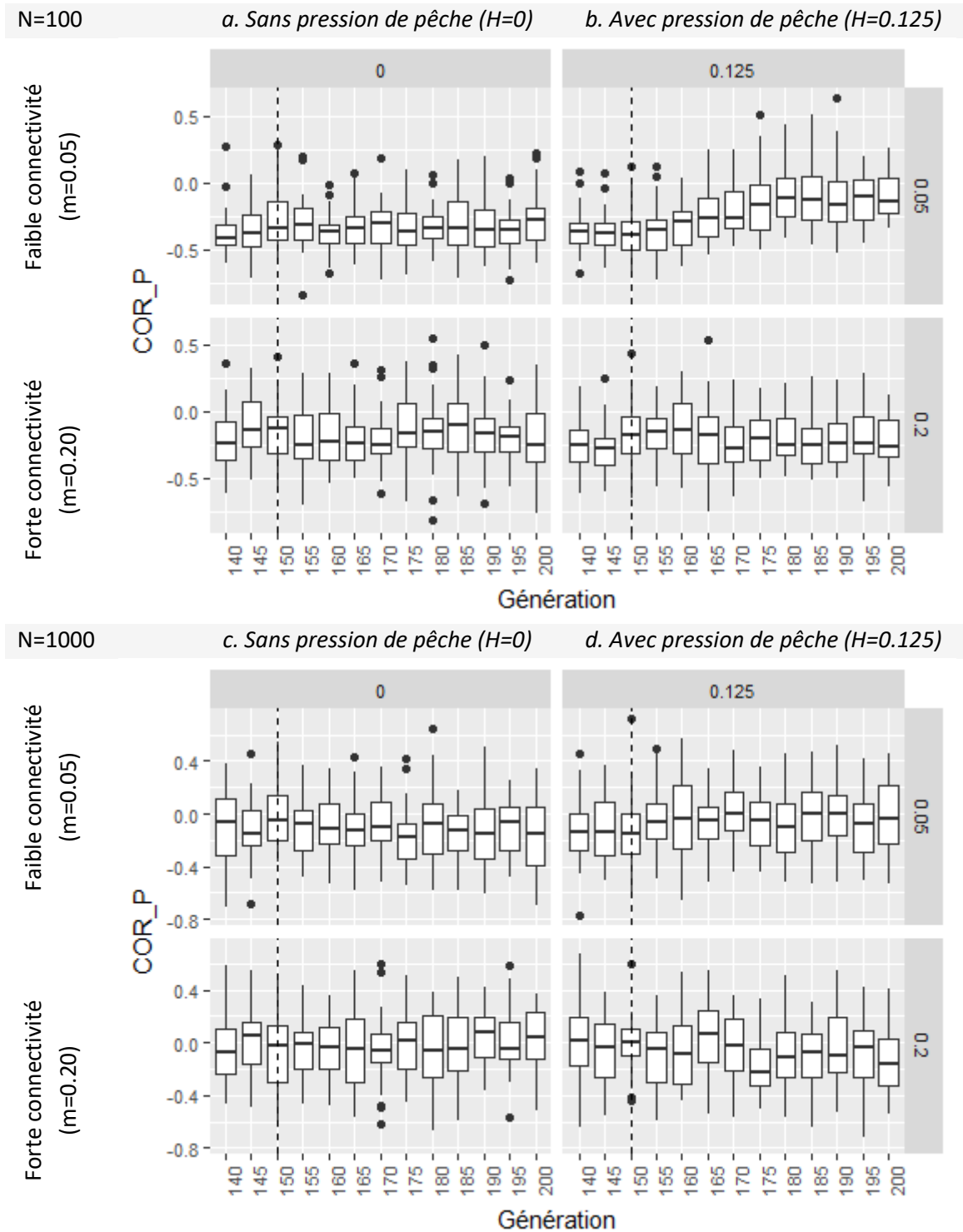


Figure C3.4: Evolution de COR_P , au cours des générations n°140 à 200, pour chacune des huit combinaisons de paramètres explorées durant cette procédure de simulations d'une métapopulation de type dendritique (taille des patch $N=100$ ou 1000 ; probabilités de migration $m=0,05$ ou $0,20$; proportion d'individus éliminés dans patchs centraux $H=0$ ou $0,125$). Chaque boxplot représente l'ensemble des 30 réplicas simulés. La ligne verticale en pointillés représente la génération à laquelle la valeur H intervient (H est donc toujours nul avant la génération 150). La taille des patchs varie au cours du temps au gré de la contrainte imposée par H et des flux de migrants.

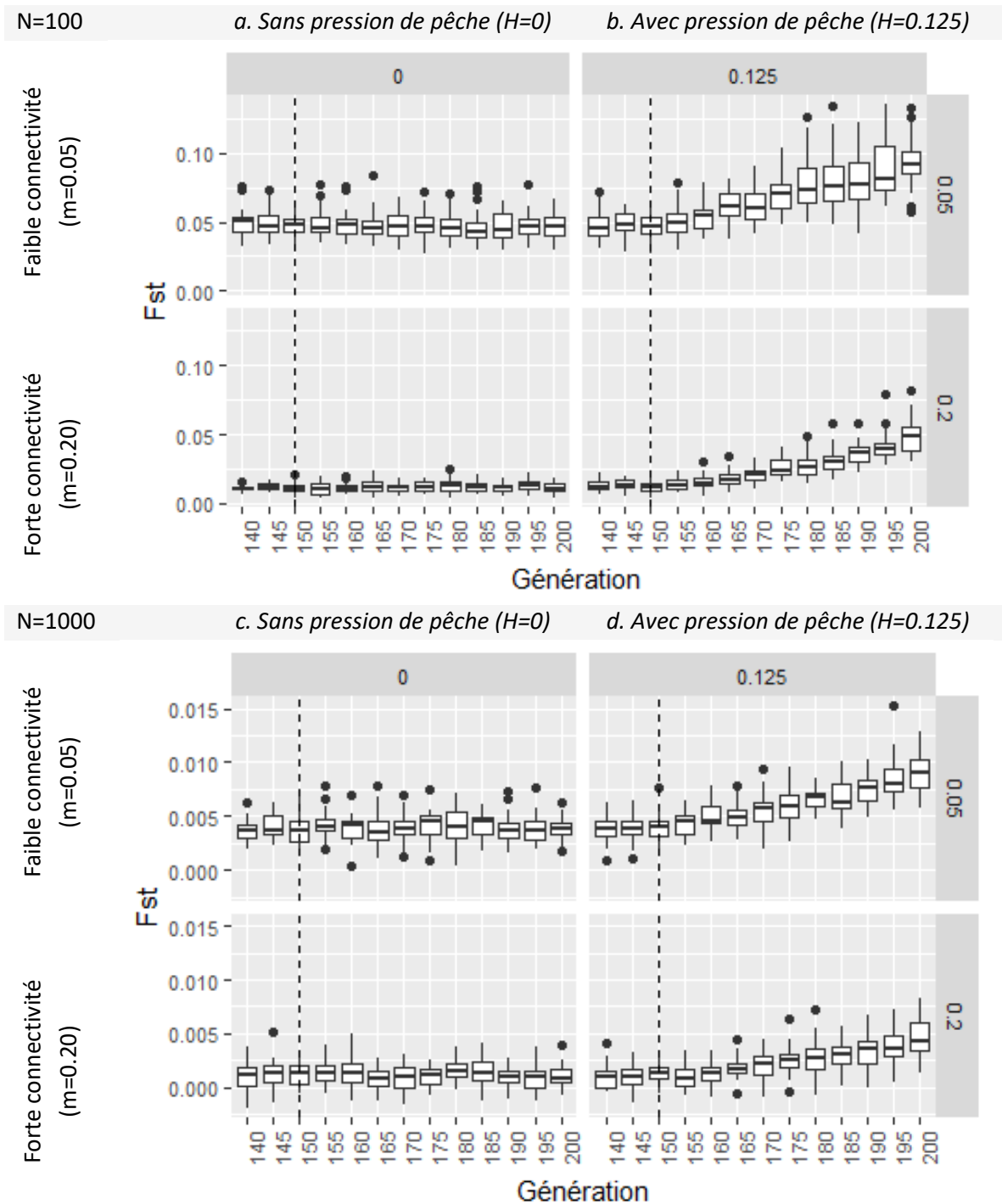


Figure C3.5: Evolution du F_{ST} global au cours des générations n°140 à 200, pour chacune des huit combinaisons de paramètres explorées durant cette procédure de simulations d'une métapopulation de type dendritique (taille des patch $N=100$ ou 1000 ; probabilités de migration $m=0,05$ ou $0,20$; proportion d'individus éliminés dans patchs centraux $H=0$ ou $0,125$). Chaque boxplot représente l'ensemble des 30 réplicas simulés. La ligne verticale en pointillés représente la génération à laquelle la valeur H intervient (H est donc toujours nul avant la génération 150). La taille des patchs varie au cours du temps au gré de la contrainte imposée par H et des flux de migrants.

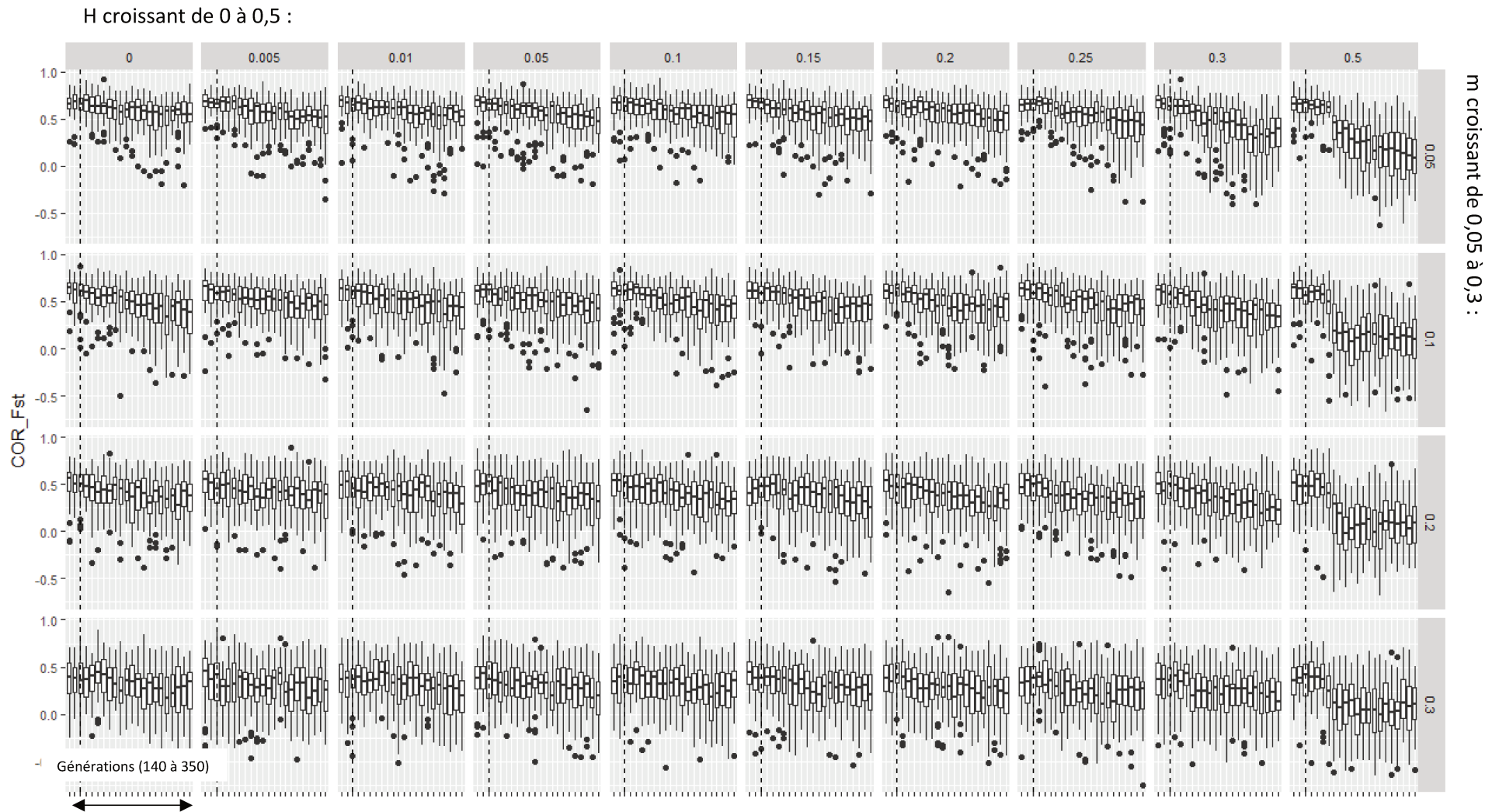


Figure C3.6 : Evolution de COR_{FST} , au cours des générations 140 à 350 pour plusieurs combinaisons de paramètres explorées durant cette procédure de simulations d'une métapopulation dendritique (taille des patches $N=100$; probabilités de migration $m=0,05$; $0,10$; $0,20$ ou $0,30$; proportion d'individus éliminés dans patches centraux $H=0$ à $0,50$). Chaque boxplot représente l'ensemble des 50 réplicas simulés et l'incrément est de 10 générations. La ligne verticale en pointillés représente la génération 160 à laquelle la valeur H intervient (H est toujours nul avant la génération 160). La taille des patches est maintenue stable au cours du temps.

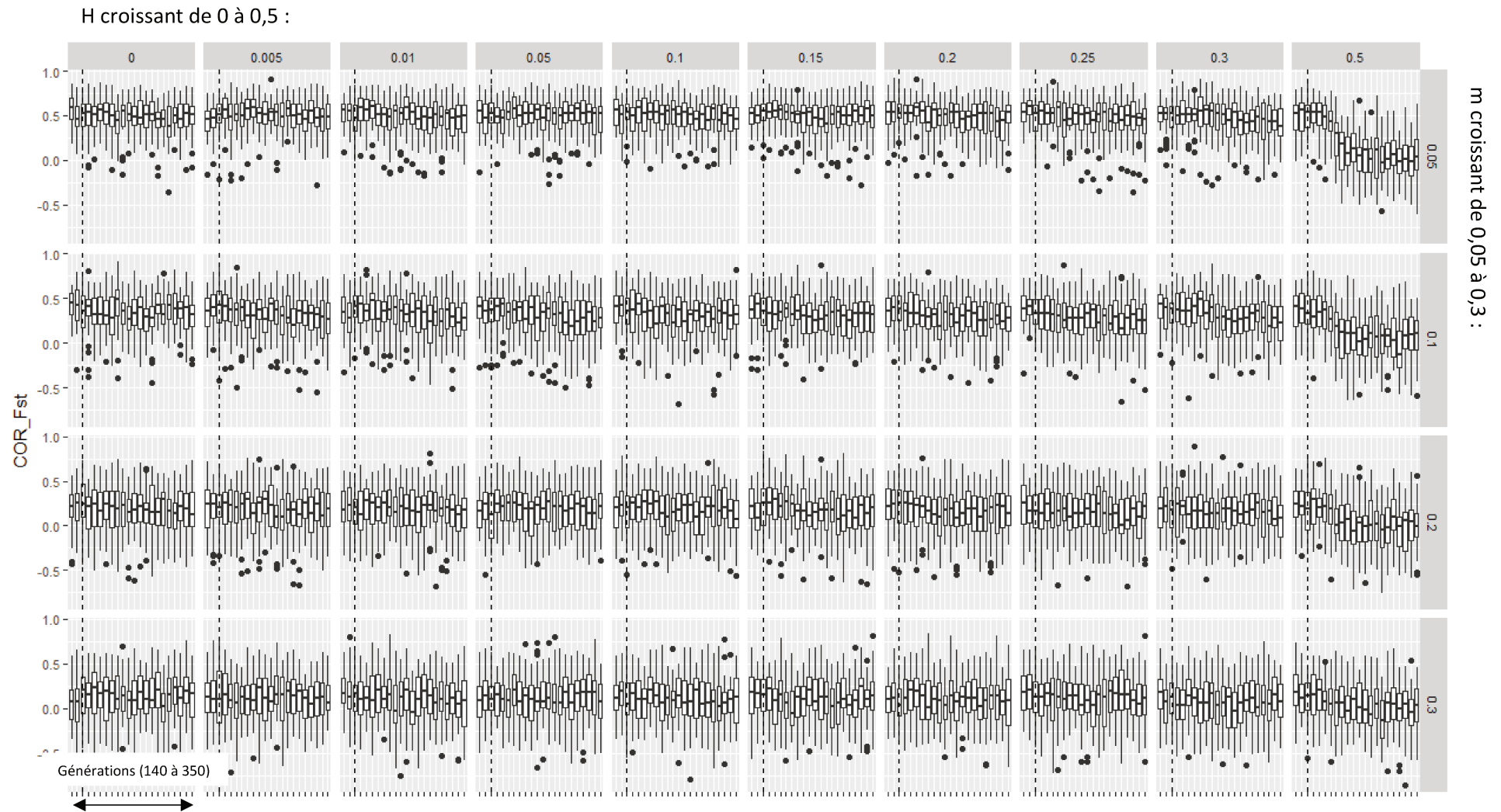


Figure C3.7 : Evolution de COR_{FST} , au cours des générations 140 à 350 pour plusieurs combinaisons de paramètres explorées durant cette procédure de simulations d'une métapopulation dendritique (taille des patches $N=1000$; probabilités de migration $m=0,05$; $0,10$; $0,20$ ou $0,30$; proportion d'individus éliminés dans patches centraux $H=0$ à $0,50$). Chaque boxplot représente l'ensemble des 50 répliquas simulés et l'incrément est de 10 générations. La ligne verticale en pointillés représente la génération 160 à laquelle la valeur H intervient (H est toujours nul avant la génération 160). La taille des patches est maintenue stable au cours du temps.

H croissant de 0 à 0,5 :

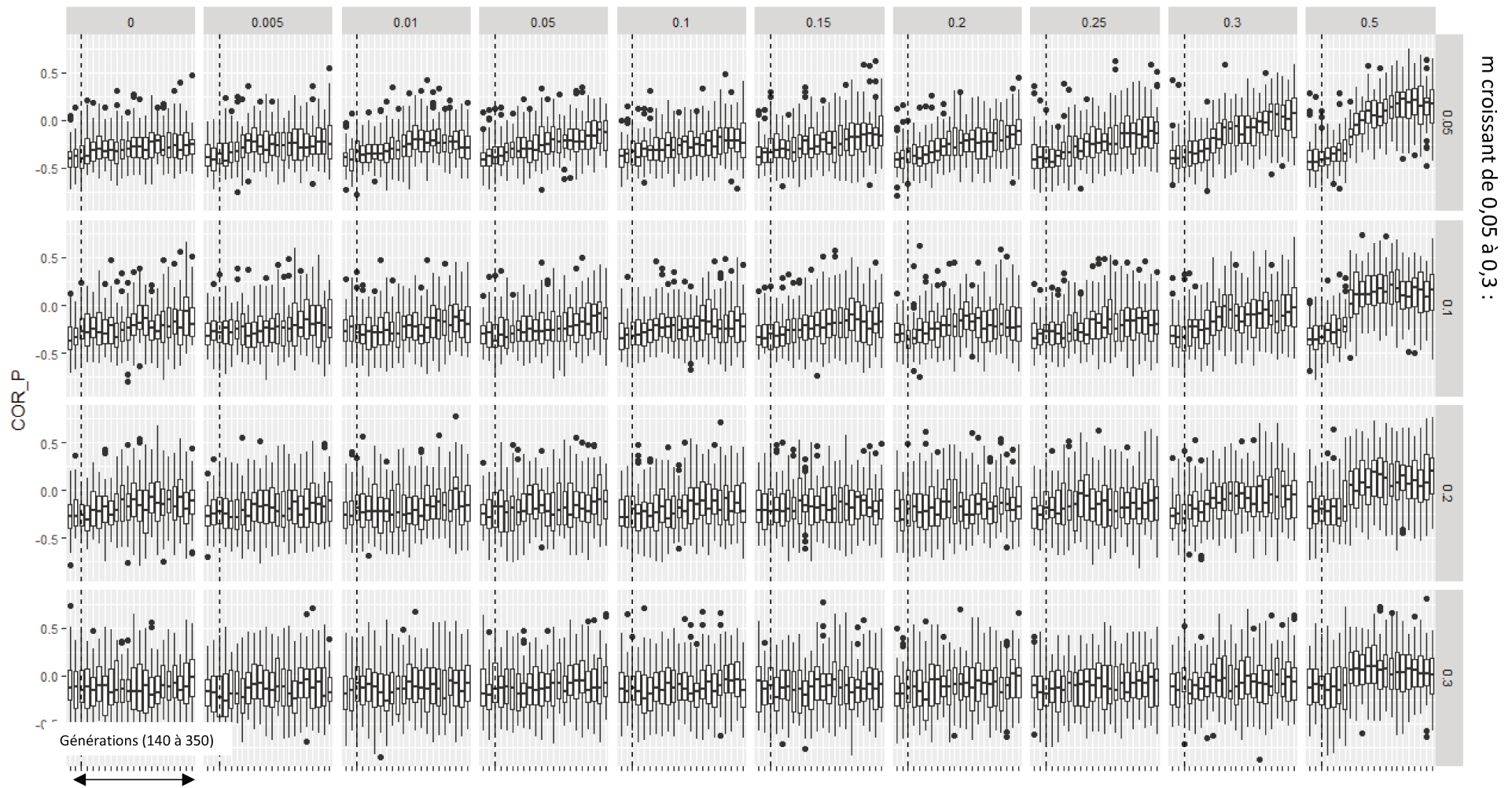


Figure C3.8 : Evolution de COR_P , au cours des générations 140 à 350 pour plusieurs combinaisons de paramètres explorées durant cette procédure de simulations d'une métapopulation dendritique (taille des patches $N=100$; probabilités de migration $m=0,05$; $0,10$; $0,20$ ou $0,30$; proportion d'individus éliminés dans patches centraux $H=0$ à $0,50$). Chaque boxplot représente l'ensemble des 50 répliquas simulés et l'incrément est de 10 générations. La ligne verticale en pointillés représente la génération 160 à laquelle la valeur H intervient (H est toujours nul avant la génération 160). La taille des patches est maintenue stable au cours du temps.

H croissant de 0 à 0,5 :

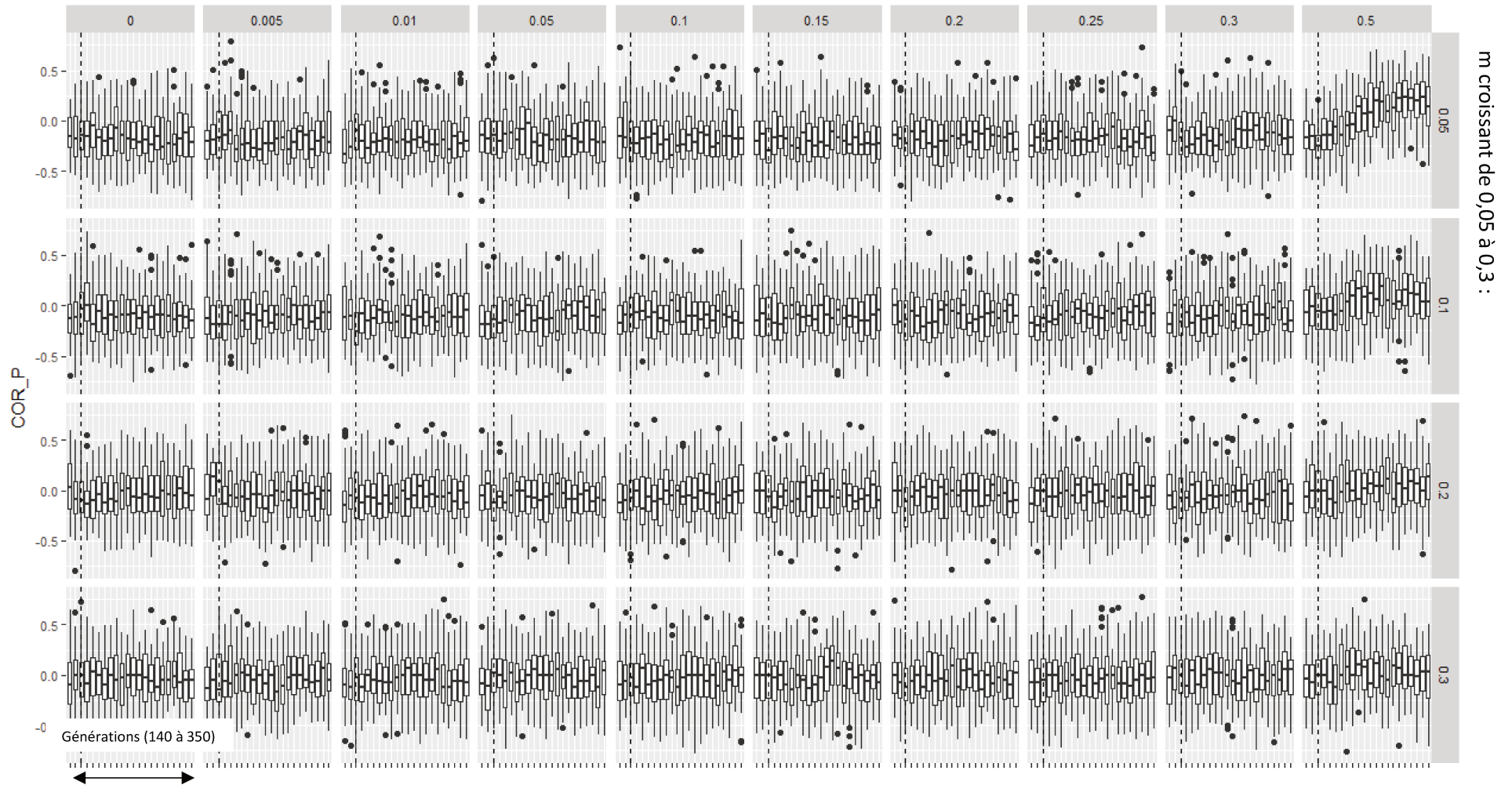


Figure C3.9 : Evolution de COR_P , au cours des générations 140 à 350 pour plusieurs combinaisons de paramètres explorées durant cette procédure de simulations d'une métapopulation dendritique (taille des patches $N=1000$; probabilités de migration $m=0,05$; $0,10$; $0,20$ ou $0,30$; proportion d'individus éliminés dans patches centraux $H=0$ à $0,50$). Chaque boxplot représente l'ensemble des 50 réplicas simulés et l'incrément est de 10 générations. La ligne verticale en pointillés représente la génération 160 à laquelle la valeur H intervient (H est toujours nul avant la génération 160). La taille des patches est maintenue stable au cours du temps.

Discussion générale & perspectives:

Ce projet de thèse s'est inscrit, dès sa première élaboration, dans une dimension fortement appliquée faisant écho à un double constat. Premièrement, les caractéristiques de la pêche de subsistance sur le bassin du Haut-Maroni ont subi de significatifs bouleversements à différents niveaux au cours des décennies passées, interrogeant pêcheurs et gestionnaires sur l'état de santé actuel des populations de poissons exploitées. Deuxièmement, le manque de données biologiques et écologiques disponibles sur ces ressources permet difficilement d'apporter des réponses précises à ces questionnements. Chaque étape prévue dans le déroulement de cette thèse visait donc à fournir des outils et des connaissances biologiques complémentaires dans l'objectif ultime de contribuer, directement ou non, à la gestion et la conservation des populations de poissons qui gardent une grande importance socio-culturelle pour les communautés locales Amérindiennes et Bushinengués.

Durant cette thèse, j'ai tout d'abord eu l'occasion de contribuer à la mise au point d'un protocole de développement rapide de ressources moléculaires dédiées à l'étude de la variabilité génétique neutre chez plusieurs espèces non-modèles, simultanément et à faible coût. Grâce à une importante collecte de matériel biologique sur le terrain, permise par un déploiement logistique et humain de grande envergure, il m'a été possible de créer un jeu de données génétique regroupant 18 espèces co-échantillonnées sur 12 localités du Haut-Maroni et génotypées pour 111 à 188 marqueurs SNP. L'analyse de ces données a permis de documenter, pour la première fois en Guyane, les patrons de structure génétique de plusieurs espèces à l'échelle d'un même bassin versant. A l'aide d'une approche comparative entre espèces, il a été confirmé que la structuration génétique des espèces à faible dispersion répondait à la configuration physique du réseau hydrographique, en accord avec le modèle théorique des réseaux écologiques dendritiques. Il a également été détecté, chez plusieurs espèces, une diversité génétique plus faible au niveau de certains tributaires du Haut-Maroni, vraisemblablement liée à des contraintes environnementales localisées comme la qualité de l'habitat. Cependant, nous avons observé peu de cas d'appauvrissement de la diversité génétique locaux en ce qui concerne les sites où la pression de pêche est forte comparativement aux sites de référence. Néanmoins certains sites, comme le village de Papaïchton, ou certaines espèces comme l'emblématique *Hoplias aimara*, font exception et méritent une vigilance particulière. L'approche comparative suggère que les espèces présentent différents patrons de variabilité génétique spatiale en fonction de leur stratégies de reproduction et leur comportement migrateur. Cependant, certains contrastes observés entre espèces à traits d'histoire de vie similaires mais d'intérêt halieutique différents soulèvent des interrogations. En outre, l'utilisation à des fins exploratoires des outils de simulation de données génétiques a suggéré qu'une pression de pêche localisée sur les principales confluences d'un bassin versant pouvait modifier les patrons de structure génétique globaux à grande échelle au cours des générations. Il ne faut donc en aucun cas exclure la possibilité qu'une pression de pêche, éventuellement conjuguée à d'autres contraintes

environnementales, influence les patrons de variabilité génétique à échelle régionale, plutôt que locale, chez les espèces du Haut-Maroni.

Les résultats scientifiques obtenus au cours de ce travail ont été largement discutés dans les chapitres précédents. Dans les pages qui vont suivre, je m'attacherai à synthétiser ces résultats afin de présenter les apports de ce travail au regard des enjeux scientifiques soulevés en introduction. Dans un deuxième temps et pour conclure ce manuscrit, je tenterai de formuler des recommandations et perspectives pour la gestion des populations de poissons exploités du Haut-Maroni.

1. Réponse aux enjeux méthodologiques de l'étude

1.1. Le développement des ressources moléculaires

1.1.1. Des SNP disponibles pour les espèces de poissons du Haut-Maroni

Dans la première partie de ce travail de thèse (Chapitre 1), nous avons montré qu'il était possible d'obtenir des ressources moléculaires pour de nombreuses espèces en simultanément grâce à un unique protocole RADSeq classique et très facile à mettre en œuvre en routine par des plateformes spécialisées.

Le protocole a été mis en place pour que les marqueurs développés soient aussi représentatifs que possible de la diversité génétique au sein de la zone d'étude. En effet, les ressources moléculaires des différentes espèces ont été développées sur une fenêtre spatio-temporelle commune et selon un protocole respectant des étapes identiques, dans le but d'éviter au maximum les biais d'incertitude (*ascertainment bias*, AB) liés à l'application de marqueurs développés dans très différents contextes (par exemple par cross-amplification de marqueurs moléculaires issus d'autres espèces ou développés à partir d'échantillons de zones géographiques différentes). L'analyse relativement aisée des données brutes de spectrométrie de masse et le format routinier des données génotypiques finales contribuent à la simplicité et à la reproductibilité de notre approche. Ainsi, nous avons développé un jeu de marqueurs SNP pour 18 espèces de poissons de Guyane, génotypables de façon simple et reproductible, permettant des études classiques de génétique des populations. Ces ressources sont désormais disponibles et réutilisables à l'échelle de la zone d'étude.

1.1.2. Sources possibles de biais et recommandations

Malgré la mise en place d'un protocole le plus standardisé possible entre espèces focales, il reste cependant impossible d'exclure avec certitude l'introduction d'AB au cours de l'approche développée en Chapitre 1. Si les SNP sélectionnés par cette approche représentent un échantillonnage non-représentatif du génome, les spectres de fréquences alléliques fournis par ces loci peuvent conduire à une sur-, ou sous-estimation de la diversité moyenne réelle pour les espèces à l'échelle de la zone d'étude (Nielsen 2000; Morin et al. 2004). Dans le cas des SNP, l'AB résulte généralement d'un sous-échantillonnage des loci à fréquences alléliques déséquilibrées, c'est-à-dire portant un allèle rare à l'échelle de la zone d'étude considérée (faible fréquence allélique minoritaire, *MAF*), de tels loci étant plus difficiles à détecter dans les populations naturelles. Le spectre de fréquences alléliques est alors décalé en faveur de *MAF* plus élevées, ce qui peut par exemple imiter la signature d'une chute brutale de l'effectif d'une population (goulot d'étranglement) de manière artefactuelle (Polanski & Kimmel 2003). Une sous-représentation de loci à faibles *MAF* peut également entraîner un biais dans l'estimation d'autres paramètres à l'échelle de la zone d'étude, comme le taux d'hétérozygotie global (Nielsen 2000; Wakeley et al. 2001), le déséquilibre de liaison (Akey et al. 2003; Nielsen & Signorovitch 2003) ou le degré de différenciation génétique entre populations (Nielsen 2004). Ce dernier point est particulièrement vrai si la différenciation génétique est de faible magnitude et s'effectue préférentiellement sur les loci à faibles *MAF* (Lachance et al. 2013). Dans une telle situation, l'AB peut alors conduire à un manque de puissance statistique à détecter de faibles niveaux de différenciation. Dans notre protocole, une perte de loci à faible *MAF* peut survenir à plusieurs niveaux comme indiqué ci-dessous (Fig.D.1).

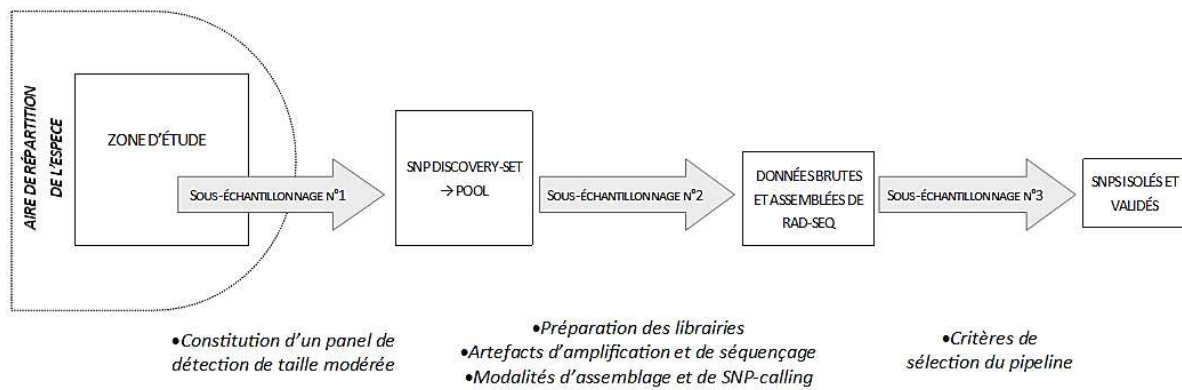


Figure D.1 : Schéma décrivant les étapes successives de sélection de polymorphisme génétique pour les espèces focales du Haut-Maroni, depuis la première étape de la constitution du panel de détection de SNP jusqu'à obtention du set de marqueurs final ayant servi de base au génotypage. Les différents sous-échantillonnages correspondent à tout autant de «goulots d'étranglement», à l'issue desquels le nombre de marqueurs conservés se trouve de plus en plus réduit.

Chronologiquement, la constitution du panel de détection de SNP (c'est-à-dire, la sélection des individus utilisés pour créer le pool d'ADN à partir desquels la sélection de marqueurs candidats sera effectuée) est la première étape critique en ce qui concerne l'AB (Nielsen 2004) (Fig.D.1 « Sous-échantillonnage 1 »). Les SNP à faibles MAF à l'échelle de la zone de collecte présentent en effet une probabilité plus faible d'être échantillonnés. Il faudrait alors échantillonner davantage d'individus pour pouvoir observer les deux états alléliques de tels loci, et donc pour les détecter comme étant polymorphes. Il faut cependant prendre garde à ce que l'augmentation du nombre d'individus dans chaque panel ne se fasse pas au détriment de la profondeur de couverture moyenne de chacun. Pour construire nos panels de détection de SNP, nous nous sommes appuyés sur les recommandations de Morin et al. (2009), qui suggéraient l'intégration d'au moins 10 individus dans chaque panel. En outre, nous avons sélectionné aléatoirement cette dizaine d'individus sur l'ensemble de la zone géographique d'intérêt, ce qui a été évaluée comme un bon moyen d'éviter au mieux l'AB à ce niveau (Akey et al. 2003; Rosenblum & Novembre 2007). Au final, en comparant, pour chaque espèce, les fréquences alléliques des SNP génotypés parmi les individus constitutifs du panel de détection et parmi l'ensemble des individus collectés sur la zone d'étude, j'ai constaté une bonne correspondance entre les fréquences alléliques à ces deux échelles (Annexe XII, Table XII.1A et Fig.XII.2). Cette observation soutient la pertinence de notre démarche et son caractère aléatoire dans la constitution du panel de détection, qui a permis de capturer au minimum la variabilité génétique la plus évidente à l'échelle de la zone d'étude.

L'AB peut également intervenir lors du processus de séquençage de l'ADN, puis de l'assemblage des reads obtenus (Fig.D.1 « Sous-échantillonnage 2 »). La préparation des librairies avec des pools issus d'espèces différentes (et donc des tailles ou caractéristiques de génome potentiellement variables) pourrait diminuer la profondeur de couverture moyenne par espèce, au détriment des allèles rares pour certaines d'entre elles. De même, les individus constitutifs de chaque pool peuvent être représentés de façon déséquilibrée lors de la procédure d'amplification et de séquençage, auquel cas un polymorphisme rare risquerait d'être non-déTECTÉ s'il est porté par un seul individu peu représenté dans ce séquençage. Ces biais aléatoires, inhérents au processus de séquençage, ne peuvent pas être directement contrôlés. Nous pouvons cependant les illustrer partiellement. Ainsi, pour les SNP retenus, j'ai

constaté une équivalence correcte entre les fréquences alléliques dans le pool d'ADN séquencé (en moyenne sur les deux réplicats), et les fréquences alléliques obtenues à partir des génotypes des individus ayant servi à créer ces pools (Annexe XII, Table XII.1B et Fig.XII.3). Ceci suggère une représentation correcte de chaque individu pour ces loci en utilisant les deux réplicats. Par contre, l'équivalence des fréquences alléliques entre réplicats eux-mêmes (comparaison des profondeurs de couverture de chaque allèle entre les deux pools d'une espèce donnée, Annexe XII, Table XII.1C et D, Fig.XII.4) était moins marquée, indiquant une représentation fluctuante de chaque allèle d'un réplicat à l'autre. En outre, les algorithmes de *SNP-calling* (détection des sites polymorphes) comme celui implémenté dans le logiciel Stacks ont généralement pour objectif de distinguer les polymorphismes réels de simples erreurs de séquençage en analysant notamment les profondeurs de couverture (Emerson et al. 2010; Raineri et al. 2012; Catchen et al. 2013). La distinction entre erreurs de séquençage et allèles rares à faible profondeur de couverture devient alors souvent délicate, et peut conduire à ignorer des loci à faible MAF. L'occurrence de contaminations croisées au cours du protocole de séquençage peut renforcer ce biais en augmentant la proportion de marqueurs faux-positifs. La question des contaminations croisées (comment les identifier et comment y faire face) a été abordée dans le Chapitre 1 (voir aussi Annexe V).

Pour finir, les critères de filtrage de notre propre pipeline peuvent aussi agir en défaveur des loci à faibles MAF (Fig.D.1 « Sous-échantillonnage 3 »). Par exemple, le filtre SQL (qui permet la première sélection de SNP candidats à partir du catalogue de RADtags de Stacks) agit sur la profondeur de couverture minimale d'un allèle dans chaque réplicat, et force un SNP à être présent dans les deux réplicats d'une espèce pour être considéré comme candidat valide – ce qui serait a priori moins probable pour un SNP portant un allèle rare. Pour certaines espèces, il a été nécessaire de rendre ces filtres légèrement plus ou moins stringents pour, respectivement, éviter l'effet de contaminations croisées ou augmenter le nombre de SNP candidats de départ. En l'absence d'*a priori* disponible sur le spectre de fréquences alléliques réel à l'échelle du génome de l'espèce sur la zone d'étude considérée, il est impossible d'ajuster ces paramètres pour orienter la sélection de SNP candidats en fonction de leur MAF dans les RADtags pour approcher une représentation correcte de la réalité. C'est pourquoi le pipeline présenté en Chapitre 1 favorise la sélection de SNP ayant la plus forte probabilité d'être facilement génotypés, plutôt que sur la sélection (très ardue à mettre en œuvre) de SNP réellement représentatifs du spectre de fréquences alléliques réelles. Il faudrait de plus, pour cela, faire l'hypothèse forte que les deux niveaux de « sous-échantillonnage » précédents n'ont pas déjà préalablement introduit d'AB.

Si nécessaire, il est possible de minimiser l'impact de l'AB en corrigeant directement le spectre de fréquences alléliques au niveau des données brutes avant de s'en servir pour estimer des paramètres biologiques (Nielsen et al. 2004; Rosenblum & Novembre 2007). Par exemple, l'utilisation des ressources moléculaires développées en vue d'estimer la croissance d'une métapopulation (et mettre en évidence un phénomène d'expansion ou de réduction globale) rendrait indispensable une correction des spectres de fréquences alléliques (ex. Polanski & Kimmel 2003). De telles corrections permettraient également d'estimer diverses mesures de variabilité génétique (ex. Ramirez-Soriano et al. 2008). Cependant, tester la significativité de ces estimations nécessite alors le recours à des méthodes analytiques laborieuses pour constituer des intervalles de confiance adaptés (Rosenblum & Novembre 2007). En outre, de

telles corrections ne sont pas disponibles pour l'application de méthodes basées sur les données génotypiques (Helyar et al. 2011), comme l'étude du déséquilibre de liaison pour l'estimation des tailles efficaces par exemple. Ainsi dans notre cas, l'effort déployé pour tenter de corriger l'AB sur chacune des 18 espèces n'aurait été que peu pertinent au regard des résultats partiels que nous aurions été en mesure d'obtenir. Nous ne l'avons donc pas tenté.

La question de l'*ascertainment bias* lors de l'utilisation de panels de SNP reste toujours d'actualité bien qu'elle soit peu à peu minimisée avec le développement des techniques de génotypage-par-séquençage (Lachance et al. 2013; Malomane et al. 2018). Au cours de ce travail, il a fallu trouver un compromis entre sa minimisation lorsque cela était possible (notamment lors de la constitution du panel de détection de SNP) et l'évitement de marqueurs faux-positifs (comme la sélection d'un SNP à faible MAF qui aurait pu résulter d'une simple erreur de séquençage). Dans notre cas, la principale limitation liée à l'AB est le risque de non-détection d'une différenciation génétique de faible amplitude. Il s'agit donc essentiellement d'une question de puissance statistique, qui ne remet pas en question la validité des patrons de variabilité génétique détectés. Les marqueurs que nous avons développés par cette méthode nous ont bien permis de mettre en évidence des patrons de variation génétique en cohérence avec les connaissances biologiques générales disponibles.

1.1.3. Recommandations pratiques

A l'heure actuelle où le développement de ressources moléculaires reste nécessaire pour caractériser la variabilité génétique de nombreuses populations naturelles, de nouveaux protocoles de mise au point de marqueurs génétiques se développent en lien avec les technologies NGS, y compris en ce qui concerne les espèces néotropicales (Martínez et al. 2016; 2017; Pimentel et al. 2018). A l'issue de cette première étape de travail, il est possible de formuler quelques recommandations à destination des utilisateurs qui souhaiteraient reproduire la méthodologie que nous proposons.

Dans notre cas, l'objectif était de développer des ressources moléculaires utilisables pour comprendre la diversité génétique à petite échelle spatiale. D'autres applications peuvent nécessiter le développement de marqueurs diagnostiques (c'est-à-dire permettant de réassigner un individu à une population d'origine), ou permettant de décrire la variabilité génétique à des échelles spatiales ou temporelles différentes.

La définition des objectifs conditionne la suite du processus et, notamment, la façon de constituer le panel de détection de SNP. Nous pourrions par exemple développer des marqueurs moléculaires applicables à différentes échelles géographiques en Guyane française. En effet, l'utilisation des ressources moléculaires que nous avons développées pour étudier, par exemple, un autre bassin versant que le Maroni, pourrait induire de l'AB spatial (Bradbury et al. 2011). Ce biais n'invalide pas la possibilité de réutiliser les ressources génétiques développées au cours de cette thèse, mais pourrait si besoin être évité en développant de nouvelles ressources génétiques propres à la zone géographique considérée. Il faudrait alors constituer différents pools d'ADN incluant des individus couvrant au mieux l'ensemble de la zone géographique considérée. L'utilisation de plusieurs pools d'origines géographiques différentes, pour une même espèce, permettrait la mise au point de SNP communs d'une région à l'autre ou au contraire de SNP diagnostiques, sous réserve de modifier notre pipeline

informatique à cet effet. Dans ce dernier cas, optimiser la détection de loci à faibles MAF pourrait se révéler pertinente. Il faudrait alors accroître le nombre d'individus présents dans chaque pool afin d'augmenter la probabilité d'échantillonner de tels loci. Le meilleur compromis entre une profondeur de couverture suffisante pour que chaque individu soit séquencé de façon équivalente, et le nombre total de pools séquencés devra être recherché, tout en gardant des répliquats pour chacun d'entre eux.

Dans le cadre d'un protocole mêlant des espèces morphologiquement semblables, nous avons souligné l'intérêt de vérifier l'identité des spécimens incorporés dans chaque panel de détection à l'aide du barcoding (Annexe IV). Ceci permet de repérer d'éventuelles erreurs d'étiquetage d'échantillon, qui peuvent arriver lorsque de nombreuses espèces sont collectées lors d'une même campagne de terrain. En outre, une comparaison systématique, à l'issue du séquençage, des RAD-tags des différentes espèces permet de détecter de manière précoce d'éventuelles contaminations croisées et d'éviter le génotypage de marqueurs faux-positifs.

A ce stade, nous pouvons souligner de nouveau la nécessité des répliquats dans notre manipulation. Ils permettent en effet de fournir une double information sur chaque SNP détecté et de renforcer la probabilité d'éviter les marqueurs faux-positifs. En outre, l'augmentation du nombre de répliquats augmente la probabilité de détection de loci à faible MAF si ceux-ci sont nécessaires, par exemple pour la mise au point de SNP diagnostiques.

Enfin, l'outil de génotypage choisi pour caractériser et valider les ressources moléculaires obtenues dépend également du nombre d'individus ou d'espèces ciblés et du coût induit. Dans notre travail, la spectrométrie de masse a été retenue car le nombre d'individus collectés par espèce était relativement faible, alors que le nombre total de loci à génotyper était élevé. En prenant en compte le coût d'achat d'environ 150 couples d'amorces pour chaque espèce, c'est la méthode qui nous offrait le coût le plus bas par individu. En comparaison, une méthode de génotypage combinant PCR en multiplex et séquençage haut débit, telle que les kits TruSeq (Illumina) ou encore le GT-Seq (Campbell et al. 2011) aurait coûté 2 à 3 fois plus cher par individu, du fait d'un surcoût des amorces. Cependant, ces méthodes peuvent se révéler compétitives si le nombre d'individus augmente, du fait de l'étalement des coûts des amorces par individu. De nouvelles méthodes de génotypage similaires, encore plus simples et financièrement compétitives, ont été développées puis améliorées depuis le début de ma thèse (ex. Hi-Plex, Pope et al. 2018) et peuvent aussi constituer des alternatives intéressantes.

1.2. Appliquer une approche comparative entre espèces

Dans la seconde partie de ce travail de thèse (Chapitre 2), une approche de génétique comparative des populations a été mise en œuvre pour documenter les tendances générales et particulières montrées par les espèces focales du Haut-Maroni.

Comme évoqué en introduction, la plupart des études multi-espèces doivent souvent trouver un compromis entre l'étude de nombreuses espèces pour lesquelles les données disponibles sont de nature et d'origines variables (approche méta-analytique), ou un nombre d'espèces plus réduit mais étudiées sur une zone géographique commune (co-échantillonnage). Ici, nous présentons un travail d'une envergure inédite portant sur 18 espèces co-échantillonnées au niveau de sites de collectes identiques, sur une période de temps commune et analysées de manière identique.

Comme espéré, cette approche a permis de mettre en évidence différents patrons de variabilité génétique spatiale dont certains se retrouvaient chez plusieurs espèces. Nous avons par exemple mis en évidence, chez plusieurs espèces, l'influence de la configuration dendritique sur les patrons de variabilité génétique. L'approche comparative a également permis de mettre en évidence les particularités de certaines zones géographiques du Haut-Maroni, et notamment des tributaires du Tampok et de la Waki. A l'inverse, elle a mis en lumière des cas particuliers, c'est-à-dire des espèces présentant des patrons de variation uniques comme par exemple *Hoplias aimara* ou encore *Hypostomus gymnorhynchus*.

J'ai également évoqué en introduction les biais pouvant être introduits dans une étude de génétique comparative en utilisant des marqueurs d'origine différente. Afin de s'en affranchir au maximum, nous avons développé des marqueurs spécifiques, en suivant un protocole le plus standardisé possible entre les espèces. Cependant, nous avons vu précédemment que les problématiques d'AB ne peuvent être totalement écartées et pourraient également bruite des comparaisons quantitatives. En effet, si l'influence de l'AB diffère légèrement entre deux espèces, la capacité d'un des jeux de données à détecter un certain patron de variation spatial peut être moins marquée et donc, plus difficile à comparer quantitativement avec une seconde espèce. Il est actuellement impossible de quantifier ce biais pour chaque espèce, même s'il est vraisemblable que son effet reste mineur.

1.3. Simuler des processus et explorer des hypothèses

Dans la troisième partie de ce travail de thèse (Chapitre 3), j'ai utilisé les outils de simulation afin de mieux visualiser de possibles évolution des patrons de variabilité génétique au sein d'une métapopulation subissant une pression de pêche.

Cette étape m'a surtout permis de mieux appréhender les effets confondants d'une pression de pêche localisée et de flux de gènes à plus grande échelle, qui peuvent bruite une comparaison spatiale entre zones soumises à des pressions de pêche variable. Ceci souligne l'utilité de considérer, en plus des analyses de variabilité génétique spatiales indispensables telles que celles effectuées en Chapitre 2, d'autres approches complémentaires. Dans le cadre plus appliqué de ce projet de thèse, l'idéal serait de pouvoir par la suite combiner les données simulées aux données observées du Chapitre 2. Via une approche inférentielle, il s'agirait d'évaluer quelle serait la situation actuelle la plus probable pour chacune des espèces étudiées dans ce projet. Cela reviendrait à se poser la question suivante : les patrons de variabilité génétique observés, qu'ils semblent en cohérence avec la configuration dendritique du milieu hydrographique ou qu'au contraire ils s'en éloignent, sont-ils d'origine naturelle – ex. flux de gènes et tailles efficaces locales naturellement très limités; ou bien ont-ils été, au moins en partie, générés ou renforcés par la pression de pêche (ou toute autre contrainte, ex. dégradation des milieux autour des zones aval plus urbanisées) ?

Cependant à ce stade, une approche inférentielle semble complexe à mettre en œuvre. Nous avons évoqué en Chapitre 3 l'absence, pour le moment, de statistiques résumées qui auraient pu aider à la distinction de différentes combinaisons de tailles locales de patches (N), de probabilité de migration (m) et de proportion d'individus éliminés aléatoirement (H) dans certains patches. Il faudrait alors pouvoir documenter plus finement certains paramètres démographiques des populations focales du Haut-Maroni afin d'alimenter les modèles simulés

et de les rendre aussi proches que possible de la réalité biologique pour améliorer leur pouvoir inférentiel : tailles efficaces locales, sens et intensité des flux de migration par exemple. Cependant, et comme souligné en Chapitre 2, les estimations précises de tailles efficaces restent complexes à obtenir. Par ailleurs et malgré la mobilisation d'outils analytiques comme le programme divMigrate (Sundqvist et al. 2016) et BayesAss (Wilson & Rannala 2003), il n'a pas été possible de détecter une éventuelle asymétrie des flux de migration. Ainsi, une autre option pour discriminer différents scénarii évolutifs en cours dans une population naturelle, serait de comparer des patrons de variabilité génétique entre deux pas de temps afin de voir leurs évolutions. En effet si, ponctuellement, un patron donné de variabilité génétique peut être expliqué par plusieurs scénarii, nous avons vu que ces différents scénarii ne montraient pas toujours la même évolution au cours du temps. Dans cette optique, les outils de simulation pourraient se révéler utiles pour affiner les attendus théoriques quant à l'évolution temporelle de différents scénarii, mais aussi pour évaluer la capacité de différents plans d'échantillonnage temporel à détecter un changement significatif d'indices tels que la diversité génétique globale. Par exemple, Hoban et al. (2014) ont estimé, à l'aide du logiciel de simulations Nemo (Guillaume & Rougemont 2006), que 10 générations d'écart étaient nécessaires entre deux échantillonnages successifs pour détecter, via l'étude de la richesse allélique de 20 loci microsatellites, une réduction d'effectif de population de $N_e = 2000$ à $N_e = 50$.

En outre et à défaut d'une approche inférentielle, les simulations peuvent être poursuivies pour explorer plus finement l'évolution de différents scénarii théoriques sur l'évolution des patrons de variabilité génétique. Elles resteraient alors ancrées dans un cadre théorique et exploratoire. Cependant, il est toujours utile de comprendre à la fois les impacts locaux et globaux de la pêche (ou de tout autre impact anthropique, comme par exemple une dégradation de l'habitat) dans une métapopulation, car ils conditionnent sa persistance au cours du temps. La capacité de persistance d'une métapopulation peut ainsi être conditionnée par différents phénomènes comme l'asynchronie des dynamiques locales entre patchs (ex. effet portfolio, Figge 2004; Schindler et al. 2010), les patchs d'une métapopulation agissant comme sources ou comme puits de diversité, ou encore le fait que la pêche agisse plutôt sur les sources ou plutôt sur les puits. Ainsi, la productivité de la métapopulation (en termes d'abondance), sa diversité génétique globale, sa taille efficace totale (*meta- N_e* , ex. Gomez-Uchida et al. 2013) en comparaison de la somme des tailles efficaces locales, sont autant d'éléments qui peuvent évoluer sous l'influence de la pêche. Cette évolution peut en outre varier en fonction du fonctionnement naturel de cette métapopulation (intensité et sens des flux de migration, tailles efficaces locales et leur variabilité spatiale) et nécessiter un réajustement de l'échelle des mesures de gestion et de conservation des populations à mettre en place.

2. Pratiques de pêche et conservation des populations de poissons du Maroni

Les travaux menés par Guillaume Longin, au cours de la première partie du Programme Pêche, ont mis en lumière l'ambivalence qui caractérise en partie les pêcheries artisanales liées aux communautés natives du Haut-Maroni. D'abord, nous pouvons souligner le fait que les modes de vie traditionnels ne sont originellement pas ancrés dans une logique d'exploitation intensive et productiviste des ressources. La possibilité logistique limitée d'accéder aux zones de pêche lointaines, conjuguée à l'accessibilité plus importante aux produits d'importation via

l'implantation de magasins « chinois » sur les rives surinamaises du Maroni, pourrait même être susceptible de freiner, à long terme, la demande en produits issus du fleuve et donc de réduire l'impact de la pêche. Il semble ainsi que nous restons loin, pour l'heure, des caractéristiques classiques d'une pêcherie commerciale dont les objectifs seraient de maximiser le rendement annuel des captures. A priori, même les captures à destination de la vente se limitent à des événements ponctuels - par exemple lors de la capture de *Tometes lebaili* pendant les nivrées ou les explorations non-systématiques des zones amont du Maroni par les communautés Amérindiennes.

Pourtant, l'utilisation d'engins de pêche plus efficaces et moins sélectifs depuis quelques décennies, la concentration locale dans l'espace des activités de pêche et leur étalement sur l'année, y compris en période de reproduction, est avérée (Longin 2016) et le sentiment d'une baisse d'abondance de la ressource est partagé par l'ensemble des communautés du Haut-Maroni. Alors, la modernisation des pratiques de pêche s'est-elle accompagnée d'une pression de pêche réellement plus forte sur les ressources halieutiques que ce qu'elle était par le passé ? Ne constitue-t-elle qu'une augmentation transitoire de la pression de pêche, menée à s'annuler avec la persistance de garde-fous culturels empêchant de considérer la ressource comme illimitée (Longin 2016), et la diversification des sources alimentaires ? A-t-elle malgré tout suffi à générer une baisse de la ressource halieutique et une érosion de sa diversité ?

Ce travail de thèse, approche complémentaire à la description des pratiques de pêche menée par le Parc Amazonien de Guyane durant le premier volet du Programme Pêche, a permis d'apporter des éléments de réflexion supplémentaires.

2.1. Variabilité génétique et paysage aquatique

2.1.1. Influence de l'environnement naturel

Outre l'influence de la configuration du réseau dendritique sur la variabilité génétique de certaines espèces, nous avons soulevé au cours de ces travaux de possibles effets locaux exercés par le milieu, surtout au niveau des tributaires du Tampok et de la Waki. Des effets locaux sont également susceptibles d'expliquer la forte structuration génétique observés chez les espèces *Geophagus harreri* et *Serrasalmus eigenmanni* (isolement génétique marqué même entre des sites proches). Cependant, sans descripteurs précis des habitats échantillonnés, il est difficile d'aller plus loin dans ces interprétations.

Nous pouvons cependant nous focaliser sur l'hypothèse d'un impact de l'orpaillage comme facteurs explicatif de la faible diversité génétique observée chez plusieurs espèces sur le Tampok et la Waki. La carte ci-dessous (Fig.D.2), élaborée dans le cadre d'un projet d'étude des populations de loutres en Guyane française (P. Joubert, en cours - non publié), indique le nombre de sites d'orpaillage répertoriés sur les différents tronçons du Haut-Maroni, ainsi que l'ancienneté de présence d'orpaillage d'une manière générale. Le fort impact sur la Waki et le Tampok (surtout en aval) est immédiatement visible. Les collectes de terrain menées en Décembre 2015 avaient par ailleurs permis de constater la forte turbidité de l'eau au niveau du village de Kayodé, sur le bas-Tampok (R. Covain, pers. comm.). Il est donc vraisemblable que les activités d'orpaillage puissent avoir un impact négatif sur la qualité de l'habitat et agir comme une contrainte influençant la variabilité génétique des populations, en plus d'autres variables relatives à l'environnement naturel.

Cependant, par manque de données descriptives du milieu, l'influence plus précise des variables environnementales locales reste difficile à définir. Une description plus précise des paramètres environnementaux associés aux 12 localités échantillonnées serait utile. L'échantillonnage d'individus sur un plus grand nombre de sites répartis sur d'autres tributaires du Maroni – dont certains, qui n'ont pas été intégrés au présent travail car localisés plus en aval sur le bassin versant, sont connus comme fortement impactés par certaines activités humaines et notamment l'orpaillage ; pourrait permettre définir les hotspots de diversité génétique en lien avec l'environnement local. En parallèle, l'échantillonnage de sites plus en aval des Abattis Cottica permettrait d'analyser comment la variabilité génétique évolue à mesure sur l'on se rapproche de l'embouchure du Maroni. En particulier, nous pourrions voir si elle se maintient à des niveaux équivalents, voire supérieurs, à ceux observés au niveau des villages du Haut-Maroni, en relation avec la théorie d'augmentation de la diversité génétique vers l'aval des bassins versants (Paz-Vinas et al. 2015), ou si au contraire elle varie de façon inattendue (par exemple sous la pression croissante des activités anthropiques en direction des zones côtières).

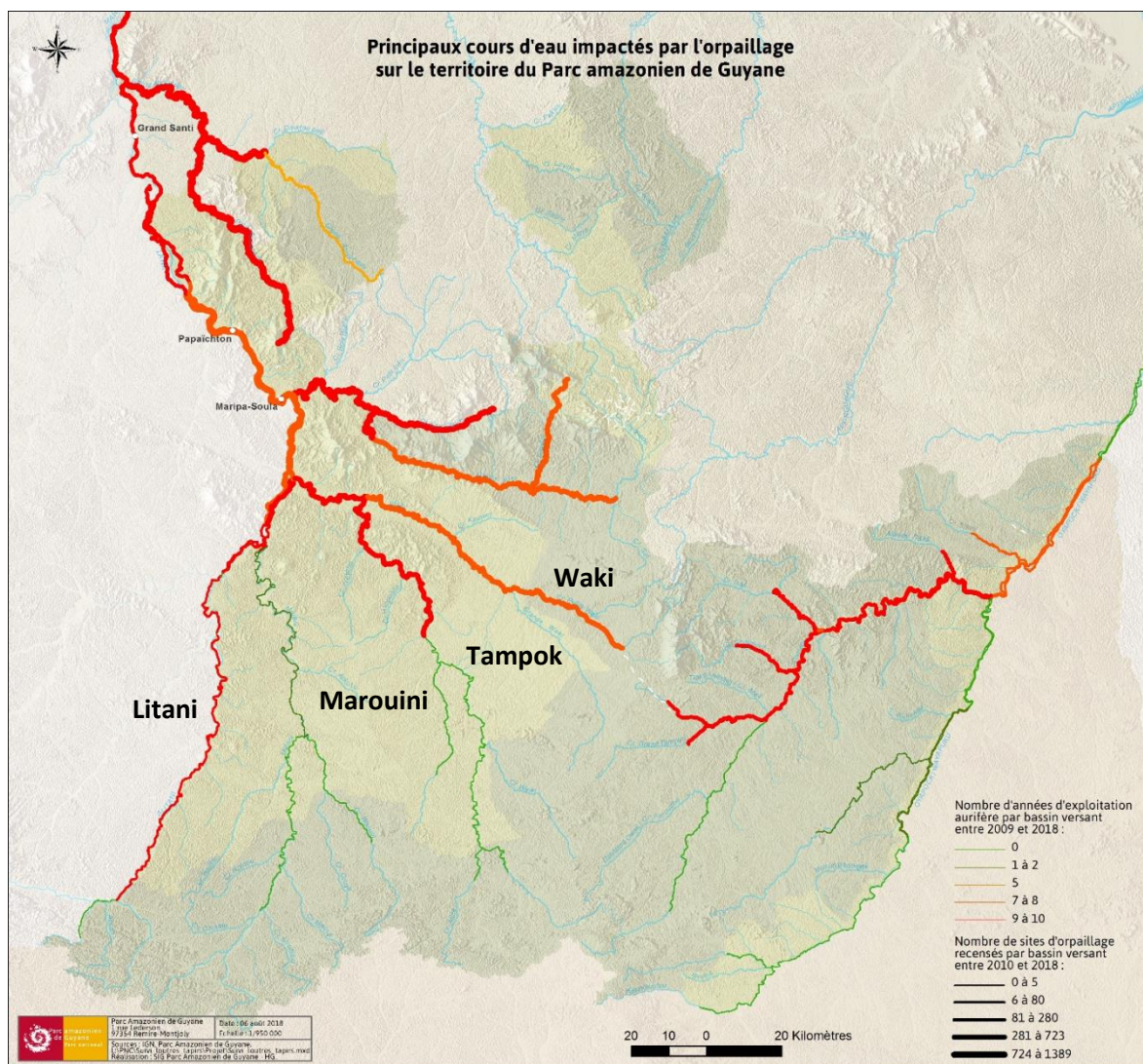


Figure D.2 : Principaux cours d'eau impactés par l'orpaillage sur le territoire du Parc Amazonien de Guyane (P. Joubert, communication personnelle)

2.1.2. Contrastes géographiques des rendements de captures

Guillaume Longin soulève, dans son travail, l'inégalité des rendements de captures entre les différents villages du Haut-Maroni. Il souligne principalement trois cas : 1) Les captures par unité d'effort de pêche effectuées au niveau des Abattis Cottica sont faibles et entraînent la mobilisation d'un effort de pêche important, 2) Le village de Kayodé, dans une moindre mesure, présente aussi des rendements de captures plus faibles comparé à d'autres villages Amérindiens du HM. Dans ce second cas, G. Longin fait l'hypothèse d'une accessibilité aux espèces diminuées à cause des fortes variations saisonnières de la hauteur d'eau dans cette zone, qui expliquerait ces faibles rendements par simple effet de dilution. 3) Enfin, le village de Taluen (et la zone de Twenké associée) présente aussi un rendement bas, à la fois en nombre de captures et en biomasse (Longin 2016). On peut donc s'interroger sur l'état des ressources halieutiques au niveau de ces trois zones géographiques.

Certains de nos résultats en termes de diversité génétique concordent avec ces observations. A Kayodé, il est vraisemblable que les niveaux de diversité génétique, légèrement inférieurs à ceux des autres villages pour certaines espèces, soient également liés aux conditions environnementales particulières rencontrées sur le Tampok (fortes variations saisonnières des hauteurs d'eau et forte turbidité) plutôt qu'à la pression d'exploitation. Ces observations conjointes, quelles que soient leurs causes réelles, des rendements de captures et des niveaux de diversité génétique plus faibles sur le Tampok et la Waki justifient une certaine vigilance à l'égard des ressources exploitées à cet endroit. Le cas du village de Papaïchton, qui exploite les Abattis Cottica, est plus complexe. Les niveaux de diversité génétique, là encore, sont un peu plus faibles chez certaines espèces qu'au sein des autres villages. D'autre part, les analyses de tailles efficaces indiquent des valeurs pouvant être très faibles (≤ 60 pour plusieurs cas).

Ces résultats sont cependant à prendre avec précautions au vu de la difficulté d'estimation précise de N_e . En outre, une réduction de taille de populations n'est pas le seul facteur susceptible de générer un déséquilibre de liaison d'allèles, et des facteurs confondants peuvent donc brouter l'estimation de la taille efficace. A ce stade, nous pouvons donc situer Papaïchton comme un site méritant une attention particulière. D'une manière générale, les faibles rendements en captures observés dans ce bassin de vie et autour de la zone des Abattis Cottica, leur situation géographique en aval de la zone fortement urbanisée de Maripasoula et nos propres résultats génétiques justifient une certaine vigilance à l'égard de cette portion géographique du Haut-Maroni.

Nous n'avons, en revanche, pas observé de plus faible diversité génétique au niveau du village de Taluen et Twenké. Au contraire, cette zone située sur la partie haute du cours principal de la Lawa abrite globalement les niveaux de diversité génétique les plus élevés. Ici, les faibles rendements observés par Guillaume Longin ne semblent donc pas liés à une vulnérabilité de populations associée à une diversité génétique plus faible. On peut donc supposer que d'autres mécanismes expliquent les faibles rendements observés, comme un déplacement des populations vers des zones plus amont ou au niveau de la confluence entre le Marouini et le Litani, ou encore un effet de dilution de la ressource dû à un étalement important de la surface du cours d'eau ou à des habitats moins diversifiés.

En conclusion, la disparité des rendements en captures constatée entre villages au cours de la première partie du Programme Pêche, et les analyses de variabilité génétique effectuées apparaissent surtout concordantes – et potentiellement liées à la pression de pêche, au niveau du bassin de vie de Papaïchton. Dans tous les cas, il est fortement probable que cette pression n’agisse pas seule mais en conjonction avec d’autres variables environnementales ou d’origines anthropiques extérieures à la pêche. Cela pourrait par exemple être le cas à Kayodé, où il n’est pas exclu qu’un milieu naturellement moins favorable puisse rendre les populations locales plus vulnérables face à la pression de pêche à moyen terme. Si les contraintes naturelles et anthropiques peuvent générer des effets locaux, détectables ou non; leurs possibles impacts globaux sur l’ensemble des métapopulations du Haut-Maroni ne sont pas à négliger.

2.2. Variabilité génétique et traits d’histoire de vie

2.2.1. Relation entre les caractéristiques biologiques et la structure génétique des populations

Lorsque les comparaisons spatiales entre zones géographiques soumises à des pressions de pêche différentes, ou encore la caractérisation de la dynamique des métapopulations étudiées, se révèlent complexes; les comparaisons de la différenciation génétique globale entre espèces peuvent apporter des éléments de réflexion supplémentaires.

Outre la caractérisation de patrons de variabilité génétique spatiale plus ou moins partagés entre espèces, une autre application de la génétique comparative des populations est la mise en relation entre patrons de variabilité génétique et traits d’histoire de vie. Dans le Chapitre 2, nous avons vu qu’en classifiant les espèces en trois grandes catégories en fonction de leur capacité de dispersion (mobilité passive des œufs/larves et mobilité active des adultes), on observe des patrons de différenciation génétique spatiale relativement similaires pour les espèces d’une même catégorie. Nous pourrions ainsi espérer prédire les patrons globaux de variation génétique d’espèces avec des traits d’histoire de vie identiques et pour lesquelles on ne dispose pas de données génétiques. En outre, nous pouvons également comparer les patrons de variabilité génétique chez des espèces présentant des traits d’histoire de vie identiques mais des intérêts halieutiques contrastés.

Certaines espèces ont pourtant montré des niveaux de différenciation génétique inattendus au regard de leur catégorie. Ceci pourrait s’expliquer par deux raisons principales. La première serait qu’un des critères choisis pour évaluer la capacité de dispersion, c’est-à-dire la stratégie de reproduction ou le comportement migrateur, a été mal évalué pour certaines espèces ou encore que d’autres traits d’histoire de vie, moins précisément connus, influencent davantage la variabilité génétique. La seconde raison serait qu’une contrainte extérieure particulière agissant uniquement sur certaines espèces, comme la pression de pêche, modifie leurs patrons naturels de variabilité génétique. Dans le Chapitre 2, nous soupçonnons ainsi que le degré de spécialisation écologique constitue l’explication la plus probable pour les différences observées parmi les six espèces de Serrasalmidés, présentant des capacités de dispersion théoriques identiques mais dont trois montrent une structure génétique plus prononcée que les trois autres, et ce malgré la présence d’espèces fortement ciblées par la pêche dans les deux cas. Cependant, les analyses préliminaires du Chapitre 3 ont suggéré la possibilité que la

pression de pêche puisse renforcer le patron de variabilité génétique qui est attendu au sein d'une métapopulation de type dendritique. A ce stade, nous ne pouvons donc identifier clairement l'origine des patrons de variabilité génétique contrastés que l'on observe chez les Serrasalmidés. Un phénomène similaire est observé pour les Loricariidae où l'espèce *Pseudancistrus barbatus*, très abondante et fortement ciblée par la pêche, présente une structure génétique plus marquée que les espèces voisines *Hypostomus gymnorhynchus* ou *Harttia guianensis* qui présentent pourtant une écologie très similaire mais sont, respectivement, moins abondante et moins recherchée par les pêcheurs.

Ainsi, la définition préalable de catégories d'espèces suivant leurs traits de vie s'est révélée utile pour définir des attendus généraux. Cependant, pour comprendre plus précisément, au sein d'une même catégorie, l'origine de patrons de variabilité génétique contrastés, des connaissances biologiques complémentaires sont nécessaires. Il serait notamment important de tirer parti des savoirs empiriques chez les communautés humaines locales, qui connaissent en grande partie les éléments de leurs cycles de vie. Une telle démarche de collecte des savoirs locaux n'a pas encore pu être mise en place à ce jour mais son importance a déjà été soulignée au cours des travaux développés durant le premier volet du Programme Pêche. Des données plus précises sur la taille à maturité des espèces pourraient également être utiles. Par exemple, l'espèce *Tometes lebaili* n'atteint sa maturité sexuelle qu'à la moitié de sa vie d'après Lord et al. (2007), ce qui correspond à une taille standard de plus de 20 cm de longueur et environ 10 cm de hauteur du corps, tailles au-dessous desquelles les individus entrent déjà dans la phase capturable par utilisation du filet tramail classiquement utilisé par les pêcheurs du Haut-Maroni. D'après ces éléments, il est ainsi possible que les individus *Tometes lebaili* puissent être capturés avant même le début de la reproduction, ce qui peut augmenter sa vulnérabilité face à la pression de pêche.

2.2.2. Composition spécifique des captures sur le Haut-Maroni

Sur l'ensemble des périodes de pêche suivies par Guillaume Longin au cours de la première partie du Programme Pêche, il a été possible d'identifier les espèces contribuant le plus aux captures effectuées à l'aide du filet tramail monofilament. Certaines espèces, comme *Doras micropoeus*, *Myloplus rubripinnis* ou *Pseudancistrus barbatus*, font partie des espèces capturées quotidiennement quel que soit le village considéré. D'autres, comme *Leporinus friderici*, *Leporinus lebaili* ou *Tometes lebaili*, ont été débarquées de manière plus sporadique et n'apparaissaient de façon régulière que dans certains des villages. Ce phénomène était généralement cohérent avec la nature des biotopes environnants, qui conditionnent leur abondance locale. Cependant, il peut aussi souligner la rareté naturelle de certaines espèces.

Considérant l'impossibilité d'obtenir des estimations précises de l'abondance des différentes espèces à l'échelle locale, le renouvellement de ce type d'observation à l'avenir pourrait aider à appréhender la rareté relative de certaines espèces et constituer une information additionnelle à celle fournie par les traits d'histoire de vie. Dans le cas des Serrasalmidés par exemple, les trois espèces pour lesquelles nous avons mis en évidence les patrons de structure génétique les plus ténus (*Serrasalmus rhombeus*, *Myloplus rubripinnis* et *Acnodon oligacanthus*), font toutes partie des espèces les plus capturées dans les cinq villages étudiés (trois villages dans le cas d'*A. oligacanthus*). En revanche, *Myloplus rhomboidalis*, *Tometes lebaili* et *Serrasalmus eigenmanni*, pour lesquelles le degré de structuration génétique était

plus important, font partie des espèces les plus débarquées dans respectivement trois, deux et aucun des villages (Longin 2016) ce qui souligne leur plus faible accessibilité, quelle qu'en soit la raison. Chez ces dernières espèces, de telles caractéristiques, si elles sont observées de façon répétée au cours du temps, pourraient préciser l'influence du degré de spécialisation écologique sur le degré de variabilité génétique spatiale et/ou souligner encore davantage leur possible vulnérabilité à la pression de pêche (voir Chapitre 2).

Les informations de captures et de biomasses spécifiques observées dans chacun des villages ont également été très informatives dans le cas de l'espèce *Hoplias aimara*. Elles ont confirmé que cette espèce était particulièrement ciblée dans le village de Papaïchton, étant la seconde espèce la plus capturée et représentant plus de la moitié de la biomasse débarquée, tout du moins au cours de la période durant laquelle les débarquements ont pu être suivis. De plus, il s'agit d'une espèce dont l'exploitation est orientée vers un commerce de petite échelle plutôt qu'une consommation quotidienne de subsistance. Or, nous avons vu qu'il s'agit de la seule espèce montrant à la fois une forte discontinuité génétique entre Papaïchton et les autres sites, une baisse de polymorphisme et une très faible taille efficace -alors que l'hypothèse d'un biais causé par une sous-structure temporelle sur l'estimation de N_e ne se vérifie pas pour cette espèce. Cela renforce l'hypothèse d'une sensibilité particulière à la pression de pêche pour cette espèce particulièrement emblématique et la nécessité de formuler des recommandations de gestion particulières dans le village de Papaïchton.

2.3. Conclusion et recommandations

A l'issue de ce travail de thèse, il n'a pas été observé de phénomène d'érosion génétique au niveau des aires géographiques exploitées par la pêche, tout du moins pour la majorité des espèces étudiées. Cependant, nous suggérons de développer une vigilance sur le suivi des populations, notamment en ce qui concerne certains sites comme Kayodé ou Papaïchton, mais également en ce qui concerne certaines espèces, comme *Hoplias aimara* ou *Tometes lebaili*. Cependant, afin de mieux comprendre le fonctionnement de l'ensemble des espèces étudiées et l'évolution de leur population face à la pression de pêche -mais aussi d'autres variables environnementales ou anthropiques (comme la pression d'orpaillage), nous pouvons suggérer plusieurs perspectives de travail à plus ou moins long terme:

- Documenter plus précisément les variables environnementales (physico-chimique, topographiques, etc) des sites de collecte étudiés pour affiner certaines des interprétations proposées au cours de ce travail.
- Compléter la stratégie de collecte d'échantillons en intégrant un plus grand nombre de sites couvrant des caractéristiques environnementales contrastées, si possible sur l'ensemble du bassin du Maroni afin de visualiser l'influence de différents forçages environnementaux lorsqu'ils sont connus.
- Approfondir nos connaissances sur les traits d'histoire de vie des différentes espèces, en confrontant les sources d'informations et données bibliographiques actuelles et à venir avec les connaissances empiriques des pêcheurs du Haut-Maroni.
- A plus long terme, étendre la stratégie d'analyse comparative à d'autres espèces pour étayer l'influence des traits d'histoire de vie sur la diversité génétique intra et interpopulations.

Les deux premiers points visent à tenir compte des autres variables environnementales susceptibles d'agir, au même titre que la pression de pêche, sur les niveaux de variabilité génétique locaux et globaux des espèces. Les deux derniers points permettraient d'avoir un point de vue plus argumenté et précis sur l'influence des traits d'histoire de vie sur la diversité génétique des espèces, afin de révéler le cas échéant des espèces présentant des patrons de variation inattendus pour mieux les intégrer dans les plans de gestion.

Si ces points de perspective se révèlent complexes à mettre en œuvre, une autre option pour discriminer différents scénarii évolutifs en cours dans une population naturelle réside dans le suivi et la comparaison des patrons de variabilité génétique de différentes espèces au cours du temps. En génétique du paysage, l'approche temporelle s'avère souvent d'une grande utilité pour discriminer des processus anciens liés à la démographie naturelle des populations, des processus contemporains (Epps & Keyghobadi 2015) d'origine anthropique par exemple. Dans certains cas, des échantillons anciens, issus de collections ou d'archives de musées, peuvent également fournir une référence temporelle ('monitoring rétrospectif', ex. Poulsen et al. 2006). Malheureusement, en Guyane française, l'absence de ce type de ressources pour les espèces et la zone géographique étudiées ne permettait pas d'envisager de monitoring rétrospectif. La mise en place de programmes de monitoring génétique (suivi, au cours du temps, de l'évolution de plusieurs variables génétiques chez une ou plusieurs populations via des collectes régulières sur le terrain), constitue néanmoins un outil précieux en biologie de la conservation (Schwartz et al. 2007). En s'appuyant sur une série temporelle de quelques années seulement, Mathieu-Bégné et al. (2018) ont par exemple mis en évidence une décroissance significative de la richesse allélique moyenne, une augmentation significative du déficit en hétérozygotes et une taille efficace maintenue très petite chez des populations de *Leuciscus burdigalensis* subissant une chute démographique graduelle. Dans notre cas, l'utilisation d'échantillonnages temporels permettrait ainsi par exemple de suivre les niveaux globaux et locaux de diversité génétique ou encore d'estimer plus précisément la taille efficace des populations qu'avec un échantillonnage unique (Gilbert & Whitlock 2015), d'autant plus si les contraintes naturelles et logistiques restreignent la taille des échantillons collectés (<50 individus) (Antao et al. 2011). Ceci permettrait de disposer d'indicateurs robustes de l'effet de la pêche sur les ressources halieutiques afin d'orienter efficacement les plans de gestion qui les concernent.

En parallèle et pour mieux appréhender l'impact de la pêche, le monitoring temporel pourrait également concerner le suivi de la composition spécifique des captures ou encore la quantité de protéines de poissons accessible par habitant. Nous pourrions également, sur la base des données biométriques déjà collectées au cours de ce projet (Longin 2016), envisager de comparer le coefficient de condition physique actuel des individus avec ceux disponibles via la bibliographie pour certaines espèces (ex. Lecomte et al. 1993 pour *Myloplus rhomboidalis*, Lord et al. 2007 pour *Tometes lebaili*). Nous pourrions ainsi voir, par exemple, si le coefficient de condition de certaines espèces diminue au cours du temps, révélant un état physiologique dégradé et donc une vulnérabilité accrue à certaines pressions anthropiques. Pour améliorer la comparabilité des données collectées au cours d'une telle procédure de monitoring « multi-paramétrique », nous pourrions nous appuyer sur des collectes communes de matériel biologique et biométrique issu par exemple des débarquements dans les villages.

Les ressources moléculaires développées pour ce travail de thèse rassemblent les caractéristiques idéales pour une approche de monitoring génétique : elles peuvent être génotypées par des méthodes robustes et routinières, y compris pour de l'ADN dégradé. Les génotypes obtenus peuvent être analysés à l'aide des outils analytiques et informatiques classiques de génétique des populations. Cela présente un avantage certain dans le cadre d'une collaboration entre différents acteurs de la conservation des ressources naturelles (chercheurs, gestionnaires, pêcheurs). L'intégration d'études de variabilité génétique au cœur de programmes de conservation des populations sauvages tend à se développer à l'heure actuelle. Un des bénéfices de cette approche intégrée réside dans la construction de jeux de données faciles à appréhender (voire à traiter) pour les gestionnaires mais ayant souvent, également, finalité à être standardisés et mis en commun entre différents acteurs de la conservation. Ces caractéristiques contribuent ainsi à unifier les aspects théoriques et appliqués de la génétique de la conservation (Hoban et al. 2013a; 2013b).

Le 18 décembre 2018

Références bibliographiques

- Abell R. (2002). Conservation Biology for the Biodiversity Crisis: a Freshwater Follow-up. *Conservation Biology*, 16(5), 1435-1437. <https://doi.org/10.1046/j.1523-1739.2002.01532.x>
- Abell R., Thieme M. L., Revenga C., Bryer M., Kottelat M., Bogutskaya N. *et al.* (2008). Freshwater Ecoregions of the World: A New Map of Biogeographic Units for Freshwater Biodiversity Conservation. *BioScience*, 58(5), 403-414. <https://doi.org/10.1641/B580507>
- Ackerman M. & Campbell M. (2012). Chinook and steelhead genotyping for genetic stock identification at lower granite dam. Annual Progress Report July 1, 2010 - June 30, 2011. Submitted to Bonneville Power Administration. Contract #48347; Project # 2010-026-00
- Agostinho A. A., Gomes L. C., Santos N. C. L., Ortega J. C. G. & Pelicice F. M. (2016). Fish assemblages in Neotropical reservoirs: Colonization patterns, impacts and management. *Fisheries Research*, 173, 26-36. <https://doi.org/10.1016/j.fishres.2015.04.006>
- Aguirre W. E., Shervette V. R., Navarrete R., Calle P. & Agorastos S. (2013). Morphological and Genetic Divergence of *Hoplias microlepis* (Characiformes: Erythrinidae) in Rivers and Artificial Impoundments of Western Ecuador. *Copeia*, 2013(2), 312-323. <https://doi.org/10.1643/CI-12-083>
- Akey J. M. (2002). Interrogating a High-Density SNP Map for Signatures of Natural Selection. *Genome Research*, 12(12), 1805-1814. <https://doi.org/10.1101/gr.631202>
- Akey J. M., Zhang K., Xiong M. & Jin L. (2003). The Effect of Single Nucleotide Polymorphism Identification Strategies on Estimates of Linkage Disequilibrium. *Molecular Biology and Evolution*, 20(2), 232-242. <https://doi.org/10.1093/molbev/msg032>
- Allan J. D., Abell R., Hogan Z., Revenga C., Taylor B. W., Welcomme R. L. & Winemiller K. (2005). Overfishing of Inland Waters. *BioScience*, 55(12), 1041. [https://doi.org/10.1641/0006-3568\(2005\)055\[1041:OOIW\]2.0.CO;2](https://doi.org/10.1641/0006-3568(2005)055[1041:OOIW]2.0.CO;2)
- Allard L. (2014) *Elaboration d'un indice de qualité des eaux basé sur la structure taxonomique et fonctionnelle des assemblages de poissons des petits cours d'eau de Guyane*. Thèse de doctorat, Université de Toulouse.
- Allard L., Brosse S., Covain R., Gozlan R., Le Bail P.-Y. *et al.* La Liste rouge des espèces menacées en France - Chapitres de la Faune vertébrée de Guyane. La Liste rouge des espèces menacées en France, Comité Français de l'UICN, MNHN Museum National d'Histoire Naturelle, pp.32-34, 2017.
- Allard L., Grenouillet G., Khazraie K., Tudesque L., Vigouroux R. & Brosse S. (2014). Electrofishing efficiency in low conductivity neotropical streams: towards a non-destructive fish sampling method. *Fisheries Management and Ecology*, 21(3), 234-243. <https://doi.org/10.1111/fme.12071>
- Allard L., Popée M., Vigouroux R. & Brosse S. (2016). Effect of reduced impact logging and small-scale mining disturbances on Neotropical stream fish assemblages. *Aquatic Sciences*, 78(2), 315-325. <https://doi.org/10.1007/s00027-015-0433-4>
- Allard L., Toussaint A., Vigouroux R. & Brosse, S. (2015). Length-weight relationships of 58 fish species in French Guiana streams. *Journal of Applied Ichthyology*, 31(3), 567-570. <https://doi.org/10.1111/jai.12712>
- Allendorf F. W., England P., Luikart G., Ritchie P. & Ryman N. (2008). Genetic effects of harvest on wild animal populations. *Trends in Ecology & Evolution*, 23(6), 327-337. <https://doi.org/10.1016/j.tree.2008.02.008>
- Allendorf F. W., Hohenlohe P. A. & Luikart, G. (2010). Genomics and the future of conservation genetics. *Nature Reviews Genetics*, 11(10), 697-709. <https://doi.org/10.1038/nrg2844>
- Alp M., Keller I., Westram A. M. & Robinson C. T. (2012). How river structure and biological traits influence gene flow: a population genetic study of two stream invertebrates with differing dispersal

- abilities: Biological traits and gene flow in stream invertebrates. *Freshwater Biology*, 57(5), 969-981. <https://doi.org/10.1111/j.1365-2427.2012.02758.x>
- Andrew R. L., Bernatchez L., Bonin A., Buerkle C. A., Carstens B. C., Emerson B. C., ... Rieseberg L. H. (2013). A road map for molecular ecology. *Molecular Ecology*, 22(10), 2605-2626. <https://doi.org/10.1111/mec.12319>
- Angeloni F., Wagemaker N., Vergeer P. & Ouborg, J. (2012). Genomic toolboxes for conservation biologists: Genomic toolboxes for conservation biologists. *Evolutionary Applications*, 5(2), 130-143. <https://doi.org/10.1111/j.1752-4571.2011.00217.x>
- Antao T., Pérez-Figueroa A. & Luikart G. (2011). Early detection of population declines: high power of genetic monitoring using effective population size estimators: Early detection of population declines. *Evolutionary Applications*, 4(1), 144-154. <https://doi.org/10.1111/j.1752-4571.2010.00150.x>
- Anthony W. J. B., Dreslik M. J., Douglas M. R., Thompson D., Klut G. M., Kuhns A. R., ... Phillips C. A. (2018). Population genetic evaluations within a co-distributed taxonomic group: a multi-species approach to conservation planning. *Animal Conservation*, 21(2), 137-147. <https://doi.org/10.1111/acv.12365>
- Antonio R. R., Agostinho A. A., Pelicice F. M., Bailly D., Okada E. K. & Dias, J. H. P. (2007). Blockage of migration routes by dam construction: can migratory fish find alternative routes? *Neotropical Ichthyology*, 5(2), 177-184. <https://doi.org/10.1590/S1679-62252007000200012>
- Antunes, A. P., Fewster, R. M., Venticinque, E. M., Peres, C. A., Levi, T., Rohe, F., & Shepard, G. H. (2016). Empty forest or empty rivers? A century of commercial hunting in Amazonia. *Science Advances*, 2(10), e1600936-e1600936. <https://doi.org/10.1126/sciadv.1600936>
- Arce H. M., Reis R. E., Geneva A. J. & Sabaj Pérez M. H. (2013). Molecular phylogeny of thorny catfishes (Siluriformes: Doradidae). *Molecular Phylogenetics and Evolution*, 67(3), 560-577. <https://doi.org/10.1016/j.ympev.2013.02.021>
- Archambault B., Rivot E., Savina M. & Le Pape O. (2018). Using a spatially structured life cycle model to assess the influence of multiple stressors on an exploited coastal-nursery-dependent population. *Estuarine, Coastal and Shelf Science*, 201, 95-104. <https://doi.org/10.1016/j.ecss.2015.12.009>
- Ardura A., Planes S. & Garcia-Vazquez E. (2011). Beyond Biodiversity: Fish Metagenomes. *PLoS ONE*, 6(8), e22592. <https://doi.org/10.1371/journal.pone.0022592>
- Arrington D. A., Davidson B. K., Winemiller K. O. & Layman C. A. (2006). Influence of life history and seasonal hydrology on lipid storage in three neotropical fish species. *Journal of Fish Biology*, 68(5), 1347-1361. <https://doi.org/10.1111/j.0022-1112.2006.00996.x>
- Azevedo M. A. (2010). Reproductive characteristics of characid fish species (Teleostei, Characiformes) and their relationship with body size and phylogeny. *Iheringia. Série Zoologia*, 100(4), 469-482. <https://doi.org/10.1590/S0073-47212010000400020>
- Baird N. A., Etter P. D., Atwood T. S., Currey M. C., Shiver A. L., Lewis Z. A., ... Johnson E. A. (2008). Rapid SNP Discovery and Genetic Mapping Using Sequenced RAD Markers. *PLoS ONE*, 3(10), e3376. <https://doi.org/10.1371/journal.pone.0003376>
- Balland P. & Roux A. (2005). *Rapport de l'Inspection générale de l'environnement: Pêche et gestion piscicole en Guyane*. Ministère de l'Ecologie et du Développement durable.
- Barletta M., Jaureguizar A. J., Baigun C., Fontoura N. F., Agostinho A. A., Almeida-Val V. M. F., ... Corrêa M. F. M. (2010). Fish and aquatic habitat conservation in South America: a continental overview with emphasis on neotropical systems. *Journal of Fish Biology*, 76(9), 2118-2176. <https://doi.org/10.1111/j.1095-8649.2010.02684.x>
- Batista J. S. & Alves-Gomes J. A. (2006). Phylogeography of *Brachyplatystoma rousseauxii* (Siluriformes - Pimelodidae) in the Amazon Basin offers preliminary evidence for the first case of "homing" for an Amazonian migratory catfish. *Genetics and Molecular Research*, 5(3), 723-740.

- Bayerl H., Kraus R. H. S., Nowak C., Foerster D. W., Fickel J. & Kuehn R. (2018). Fast and cost-effective single nucleotide polymorphism (SNP) detection in the absence of a reference genome using semideep next-generation Random Amplicon Sequencing (RAMseq). *Molecular Ecology Resources*, 18(1), 107-117. <https://doi.org/10.1111/1755-0998.12717>
- Bayley P. & Petrere M. (1989) Amazon fisheries: assessment methods, current status and management options. p. 385-398. In D.P. Dodge [ed.] Proceedings of the International Large River Symposium. Can. Spec. Publ. Fish. Aquat. Sci. 106.
- Begon M., Harper J. L. & Trownsend C. R. (1996) Ecology: Individuals, Populations & Communities. Blackwell 1996, 1068 p. Edition 3.
- Blanchet S., Prunier J. G. & De Kort H. (2017). Time to Go Bigger: Emerging Patterns in Macrogenetics. *Trends in Genetics*, 33(9), 579-580. <https://doi.org/10.1016/j.tig.2017.06.007>
- Blanchet S., Rey O., Etienne R., Lek S. & Loot G. (2010). Species-specific responses to landscape fragmentation: implications for management strategies. *Evolutionary Applications*, 3(3), 291-304. <https://doi.org/10.1111/j.1752-4571.2009.00110.x>
- Bonfanti J., Hedde M., Joimel S., Krogh P. H., Violle C., Nahmani J. & Cortet J. (2018). Intraspecific body size variability in soil organisms at a European scale: Implications for functional biogeography. *Functional Ecology*, 32(11), 2562-2570. <https://doi.org/10.1111/1365-2435.13194>
- Boudou A., Maury-Brachet R., Coquery M., Durrieu G. & Cossa D. (2005). Synergic Effect of Gold Mining and Damming on Mercury Contamination in Fish. *Environmental Science & Technology*, 39(8), 2448-2454. <https://doi.org/10.1021/es049149r>
- Boujard T., Pascal M., Meunier J.F. & Le Bail P.-Y. (1997) *Guide écologique de l'Approuague et de la réserve des Nouragues*. INRA, Paris.
- Boujard T., Lecomte F., Renno J.-F., Meunier F. & Neveu P. (1991). Growth in four populations of *Leporinus friderici* (Bloch, 1794) (Anostomidae, Teleostei) in French Guiana. *Journal of Fish Biology*, 38(3), 387-397. <https://doi.org/10.1111/j.1095-8649.1991.tb03128.x>
- Boujard T. (1992). Space-time organization of riverine fish communities in French Guiana. *Environmental Biology of Fishes*, 34(3), 235-246. <https://doi.org/10.1007/BF00004769>
- Boujard T., Le Bail P.-Y. & Planquette P. (1988). Données biologiques sur quelques espèces continentales de Guyane Française d'intérêt piscicole. *Aquatic Living Resources*, 1(2), 107-113. <https://doi.org/10.1051/alr:1988013>
- Boujard T. & Tito de Morais L. (1990). Biogéographie des poissons d'eau douce de Guyane française (p. 12). Présenté à Symposium International sur l'Evolution des Littoraux des Guyanes et de la Zone Caraïbe Méridionale pendant le Quaternaire, Cayenne (GUF), Cayenne, Orstom.
- Boulenger C., Acou A., Gimenez O., Charrier F., Tremblay J. & Feunteun E. (2016). Factors determining survival of European eels in two unexploited sub-populations. *Freshwater Biology*, 61(6), 947-962. <https://doi.org/10.1111/fwb.12759>
- Bradbury I. R., Hubert S., Higgins B., Bowman S., Paterson I. G., Snelgrove P. V. R., ... Bentzen P. (2011). Evaluating SNP ascertainment bias and its impact on population assignment in Atlantic cod, *Gadus morhua*: SNP GENOTYPING AND APPLICATIONS. *Molecular Ecology Resources*, 11, 218-225. <https://doi.org/10.1111/j.1755-0998.2010.02949.x>
- Brauer C. J., Hammer M. P. & Beheregaray L. B. (2016). Riverscape genomics of a threatened fish across a hydroclimatically heterogeneous river basin. *Molecular Ecology*, 25(20), 5093-5113. <https://doi.org/10.1111/mec.13830>
- Brito P. M., Meunier F. J. & Leal M. E. C. (2007). Origine et diversification de l'ichtyofaune néotropicale : une revue. *Cybium* 31(2), 139-153.
- Brosse S., Montoya-Burgos J. I., Grenouillet G. & Surugue N. (2013). Determinants of fish assemblage structure in Mount Itoupé mountain streams (French Guiana). *Annales de Limnologie - International Journal of Limnology*, 49(1), 43-49. <https://doi.org/10.1051/limn/2013037>

- Brumfield R. T., Beerli P., Nickerson D. A. & Edwards S. V. (2003). The utility of single nucleotide polymorphisms in inferences of population history. *Trends in Ecology & Evolution*, 18(5), 249-256. [https://doi.org/10.1016/S0169-5347\(03\)00018-1](https://doi.org/10.1016/S0169-5347(03)00018-1)
- Caballero A. & Toro M. A. (2002). Analysis of genetic diversity for the management of conserved subdivided populations. *Conservation Genetics*, 3, 289-299.
- Cahill A. E., De Jode A., Dubois S., Bouzaza Z., Aurelle D., Boissin E., ... Chenuil A. (2017). A multispecies approach reveals hot spots and cold spots of diversity and connectivity in invertebrate species with contrasting dispersal modes. *Molecular Ecology*, 26(23), 6563-6577. <https://doi.org/10.1111/mec.14389>
- Camargo M., Giarrizzo T. & Isaac V. (2015). Population and biological parameters of selected fish species from the middle Xingu River, Amazon Basin. *Brazilian Journal of Biology*, 75(3 suppl 1), 112-124. <https://doi.org/10.1590/1519-6984.01914BM>
- Cammen K. M., Schultz T. F., Rosel P. E., Wells R. S. & Read A. J. (2015). Genomewide investigation of adaptation to harmful algal blooms in common bottlenose dolphins (*Tursiops truncatus*). *Molecular Ecology*, 24(18), 4697-4710. <https://doi.org/10.1111/mec.13350>
- Campbell E. O., Davis C. S., Dupuis J. R., Muirhead K. & Sperling F. A. H. (2017). Cross-platform compatibility of *de novo* -aligned SNPs in a nonmodel butterfly genus. *Molecular Ecology Resources*, 17(6), e84-e93. <https://doi.org/10.1111/1755-0998.12695>
- Campbell N. R., Harmon S. A. & Narum S. R. (2015). Genotyping-in-Thousands by sequencing (GT-seq): A cost effective SNP genotyping method based on custom amplicon sequencing. *Molecular Ecology Resources*, 15(4), 855-867. <https://doi.org/10.1111/1755-0998.12357>
- Campbell Grant E. H., Lowe W. H. & Fagan W. F. (2007). Living in the branches: population dynamics and ecological processes in dendritic networks. *Ecology Letters*, 10(2), 165-175. <https://doi.org/10.1111/j.1461-0248.2006.01007.x>
- Cañedo-Argüelles M., Boersma K. S., Bogan M. T., Olden J. D., Phillipsen I., Schriever T. A. & Lytle D. A. (2015). Dispersal strength determines meta-community structure in a dendritic riverine network. *Journal of Biogeography*, 42(4), 778-790. <https://doi.org/10.1111/jbi.12457>
- Cariou M., Duret L. & Charlat S. (2016). How and how much does RAD-seq bias genetic diversity estimates? *BMC Evolutionary Biology*, 16(1). <https://doi.org/10.1186/s12862-016-0791-0>
- Carolsfeld J., Harvey B., Ross C. & Baer A. (2004). *Migratory Fishes of South America: Biology, Fisheries and Conservation Status*. (World Fisheries Trust, The World Bank, International Development Research Centre, Éd.). The World Bank. <https://doi.org/10.1596/1-5525-0114-0>
- Carranza M. L., D'Alessandro E., Saura S. & Loy A. (2012). Connectivity providers for semi-aquatic vertebrates: the case of the endangered otter in Italy. *Landscape Ecology*, 27(2), 281-290. <https://doi.org/10.1007/s10980-011-9682-3>
- Carrara F., Altermatt F., Rodriguez-Iturbe I. & Rinaldo A. (2012). Dendritic connectivity controls biodiversity patterns in experimental metacommunities. *Proceedings of the National Academy of Sciences*, 109(15), 5761-5766. <https://doi.org/10.1073/pnas.1119651109>
- Carrara F., Rinaldo A., Giometto A. & Altermatt F. (2014). Complex Interaction of Dendritic Connectivity and Hierarchical Patch Size on Biodiversity in River-Like Landscapes. *The American Naturalist*, 183(1), 13-25. <https://doi.org/10.1086/674009>
- Carvalho R. A. & Tejerina-Garro F. L. (2015). The influence of environmental variables on the functional structure of headwater stream fish assemblages: a study of two tropical basins in Central Brazil. *Neotropical Ichthyology*, 13(2), 349-360. <https://doi.org/10.1590/1982-0224-20130148>
- Castello L. (2004). A Method to Count Pirarucu *Arapaima gigas*: Fishers, Assessment, and Management. *North American Journal of Fisheries Management*, 24(2), 379-389. <https://doi.org/10.1577/M02-024.1>

- Castello L., McGrath D. G., Hess L. L., Coe M. T., Lefebvre P. A., Petry P., ... Arantes C. C. (2013). The vulnerability of Amazon freshwater ecosystems: Vulnerability of Amazon freshwater ecosystems. *Conservation Letters*, 6(4), 217-229. <https://doi.org/10.1111/conl.12008>
- Castello L., Viana J. P., Watkins G., Pinedo-Vasquez M. & Luzadis V. A. (2009). Lessons from Integrating Fishers of Arapaima in Small-Scale Fisheries Management at the Mamirauá Reserve, Amazon. *Environmental Management*, 43(2), 197-209. <https://doi.org/10.1007/s00267-008-9220-5>
- Catchen J., Hohenlohe P. A., Bassham S., Amores A. & Cresko W. A. (2013). Stacks: an analysis tool set for population genomics. *Molecular Ecology*, 22(11), 3124-3140. <https://doi.org/10.1111/mec.12354>
- Catchen J. M., Amores A., Hohenlohe P. A., Cresko W. & Postlethwait J. H. (2011). *Stacks* : Building and Genotyping Loci De Novo From Short-Read Sequences. *G3:Genes/Genomes/Genetics*, 1(3), 171-182. <https://doi.org/10.1534/g3.111.000240>
- Chaput-Bardy A., Fleurant C., Lemaire C. & Secondi J. (2009). Modelling the effect of in-stream and overland dispersal on gene flow in river networks. *Ecological Modelling*, 220(24), 3589-3598. <https://doi.org/10.1016/j.ecolmodel.2009.06.027>
- Charlesworth D. & Wright S. I. (2001). Breeding systems and genome evolution. *Current Opinion in Genetics & Development*, 11(6), 685-690. [https://doi.org/10.1016/S0959-437X\(00\)00254-9](https://doi.org/10.1016/S0959-437X(00)00254-9)
- Cheung W., Lam V. & Pauly D. (2008) Dynamic bioclimate envelope model to predict climate-induced changes in distribution of marine fishes and invertebrates. In: Modelling Present and Climate-Shifted Distribution of Marine Fishes and Invertebrates. W. Cheung, V. Lam & D. Pauly (eds) Fisheries Centre Research Reports 16:5–50.
- Chust G., Villarino E., Chenuil A., Irigoien X., Bizsel N., Bode A., ... Borja A. (2016). Dispersal similarly shapes both population genetics and community patterns in the marine realm. *Scientific Reports*, 6(1). <https://doi.org/10.1038/srep28730>
- Chutimanitsakun Y., Nipper R. W., Cuesta-Marcos A., Cistué L., Corey A., Filichkina T., ... Hayes P. M. (2011). Construction and application for QTL analysis of a Restriction Site Associated DNA (RAD) linkage map in barley. *BMC Genomics*, 12(1). <https://doi.org/10.1186/1471-2164-12-4>
- Cilleros K., Allard L., Vigouroux R. & Brosse S. (2017). Disentangling spatial and environmental determinants of fish species richness and assemblage structure in Neotropical rainforest streams. *Freshwater Biology*, 62(10), 1707-1720. <https://doi.org/10.1111/fwb.12981>
- Cilleros K., Valentini A., Allard L., Dejean T., Etienne R., Grenouillet G., ... Brosse S. (2018). Unlocking biodiversity and conservation studies in high-diversity environments using environmental DNA (eDNA): A test with Guianese freshwater fishes. *Molecular Ecology Resources*. <https://doi.org/10.1111/1755-0998.12900>
- Cissé A. A., Blanchard F. & Guyader O. (2014). Sustainability of tropical small-scale fisheries: Integrated assessment in French Guiana. *Marine Policy*, 44, 397-405. <https://doi.org/10.1016/j.marpol.2013.10.003>
- Cohen M. A., Beaumont R. C. B. & Thorp N. C. (1999). Movement and activity patterns of the black piranha. *Environmental Biology of Fishes*, 54, 45-52.
- Collen B., Whitton F., Dyer E. E., Baillie J. E. M., Cumberlidge N., Darwall W. R. T., ... Böhm M. (2014). Global patterns of freshwater species diversity, threat and endemism: Global freshwater species congruence. *Global Ecology and Biogeography*, 23(1), 40-51. <https://doi.org/10.1111/geb.12096>
- Comte L., Buisson L., Daufresne M. & Grenouillet G. (2013). Climate-induced changes in the distribution of freshwater fish: observed and predicted trends: *Climate change and freshwater fish*. *Freshwater Biology*, 58(4), 625-639. <https://doi.org/10.1111/fwb.12081>
- Comte L. & Olden J. D. (2018a). Evidence for dispersal syndromes in freshwater fishes. *Proceedings of the Royal Society B: Biological Sciences*, 285(1871), 20172214.

- <https://doi.org/10.1098/rspb.2017.2214>
- Comte L. & Olden J. D. (2018b). Fish dispersal in flowing waters: A synthesis of movement- and genetic-based studies. *Fish and Fisheries*. <https://doi.org/10.1111/faf.12312>
- Cooke S. J., Midwood J. D., Thiem J. D., Klimley P., Lucas M. C., Thorstad E. B., ... Ebner B. C. (2013). Tracking animals in freshwater with electronic tags: past, present and future. *Animal Biotelemetry*, 1(1), 5. <https://doi.org/10.1186/2050-3385-1-5>
- Cooper D. N. & Krawczak M. (1989). Cytosine methylation and the fate of CpG dinucleotides in vertebrate genomes. *Human Genetics*, 83(2), 181–188.
- Costantini F., Ferrario F. & Abbiati M. (2018). Chasing genetic structure in coralligenous reef invertebrates: patterns, criticalities and conservation issues. *Scientific Reports*, 8(1). <https://doi.org/10.1038/s41598-018-24247-9>
- Covain R., Fisch-Muller S., Montoya-Burgos J. I. & Mol J. H. (2012). The Harttiini (Siluriformes, Loricariidae) from the Guianas: a multi-table approach to assess their diversity, evolution, and distribution, *Cybium* 36(1), 115-161.
- Covain R., Fisch-Muller S., Oliveira C., Mol J. H., Montoya-Burgos J. I. & Dray S. (2016). Molecular phylogeny of the highly diversified catfish subfamily Loricariinae (Siluriformes, Loricariidae) reveals incongruences with morphological classification. *Molecular Phylogenetics and Evolution*, 94, 492-517. <https://doi.org/10.1016/j.ympev.2015.10.018>
- Crutsinger G. M. (2016). A community genetics perspective: opportunities for the coming decade. *New Phytologist*, 210(1), 65-70. <https://doi.org/10.1111/nph.13537>
- Csilléry K., Blum M. G. B., Gaggiotti O. E. & François O. (2010). Approximate Bayesian Computation (ABC) in practice. *Trends in Ecology & Evolution*, 25(7), 410-418. <https://doi.org/10.1016/j.tree.2010.04.001>
- Da Fonseca R. R., Albrechtsen A., Themudo G. E., Ramos-Madriral J., Sibbesen J. A., Maretty L., ... Pereira R. J. (2016). Next-generation biology: Sequencing and data analysis approaches for non-model organisms. *Marine Genomics*, 30, 3-13. <https://doi.org/10.1016/j.margen.2016.04.012>
- Davey J. W., Hohenlohe P. A., Etter P. D., Boone J. Q., Catchen J. M. & Blaxter M. L. (2011). Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics*, 12(7), 499-510. <https://doi.org/10.1038/nrg3012>
- Dawson M. N., Hays C. G., Grosberg R. K. & Raimondi P. T. (2014). Dispersal potential and population genetic structure in the marine intertidal of the eastern North Pacific. *Ecological Monographs*, 84(3), 435-456. <https://doi.org/10.1890/13-0871.1>
- De Barba M., Miquel C., Lobréaux S., Quenette P. Y., Swenson J. E. & Taberlet P. (2017). High-throughput microsatellite genotyping in ecology: improved accuracy, efficiency, standardization and success with low-quantity and degraded DNA. *Molecular Ecology Resources*, 17(3), 492-507. <https://doi.org/10.1111/1755-0998.12594>
- De Mérona B., Huguény B., Tejerina-Garro F. L. & Gautheret E. (2008). Diet-morphology relationship in a fish assemblage from a medium-sized river of French Guiana: the effect of species taxonomic proximity. *Aquatic Living Resources*, 21(2), 171-184. <https://doi.org/10.1051/alr:2008032>
- De Mérona B., Tejerina-Garro F. L. & Vigouroux R. (2012). Fish-habitat relationships in French Guiana rivers: a review, 10.
- De Mérona B. & Vigouroux R. (2012). The role of ecological strategies in the colonization success of pelagic fish in a large tropical reservoir (Petit-Saut Reservoir, French Guiana). *Aquatic Living Resources*, 25(1), 41-54. <https://doi.org/10.1051/alr/2011153>
- Deck J., Gaither M. R., Ewing R., Bird C. E., Davies N., Meyer C., ... Crandall E. D. (2017). The Genomic Observatories Metadatabase (GeOME): A new repository for field and sampling event metadata associated with genetic samples. *PLOS Biology*, 15(8), e2002925. <https://doi.org/10.1371/journal.pbio.2002925>

- Dedieu N., Allard L., Vigouroux R., Brosse S. & Céréghino R. (2014). Physical habitat and water chemistry changes induced by logging and gold mining in French Guiana streams. *Knowledge and Management of Aquatic Ecosystems*, (415), 02. <https://doi.org/10.1051/kmae/2014026>
- Delord C., Lassalle G., Oger A., Barloy D., Coutellec M.-A., Delcamp A., ... Petit E. J. (2018). A cost-and-time effective procedure to develop SNP markers for multiple species: A support for community genetics. *Methods in Ecology and Evolution*, 9(9), 1959-1974. <https://doi.org/10.1111/2041-210X.13034>
- Delord C., Lassalle G., Oger A., Barloy D., Coutellec M., Delcamp A., Evanno G., Genthon C., Guichoux E., Le Bail P. Y., Le Quilliec P., Longin G., Lorvelec O., Massot M., Reveillac E., Rinaldo R., Roussel J.-M., Vigouroux R., Launey S., Petit E. J. Data from: A cost-and-time effective procedure to develop SNP markers for multiple species: a support for community genetics. Dryad Digital Repository. <https://doi.org/10.5061/dryad.2b6b43k>
- Del Re A. C. (2013). *compute.es*: Compute Effect Sizes. R package version 0.2-2. <http://cran.r-project.org/web/packages/compute.es>
- Devictor V., Mouillot D., Meynard C., Jiguet F., Thuiller W. & Mouquet N. (2010). Spatial mismatch and congruence between taxonomic, phylogenetic and functional diversity: the need for integrative conservation strategies in a changing world: Spatial mismatch between diversity facets. *Ecology Letters*, no-no. <https://doi.org/10.1111/j.1461-0248.2010.01493.x>
- DeWoody J. & Avise J. C. (2000). Microsatellite variation in marine, freshwater and anadromous fishes compared with other animals. *Journal of Fish Biology*, 56(3), 461-473. <https://doi.org/10.1006/jfbi.1999.1210>
- Do C., Waples R. S., Peel D., Macbeth G. M., Tillett B. J. & Ovenden J. R. (2014). NeEstimator v2: re-implementation of software for the estimation of contemporary effective population size (Ne) from genetic data. *Molecular Ecology Resources*, 14(1), 209-214. <https://doi.org/10.1111/1755-0998.12157>
- Dobzhansky T. (1965). Genetic diversity and fitness. *Genetics today*, 3, 541-552.
- Durrieu G., Maury-Brachet R. & Boudou A. (2005). Goldmining and mercury contamination of the piscivorous fish *Hoplias aimara* in French Guiana (Amazon basin). *Ecotoxicology and Environmental Safety*, 60(3), 315-323. <https://doi.org/10.1016/j.ecoenv.2004.05.004>
- Eble J. A., Toonen R. J. & Bowen B. W. (2009). Endemism and dispersal: comparative phylogeography of three surgeonfishes across the Hawaiian Archipelago. *Marine Biology*, 156(4), 689-698. <https://doi.org/10.1007/s00227-008-1119-4>
- Eklblom R. & Galindo J. (2011). Applications of next generation sequencing in molecular ecology of non-model organisms. *Heredity*, 107(1), 1-15. <https://doi.org/10.1038/hdy.2010.152>
- Elith J., Graham C., Anderson R., Dudík M., Ferrier S., Guisan A., ...Zimmermann N. (2006). Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, 29(2), 129-151. <https://doi.org/10.1111/j.2006.0906-7590.04596.x>
- Ellegren H. (2004). Microsatellites: simple sequences with complex evolution. *Nature Reviews Genetics*, 5(6), 435-445. <https://doi.org/10.1038/nrg1348>
- Ellegren H. & Galtier N. (2016). Determinants of genetic diversity. *Nature Reviews Genetics*, 17(7), 422-433. <https://doi.org/10.1038/nrg.2016.58>
- Emerson K. J., Merz C. R., Catchen J. M., Hohenlohe P. A., Cresko W. A., Bradshaw W. E. & Holzapfel C. M. (2010). Resolving postglacial phylogeography using high-throughput sequencing. *Proceedings of the National Academy of Sciences*, 107(37), 16196-16200. <https://doi.org/10.1073/pnas.1006538107>
- Epps C. W. & Keyghobadi N. (2015). Landscape genetics in a changing world: disentangling historical and contemporary influences and inferring change. *Molecular Ecology*, 24(24), 6021-6040. <https://doi.org/10.1111/mec.13454>

- Estoup A., & Angers B. (1998). Microsatellites and minisatellites for molecular ecology: theoretical and empirical considerations. In *Advances in molecular ecology* (Carvalho GR, p. 55-86). IOS press, Amsterdam.
- Evanno G., Castella E., Antoine C., Paillat G. & Goudet J. (2009). Parallel changes in genetic diversity and species diversity following a natural disturbance. *Molecular Ecology*, 18(6), 1137-1144. <https://doi.org/10.1111/j.1365-294X.2009.04102.x>
- Fagan W. F. (2002). Connectivity, Fragmentation, and Extinction Risk in Dendritic Metapopulations. *Ecology*, 83(12), 3243-3249. [https://doi.org/10.1890/0012-9658\(2002\)083\[3243:CFAERI\]2.0.CO;2](https://doi.org/10.1890/0012-9658(2002)083[3243:CFAERI]2.0.CO;2)
- Faivre-Rampant P., Zaina G., Jorge V., Giacomello S., Segura V., Scalabrin S., ... Morgante M. (2016). New resources for genetic studies in *Populus nigra* : genome-wide SNP discovery and development of a 12k Infinium array. *Molecular Ecology Resources*, 16(4), 1023-1036. <https://doi.org/10.1111/1755-0998.12513>
- Fazzi-Gomes P., Guerreiro S., Palheta G. D. A., Melo N. F. A. C., de Santos S. & Hamoy I. (2017). High genetic diversity and connectivity in *Colossoma macropomum* in the Amazon basin revealed by microsatellite markers. *Genetics and Molecular Biology*, 40(1), 142-146. <https://doi.org/10.1590/1678-4685-gmb-2015-0222>
- Figge F. (2004). Bio-folio: applying portfolio theory to biodiversity. *Biodiversity and Conservation*, 13(4), 827-849. <https://doi.org/10.1023/B:BIOC.0000011729.93889.34>
- Finn D. S. & Poff N. L. (2011). Examining spatial concordance of genetic and species diversity patterns to evaluate the role of dispersal limitation in structuring headwater metacommunities. *Journal of the North American Benthological Society*, 30(1), 273-283. <https://doi.org/10.1899/10-035.1>
- Fisher F. R. S. (1934). The effects of methods of ascertainment upon the estimation of frequencies. *Annals of human genetics*, 6(1), 13-25.
- Fleury M. (1998). Les populations du Haut-Maroni et le projet de Parc national de la Guyane. *Journal d'agriculture traditionnelle et de botanique appliquée*, 40(1-2), 577-610. <https://doi.org/10.3406/jatba.1998.3693>
- Fleury M., Alupki T., Opoya A. & Aloïké W. (2016). Les Wayana de Guyane française sur les traces de leur histoire: Cartographie participative sur le Litani (Aletani) et mémoire orale. *Revue d'ethnoécologie*, (9). <https://doi.org/10.4000/ethnoecologie.2711>
- Fourte L., Paz-Vinas I., Loot G., Prunier J. G. & Blanchet S. (2016). Lessons from the fish: a multi-species analysis reveals common processes underlying similar species-genetic diversity correlations. *Freshwater Biology*, 61(11), 1830–1845. <https://doi.org/10.1111/fwb.12826>
- Frankel O. H. (1974) Genetic conservation: our evolutionary responsibility. *Genetics*, 78, 53-65.
- Frankham R. (1995a). Conservation Genetics. *Annual Review of Genetics*, (29), 305-327.
- Frankham R. (1995b). Effective population size/adult population size ratios in wildlife: a review. *Genetical Research*, 66(02), 95. <https://doi.org/10.1017/S0016672300034455>
- Frankham R. (1996). Relationship of Genetic Variation to Population Size in Wildlife. *Conservation Biology*, 10(6), 1500-1508. <https://doi.org/10.1046/j.1523-1739.1996.10061500.x>
- Frankham R. (2005). Genetics and extinction. *Biological Conservation*, 126(2), 131-140. <https://doi.org/10.1016/j.biocon.2005.05.002>
- Fréry N., Maury-Brachet R., Maillot E., Deheeger M., de Mérona B. & Boudou A. (2001). Gold-Mining Activities and Mercury Contamination of Native Amerindian Communities in French Guiana: Key Role of Fish in Dietary Uptake. *Environmental Health Perspectives*, 109(5), 8.
- Frey D., Arrigo N., Granereau G., Sarr A., Felber F. & Kozłowski G. (2016). Parallel declines in species and genetic diversity driven by anthropogenic disturbance: a multispecies approach in a French Atlantic dune system. *Evolutionary Applications*, 9(3), 479-488. <https://doi.org/10.1111/eva.12351>
- Fromentin J.-M. & Powers J. E. (2005) Atlantic Bluefin Tuna: Population Dynamics, Ecology, Fisheries and Management. *Fish and Fisheries* 6(4): 281-306.

- Futschik A. & Schlötterer C. (2010). The Next Generation of Molecular Markers From Massively Parallel Sequencing of Pooled DNA Samples. *Genetics*, 186(1), 207-218.
<https://doi.org/10.1534/genetics.110.114397>
- Galindo B. A., Ferreira D. G., Almeida F. S., Carlsson J. & Sofia S. H. (2015). Isolation and characterization of 13 polymorphic microsatellite loci in *Hypostomus ancistroides* (Teleostei, Loricariidae) and cross-amplification in related species: microsatellites in *hypostomus ancistroides*. *Journal of Fish Biology*, 86(6), 1860-1866. <https://doi.org/10.1111/jfb.12675>
- Galtier N., Nabholz B., Glémin S. & Hurst G. D. D. (2009). Mitochondrial DNA as a marker of molecular diversity: a reappraisal. *Molecular Ecology*, 18(22), 4541-4550. <https://doi.org/10.1111/j.1365-294X.2009.04380.x>
- Garvin M. R., Saitoh K. & Gharrett A. J. (2010). Application of single nucleotide polymorphisms to non-model species: a technical review: INVITED TECHNICAL REVIEW. *Molecular Ecology Resources*, 10(6), 915-934. <https://doi.org/10.1111/j.1755-0998.2010.02891.x>
- Gaston K. J. (2000). Global patterns in biodiversity. *Nature*, 405(6783), 220-227.
<https://doi.org/10.1038/35012228>
- Gentili J. (1949) Foundation of Australian bird geography. *Emu*, 49, 85-130.
- Géry J. (1972). Poissons characoïdes des Guyanes. I. Généralités. II. Famille des Serrasalminidae. Zoologische Verhandelingen (Leiden) No. 122: 1-250, Pls. 1-16.
- Gilbert K. J. & Whitlock M. C. (2015). Evaluating methods for estimating local effective population size with and without migration: Estimating N_e in the presence of migration. *Evolution*, 69(8), 2154-2166. <https://doi.org/10.1111/evo.12713>
- Giron Y., Diallo-Bourguignon G. & Le Bail P.-Y. (2001). *Etude de la faisabilité d'une pisciculture vivrière à Papaïchton, Haut-Maroni, Guyane française* (Rapport Final Cofrepeche) (p. 100). Plouzané: Cofrepeche, Centre Ifremer de Brest.
- Gomez-Uchida D., Knight T. W. & Ruzzante D. E. (2009). Interaction of landscape and life history attributes on genetic diversity, neutral divergence and gene flow in a pristine community of salmonids. *Molecular Ecology*, 18(23), 4854-4869. <https://doi.org/10.1111/j.1365-294X.2009.04409.x>
- Gomez-Uchida D., Palstra F. P., Knight T. W. & Ruzzante D. E. (2013). Contemporary effective population and metapopulation size (N_e and meta- N_e): comparison among three salmonids inhabiting a fragmented system and differing in gene flow and its asymmetries. *Ecology and Evolution*, 3(3), 569-580. <https://doi.org/10.1002/ece3.485>
- Gordon T. A. C., Harding H. R., Clever F. K., Davidson I. K., Davison W., Montgomery D. W., ... Santos E. M. (2018). Fishes in a changing world: learning from the past to promote sustainability of fish populations. *Journal of Fish Biology*, 92(3), 804-827. <https://doi.org/10.1111/jfb.13546>
- Goudet, J. (1995). FSTAT (Version 1.2): A computer program to calculate F-Statistics. *Journal of Heredity*, 86(6), 485-486.
- Goudet J. (2005). *hierfstat*, a package for R to compute and test hierarchical F-statistics. *Molecular Ecology Notes*, 5(1), 184-186. <https://doi.org/10.1111/j.1471-8286.2004.00828.x>
- Goudet J. & Jombart T. (2017). *hierfstat*: Estimation and Tests of Hierarchical F-Statistics. R package version 0.04-22. <http://www.r-project.org>, <http://github.com/jgx65/hierfstat>.
- Goulding M. (1980) *The Fishes and the Forest*. Los Angeles and London: University of California Press. 279 pp.
- Grenouillet G., Buisson L., Casajus N. & Lek S. (2011). Ensemble modelling of species distribution: the effects of geographical and environmental ranges. *Ecography*, 34(1), 9-17.
<https://doi.org/10.1111/j.1600-0587.2010.06152.x>
- Guillaume F., & Rougemont, J. (2006). Nemo: an evolutionary and population genetics programming framework. *Bioinformatics*, 22(20), 2556-2557. <https://doi.org/10.1093/bioinformatics/btl415>

- Guisan A. & Zimmermann N. E. (2000). Predictive habitat distribution models in ecology. *Ecological Modelling*, 135(2-3), 147-186. [https://doi.org/10.1016/S0304-3800\(00\)00354-9](https://doi.org/10.1016/S0304-3800(00)00354-9)
- Gutiérrez J. P., Royo L. J., Álvarez I. & Goyache F. (2005). MolKin v2.0: A Computer Program for Genetic Analysis of Populations Using Molecular Coancestry Information. *Journal of Heredity*, 96(6), 718-721. <https://doi.org/10.1093/jhered/esi118>
- Haffer J. R. (1982). General aspects of the refuge theory. In *Biological diversification in the Tropics* (New-York, Columbia University Press, p. 6-26).
- Haffer J. R. (1997). Alternative models of vertebrate speciation in Amazonia: an overview. *Biodiversity and Conservation*, 6, 451-476.
- Hansen E. (2009). Réglementation de la chasse en Guyane, une situation particulière, 3.
- Hasler A. D. (1947). Eutrophication of lakes by domestic drainage. *Ecology*, 28, 383-395.
- Hauser L., Adcock G. J., Smith P. J., Bernal Ramirez J. H. & Carvalho G. R. (2002). Loss of microsatellite diversity and low effective population size in an overexploited population of New Zealand snapper (*Pagrus auratus*). *Proceedings of the National Academy of Sciences*, 99(18), 11742-11747. <https://doi.org/10.1073/pnas.172242899>
- Hebert P. D. N., Cywinska A., Ball S. L. & DeWaard J. R. (2003). Biological identifications through DNA barcodes. *Proceedings of the Royal Society B: Biological Sciences*, 270(1512), 313-321. <https://doi.org/10.1098/rspb.2002.2218>
- Helyar S. J., Hemmer-Hansen J., Bekkevold D., Taylor M. I., Ogden R., Limborg M. T., ... Nielsen E. E. (2011). Application of SNPs for population genetics of nonmodel organisms: new opportunities and challenges: ANALYTICAL APPROACHES. *Molecular Ecology Resources*, 11, 123-136. <https://doi.org/10.1111/j.1755-0998.2010.02943.x>
- Herrera S., Reyes-Herrera P. H. & Shank T. M. (2015). Predicting RAD-seq Marker Numbers across the Eukaryotic Tree of Life. *Genome Biology and Evolution*, 7(12), 3207-3225. <https://doi.org/10.1093/gbe/evv210>
- Hilsdorf A. W. S. & Hallerman E. M. (2017). *Genetic Resources of Neotropical Fishes*. Cham: Springer International Publishing. <https://doi.org/10.1007/978-3-319-55838-7>
- Hindar K., Tufto J., Sættem L. & Balstad T. (2004). Conservation of genetic variation in harvested salmon populations. *ICES Journal of Marine Science*, 61(8), 1389-1397. <https://doi.org/10.1016/j.icesjms.2004.08.011>
- Hoban S., Arntzen J. W., Bertorelle G., Bryja J., Fernandes M., Frith K., ... Bruford M. W. (2013a). Conservation Genetic Resources for Effective Species Survival (ConGRESS): Bridging the divide between conservation research and practice. *Journal for Nature Conservation*, 21(6), 433-437. <https://doi.org/10.1016/j.jnc.2013.07.005>
- Hoban S., Arntzen J. W., Bruford M. W., Godoy J. A., Rus Hoelzel A., Segelbacher G., ... Bertorelle G. (2014). Comparative evaluation of potential indicators and temporal sampling protocols for monitoring genetic erosion. *Evolutionary Applications*, 7(9), 984-998. <https://doi.org/10.1111/eva.12197>
- Hoban S., Bertorelle G. & Gaggiotti O. E. (2012). Computer simulations: tools for population and evolutionary genetics. *Nature Reviews Genetics*, 13(2), 110-122. <https://doi.org/10.1038/nrg3130>
- Hoban S., Haufler H. C., Pérez-Espona S., Arntzen J. W., Bertorelle G., Bryja J., ... Bruford M. W. (2013b). Bringing genetic diversity to the forefront of conservation policy and management. *Conservation Genetics Resources*, 5(2), 593-598. <https://doi.org/10.1007/s12686-013-9859-y>
- Hohenlohe P. A., Bassham S., Etter P. D., Stiffler N., Johnson E. A. & Cresko W. A. (2010). Population Genomics of Parallel Adaptation in Threespine Stickleback using Sequenced RAD Tags. *PLoS Genetics*, 6(2), e1000862. <https://doi.org/10.1371/journal.pgen.1000862>
- Horne J. B. (2014). Emerging patterns and emerging challenges of comparative phylogeography. *Frontiers of biogeography*, 6(4), 166-168.

- Hrbek T., Crossa, M., & Farias, I. (2007). Conservation strategies for *Arapaima gigas* (Schinz, 1822) and the Amazonian várzea ecosystem. *Brazilian Journal of Biology*, 67(4 suppl), 909-917. <https://doi.org/10.1590/S1519-69842007000500015>
- Huang H. & Knowles L. L. (2016). Unforeseen Consequences of Excluding Missing Data from Next-Generation Sequences: Simulation Study of RAD Sequences. *Systematic Biology*, 65(3), 357-365. <https://doi.org/10.1093/sysbio/syu046>
- Hughes J. M., Huey J. A. & Schmidt D. J. (2013). Is realised connectivity among populations of aquatic fauna predictable from potential connectivity?: Connectivity among populations of aquatic fauna. *Freshwater Biology*, 58(5), 951-966. <https://doi.org/10.1111/fwb.12099>
- Hughes J. M., Schmidt D. J. & Finn D. S. (2009). Genes in Streams: Using DNA to Understand the Movement of Freshwater Fauna and Their Riverine Habitat. *BioScience*, 59(7), 573-583. <https://doi.org/10.1525/bio.2009.59.7.8>
- Hung C.-M., Drovetski S. V. & Zink R. M. (2016). Matching loci surveyed to questions asked in phylogeography. *Proceedings of the Royal Society B: Biological Sciences*, 283(1826), 20152340. <https://doi.org/10.1098/rspb.2015.2340>
- Hung C.-M., Drovetski S. V. & Zink R. M. (2017). The roles of ecology, behaviour and effective population size in the evolution of a community. *Molecular Ecology*, 26(14), 3775-3784. <https://doi.org/10.1111/mec.14152>
- Hurault J. (1965), La vie matérielle des Noirs réfugiés Boni et des Indiens Wayana du Haut-Maroni (Guyane Française) : agriculture, économie et habitat, ORSTOM, Paris, 154 p.
- Hutchings J. A. & Baum J. K. (2005). Measuring marine fish biodiversity: temporal changes in abundance, life history and demography. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 360(1454), 315-338. <https://doi.org/10.1098/rstb.2004.1586>
- Iervolino F., de Resende E. K. & Hilsdorf A. W. S. (2010). The lack of genetic differentiation of pacu (*Piaractus mesopotamicus*) populations in the Upper-Paraguay Basin revealed by the mitochondrial DNA D-loop region: Implications for fishery management. *Fisheries Research*, 101(1-2), 27-31. <https://doi.org/10.1016/j.fishres.2009.09.003>
- Jégu M., Keith P. & Belmont-Jégu E. (2002). Une nouvelle espèce de Tometes (Teleostei : Characidae : Serrasalminae) du bouclier guyanais, *Tometes lebaili* n. Sp. *Bulletin Français de la Pêche et de la Pisciculture*, 364, 23-48. <https://doi.org/10.1051/kmae:2002002>
- Jégu M., & Keith P. (2005). Threatened fishes of the world: *Tometes lebaili* (Jégu, Keith & Belmont-Jégu 2002) (Characidae: Serrasalminae). *Environmental Biology of Fishes*, 72(4), 378-378. <https://doi.org/10.1007/s10641-004-4126-4>
- Jenkins T. L., Castilho R. & Stevens J. R. (2018). Meta-analysis of northeast Atlantic marine taxa shows contrasting phylogeographic patterns following post-LGM expansions. *PeerJ*, 6, e5684. <https://doi.org/10.7717/peerj.5684>
- Jombart T. (2008). *adegenet*: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, 24(11), 1403-1405. <https://doi.org/10.1093/bioinformatics/btn129>
- Jones O. R., & Wang J. (2010). COLONY: a program for parentage and sibship inference from multilocus genotype data. *Molecular Ecology Resources*, 10(3), 551-555. <https://doi.org/10.1111/j.1755-0998.2009.02787.x>
- Jucker T., Wintle B., Shackelford G., Bocquillon P., Geffert J. L., Kasoar T., ... Mukherjee N. (2018). Ten-year assessment of the 100 priority questions for global biodiversity conservation: Knowledge Gaps. *Conservation Biology*, 32(6), 1457-1463. <https://doi.org/10.1111/cobi.13159>
- Kamvar Z. N., Tabima J. F. & Grünwald N. J. (2014). *poppr*: An R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ*, 2, e281. <https://doi.org/10.7717/peerj.281>

- Kantoussan J., Laë R. & Tine M. (2018). Review of the Fisheries Indicators for Monitoring the Impacts of Fishing on Fish Communities. *Reviews in Fisheries Science & Aquaculture*, 26(4), 460-478. <https://doi.org/10.1080/23308249.2018.1458282>
- Karr J. R. 1981. Assessment of biotic integrity using fish communities. *Fisheries*, 6, 21-27.
- Keenan K., McGinnity P., Cross T. F., Crozier W. W. & Prodöhl P. A. (2013). *diveRsity*: An R package for the estimation and exploration of population genetics parameters and their associated errors. *Methods in Ecology and Evolution*, 4(8), 782-788. <https://doi.org/10.1111/2041-210X.12067>
- Keith P., Le Bail P.-Y. & Planquette P. (2000). *Atlas des poissons d'eau douce de Guyane - Tome 2, Fascicule I*. Paris.
- Keith P., Persat H., Feunteun E. & Allardi J. (coord.), *Les poissons d'eau douce de France*, 2011, co-édition Biotope, Mèze-Publications Scientifiques du Muséum national d'Histoire naturelle, Paris (Collection Inventaires et Biodiversité), 552 p.
- Keller I., Bensasson D. & Nichols R. A. (2007). Transition-Transversion Bias Is Not Universal: A Counter Example from Grasshopper Pseudogenes. *PLoS Genetics*, 3(2), e22. <https://doi.org/10.1371/journal.pgen.0030022>
- Kelly R. P. & Palumbi S. R. (2010) Genetic structure among 50 species of the northeastern Pacific rocky intertidal community. *PLoS One*, 5(1), e8594. doi: 10.1371/journal.pone.0008594.
- Khimoun A., Eraud C., Ollivier A., Arnoux E., Rocheteau V., Bely M., ... Garnier S. (2016). Habitat specialization predicts genetic response to fragmentation in tropical birds. *Molecular Ecology*, 25(16), 3831-3844. <https://doi.org/10.1111/mec.13733>
- Kimura M. & Ohta T. (1978). Stepwise mutation model and distribution of allelic frequencies in a finite population. *Proceedings of the National Academy of Sciences*, 75(6), 2868-2872. <https://doi.org/10.1073/pnas.75.6.2868>
- Kingman J.F.C. (1982). The coalescent. *Stochastic Process. Appl.* 13 (1982), no.3,235-248.
- Kullander S. O. (2003). Family Cichlidae. In *Checklist of the Freshwater fishes of South and Central America* (Edipucrs). Consulté à l'adresse <http://www.bioone.org/doi/abs/10.1643/OT-04-142>
- Kunin W. E. (1998). Biodiversity at the edge: A test of the importance of spatial « mass effects » in the Rothamsted Park Grass experiments. *Proceedings of the National Academy of Sciences*, 95(1), 207-212. <https://doi.org/10.1073/pnas.95.1.207>
- Kuparinen A. & Merilä J. (2007) Detecting and managing fisheries-induced evolution. *Trends in Ecology and Evolution*, 22(12), 652-659. doi:10.1016/j.tree.2007.08.011.
- Labonne J., Ravigné V., Parisi B. & Gaucherel C. (2008). Linking dendritic network structures to population demogenetics: The downside of connectivity. *Oikos*, 117(10), 1479-1490. <https://doi.org/10.1111/j.0030-1299.2008.16976.x>
- Lachance J. & Tishkoff S. A. (2013). SNP ascertainment bias in population genetic analyses: Why it is important, and how to correct it: Prospects & Overviews. *BioEssays*, 35(9), 780-786. <https://doi.org/10.1002/bies.201300014>
- Lamy T., Laroche F., David P., Massol F. & Jarne P. (2017). The contribution of species-genetic diversity correlations to the understanding of community assembly rules. *Oikos*, 126(6), 759-771. <https://doi.org/10.1111/oik.03997>
- Laperche V., Hellal J., Maury-Brachet R., Joseph B., Laporte P., Breeze D. & Blanchard F. (2014). Regional distribution of mercury in sediments of the main rivers of French Guiana (Amazonian basin). *SpringerPlus*, 3(1), 322. <https://doi.org/10.1186/2193-1801-3-322>
- Le Bail P.-Y., Covain R., Jégu M., Fisch-Muller S., Vigouroux R. & Keith P. (2012). Updated checklist of the freshwater and estuarine fishes of French Guiana, *Cybium*, 36(1), 293-319.
- Le Bail P.-Y., Keith P. & Planquette P. (2000). *Atlas des poissons d'eau douce de Guyane - Tome 2, Fascicule II*. Paris.

- Lê S., Josse J. & Husson F. (2008). *FactoMineR*: An R Package for Multivariate Analysis. *Journal of Statistical Software*, 25(1), pp. 1-18.
- Lean J., Hammer M. P., Unmack P. J., Adams M. & Beheregaray L. B. (2017). Landscape genetics informs mesohabitat preference and conservation priorities for a surrogate indicator species in a highly fragmented river system. *Heredity*, 118(4), 374-384. <https://doi.org/10.1038/hdy.2016.111>
- Lebreton J. D., Burnham K. P., Clobert J. & Anderson D. R. (1992). Modeling survival and testing biological hypotheses using marked animals: a unified approach with case studies, *Ecological Monographs*, 62, 67-118.
- Lebreton J., Nichols J. D., Barker R. J., Pradel R. & Spendelov J. A. (2009) Chapter 3, Modeling Individual Animal Histories with Multistate Capture–Recapture Models. In: *Advances in Ecological Research*. (Ed. Hal Caswell), pp. 87-173. Academic Press.
- Lecomte F., Boujard T., Meunier F. J., Renno J. F. & Rojas-Beltran R. (1993) The growth of *Myleus rhomboidalis* (Cuvier, 1817) in two rivers of French Guiana. *Rev. Evol. (Terre Vie)*, 48, 431-444.
- Lefkovich L.P. (1965) The study of population growth in organisms grouped by stages. *Biometrics*, 21(1), 1-18.
- Lemopoulos A. & Covain R. (2018). Biogeography of the freshwater fishes of the Guianas using a partitioned parsimony analysis of endemism with reappraisal of ecoregional boundaries. *Cladistics*, 35(1), 106-124. <https://doi.org/10.1111/cla.12341>
- Lenth R. (2018). *emmeans*: Estimated Marginal Means, aka Least-Squares Means. R package version 1.2.2. <https://CRAN.R-project.org/package=emmeans>
- Lepoittevin C., Bodénès C., Chancerel E., Villate L., Lang T., Lesur I., ... Kremer A. (2015). Single-nucleotide polymorphism discovery and validation in high-density SNP array for genetic analysis in European white oaks. *Molecular Ecology Resources*, 15(6), 1446-1459. <https://doi.org/10.1111/1755-0998.12407>
- Leslie P. H. (1945) On the use of matrices in certain population mathematics. *Biometrika*, 33, 183-212.
- Lévêque C., Oberdorff T., Paugy D., Stiassny M. L. J. & Tedesco P. A. (2008). Global diversity of fish (Pisces) in freshwater. In E. V. Balian, C. Lévêque, H. Segers, & K. Martens (Éd.), *Freshwater Animal Diversity Assessment* (Vol. 198, p. 545-567). Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-1-4020-8259-7_53
- Lézy E. (2000), *Guyane, Guyanes : une géographie sauvage de l'Orénoque à l'Amazone*, Belin, 352 pp.
- Li B. & Kimmel M. (2013). Factors Influencing Ascertainment Bias of Microsatellite Allele Sizes: Impact on Estimates of Mutation Rates. *Genetics*, 195(2), 563-572. <https://doi.org/10.1534/genetics.113.154161>
- Li H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *ArXiv:1303.3997 [q-Bio]*. Consulté à l'adresse <http://arxiv.org/abs/1303.3997>
- Li H., Handsaker B., Wysoker A., Fennell T., Ruan J., Homer N., ... 1000 Genome Project Data Processing Subgroup. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25(16), 2078-2079. <https://doi.org/10.1093/bioinformatics/btp352>
- Liggins L., Tremblay E. A., Possingham H. P. & Riginos C. (2016). Seascape features, rather than dispersal traits, predict spatial genetic patterns in co-distributed reef fishes. *Journal of Biogeography*, 43(2), 256-267. <https://doi.org/10.1111/jbi.12647>
- Lima de Melo T., Tejerina-Garro F. L. & Melo C. E. de. (2009). Influence of environmental parameters on fish assemblage of a neotropical river with a flood pulse regime, Central Brazil. *Neotropical Ichthyology*, 7(3), 421-428. <https://doi.org/10.1590/S1679-62252009000300009>
- Lino A., Fonseca C., Rojas D., Fischer E. & Ramos Pereira M. J. (2018). A meta-analysis of the effects of habitat loss and fragmentation on genetic diversity in mammals. *Mammalian Biology*, 94, 69-76. <https://doi.org/10.1016/j.mambio.2018.09.006>

- Liu C., Comte L. & Olden J. D. (2016). Heads you win, tails you lose: Life-history traits predict invasion and extinction risk of the world's freshwater fishes: TRAITS PREDICT GLOBAL FRESHWATER FISH INVASION AND EXTINCTION. *Aquatic Conservation: Marine and Freshwater Ecosystems*, 27(4), 773-779. <https://doi.org/10.1002/aqc.2740>
- Longin G. (2016). *La pêche chez les Wayana, Teko et Aluku en 2014 sur le Haut-Maroni (Guyane française), complémentarité des enquêtes halieutiques et des cartes cognitives*. Parc Amazonien de Guyane. <https://dumas.ccsd.cnrs.fr/dumas-01552857/document>.
- Lord C., Fermon Y., Meunier F. J., Jégu M. & Keith P. (2007). Croissance et longévité du Watau yaïke, *Tometes lebaili* (Serrasalminae), dans le bassin du Haut Maroni (Guyane française). Résultats préliminaires, *Cybiu*, 31(3), 359-367.
- Lowe A. J., Boshier D., Ward M., Bacles C. F. E. & Navarro C. (2005). Genetic resource impacts of habitat loss and degradation; reconciling empirical evidence and predicted theory for neotropical trees. *Heredity*, 95(4), 255-273. <https://doi.org/10.1038/sj.hdy.6800725>
- Lowe-McConnell R. (1987). *Ecological studies in tropical fish communities*. Cambridge [Cambridgeshire] ; New York: Cambridge University Press.
- Luikart G., England P. R., Tallmon D., Jordan S. & Taberlet P. (2003). The power and promise of population genomics: from genotyping to genome typing. *Nature Reviews Genetics*, 4(12), 981-994. <https://doi.org/10.1038/nrg1226>
- Lujan N. K., Winemiller K. O. & Armbruster J. W. (2012). Trophic diversity in the evolution and community assembly of loricariid catfishes. *BMC Evolutionary Biology*, 12(1), 124. <https://doi.org/10.1186/1471-2148-12-124>
- Lundberg J. G., Marshall L. G., Guerrero J., Horton B., Malabarba M. C. S. & Wesselingh F. (1998). The Stage for Neotropical fish diversification: A history. In *Phylogeny and Classification of Neotropical Fishes*.
- Lyell C. (1832). *Principles of geology* (Vol. 2). London: John Murray.
- Machado C. A., Kliman R. M., Markert J. A. & Hey J. (2002). Inferring the History of Speciation from Multilocus DNA Sequence Data: The Case of *Drosophila pseudoobscura* and close relatives. *Molecular Biology and Evolution*, 19(4), 472-488. <https://doi.org/10.1093/oxfordjournals.molbev.a004103>
- Malomane D. K., Reimer C., Weigend S., Weigend A., Sharifi A. R. & Simianer H. (2018). Efficiency of different strategies to mitigate ascertainment bias when using SNP panels in diversity studies. *BMC Genomics*, 19(1). <https://doi.org/10.1186/s12864-017-4416-9>
- Manel S., Schwartz M. K., Luikart G. & Taberlet P. (2003). Landscape genetics: combining landscape ecology and population genetics. *Trends in Ecology & Evolution*, 18(4), 189-197. [https://doi.org/10.1016/S0169-5347\(03\)00008-9](https://doi.org/10.1016/S0169-5347(03)00008-9)
- Martin E. (2014). *La proie, l'animal personne ou l'ennemi des hommes: Nommer, classer, penser et se nourrir d'animaux sur le haut-Maroni des Wayana (Guyane française)*. Paris Ouest Nanterre La Défense.
- Martínez J. G., Caballero-Gaitán S. J., Sánchez-Bernal D., de Assunção E. N., Astolfi-Filho S., Hrbek T., & Farias I. P. (2016). De novo SNP markers development for the Neotropical gilded catfish *Brachyplatystoma rousseauxii* using next-generation sequencing-based genotyping. *Conservation Genetics Resources*, 8(4), 415-418. <https://doi.org/10.1007/s12686-016-0584-1>
- Martínez J. G., Machado V. N., Caballero-Gaitán S. J., Santos M. da C. F., Alencar R. M., Escobar L. M. D., ... Farias I. P. (2017). SNPs markers for the heavily overfished tambaqui *Colossoma macropomum*, a Neotropical fish, using next-generation sequencing-based de novo genotyping. *Conservation Genetics Resources*, 9(1), 29-33. <https://doi.org/10.1007/s12686-016-0610-3>

- Marty L., Dieckmann U. & Ernande B. (2015). Fisheries-induced neutral and adaptive evolution in exploited fish populations and consequences for their adaptive potential. *Evolutionary Applications*, 8(1), 47-63. <https://doi.org/10.1111/eva.12220>
- Mastretta-Yanes A., Arrigo N., Alvarez N., Jorgensen T. H., Piñero D. & Emerson B. C. (2015). Restriction site-associated DNA sequencing, genotyping error estimation and *de novo* assembly optimization for population genetic inference. *Molecular Ecology Resources*, 15(1), 28-41. <https://doi.org/10.1111/1755-0998.12291>
- Mathieu-Bégné E., Loot G., Chevalier M., Paz-Vinas I. & Blanchet S. (2018). Demographic and genetic collapses in spatially structured populations: insights from a long-term survey in wild fish metapopulations. *Oikos*. <https://doi.org/10.1111/oik.05511>
- Maury-Brachet R., Durrieu G., Dominique Y. & Boudou A. (2006). Mercury distribution in fish organs and food regimes: Significant relationships from twelve species collected in French Guiana (Amazonian basin). *Science of The Total Environment*, 368(1), 262-270. <https://doi.org/10.1016/j.scitotenv.2005.09.077>
- McAllister D. E. (1991) What is the status of the world's coral reef fishes? *Sea Wind* 5, 14–18.
- McRae BH (2006) Isolation by resistance. *Evolution*, 60, 1551–1561.
- Meffe G. K. & Vrijenhoek R. C. (1988). Conservation Genetics in the Management of Desert Fishes. *Conservation Biology*, 2(2), 157-169. <https://doi.org/10.1111/j.1523-1739.1988.tb00167.x>
- Meirmans P. G., Goudet J., IntraBioDiv Consortium & Gaggiotti O. E. (2011). Ecology and life history affect different aspects of the population structure of 27 high-alpine plants: LIFE HISTORY, ECOLOGY AND GENETIC STRUCTURE. *Molecular Ecology*, 20(15), 3144-3155. <https://doi.org/10.1111/j.1365-294X.2011.05164.x>
- Mendisco F., Keyser C., Hollard C., Seldes V., Nielsen A. E., Crubézy E. & Ludes B. (2011). Application of the iPLEXTM Gold SNP genotyping method for the analysis of Amerindian ancient DNA samples: Benefits for ancient population studies. *ELECTROPHORESIS*, 32(3-4), 386-393. <https://doi.org/10.1002/elps.201000483>
- Mérigoux S., Dolédec S. & Statzner B. (2001). Species traits in relation to habitat variability and state: neotropical juvenile fish in floodplain creeks. *Freshwater Biology*, 46(9), 1251-1267. <https://doi.org/10.1046/j.1365-2427.2001.00744.x>
- Mérigoux S. M., Ponton D. & De Mérona B. (1998). Fish richness and species-habitat relationships in two coastal streams of French Guiana, South America. *Environmental Biology of Fishes*, 51, 25-39.
- Meunier F. J. (coord.), (2004). *Piranhas enivrés, des poissons et des hommes en Guyane*. SFI/RMN éditeurs, 128 pp.
- Meunier F. J., Rojas-Beltran R., Boujard T. & Lecomte F. (1994). Rythmes saisonniers de la croissance chez quelques Téléostéens de Guyane française. *Revue d'hydrobiologie tropicale*, 27(4), 423-440.
- Meunier F. J., Keith P. & Le Bail P.-Y. (1998). Conserver les spécificités de l'ichtyofaune guyanaise. *Journal d'agriculture traditionnelle et de botanique appliquée*, 40(1), 239-259. <https://doi.org/10.3406/jatba.1998.3672>
- Miraldo A., Li S., Borregaard M. K., Florez-Rodriguez A., Gopalakrishnan S., Rizvanovic M., ... Noguez-Bravo D. (2016). An Anthropocene map of genetic diversity. *Science*, 353(6307), 1532-1535. <https://doi.org/10.1126/science.aaf4381>
- Mol J. H., Vari R. P., Covain R., Willink P. W. & Fisch-Muller S. (2012). Annotated checklist of the freshwater fishes of Suriname, *Cybium*, 36(1), 263-292.
- Montorio L., Evanno G. & Nevoux M. (2018). Intra- and interspecific densities shape life-history traits in a salmonid population. *Oecologia*, 188(2), 451-464. <https://doi.org/10.1007/s00442-018-4213-4>
- Morin P. A., Luikart G., Wayne R. K. & the SNP workshop group. (2004). SNPs in ecology, evolution and conservation. *Trends in Ecology & Evolution*, 19(4), 208-216. <https://doi.org/10.1016/j.tree.2004.01.009>

- Morin P. A., Martien K. K. & Taylor B. L. (2009). Assessing statistical power of SNPs for population structure and conservation studies. *Molecular Ecology Resources*, 9(1), 66-73. <https://doi.org/10.1111/j.1755-0998.2008.02392.x>
- Morin P. A. & McCarthy M. (2007). Highly accurate SNP genotyping from historical and low-quality samples. *Molecular Ecology Notes*, 7(6), 937-946. <https://doi.org/10.1111/j.1471-8286.2007.01804.x>
- Morrissey M. B. & de Kerckhove D. T. (2009). The Maintenance of Genetic Variation Due to Asymmetric Gene Flow in Dendritic Metapopulations. *The American Naturalist*, 174(6), 875-889. <https://doi.org/10.1086/648311>
- Mouillot D., Bellwood D. R., Baraloto C., Chave J., Galzin R., Harmelin-Vivien M., ... Thuiller W. (2013). Rare Species Support Vulnerable Functions in High-Diversity Ecosystems. *PLoS Biology*, 11(5), e1001569. <https://doi.org/10.1371/journal.pbio.1001569>
- Narum S. R., Buerkle C. A., Davey J. W., Miller M. R. & Hohenlohe P. A. (2013). Genotyping-by-sequencing in ecological and conservation genomics. *Molecular Ecology*, 22(11), 2841-2847. <https://doi.org/10.1111/mec.12350>
- Nascimento P. R. M., Santos C. H. A., Sousa C. F. S., Climaco G. T., Paula-Silva M. N., & Almeida-Val V. M. F. (2011). Isolation and development microsatellite markers in the *Pygocentrus nattereri* (Kner, 1858) (Characiformes, Serrasalminae), an important freshwater fish in the Amazon. *Conservation Genetics Resources*, 4(2), 1-4.
- Neel M. C., McKelvey K., Ryman N., Lloyd M. W., Short Bull R., Allendorf F. W., ... Waples R. S. (2013). Estimation of effective population size in continuously distributed populations: there goes the neighborhood. *Heredity*, 111(3), 189-199. <https://doi.org/10.1038/hdy.2013.37>
- Neuhauser C., Andow D. A., Heimpel G. E., May G., Shaw R. G. & Wagenius S. (2003). Community Genetics: Expanding the synthesis of Ecology and Genetics. *Ecology*, 84(3), 545-558. [https://doi.org/10.1890/0012-9658\(2003\)084\[0545:CGETSO\]2.0.CO;2](https://doi.org/10.1890/0012-9658(2003)084[0545:CGETSO]2.0.CO;2)
- Nielsen E. S., Beger M., Henriques R., Selkoe K. A., & von der Heyden S. (2016). Multispecies genetic objectives in spatial conservation planning: Multispecies Genetic Planning. *Conservation Biology*, 31(4), 872-882. <https://doi.org/10.1111/cobi.12875>
- Nielsen R. (2000). Estimation of Population Parameters and Recombination Rates From Single Nucleotide Polymorphisms. *Genetics*, 154, 931-942.
- Nielsen R. (2004). Population genetic analysis of ascertained SNP data. *Human Genomics*, 1(3), 218. <https://doi.org/10.1186/1479-7364-1-3-218>
- Nielsen R., Hubisz M. J. & Clark A. G. (2004). Reconstituting the Frequency Spectrum of Ascertained Single-Nucleotide Polymorphism Data. *Genetics*, 168(4), 2373-2382. <https://doi.org/10.1534/genetics.104.031039>
- Nielsen R. & Signorovitch J. (2003). Correcting for ascertainment biases when analyzing SNP data: applications to the estimation of linkage disequilibrium. *Theoretical Population Biology*, 63(3), 245-255. [https://doi.org/10.1016/S0040-5809\(03\)00005-4](https://doi.org/10.1016/S0040-5809(03)00005-4)
- Nogueira-Machado V., Willis S. C., Teixeira A. S., Hrbek T. & Farias I. P. (2016). Population genetic structure of the Amazonian black flannelmouth characin (Characiformes, Prochilodontidae: *Prochilodus nigricans* Spix & Agassiz, 1829): contemporary and historical gene flow of a migratory and abundant fishery species. *Environmental Biology of Fishes*, 100(1), 1-16. <https://doi.org/10.1007/s10641-016-0547-0>
- Oberdorff T., Tedesco P. A., Hugueny B., Leprieur F., Beauchard O., Brosse S. & Dürr H. H. (2011). Global and Regional Patterns in Riverine Fish Species Richness: A Review. *International Journal of Ecology*, 2011, 1-12. <https://doi.org/10.1155/2011/967631>

- Oksanen J., Blanchet G., Friendly M., Kindt R., Legendre P., McGlinn D., Minchin P. R., O'Hara R. B., Simpson G. L., Solymos P., Stevens M. H. H., Szoecs E. & Wagner H. (2018). *vegan*: Community Ecology Package. R package version 2.4-6. <https://CRAN.R-project.org/package=vegan>
- Oliveira C., Foresti F. & Hilsdorf A. W. S. (2009). Genetics of neotropical fish: from chromosomes to populations. *Fish Physiology and Biochemistry*, 35(1), 81-100. <https://doi.org/10.1007/s10695-008-9250-1>
- Ortmann A. E. (1909). The destruction of the fresh-water fauna in western Pennsylvania. *Proceedings of the American Philosophical Society*, 48, 90–110.
- Pagezy H. & Jégu M. (2002). Valeur patrimoniale de Serrasalminae herbivores du Haut-Maroni (Guyane française): approches biologiques et socioculturelle en pays Wayana. *Bulletin Français de la Pêche et de la Pisciculture*, 364, 49-69. <https://doi.org/10.1051/kmae:2002003>
- Palsbøll P., Berube M. & Allendorf F. W. (2007). Identification of management units using population genetic data. *Trends in Ecology & Evolution*, 22(1), 11-16. <https://doi.org/10.1016/j.tree.2006.09.003>
- Palstra F. P. & Ruzzante D. E. (2011). Demographic and genetic factors shaping contemporary metapopulation effective size and its empirical estimation in salmonid fish. *Heredity*, 107(5), 444-455. <https://doi.org/10.1038/hdy.2011.31>
- Pante E., Abdelkrim J., Viricel A., Gey D., France S. C., Boisselier M. C. & Samadi S. (2015). Use of RAD sequencing for delimiting species. *Heredity*, 114(5), 450-459. <https://doi.org/10.1038/hdy.2014.105>
- Papadopoulou A., Anastasiou I., Spagopoulou F., Stalimerou M., Terzopoulou S., Legakis A. & Vogler A. P. (2011). Testing the Species–Genetic Diversity Correlation in the Aegean Archipelago: Toward a Haplotype-Based Macroecology? *The American Naturalist*, 178(2), 241-255. <https://doi.org/10.1086/660828>
- Paradis E. *pegas*: an R package for population genetics with an integrated-modular approach. *Bioinformatics, Oxford University Press (OUP)*, 26 (3), 419 - 420. [10.1093/bioinformatics/btp696](https://doi.org/10.1093/bioinformatics/btp696)
- Paris J. R., Sherman K. D., Bell E., Boulenger C., Delord C., El-Mahdi M. B. M., ... Stevens J. R. (2018). Understanding and managing fish populations: keeping the toolbox fit for purpose. *Journal of Fish Biology*, 92(3), 727-751. <https://doi.org/10.1111/jfb.13549>
- Pauls S. U., Alp M., Bálint M., Bernabò P., Čiampor F., Čiamporová-Zaťovičová Z., ... Monaghan M. T. (2014). Integrating molecular tools into freshwater ecology: developments and opportunities. *Freshwater Biology*, 59(8), 1559-1576. <https://doi.org/10.1111/fwb.12381>
- Paz-Vinas I. & Blanchet S. (2015). Dendritic connectivity shapes spatial patterns of genetic diversity: a simulation-based study. *Journal of Evolutionary Biology*, 28(4), 986-994. <https://doi.org/10.1111/jeb.12626>
- Paz-Vinas I., Loot G., Hermoso V., Veyssiére C., Poulet N., Grenouillet G. & Blanchet S. (2018). Systematic conservation planning for intraspecific genetic diversity. *Proceedings of the Royal Society B: Biological Sciences*, 285:20172746, 10.
- Paz-Vinas I., Loot G., Stevens V. M., & Blanchet S. (2015). Evolutionary processes driving spatial patterns of intraspecific genetic diversity in river ecosystems. *Molecular Ecology*, 24(18), 4586-4604. <https://doi.org/10.1111/mec.13345>
- Pecuchet L., Reygondeau G., Cheung W. W. L., Licandro P., van Denderen P. D., Payne M. R. & Lindegren M. (2018). Spatial distribution of life-history traits and their response to environmental gradients across multiple marine taxa. *Ecosphere*, 9(10), e02460. <https://doi.org/10.1002/ecs2.2460>
- Pelc R. A., Warner R. R. & Gaines S. D. (2009). Geographical patterns of genetic structure in marine species with contrasting life histories. *Journal of Biogeography*, 36(10), 1881-1890. <https://doi.org/10.1111/j.1365-2699.2009.02138.x>

- Pelletier T. A. & Carstens B. C. (2018). Geographical range size and latitude predict population genetic structure in a global survey. *Biology Letters*, 14(1), 20170566.
<https://doi.org/10.1098/rsbl.2017.0566>
- Peng B. & Amos C. I. (2008). Forward-time simulations of non-random mating populations using simuPOP. *Bioinformatics*, 24(11), 1408-1409. <https://doi.org/10.1093/bioinformatics/btn179>
- Peng B., Chen H.-S., Mechanic L. E., Racine B., Clarke J., Gillanders E. & Feuer E. J. (2014). Genetic Data Simulators and their Applications: An Overview. *Genetic Epidemiology*, 39(1), 2-10.
<https://doi.org/10.1002/gepi.21876>
- Pereira H. M., Ferrier S., Walters M., Geller G. N., Jongman R. H. G., Scholes R. J., ... Wegmann M. (2013). Essential Biodiversity Variables. *Science*, 339(6117), 277-278.
<https://doi.org/10.1126/science.1229931>
- Pereira L. H. G., Foresti F. & Oliveira C. (2009). Genetic structure of the migratory catfish *Pseudoplatystoma corruscans* (Siluriformes: Pimelodidae) suggests homing behaviour. *Ecology of Freshwater Fish*, 18(2), 215-225. <https://doi.org/10.1111/j.1600-0633.2008.00338.x>
- Petitgas P., Rijnsdorp A. D., Dickey-Collas M., Engelhard G. H., Peck M. A., Pinnegar J. K., ... Nash R. D. M. (2013). Impacts of climate change on the complex life cycles of fish. *Fisheries Oceanography*, 22(2), 121-139. <https://doi.org/10.1111/fog.12010>
- Pflüger F. J., & Balkenhol N. (2014). A plea for simultaneously considering matrix quality and local environmental conditions when analysing landscape impacts on effective dispersal. *Molecular Ecology*, 23(9), 2146-2156. <https://doi.org/10.1111/mec.12712>
- Pilger T. J., Gido K. B., Propst D. L., Whitney J. E. & Turner T. F. (2017). River network architecture, genetic effective size and distributional patterns predict differences in genetic structure across species in a dryland stream fish community. *Molecular Ecology*, 26(10), 2687-2697.
<https://doi.org/10.1111/mec.14079>
- Pimentel J. S. M., Carmo A. O., Rosse I. C., Martins A. P. V., Ludwig S., Facchin S., ... Kalapothakis E. (2018). High-Throughput Sequencing Strategy for Microsatellite Genotyping Using Neotropical Fish as a Model. *Frontiers in Genetics*, 9. <https://doi.org/10.3389/fgene.2018.00073>
- Pinsky M. L., & Palumbi S. R. (2014). Meta-analysis reveals lower genetic diversity in overfished populations. *Molecular Ecology*, 23(1), 29-39. <https://doi.org/10.1111/mec.12509>
- Piorski N., Sanches A., Carvalho-Costa L., Hatanaka T., Carrillo-Avila M., Freitas P. & Galetti Jr. P. (2008). Contribution of conservation genetics in assessing neotropical freshwater fish biodiversity. *Brazilian Journal of Biology*, 68(4 suppl), 1039-1050. <https://doi.org/10.1590/S1519-69842008000500011>
- Pitcher T. J. & Preikshot D. (2001). RAPFISH: a rapid appraisal technique to evaluate the sustainability status of fisheries. *Fisheries Research*, 49, 255-270.
- Planque B., Fromentin J.-M., Cury P., Drinkwater K. F., Jennings S., Perry R. I. & Kifani S. (2010). How does fishing alter marine populations and ecosystems sensitivity to climate? *Journal of Marine Systems*, 79(3-4), 403-417. <https://doi.org/10.1016/j.jmarsys.2008.12.018>
- Planquette P., Le Bail P.-Y. & Keith P. (1996). *Atlas des poissons d'eau douce de Guyane*. Paris.
- Pluzhnikov A., Rienzo A. D. & Hudson R. R. (2002). Inferences About Human Demography Based on Multilocus Analyses of Noncoding Sequences. *Genetics*, 161, 1209-1218.
- Polanski A. & Kimmel M. (2003). New Explicit Expressions for Relative Frequencies of Single-Nucleotide Polymorphisms With Application to Statistical Inference on Population Growth, 10.
- Pont D., Huguény B., Beier U., Goffaux D., Melcher A., Noble R., ... Schmutz S. (2006). Assessing river biotic condition at a continental scale: a European approach using functional metrics and fish assemblages: Biotic integrity assessment of European rivers. *Journal of Applied Ecology*, 43(1), 70-80. <https://doi.org/10.1111/j.1365-2664.2005.01126.x>

- Ponton D. & Copp G. H. (1997). Early dry-season community structure and habitat use of young fish in tributaries of the River Sinnamary (French Guiana, South America) before and after hydrodam operation. *Environmental Biology of Fishes*, 50, 235-256.
- Ponton D., & De Mérona B. (1998). Fish life-history tactics in a neotropical river with a highly stochastic hydrological regime : the Sinnamary river, French Guiana, South America, 24.
- Ponton D. & Tito de Morais L. (1994). Les stratégies de reproduction et les premiers stades de vie des poissons du fleuve Sinnamary (Guyane française): une revue bibliographique. *Revue d'hydrobiologie tropicale*, 27(4), 441-465.
- Pope B. J., Hammet F., Nguyen-Dumont T. & Park D. J. (2018). Hi-Plex for Simple, Accurate, and Cost-Effective Amplicon-based Targeted DNA Sequencing. In *Next Generation Sequencing*.
- Pope L. C., Liggins L., Keyse J., Carvalho S. B. & Riginos C. (2015). Not the time or the place: the missing spatio-temporal link in publicly available genetic data. *Molecular Ecology*, 24(15), 3802-3809. <https://doi.org/10.1111/mec.13254>
- Poulet N., Beaulaton L. & Dembski S. (2011). Time trends in fish populations in metropolitan France: insights from national monitoring data. *Journal of Fish Biology*, 79(6), 1436-1452. <https://doi.org/10.1111/j.1095-8649.2011.03084.x>
- Poulsen N. A., Nielsen E. E., Schierup M. H., Loeschcke V. & GrønkJaer P. (2005). Long-term stability and effective population size in North Sea and Baltic Sea cod (*Gadus morhua*): Effective population size in Atlantic cod. *Molecular Ecology*, 15(2), 321-331. <https://doi.org/10.1111/j.1365-294X.2005.02777.x>
- Prunier J. G., Dubut V., Chikhi L. & Blanchet S. (2017). Contribution of spatial heterogeneity in effective population sizes to the variance in pairwise measures of genetic differentiation. <https://doi.org/10.1101/031633>
- Prunier J. G., Dubut V., Loot G., Tudesque L. & Blanchet S. (2017). The relative contribution of river network structure and anthropogenic stressors to spatial patterns of genetic diversity in two freshwater fishes: A multiple-stressors approach. *Freshwater Biology*, 63(1), 6-21. <https://doi.org/10.1111/fwb.13034>
- Puyo J. (1949). *Poissons de la Guyane française. Faune de l'empire français, XII* (Office de la recherche scientifique d'Outre-Mer).
- Quéméré E., Baglinière J.-L., Roussel J.-M., Evanno G., McGinnity P. & Launey S. (2016). Seascape and its effect on migratory life-history strategy influences gene flow among coastal brown trout (*Salmo trutta*) populations in the English Channel. *Journal of Biogeography*, 43(3), 498-509. <https://doi.org/10.1111/jbi.12632>
- Quinn T. P., Hodgson S., Flynn L., Hilborn R. & Rogers D. E. (2007). Directional selection by fisheries and the timing of sockeye salmon (*Oncorhynchus nerka*) migrations. *Ecological Applications*, 17(3), 731-739. <https://doi.org/10.1890/06-0771>
- Radinger J. & Wolter C. (2014). Patterns and predictors of fish dispersal in rivers. *Fish and Fisheries*, 15(3), 456-473. <https://doi.org/10.1111/faf.12028>
- Raineri E., Ferretti L., Esteve-Codina A., Nevado B., Heath S. & Pérez-Enciso M. (2012). SNP calling by sequencing pooled samples. *BMC Bioinformatics*, 13(1), 239. <https://doi.org/10.1186/1471-2105-13-239>
- Ramirez-Soriano A. & Nielsen R. (2008). Correcting Estimators of and Tajima's D for Ascertainment Biases Caused by the Single-Nucleotide Polymorphism Discovery Process. *Genetics*, 181(2), 701-710. <https://doi.org/10.1534/genetics.108.094060>
- Reis R. E., Kullander S. O. & Ferraris Jr C. J. (Eds) (2003): Checklist of the freshwater fishes of South and Central America. Porto Alegre: Edipucrs.
- Reis R. E. (2013). Conserving the freshwater fishes of South America: Conserving South American Freshwater Fishes. *International Zoo Yearbook*, 47(1), 65-70. <https://doi.org/10.1111/izy.12000>

- Reis R. E., Albert J. S., Di Dario F., Mincarone M. M., Petry P. & Rocha L. A. (2016). Fish biodiversity and conservation in South America. *Journal of Fish Biology*, 89(1), 12-47.
<https://doi.org/10.1111/jfb.13016>
- Revaldaves E., Pereira L. H. G., Foresti F. & Oliveira C. (2005). Isolation and characterization of microsatellite loci in *Pseudoplatystoma corruscans* (Siluriformes: Pimelodidae) and cross-species amplification. *Molecular Ecology Notes*, 5(3), 463-465. <https://doi.org/10.1111/j.1471-8286.2005.00883.x>
- Richard-Hansen C. (2002). *Mise en place d'un observatoire de l'exploitation des ressources naturelles dans le Sud de le Guyane* (Conventions ONCFS/DIREN) (p. 111). Office national de la chasse et de la faune sauvage.
- Richmond J. Q., Backlin A. R., Galst-Cavalcante C., O'Brien J. W. & Fisher R. N. (2018). Loss of dendritic connectivity in southern California's urban riverscape facilitates decline of an endemic freshwater fish. *Molecular Ecology*, 27(2), 369-386. <https://doi.org/10.1111/mec.14445>
- Rochet M.-J. & Trenkel V. M. (2003). Which community indicators can measure the impact of fishing? A review and proposals. *Canadian Journal of Fisheries and Aquatic Sciences*, 60(1), 86-99. <https://doi.org/10.1139/f02-164>
- Rodríguez-Olarte D., Taphorn D. C. & Lobón-Cerviá J. (2011). Do protected areas conserve neotropical freshwater fishes? A case study of a biogeographic province in Venezuela. *Animal Biodiversity and Conservation*, 13.
- Romiguier J., Gayral P., Ballenghien M., Bernard A., Cahais V., Chenuil A., ... Galtier N. (2014). Comparative population genomics in animals uncovers the determinants of genetic diversity. *Nature*, 515(7526), 261-263. <https://doi.org/10.1038/nature13685>
- Rosenblum E. B. & Novembre J. (2007). Ascertainment Bias in Spatially Structured Populations: A Case Study in the Eastern Fence Lizard. *Journal of Heredity*, 98(4), 331-336. <https://doi.org/10.1093/jhered/esm031>
- Rousset F. (1997) Genetic Differentiation and Estimation of Gene Flow from F-Statistics under Isolation by Distance. *Genetics*, 145, 1219-1228.
- Salisbury S. J., McCracken G. R., Keefe D., Perry R. & Ruzzante D. E. (2016). A portrait of a sucker using landscape genetics: how colonization and life history undermine the idealized dendritic metapopulation. *Molecular Ecology*, 25(17), 4126-4145. <https://doi.org/10.1111/mec.13757>
- Santos C. H. A., de Sá Leitão C. S., Paula-Silva M. de N. & Almeida-Val V. M. F. (2016). Genetic differentiation in red-bellied piranha populations (*Pygocentrus nattereri*, Kner, 1858) from the Solimões-Amazonas River. *Ecology and Evolution*, 6(12), 4203-4213. <https://doi.org/10.1002/ece3.2195>
- Santos M. C. F., Ruffino M. L. & Farias I. P. (2007). High levels of genetic variability and panmixia of the tambaqui *Colossoma macropomum* (Cuvier, 1816) in the main channel of the Amazon River. *Journal of Fish Biology*, 71(sa), 33-44. <https://doi.org/10.1111/j.1095-8649.2007.01514.x>
- Saunders M. I., Brown C. J., Foley M. M., Febria C. M., Albright R., Mehling M. G., ... Burfeind D. D. (2015). Human impacts on connectivity in marine and freshwater ecosystems assessed using graph theory: a review. *Marine and Freshwater Research*, 67(3), 277-290. <https://doi.org/10.1071/MF14358>
- Schaefer S. A. (1998). Conflicts and resolution: impact of the new taxa on phylogenetic studies of the Neotropical cascudinhos (Siluroidei: Loricariidae). Pp. 375-400. In: Malabarba, L. R., R. E. Reis, R. P. Vari, Z. M. Lucena & C. A. S. Lucena (Eds.). *Phylogeny and classification of Neotropical fishes*. Porto Alegre, Edipucrs, 603p.
- Schick R. S. & Lindley S. T. (2007). Directed connectivity among fish populations in a riverine network: Directed connectivity. *Journal of Applied Ecology*, 44(6), 1116-1126. <https://doi.org/10.1111/j.1365-2664.2007.01383.x>

- Schindler D. E., Hilborn R., Chasco B., Boatright C. P., Quinn T. P., Rogers L. A. & Webster M. S. (2010). Population diversity and the portfolio effect in an exploited species. *Nature*, 465(7298), 609-612. <https://doi.org/10.1038/nature09060>
- Schlötterer C. (2004). The evolution of molecular markers - just a matter of fashion? *Nature Reviews Genetics*, 5(1), 63-69. <https://doi.org/10.1038/nrg1249>
- Schlötterer C., Tobler R., Kofler R. & Nolte V. (2014). Sequencing pools of individuals — mining genome-wide polymorphism data without big funding. *Nature Reviews Genetics*, 15(11), 749-763. <https://doi.org/10.1038/nrg3803>
- Schwartz M., Luikart G. & Waples R. (2007). Genetic monitoring as a promising tool for conservation and management. *Trends in Ecology & Evolution*, 22(1), 25-33. <https://doi.org/10.1016/j.tree.2006.08.009>
- Scribner K. T., Lowe W. H., Landguth E., Luikart G., Infante D. M., Whelan G. E. & Muhlfeld C. C. (2016). Applications of Genetic Data to Improve Management and Conservation of River Fishes and Their Habitats. *Fisheries*, 41(4), 174-188. <https://doi.org/10.1080/03632415.2016.1150838>
- Seeb J. E., Carvalho G., Hauser L., Naish K., Roberts S. & Seeb L. W. (2011). Single-nucleotide polymorphism (SNP) discovery and applications of SNP genotyping in nonmodel organisms: INTRODUCTION. *Molecular Ecology Resources*, 11, 1-8. <https://doi.org/10.1111/j.1755-0998.2010.02979.x>
- Selkoe K. A., D'Aloia C., Crandall E., Iacchei M., Liggins L., Puritz J., ... Toonen R. (2016b). A decade of seascape genetics: contributions to basic and applied marine connectivity. *Marine Ecology Progress Series*, 554, 1-19. <https://doi.org/10.3354/meps11792>
- Selkoe K. A., Gaggiotti O. E., ToBo Laboratory, Bowen B. W. & Toonen R. J. (2014). Emergent patterns of population genetic structure for a coral reef community. *Molecular Ecology*, 23(12), 3064-3079. <https://doi.org/10.1111/mec.12804>
- Selkoe K. A., Gaggiotti O. E., Tremblay E. A., Wren J. L. K., Donovan M. K., Hawai'i Reef Connectivity Consortium, & Toonen R. J. (2016a). The DNA of coral reef biodiversity: predicting and protecting genetic diversity of reef assemblages. *Proceedings of the Royal Society B: Biological Sciences*, 283(1829), 20160354. <https://doi.org/10.1098/rspb.2016.0354>
- Sexton J. P., Hangartner S. B. & Hoffmann A. A. (2014). Genetic Isolation by Environment or Distance: Which Pattern of Gene Flow Is Most Common? *Evolution*, 68(1), 1-15. <https://doi.org/10.1111/evo.12258>
- Seymour M., Seppälä K., Mächler E. & Altermatt F. (2016). Lessons from the macroinvertebrates: species-genetic diversity correlations highlight important dissimilar relationships. *Freshwater Biology*, 61(11), 1819-1829. <https://doi.org/10.1111/fwb.12816>
- Shafer A. B. A., Peart C. R., Tusso S., Maayan I., Brelsford A., Wheat C. W. & Wolf J. B. W. (2017). Bioinformatic processing of RAD-seq data dramatically impacts downstream population genetic inference. *Methods in Ecology and Evolution*, 8(8), 907-917. <https://doi.org/10.1111/2041-210X.12700>
- Smith N. J. H. (1985). The impact of cultural and ecological change on Amazonian fisheries. *Biological Conservation*, 32(4), 355-373. [https://doi.org/10.1016/0006-3207\(85\)90023-0](https://doi.org/10.1016/0006-3207(85)90023-0)
- Smitz N., Van Hooft P., Heller R., Cornélis D., Chardonnet P., Kraus R., ... Michaux J. (2016). Genome-wide single nucleotide polymorphism (SNP) identification and characterization in a non-model organism, the African buffalo (*Syncerus caffer*), using next generation sequencing. *Mammalian Biology - Zeitschrift Für Säugetierkunde*, 81(6), 595-603. <https://doi.org/10.1016/j.mambio.2016.07.047>
- Southwood T. R. E. (1977) Habitat, the templet for ecological strategies? *Journal of Animal Ecology*, 46, 337-365. doi:10.2307/3817.

- Spielman D., Brook B. W. & Frankham R. (2004). Most species are not driven to extinction before genetic factors impact them. *Proceedings of the National Academy of Sciences*, 101(42), 15261-15264. <https://doi.org/10.1073/pnas.0403809101>
- Stanley R. R. E., DiBacco C., Lowen B., Beiko R. G., Jeffery N. W., Wyngaarden M. V., ... Bradbury I. R. (2018). A climate-associated multispecies cryptic cline in the northwest Atlantic. *Science Advances*, 4(3), eaaq0929.
- Strayer D. L. & Dudgeon D. (2010). Freshwater biodiversity conservation: recent progress and future challenges. *Journal of the North American Benthological Society*, 29(1), 344-358. <https://doi.org/10.1899/08-171.1>
- Sundqvist L., Keenan K., Zackrisson M., Prodöhl P. & Kleinhans D. (2016). Directional genetic differentiation and relative migration. *Ecology and Evolution*, 6(11), 3461-3475. <https://doi.org/10.1002/ece3.2096>
- Swain D. P., Sinclair A. F. & Mark Hanson J. (2007). Evolutionary response to size-selective mortality in an exploited fish population. *Proceedings of the Royal Society B: Biological Sciences*, 274(1613), 1015-1022. <https://doi.org/10.1098/rspb.2006.0275>
- Taberlet P., Coissac E., Hajibabaei M. & Rieseberg L. H. (2012). Environmental DNA. *Molecular Ecology*, 21(8), 1789-1793. <https://doi.org/10.1111/j.1365-294X.2012.05542.x>
- Tejerina-Garro F. L. & De Mérona B. (2001). Spatial variability of biotic and abiotic factors of the aquatic habitat in French Guiana *Regulated Rivers: Research & Management*, 17(2), 157-169. <https://doi.org/10.1002/rrr.619>
- Tejerina-Garro F. L., De Mérona B., Oberdorff T. & Hugueny B. (2006). A fish-based index of large river quality for French Guiana (South America): method and preliminary results. *Aquatic Living Resources*, 19(1), 31-46. <https://doi.org/10.1051/alr:2006003>
- Telles M. P. C., Resende L. V., Brondani R. P. V., Collevatti R. G., Costa M. C. & Silva Junior N. J. (2010). Short Communication Isolation and characterization of microsatellite markers in the armored catfish *Hypostomus gymnorhynchus* (Loricariidae). *Genetics and Molecular Research*, 9(3), 1770-1774. <https://doi.org/10.4238/vol9-3gmr868>
- Thieme M., Lehner B., Abell R., Hamilton S. K., Kellendorfer J., Powell G. & Riveros J. C. (2007). Freshwater conservation planning in data-poor areas: An example from a remote Amazonian basin (Madre de Dios River, Peru and Bolivia). *Biological Conservation*, 135(4), 484-501. <https://doi.org/10.1016/j.biocon.2006.10.054>
- Thomaz A. T., Christie M. R. & Knowles L. L. (2016). The architecture of river networks can drive the evolutionary dynamics of aquatic populations: Brief Communication. *Evolution*, 70(3), 731-739. <https://doi.org/10.1111/evo.12883>
- Thrasher D. J., Butcher B. G., Campagna L., Webster M. S. & Lovette I. J. (2018). Double-digest RAD sequencing outperforms microsatellite loci at assigning paternity and estimating relatedness: A proof of concept in a highly promiscuous bird. *Molecular Ecology Resources*, 18(5), 953-965. <https://doi.org/10.1111/1755-0998.12771>
- Tito de Morais L., Lointier M. & Hoff M. (1995). Extent and role for fish populations of riverine ecotones along the Sinnamary river (French Guiana). *Hydrobiologia*, 303, 163-179.
- Tito de Morais L. & Raffray J. (1999). Movements of *Hoplias aimara* during the filling phase of the Petit-Saut dam, French Guyana. *Journal of Fish Biology*, 54(3), 627-635. <https://doi.org/10.1111/j.1095-8649.1999.tb00641.x>
- Toussaint A., Charpin N., Brosse S. & Villéger S. (2016). Global functional diversity of freshwater fish is concentrated in the Neotropics while functional vulnerability is widespread. *Scientific Reports*, 6(1). <https://doi.org/10.1038/srep22125>

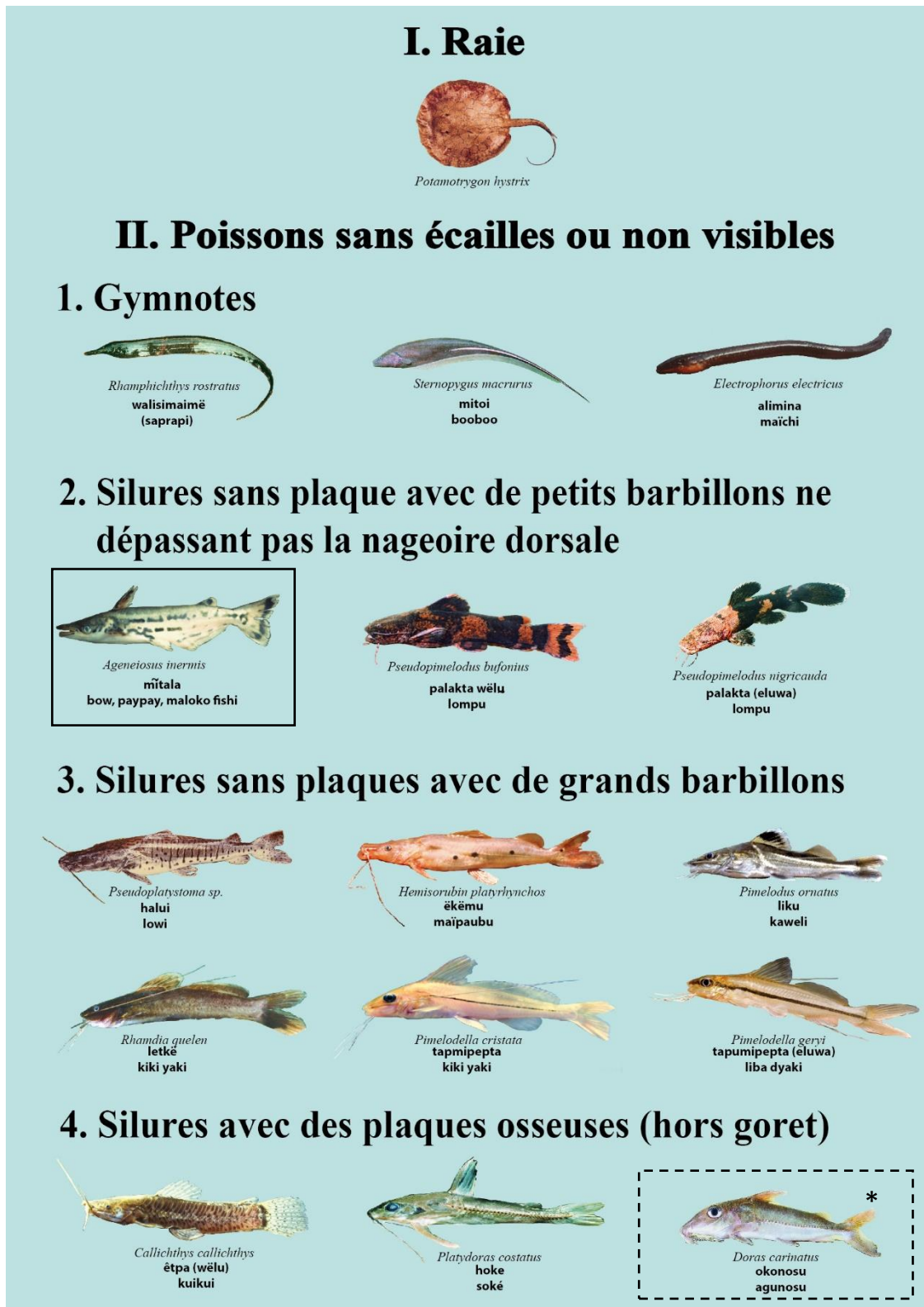
- Turak E., Harrison I., Dudgeon D., Abell R., Bush A., Darwall W., ... De Wever A. (2017). Essential Biodiversity Variables for measuring change in global freshwater biodiversity. *Biological Conservation*, 213, 272-279. <https://doi.org/10.1016/j.biocon.2016.09.005>
- Udvardy M. D. F. (1975). A classification of the biogeographical provinces of the world. IUCN Occasional Paper no. 18. Morges, Switzerland: IUCN
- Urban D. L., Minor E. S., Treml E. A. & Schick R. S. (2009). Graph models of habitat mosaics. *Ecology Letters*, 12(3), 260-273. <https://doi.org/10.1111/j.1461-0248.2008.01271.x>
- Valdes A. M., Slatkin M. & Freimert N. B. (1993). Allele Frequencies at Microsatellite Loci: The Stepwise Mutation Model Revisited. *Genetics*, 133(3), 737-749.
- Vari R. P., Ferraris C. J., Radosavljevic A. P. & Funk V. A. (2009). *Checklist of the freshwater fishes of the Guiana Shield* (Bulletin of the Biological Society of Washington).
- Vari R. P. & Malabarba L. R. (1998). Neotropical Ichthyology: An Overview. *Neotropical Ichthyology*, 12.
- Vellend M. & Geber M. A. (2005). Connections between species diversity and genetic diversity: Species diversity and genetic diversity. *Ecology Letters*, 8(7), 767-781. <https://doi.org/10.1111/j.1461-0248.2005.00775.x>
- Vignal A., Milan D., SanCristobal M. & Eggen A. (2002). A review on SNP and other types of molecular markers and their use in animal genetics. *Genetics Selection Evolution*, 34(3), 275-305. <https://doi.org/10.1051/gse:2002009>
- Villamor A., Costantini F. & Abbiati M. (2014) Genetic structuring across marine biogeographic boundaries in rocky shore invertebrates. *PLoS One*, 9(7), e101135.
- Waits E. R., Bagley M. J., Blum M. J., McCormick F. H. & Lazorchak J. M. (2008). Source-sink dynamics sustain central stonerollers (*Camptostoma anomalum*) in a heavily urbanized catchment. *Freshwater Biology*, 53(10), 2061-2075. <https://doi.org/10.1111/j.1365-2427.2008.02030.x>
- Wakeley J., Nielsen R., Liu-Cordero S. N. & Ardlie K. (2001). The Discovery of Single-Nucleotide Polymorphisms—and Inferences about Human Demographic History. *The American Journal of Human Genetics*, 69(6), 1332-1347. <https://doi.org/10.1086/324521>
- Walsh M. R., Munch S. B., Chiba S. & Conover D. O. (2006). Maladaptive changes in multiple traits caused by fishing: impediments to population recovery: Evolutionary consequences of harvesting. *Ecology Letters*, 9(2), 142-148. <https://doi.org/10.1111/j.1461-0248.2005.00858.x>
- Ward R. D., Woodrark M. & Skibinski D. O. F. (1994). A comparison of genetic diversity levels in marine, freshwater, and anadromous fishes. *Journal of Fish Biology*, 44, 213-232.
- Waters J. M. & Burrige C. P. (2016). Fine-scale habitat preferences influence within-river population connectivity: a case-study using two sympatric New Zealand *Galaxias* fish species. *Freshwater Biology*, 61(1), 51-56. <https://doi.org/10.1111/fwb.12675>
- Watson L. C., Stewart D. J. & Kretzer A. M. (2016). Genetic Diversity and Population Structure of the Threatened Giant *Arapaima* in Southwestern Guyana: Implications for Their Conservation. *Copeia*, 104(4), 864-872. <https://doi.org/10.1643/CG-15-293>
- Webb J. A. & Padgham M. (2013). How does network structure and complexity in river systems affect population abundance and persistence? *Limnologia - Ecology and Management of Inland Waters*, 43(5), 399-403. <https://doi.org/10.1016/j.limno.2013.04.006>
- Weber C., Covain R. & Fisch-Muller S. (2012). Identity of *Hypostomus plecostomus* (Linnaeus, 1758), with an overview of *Hypostomus* species from the Guianas (Teleostei: Siluriformes: Loricariidae), 34. *Cybium* 36(1):195-227.
- Weinman L. R., Solomon J. W. & Rubenstein D. R. (2015). A comparison of single nucleotide polymorphism and microsatellite markers for analysis of parentage and kinship in a cooperatively breeding bird. *Molecular Ecology Resources*, 15(3), 502-511. <https://doi.org/10.1111/1755-0998.12330>

- Weir B. S. & Cockerham C. C. (1984). Estimating F-Statistics for the Analysis of Population Structure. *Evolution*, 38(6), 1358. <https://doi.org/10.2307/2408641>
- Welcomme R. L. (1999). A review of a model for qualitative evaluation of exploitation levels in multi-species fisheries. *Fisheries Management and Ecology*, 6(1), 1-19. <https://doi.org/10.1046/j.1365-2400.1999.00137.x>
- Westgaard J.-I., Saha A., Kent M., Hansen H. H., Knutsen H., Hauser L., ... Johansen T. (2017). Genetic population structure in Greenland halibut (*Reinhardtius hippoglossoides*) and its relevance to fishery management. *Canadian Journal of Fisheries and Aquatic Sciences*, 74(4), 475-485. <https://doi.org/10.1139/cjfas-2015-0430>
- Wilson E. O. (1985) The biological diversity crisis: A challenge to Science. *Issues in science and technology*, 2, 20–29.
- Wilson G. A. & Rannala B. (2003). Bayesian Inference of Recent Migration Rates Using Multilocus Genotypes. *Genetics*, 163, 1177-1191.
- Winemiller K. O. (1989). Patterns of variation in life history among South American fishes in seasonal environments. *Oecologia*, 81(2), 225-241. <https://doi.org/10.1007/BF00379810>
- Winemiller K. O. (2005). Life history strategies, population regulation, and implications for fisheries management. *Canadian Journal of Fisheries and Aquatic Sciences*, 62(4), 872-885. <https://doi.org/10.1139/f05-040>
- Wright S. (1931). Evolution in Mendelian populations. *Genetics*, 16, 97–159.
- Wright S. (1943). Isolation by distance. *Genetics*, 28, 114–138.
- Yashima A. S. & Innan H. (2017). VARVER : a database of microsatellite variation in vertebrates. *Molecular Ecology Resources*, 17(4), 824-833. <https://doi.org/10.1111/1755-0998.12625>
- Yuan X., Miller D. J., Zhang J., Herrington D. & Wang Y. (2012). An Overview of Population Genetic Data Simulation. *Journal of Computational Biology*, 19(1), 42-54. <https://doi.org/10.1089/cmb.2010.0188>
- Zhang J., Chiodini R., Badr A. & Zhang G. (2011). The impact of next-generation sequencing on genomics. *Journal of genetics and genomics* 38(3), 95-109. <https://doi.org/10.1016/j.jgg.2011.02.003>

Annexes

Annexe I :

Planches descriptives des principales espèces d'intérêt halieutique rencontrées dans le Haut-Maroni (sont encadrées les 18 espèces étudiées au cours de cette thèse)



Dans cette thèse, c'est le très similaire *Doras micropoeus* qui a été étudié*

II. Poissons sans écailles ou non visibles

4. Gorets



Metaloricaria paucidens
lapipi takèpakem



Harttia surinamensis
tèpu lapipi
santiwipi



Ceniloricaria maculata
lapipi (eluwa)
santiwipi



Hypostomus gymmorhynchus
kabitanka, kawawa
kokooudou



Pseudancistrus barbatus
pèle
pataédéwawa



Pseudancistrus serratus
meikolo
tchobowawa, baaka wawa



Lasiancistrus brevispinis
gangu, djaka



Hemiancistrus medians
mili
gangu

III. Poissons avec écailles

1. Poissons à dorsale épineuse



Pachypops furcraeus
masau, kulekle
kubi



Plagioscion auratus
kupi
kubi



Plagioscion squamosissimus
kupi
kubi



Cichla ocellaris
matawale
aboné, tukunale



Geophagus harveri
hawa hawa
agankoy, san fishi



Geophagus surinamensis
kulisajek
agankoy, san fishi



Crenicichla albopunctata
kokopimpè
adonko koofu, alampipi



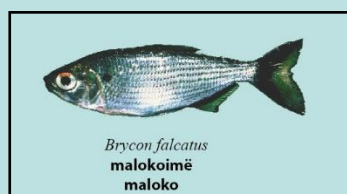
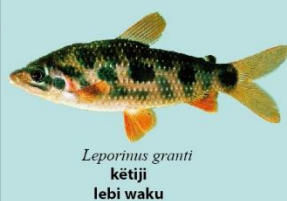
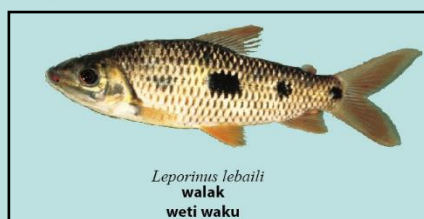
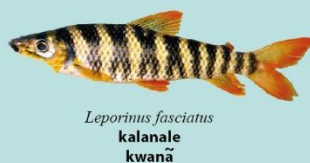
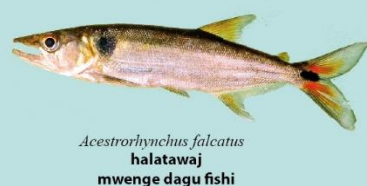
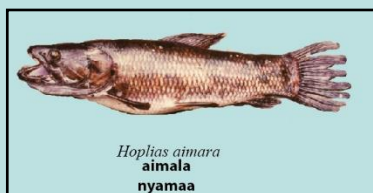
Crenicichla saxatilis
kokopimpè
adonko koofu, alampipi



Crenicichla multispinosa
kokopimpè tekupile
adonko koofu, alampipi

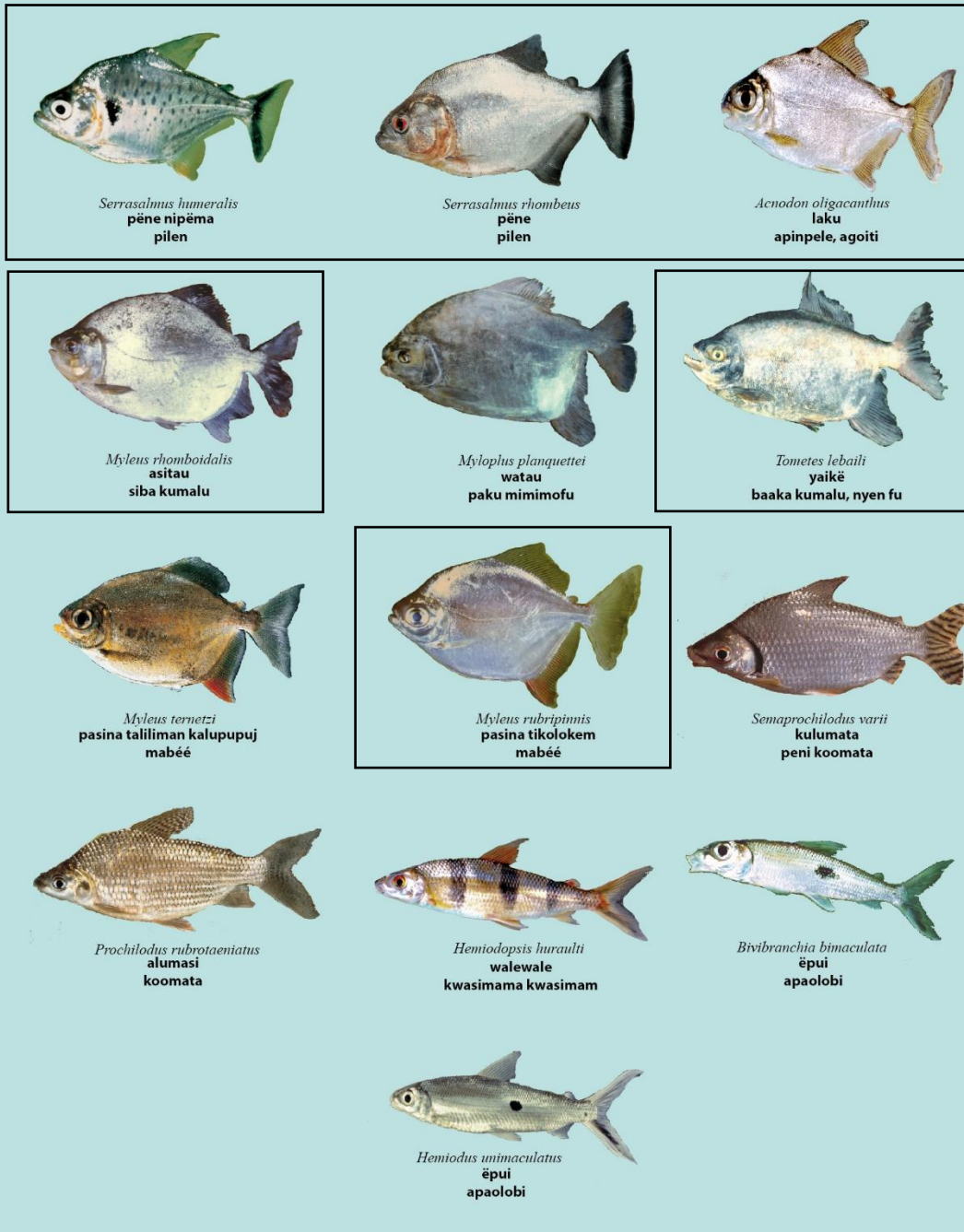
III. Poissons avec écailles

2. Poissons à dorsale non épineuse



III. Poissons avec écailles

2. Poissons à dorsale non épineuse



Pour chaque espèce, le nom scientifique est indiqué en italique. Les deux lignes suivantes en gras indiquent les noms vernaculaires des espèces, en langue Wayana puis Bushinengué. Consulter également Grenand et al. (2016), *Cybium*, 39(4): 279-300 pour des informations concernant les noms vernaculaires localement attribués aux différentes espèces de poissons d'eau douce de Guyane Française.

Annexe II :

**Understanding and managing fish populations:
keeping the toolbox fit for purpose**

Journal of Fish Biology, 92, 727-751.

Publication, dans sa version finale et paginée, produite lors du 50^{ème} Symposium de la Fisheries Society of the British Isles (FSBI) en Juillet 2017, et parue dans la revue *Journal of Fish Biology* au mois de Mars 2018.

Understanding and managing fish populations: keeping the toolbox fit for purpose

J. R. PARIS*†###, K. D. SHERMAN*###, E. BELL‡, C. BOULENGER§,
C. DELORD§, M. B. M. EL-MAHDI||, E. A. FAIRFIELD‡, A. M. GRIFFITHS*,
C. GUTMANN ROBERTS¶, R. D. HEDGER**, L. E. HOLMAN††, L. H. HOOPER*,
N. E. HUMPHRIES‡‡, I. KATSIADAKI§§, R. A. KING*, A. LEMOPOULOS|||, ¶¶,
C. J. PAYNE***, G. PEIRSON†††, K. K. RICHTER‡‡‡, M. I. TAYLOR‡,
C. N. TRUEMAN††, B. HAYDEN§§§# AND J. R. STEVENS*|||##

**Biosciences, University of Exeter, Stocker Road, Exeter EX4 4QD, U.K.*, †*School of Life Sciences, University of Sussex, Brighton, BN1 9QG, UK*, ‡*School of Biological Science, University of East Anglia, Norwich Research Park, Norwich, NR4 7UG, U.K.*, §*ESE, Ecology and Ecosystem Health, Agrocampus Ouest, INRA, 35042 Rennes, France*, ||*Faculty of Science, Laboratory of Molecular Genetics and Molecular Biology, Zoology Department, South Valley University, Qena 83523, Egypt*, ¶*Department of Life and Environmental Sciences, Faculty of Science and Technology, Bournemouth University, Bournemouth BH12 5BB, U.K.*,

Norwegian Institute for Nature Research, Trondheim, NO-7485, Norway*, ††*Ocean and Earth Science, National Oceanography Centre, University of Southampton, European Way, Southampton, SO14 3ZH, U.K.*, ‡‡*Marine Biological Association of the United Kingdom, The Laboratory, Citadel Hill, Plymouth, PL1 2PB, UK*, §§*Cefas, Weymouth Laboratory, Weymouth, DT4 8UB, U.K.*, |||*Department of Environmental and Biological Sciences, University of Eastern Finland, P.O. Box 111, FI-80101, Joensuu, Finland*, ¶¶*Department of Biology, University of Turku, FI - 20014, Turku, Finland*, *Institute of Aquaculture, Faculty of Natural Sciences, University of Stirling, Stirling, FK9 4LA, U.K.*, †††*Environment Agency, Research Analysis and Evaluation, Mance House, Kidderminster, DY11 7RA, U.K.*,

‡‡‡*BioArCh, Department of Archaeology, University of York, York, YO10 5DD, U.K.* and §§§*Biology Department, Canadian Rivers Institute, University of New Brunswick, Fredericton, New Brunswick, E3B 5A3, Canada*

Wild fish populations are currently experiencing unprecedented pressures, which are projected to intensify in the coming decades. Developing a thorough understanding of the influences of both biotic and abiotic factors on fish populations is a salient issue in contemporary fish conservation and management. During the 50th Anniversary Symposium of The Fisheries Society of the British Isles at the University of Exeter, UK, in July 2017, scientists from diverse research backgrounds gathered to discuss key topics under the broad umbrella of ‘Understanding Fish Populations’. Below, the output of one such discussion group is detailed, focusing on tools used to investigate natural fish populations. Five main groups of approaches were identified: tagging and telemetry; molecular tools; survey tools; statistical and modelling tools; tissue analyses. The appraisal covered current challenges and potential solutions for each of these topics. In addition, three key themes were identified as applicable across all tool-based applications. These included data management, public engagement, and fisheries policy and governance. The continued innovation of tools and capacity to integrate interdisciplinary approaches

|||Author to whom correspondence should be addressed. Tel.: +44 1392 723 775; email: j.r.stevens@exeter.ac.uk

##Joint first authors
#Joint senior authors

into the future assessment and management of fish populations is highlighted as an important focus for the next 50 years of fisheries research.

© 2018 The Fisheries Society of the British Isles

Key words: archaeology; genetics; modelling; stable isotopes; surveys; telemetry.

INTRODUCTION

Approximately 30% of fish species have been overexploited (FAO, 2014), representing significant losses to biodiversity, ecosystem services and socioeconomic contributions (Worm *et al.*, 2009). In light of the increasing challenges presented by climate change and other natural and anthropogenic stressors (Gordon *et al.*, 2018), an improved understanding of fish populations is critical to facilitate effective management and conservation initiatives. In July 2017, The Fisheries Society of the British Isles held its 50th Anniversary Symposium at the University of Exeter, UK, under the broad umbrella of 'Understanding Fish Populations'. To highlight key knowledge gaps and opportunities, we report the outcome of a working group convened at the symposium, which was tasked with considering the theme of tools for understanding fish populations. The scope of the discussion spanned diverse areas including spatial ecology and migration patterns, genetics and evolutionary biology, physiology, trophic ecology and developmental and population biology. In this article, we consider major advances in the use of tools across broad areas of fish biology and identify knowledge gaps and potential solutions in each area in order to guide and inform future research and to better understand and protect wild fish populations.

TAGGING AND TELEMETRY

A significant problem hampering the study of fish, marine benthic species in particular, is that of determining their geographical locations at fine scales, over long durations. Tagging and telemetry involves the application of external and or internal tags or devices to manually or passively track fish movement (Cooke *et al.*, 2013). Both forms can be particularly challenging in the marine environment, though manual tracking can work well at feeding grounds and at spawning aggregations (Murchie *et al.*, 2015), while passive tracking has valuable applications along known migration routes (Dahlgren *et al.*, 2016), for example, as anadromous and catadromous species migrate in and out of river estuaries (Lauridsen *et al.*, 2017). Suites of tools exist for such tasks [*e.g.* acoustic transmitters, passive information transponder (PIT) and Floy tags, radio, archival, *etc.*] and have been routinely used to understand the spatial ecology of a range of fish taxa (Bograd *et al.*, 2010). With technological improvements in tags and tracking equipment, the field has grown vastly in recent decades (Pine *et al.*, 2003; Jepsen *et al.*, 2015). We briefly highlight some of the tags and telemetry options commonly used by researchers along with a discussion of some of the limitations and challenges associated with these tools.

Archival data storage tags (DST), which can collect data on both the internal and external environments of fish are the only method available to assess internal states (*e.g.* bioenergetics; Cooke *et al.*, 2016). DSTs, however, currently only provide information on the environment experienced by the tagged fish if the tag is recovered, meaning these

data are lost if recapture rates are low, as is often the case in fish tagging surveys. Communication history acoustic tags (CHAT), which transmit data to nearby transponder receivers are a promising alternative. Since there have been relatively few uses of this tag type (Voegeli *et al.*, 2001; Hight & Lowe, 2007), there is potential for development in this area. Pop-off DSTs are also becoming available and may alleviate several of these concerns once problems associated with size and recoverability are resolved.

Pop-up Satellite Archival Tags (PSAT), which detach from the tagged fish after some time at sea and transmit telemetry data to overpassing satellites, are currently limited in terms of hardware, software and satellite reception. PSATs are large, so are limited in use for larger, often highly migratory individuals, and may also affect fish behaviour (Methling *et al.*, 2011). Additionally, battery failure, antenna damage, or mechanical failure may limit registration or transmission of data (Hays *et al.*, 2007; Musyl *et al.*, 2011). PSAT technology is relatively new, so future reductions in size and weight and also improvement in reliability can be expected. In terms of software, PSATs currently only transmit limited amounts of data due to transmission costs and the short time that the receiving satellite is above the horizon. Future software development is required to reduce transmission costs, optimise data transmission and provide more flexibility for users to tailor controls, in order to provide higher resolution data at the desired temporal scale. An increase in the number of satellite platforms that can receive PSAT data would help to improve reception issues. Interference on frequencies selected for tags at certain geographical locations (see Musyl *et al.*, 2011) also requires consideration.

Acoustic telemetry offers autonomous, continuous monitoring (Heupel *et al.*, 2006) and has the potential to significantly enhance our understanding of fish habitat use, activity patterns and resource partitioning (Hussey *et al.*, 2015). Acoustic arrays have been used in many studies elucidating fish movements (Papastamatiou *et al.*, 2013; Lea *et al.*, 2016) and transmitters have been used more innovatively to measure trophic interactions (Halfyard *et al.*, 2017). Issues remain however, in the significant cost and effort involved in deploying and maintaining acoustic arrays.

Organizations such as the Ocean Tracking Network (OTN; oceantrackingnetwork.org; Whoriskey, 2015) and the Integrated Marine Observing System (IMOS) animal tracking database Australian Animal Tracking Network (www.imos.org.au/facilities/animaltracking) both maintain acoustic infra-structure in the form of deployed receivers (arrays or curtains) in key ecological areas into which researchers are free to release tagged animals. These initiatives substantially reduce the cost and risk associated with acoustic tracking projects and similar approaches can be applied globally (*e.g.* a European tracking network is currently being developed). Furthermore, integration of standardised data repositories along with a comprehensive set of analytical tools to ensure rapid and sophisticated analysis of acoustic array data (Lea *et al.*, 2016) would lead to new insights into the spatial ecology of fish. Further technological developments such as the use of autonomous underwater vehicles (AUV) to perform routine data download operations, or even complement fixed acoustic receivers (Davis *et al.*, 2016), will make acoustic telemetry increasingly affordable and accessible to more researchers. Continued collaborations with established regional and international tracking networks, together with the ever-increasing sophistication, miniaturisation, durability and cost reduction of tags promises an increasingly important role for acoustic telemetry in our understanding of fish ecology.

MOLECULAR TOOLS

POPULATION GENETICS AND GENOMICS

Using molecular tools to understand fish genetic diversity and population structure has wide-ranging applications for evolutionary biology and the conservation and management of fish stocks. Until recently, molecular techniques such as mitochondrial sequencing and the analysis of microsatellite loci have been used most commonly to explore intra-specific variation in fish and many other organisms (Ferguson & Danzmann, 1998; Chistiakov *et al.*, 2006). More recently, however, the increased availability and cost efficiency of high-throughput sequencing, which is capable of producing millions of sequencing reads [*e.g.* RADseq (Davey & Blaxter, 2011) and RNAseq (Wang *et al.*, 2009)], has revolutionized the fields of population and conservation genetics (Allendorf *et al.*, 2010). It is, however, important to appreciate what extra information high-throughput sequencing data can provide, the biases involved in study design and data generation and also how its usage might be optimised. Here, we seek to identify knowledge gaps in the field of fish population genetics and contemplate how this area of research may evolve in the future.

Attaining high quality, clean DNA for large numbers of individuals is paramount for downstream sequencing processes, but in some cases can be challenging. Biological samples can often be compromised during sampling or transport, potentially rendering field efforts futile. Population genetic studies on fish frequently require sampling from river transects or remote locations at sea and so portable laboratories for sampling, storing and extracting DNA would be welcomed. Emerging technologies, *e.g.* the MinION USB sequencer (www.nanoporetech.com/products/minion), have the potential to revolutionize when and where genetic data can be generated. Most new technologies are currently restricted to sequencing small genomes, such as those of bacteria, but with on-going improvements, these technologies open up the possibility of being able to sequence DNA in real-time in the field (Hayden, 2015). Recently, the MinION technology has been used in hybrid assemblies with Illumina short reads (Austin *et al.*, 2017) and *de novo* eukaryotic genomes (including fish) are in progress (Jansen *et al.*, 2017).

Alongside population genetic studies, research based on whole genome data is emerging and the genomes of several commercially important species have now been published (*e.g.* Atlantic cod *Gadus morhua* L. 1758; Star *et al.*, 2011; Atlantic salmon *Salmo salar* L. 1758; Lien *et al.*, 2016). While the ever-reducing cost of whole genome sequencing provides opportunities to sequence and publish more fish genomes, in our view, the key priority is not simply publishing genomes, but also attaining high-quality genome annotation. Gene annotation and accurate knowledge of the function of different identified regions is of extreme importance if genomic tools are to be used reliably in conservation and management (Ekblom & Wolf, 2014). Therefore, projects such as the 'Functional Annotation of All Salmonid Genomes' (Macqueen *et al.*, 2017) should be encouraged and developed. It is also important not to underestimate or neglect the computing power and bioinformatics expertise required to produce high quality genome scaffolds and annotations, and also to recognise and account for biases in next generation sequencing data (Benestan *et al.*, 2017).

Furthermore, population genetic approaches are usually focused on a single species. Consequently, there is a mismatch between studies of a single species genotyped at high resolution, but generally at small spatial scales (*e.g.* population genetics, often

using hundreds to thousands of markers through genotyping by sequencing (GBS) or genome-wide association studies (GWAS)), and studies of multiple species at larger spatial scales but using lower resolution markers [*e.g.* phylogeography or biodiversity assessments using metabarcoding or mitochondrial (mt)DNA sequencing]. Nonetheless, the widespread application of molecular resources has led to the accumulation of rich datasets across a broad range of species, geographical regions and time periods (Blanchet *et al.*, 2017). Accordingly, we anticipate that this aggregation of data may allow the underlying processes that drive genetic variability across these regions and times to be revealed, enabling a broader testing of theories in population genetics and evolution (Pauls *et al.*, 2014; Ellegren & Galtier, 2016).

Such studies will require the combination of high genetic resolution markers across large spatial scales, which is a non-trivial task, especially when dealing with non-model species. Three challenges arise in such cases: firstly, the financial investment required to obtain reliable datasets for several species remains significant. Despite reductions in sequencing costs, it may be financially sensible to rely on more classical markers such as microsatellites or small subsets of single nucleotide polymorphisms (SNPs). Secondly, there is a need for a standardised framework in order to make datasets comparable across different species and regions. This standardization must occur when collecting samples, characterising markers (Ellis *et al.*, 2011; Helyar *et al.*, 2011) and during the subsequent data analysis to streamline user choices (Paris *et al.*, 2017), which may bias the biological interpretation of data (Rodríguez-Ezpeleta *et al.*, 2016). It is therefore important that researchers use common methods to isolate and characterise markers for entire sets of focal species and provide full access to detailed analyses when datasets are generated.

Finally, as multi-species approaches remain scarce, there is a need to define hypotheses at the beginning of such investigations. In this respect, simulation tools (Laval & Excoffier, 2004; Peng & Kimmel, 2005; Neuenschwander, 2006) are particularly useful for testing complex hypotheses and also for predictive purposes. Moreover, the integration of mathematical and statistical models with fish population genetics would be useful for revealing genotype–phenotype interactions (Ritchie *et al.*, 2015), evolutionary signatures (Stark *et al.*, 2007), functional DNA elements (Schridder & Kern, 2014), spatial dynamics (Guillot *et al.*, 2009) and species-genetic diversity correlations (SGDC; Vellend, 2003; Vellend *et al.*, 2014).

ENVIRONMENTAL DNA

The use of environmental (e)DNA to identify the presence and understand the distribution of fish has expanded rapidly in the past decade. eDNA is a polydisperse mixture (Turner *et al.*, 2014; Wilcox *et al.*, 2015) of various biological material ranging from entire cellular fragments to extracellular DNA, which is isolated from environmental samples such as water or sediment. Such techniques are used for species identification and food security purposes. Universal primers that target mtDNA can be applied for identifying species presence (Yamamoto *et al.*, 2016) or to gain information about species interactions (*e.g.* food-web construction; Sousa *et al.*, 2016).

An important component of this work is validating the results from eDNA surveys with traditional fish survey methods. In both freshwater and marine environments, eDNA has compared favourably with traditional fish survey methods (Thomsen *et al.*, 2012; Hänfling *et al.*, 2016). eDNA, however, was found to be less effective compared

with experienced snorkel surveys (Ulibarri *et al.*, 2017). This underpins the importance of validation with traditional techniques, especially in spatially heterogeneous and complex aquatic environments (Shogren *et al.*, 2017).

The development of effective PCR primers is central to the successful application of eDNA (Freeland, 2016; MacDonald & Sarre, 2017). As a result, a vast range of primer sets are available for fishes (Doi *et al.*, 2015; Clusa *et al.*, 2017). Metabarcoding primers, that simultaneously amplify eDNA from many fish species, have also been developed for monitoring entire fish communities (Miya *et al.*, 2015; Valentini *et al.*, 2016).

Beyond inferring if a fish species is present in the sampled location, researchers have begun to investigate if eDNA can provide further information regarding fish populations. The use of eDNA to infer population-level variation has been demonstrated (Uchii *et al.*, 2016; Sigsgaard *et al.*, 2016), but is still in its infancy. Similarly, although attempts to link eDNA concentration and fish biomass have shown promising results (Lacoursière-Roussel *et al.*, 2016; Yamamoto *et al.*, 2016), further development is required to improve the accuracy of these measurements. However, for applications utilising eDNA to be optimised, preexisting molecular information needs to be accessible. A number of publicly available databases [*e.g.* NCBI Genbank (www.ncbi.nlm.nih.gov/genbank/) and BOLD (www.boldsystems.org/)] hold a vast array of molecular data, but there is still a need for further mitochondrial genome sequencing to allow for optimal usage of molecular identification techniques.

MICROBIOMES

Analysis of a microbiome can provide novel insights into the health and biology of fish populations. Traditional culture-dependent tools used to map the commensal microbiota community in fish are often time-consuming, expensive and subjected to bias as only 0.1–10% of bacteria can be cultured *in vitro* (Amann *et al.*, 1995; Austin, 2006). More recently, rapid culture-independent tools such as 16S ribosomal (r)RNA targeted sequencing have been used to provide detailed profiles of the structure and diversity of the microbiota residing on the mucosal surface of fish (Ghanbari *et al.*, 2015).

The gut microbiome composition has also become an important biomarker for understanding the influence of stress in fish (Llewellyn *et al.*, 2014), as numerous stressful stimuli have been shown to alter the microbiome composition (Xia *et al.*, 2014; Gaulke *et al.*, 2016). The gut microbiome composition can provide insights into the ecology and physiology of fish in a range of areas such as ecological speciation (Sevellec *et al.*, 2014), the biology of migratory fish (Llewellyn *et al.*, 2016), trophic interactions within ecosystems (Ingerslev *et al.*, 2014) and adaptation to extreme environments (Song *et al.*, 2016).

There are a number of challenges currently facing fish microbiome research. At present, the majority of data regarding the microbiome composition in wild teleost fish originates from laboratory models (Tarnecki *et al.*, 2017). More studies are required to see if captive-reared animals provide a reliable analogue for wild populations. Standardised protocols for collecting and generating microbiome data are also lacking, which could restrict progress as several processes have the potential to introduce differential bias in microbiota profiles (Salipante *et al.*, 2014; Hart *et al.*, 2015). Adopting a framework of robust, quality-controlled protocols (*e.g.* similar to human microbiome

research; Methé *et al.*, 2012) would be of great benefit. In addition, there is currently a lack of non-invasive protocols for conducting longitudinal or repeated sampling of the gut microbial community in individual fish over time. The application of rectal swabs (Budding *et al.*, 2014) for sampling the vent of fish could provide a non-invasive strategy for collecting such data. Finally, time-series data could also enhance our knowledge in terms of the functional aspects of host lifecycles and the stability and resilience of microbiota (Goodrich *et al.*, 2014).

SURVEY TOOLS

FIELD-BASED SURVEYS

Fish population assessments are conducted using a wide range of techniques; the advantages, limitations, personnel requirements and health and safety considerations of each are presented in Table I. It is encouraging to note that even well-established methods such as hydroacoustics are continually being improved, while emerging tools such as eDNA are beginning to be included in routine monitoring. We suggest that integrating methods and data series are key priorities for future research in this field.

In large and complex habitats, it is often the case that a suite of survey methodologies has to be employed to sample different times, habitats and species effectively. Indeed, an advantage of field-based surveys is the ability to generate information from both fishery-independent (Nash *et al.*, 2016) and fishery-dependent (Shin *et al.*, 2010) data. The availability of a diversity of methodologies, however, can make the task of assessment in these habitats even more costly; issues also remain over how to use often disparate data types to develop a sound understanding of a fishery. Integrating methods represents a key means of improving data resolution from field surveys. For instance, methods such as eDNA and hydroacoustic sampling provide comparatively fast and non-invasive estimates of fish community structure and biomass. To obtain a thorough understanding of fish populations, however, this information must be combined with fish age, size and health data, typically obtained *via* destructive sampling (*e.g.* gill netting). As yet, there are no structured, universally agreed guidelines on which methods should be integrated to obtain a thorough assessment of population dynamics from a specific habitat type.

Fish-survey methodologies are typically determined at a national level, making international comparisons of data extremely challenging. In recent years, standardised protocols initiated through the European Union Water Framework Directive (E.C., 2000) have facilitated Europe-wide assessments of fish community structure. Such international standardisation is essential when assessing anthropogenic effects on fish (Gordon *et al.*, 2018) and we recommend that efforts are made to make national datasets available using standardised metadata and biodiversity information, ideally *via* open sharing platforms (*e.g.* www.freshwaterplatform.eu).

HISTORICAL RECORDS

Historical records (*e.g.* catch records) can also be useful in helping to extrapolate population data back into the recent past. Libraries and historical societies often hold picture archives and these images can in some instances be used as a form of historical

TABLE I. Summary of popular current and emerging methods used for fish surveys along with the associated advantages and limitations of each method

Method	Advantages	Limitations	Personnel requirement	Health & safety consideration
Electric fishing	Can be used in flowing and still water, among macrophytes and obstructions; relatively unselective; can be used quantitatively	Inefficient in water >1–1.5 m depth or in wide reaches; limited by water and bed conductivity; can be harmful to sensitive fish species and life stages; invasive	Significant to high	High
Seine netting	Can be used quantitatively; efficiency well-understood; relatively unselective	Limited effectiveness in very deep or very shallow water; limited effectiveness where there are macrophytes, obstructions, or soft sediment; restricted to use in low velocity water bodies; invasive	High	Significant
Trawling	Large areas of deep water can be surveyed efficiently	Restricted to use in relatively open continuous stretches of water of >2 m depth; cannot be used where there are dense growths of macrophytes, very variable bed profiles or large debris; requires sizeable boats and launching facilities; invasive	High	High
Gill netting	Can be used in a wide variety of environments among debris and macrophytes, in almost any depth	Invasive–destructive; limited ability to assess absolute fish abundance	Significant	Significant
Hydroacoustics	Huge expanses of water can be surveyed efficiently; non-invasive; quantitative abundance estimates possible	Limited effectiveness in turbulent environments; can only sample relatively open water so unsuitable to use for sampling in marginal habitats; lacks capacity to differentiate between species; cannot assess age, condition and health of fish	Significant	Significant

TABLE I. Continued

Method	Advantages	Limitations	Personnel requirement	Health & safety consideration
Fyke netting and trapping	Can be deployed in a variety of Environments; can be effective for some species difficult to sample by other methods	Very species and size-selective; limited ability to assess absolute fish abundance	Significant	Significant
Fry surveys: micromesh seine, handnet & traps	Focuses on margins of rivers and lakes, therefore less resource intensive, simple equipment only; assesses a key life stage; relatively unselective	Only assesses juvenile populations; invasive as very young fish are unlikely to survive capture	Significant	Significant
Fish counters & fixed traps: sometimes accompanied by cameras (video recorder)	Good for assessing highly mobile fish with relatively predictable migration patterns	Resource intensive, high capital costs, maintenance; quantitative assessment for migratory species only; often only operational under certain environmental conditions	High	Significant
Rod-and-line	Adaptable, can be deployed almost anywhere; amenable to volunteers and citizen-science participation	Very effort-dependent (quantity and quality); strongly influenced by conditions; very selective for species and size of fish; limited capability to assess absolute fish abundance; very noisy data	High	Low
Commercial catch monitoring	Enables large volumes of data collected over large spatial and temporal scales; relatively inexpensive as fish are being caught anyway	Can only happen where commercial fisheries exist; little control over changes in effort and methodology, which are driven by market forces; strongly influenced by conditions	Low	Low
Visual surveys: snorkelling, diving, counting from the bank	Relatively non-invasive; enables observation of fish in their surroundings	Only applicable in high water clarity and over short ranges; mostly applicable to species with distinct individual home range, typically associated with physical habitat features	Moderate	Significant to high

TABLE I. Continued

Method	Advantages	Limitations	Personnel requirement	Health & safety consideration
Methods under development				
eDNA: single-target and meta barcoding	Very adaptable, deployable anywhere; non-invasive; non-selective; low field manpower requirement	Currently can only establish fish presence and abundance of species relative to each other, absolute abundance remains a challenge; cannot assess age, size, condition or health; uncertainty around the source of eDNA in lotic environments; high laboratory time requirement	Significant	Significant
DIDSON*-ARIS#high resolution sonar	Can be used in turbid water, among obstructions; can be used in a variety of depths and flows except very turbulent water; enables visualization of target fish, species identification; quantitative estimates possible; species (some) and size of fish can be identified; observations of fish behaviour possible; non-invasive	Mobile deployment currently challenging; limited ability to assess whole water-body abundance; limited species identification capability; high data-processing requirement; cannot assess age, condition and health of fish	Significant	Significant

*Dual-frequency IDentification *SONar*.

#Adaptive Resolution Imaging Sonar.

survey data to provide information on past community composition and size distributions (McClenachan, 2009). Historical records of catch data are typically held by government agencies or can be found in local archives (*e.g.* angling club logs) and corporate records. Such data have been used successfully to reconstruct fish populations back to the late 1800s (Thurstan *et al.*, 2010; Thurstan & Roberts, 2010). Catch reconstruction approaches can also provide useful insights into fishery trends that may not be apparent from data reported only by the U.N. Food and Agriculture Organization (FAO, 2015; Smith & Zeller, 2015; Zeller *et al.*, 2015). Although limited to the information that is still available, and subject to the often-unidentifiable biases of the individuals who originally recorded the data, this can provide a unique way to extrapolate population data back in time.

STATISTICAL AND MODELLING TOOLS

BAYESIAN METHODS

Reliable estimates of demographic variables (*e.g.* abundance, survival, growth rates and fecundity) and an understanding of the processes that regulate these variables are fundamental for sustainable management of fish populations. To understand the ecological processes in order to truly inform policy, however, researchers must use multiple data sources, provide links between management actions and population responses, and also estimate uncertainty as a prerequisite to making forecasts that provide useful information. Bayesian methods in ecology and conservation biology are now increasingly being used to explore these links, for example, in stable-isotope analyses. Indeed, the Bayesian framework provides an intuitive method for estimating parameters, expressing uncertainty in these estimates and allows for the incorporation of as much or as little existing data or prior knowledge that is available (Ellison, 2004). To develop the use of this specific framework in fish ecology and management, however, there is a need to educate and train fish biologists in the use of Bayesian principles and methods.

INDIVIDUAL-BASED MODELS

Individual-based models (IBM) are process-based mechanistic computer models that simulate emergent properties of fish biology, behaviour, traits or group characteristics, based on simple heuristic functions. The use of IBMs in fish research has grown exponentially (DeAngelis & Mooij, 2005) as computational power has increased (DeAngelis & Grimm, 2014). Several IBMs were presented at the 50th Anniversary Symposium of The Fisheries Society of the British Isles, and with continued increases in computational power, IBMs look set to offer powerful new avenues for population research (DeAngelis & Grimm, 2014) in computationally challenging multifactor systems such as fish ecotoxicology (Mintram *et al.*, 2017). Additionally, a variety of tools now exist which provide for the easier creation of new models, such as various R packages (www.r-project.org) and programmable environments (*e.g.* NetLogo; www.ccl.northwestern.edu/netlogo). Programmes such as R, however, are sometimes not intuitive to new users and so additional training for fisheries scientists and collaborations between scientists from different computational and statistical backgrounds would be advantageous. For a more robust future application of IBMs within fisheries science, there

is a need for further assessment of the relative strengths and weaknesses (and potential availability and future development) of the different models.

Integration with environmental data is a pertinent issue when modelling and is becoming easier through developments in geographic information systems (GIS) and other programming environments (such as R), which now include procedures and libraries for use in ecological work. One example is the use of food-web models that integrate environmental data (Christensen & Walters, 2004) and coral-reef ecosystem modelling methods (Rogers *et al.*, 2014; Weijerman *et al.*, 2015). A hindrance to the integration of environmental data into fisheries science is that it can be difficult to find and access data sources, although availability and accessibility of such data is improving (www.worldclim.org). The existence of a central node or hub with paths to these data sources would be useful.

TISSUE ANALYSIS

STABLE-ISOTOPE ECOLOGY

Stable isotopes are now routinely used to quantify the trophic ecology (Boecklen *et al.*, 2011) and migration history (Trueman *et al.*, 2012) of fish, or to identify community level patterns in food-web structure and resource use (Layman *et al.*, 2012). Although the technique is still in its relative infancy, stable-isotope ecology (or stable-isotopes analysis, SIA) has advanced much in recent decades. Below we outline four rapidly developing areas with the potential to enhance the applicability of this tool to studies of fish biology.

BIOCHEMICAL MECHANISM

The relationship between the isotopic composition of a consumer's tissues and that of its prey is fundamental to all applications of stable isotopes in ecology. While general principles are clear [*i.e.* faster reaction rates and preferential incorporation of light isotopes into excretory metabolites a process termed trophic fractionation (DeNiro & Epstein, 1977)], the precise mechanisms leading to fractionation and, particularly, the extent of isotopic fractionation expected under differing physiological conditions cannot currently be predicted, primarily due to the complexity of amino-acid biochemistry. Uncertainties associated with the isotopic expression of tissue composition and relative rates of tissue growth and regeneration further complicate the interpretation of stable isotope values in ecology. Recent information gained from compound-specific isotope analysis (*i.e.* assessing isotopic compositions of single amino acids), however, is beginning to shed light on the fractionation process (McMahon & McCarthy, 2016).

POPULATION-LEVEL DATA

The distribution of isotopic compositions of individuals within a population (often termed the isotopic niche; Newsome *et al.*, 2007) has been proposed as a powerful comparative measure of population-level ecological characters. In addition to individual variability in consumers, however, the distribution of isotopic compositions in a population is influenced by spatial and temporal variations in the isotopic composition of primary production, temporal variability within trophic linkages and differential

rates of growth and isotopic assimilation (Gorokhova, 2017). Very few studies have attempted to combine ecological and food web theory with isotope systematics to explore the sensitivity of community isotopic metrics to changes in food web structure and function.

ISOBANK

To date, applications of stable isotopes to fish biology have predominantly focussed on analyses of specific populations or communities. The absence of a centralised, open-access repository for stable-isotope data restricts the opportunity for syntheses or meta-analyses of stable-isotope data (Pauli *et al.*, 2017). Recent efforts to address this have found broad support from the stable isotope research community (Pauli *et al.*, 2017) and would be especially beneficial to fish biologists due to the large amount of fish-isotope data currently available. Defining an ontology of stable-isotope metadata, information required to describe and interpret isotope data, for fish biologists is an immediate requirement in this regard.

MARINE ISOSCAPES

The stable-isotope ratios of a consumer's tissue encode the resources (water, air, prey *etc.*) it was using when that tissue was formed. As such, provided one has access to a suite of isotopic baseline measurements (*e.g.* water, plants and primary consumers), it is possible to trace an organisms route through space and time up to the point of capture (Trueman *et al.*, 2012). Creation of a practically useful isoscape requires relatively dense sampling of a reference organism across space (and potentially time). Bulk stable-isotope analyses are now routine, commonly available globally and relatively cheap and regional marine-isoscape models are being developed at a rapid rate (MacKenzie *et al.*, 2014; Kurlle & McWhorter, 2017). In the open ocean, sample-based isoscapes are difficult to develop, but progress is being made in isotope-enabled global biogeochemical models (Magozzi *et al.*, 2017), offering temporal and spatial models of expected isotopic variability at global scales. Improving the precision, accuracy and availability of these baseline measurements will increase the robustness and precision of isotope-based estimates animal position.

ARCHAEOLOGICAL MATERIAL

Archaeological material can allow an otherwise impossible snapshot into past populations. Traditional morphological approaches can provide age distributions and species ranges, and, with the rapid development of biomolecular archaeology in the past 20 years, many of the techniques used to explore modern fish populations can now be used to look into the past. From ancient DNA to proteomics and isotopes to lipids, a wide range of biomolecules have been recovered and explored from archaeological material (Orton, 2016). For example, compound-specific isotope analysis has the potential to track trophic level changes through time (McClelland & Montoya, 2002; Naito *et al.*, 2016). Population genetics of extinct populations have been successfully explored in terrestrial animals (Chang & Shapiro, 2016; Murray *et al.*, 2017) and these same techniques can be used on fish bones to reconstruct past genetics (Iwamoto *et al.*, 2012; Ólafsdóttir *et al.*, 2014). Ideally these data will be used to understand

environmental and anthropogenic effects on fish populations, and importantly, how modern fish populations might respond to climate change and fishing pressures.

A major barrier to the use of archaeological fish material is the fact that less than 10% of fish bones are identified to species level (Wheeler & Jones, 1989; Gobalet, 2001) and much of what is identified is buried in the grey literature of archaeological reports that are often printed in small quantities and not digitised (Linden & Webley, 2012). This makes the material relevant to an ecological question very difficult to find. Archaeologists are working towards ways to improve the amount of bones identified by better reference collections and education on fish bones (National Zooarchaeological Reference Resource, Nottingham's Archaeological Fish Resource; Vertebra@UWF) and on creating searchable databases of archaeological material (Callou, 2009; Kansa, 2010). In addition, new ZooMS (Zooarchaeology by mass spectrometry) techniques are being explored to quickly identify even small bones and scales to species using peptide mass fingerprinting (Richter *et al.*, 2011), which will allow even more material to be identified in a useful way for those working on understanding fish populations. In the near future, it should be possible for modern fish biologists, in conjunction with archaeologists, to ask direct questions of past populations (Van Neer & Ervynck, 2010).

GENERAL TOPICS IDENTIFIED AS APPLICABLE ACROSS ALL THEMES

MANAGEMENT OF DATA: INTEGRATION, CALIBRATION AND STANDARDIZATION

An integrated management framework for data classification, characterisation, storage and accessibility would be a valuable resource for fish and fisheries biologists. FishBase, which at the time of writing contains information regarding 33 600 fishes, involving 2290 collaborators, receiving over 600 000 visits per month, is an example of the potential for such a resource (www.fishbase.org; Froese & Pauly, 2017). A single database for all types of fish data (for example, DNA, tagging, isotopes, diet) is probably unworkable, but the advent of application programming interfaces (API) and analytical software, which allows automated querying across multiple databases, represents an unprecedented opportunity to access a wealth of global data. Indeed, we suggest that more data (such as those discussed here) could be integrated into FishBase. Such resources, however, require significant funding and long-term commitment from governments and trans-national organizations, *e.g.* the North Atlantic Salmon Conservation Organization (NASCO).

PUBLIC ENGAGEMENT, EDUCATION AND OUTREACH

Scientific engagement with the public is essential to effect meaningful societal change or to ensure a wider consensus is made around new discoveries or ethical considerations. Additionally, however, the power of the public as a tool in science is also being increasingly recognised. Crowdfunding, whereby a scientist requests small amounts of money from a large number of interested individuals to successfully launch a project, potentially provides a powerful new way to raise funds, overcoming some of the difficulties of raising money from traditional grant bodies, especially for early career researchers or those in developing countries (Wheat *et al.*, 2013).

In addition to funding science, the public can also actively engage in the process of research directly through citizen science projects. Whilst research conducted by non-professionals is certainly not a new concept, the numbers of projects involving citizen scientists are growing, especially in the fields of environmental science and ecology (Silvertown, 2009). Through catch records of amateur anglers and commercial net-fishery data extending back many years, research into fish and fisheries is uniquely placed to benefit from citizen science projects (Stuart-Smith *et al.*, 2013), which have effectively spanned generations of contributors. Similarly, REEF (www.reef.org) has been collecting reef-fish diversity and abundance data from trained volunteer divers for 27 years and the data have been successfully leveraged in hundreds of publications (Stallings, 2009; Serafy *et al.*, 2015). Citizen science can also help achieve important social outcomes, *e.g.* in establishing sustainable fisheries and marine protected areas (MPA) (Bonney *et al.*, 2014). As with crowdfunding, the best examples of citizen science typically encourage deeper engagement with the public and offer a pathway to the democratization of science.

FISHERIES POLICY AND GOVERNANCE

Conserving critical habitats is central to the sustainable management of fish species and populations. MPAs, networks of MPAs and marine conservation zones (MCZ) are widely accepted management tools for fish and other marine organisms that have been established in many countries (Harborne *et al.*, 2008; OSPAR Convention, 2013). The design of MPA networks could, however, benefit greatly from the integration of traditional survey data, along with modelling and connectivity data (Botsford *et al.*, 2009; Gruss *et al.*, 2014). From a social science perspective, there is a need to better understand public perceptions of marine-related conservation issues, *e.g.* fishery regulations, MPAs and MCZs and to incorporate these data into fisheries policy and governance frameworks. For example, there is high public support for MPAs, with surveys showing that people desire around 40% of the UK's marine waters to be protected (Hawkins *et al.*, 2016). But, while the public appears to realise that levels of coverage are well below 40%, there is still a substantial disconnect between perceived coverage of highly protected UK MPAs (11%) and actual MPA coverage (<0.1%); ultimately, this means that people believe the UK oceans receive a higher level of conservation than in reality (Hawkins *et al.*, 2016). Developing and implementing effective policies for fisheries management remains challenging because of the complexities of fisheries and the socio-political landscape under which they typically operate (Jentoft & Chuenpagdee, 2009). The establishment of guidelines or frameworks for fisheries policy and governance (FAO, 2015), however, have the potential to better address these challenges and provide appropriate implementable solutions.

CONCLUSIONS

Across all five of the research themes identified here, it is clear that innovative and novel tools are being employed to understand all aspects of the biology of fish populations. Notwithstanding, the authors call for the continued development of these new and emerging techniques. In particular, there is a need for better integration of these methods and resulting data, to inform scientifically sound management and conservation

of fish populations. It should be noted, however, that not infrequently, revolutionary methods have been promoted as providing the ability to offer unprecedented novel answers to long-standing practical problems. Unfortunately, the danger is that such methods can (by their novelty and the excitement surrounding them), blinker scientists into posing questions that showcase the methodology, rather than the biology [*e.g.* the plethora of papers that emerged in the early 1990s extolling the virtues of the random amplified polymorphic DNA (RAPD) technique]. The potentially reduced power of using any technique on its own (new or otherwise), in isolation of other apparently antiquated methods can turn out to be unnecessarily restrictive. Every technique has its limitations, but often the restrictions of one tool can be substantially alleviated by the inclusion of another approach (Goodwin *et al.*, 2016; Nielsen *et al.*, 2017), the marriage of which can provide a new angle for researching challenging biological problems. It is important that both traditional and emerging tools remain in the toolbox of fish biology research.

Likewise, when genetic-based assignment became popular, many researchers naively believed the days of tagging fish were over. It is now realised that due to the many stochastic drivers of population structure, genetic stock identification-based methodologies such as genetic assignment, do not always succeed. In such cases, there remains a significant role for tagging in fish and fisheries research. As tag sizes decrease and the deleterious effects of tag insertions on fish also decrease, we can anticipate that genetics and tagging will both continue to have a role to play. The importance of the relative roles of each technique will depend on the questions being addressed, the population structure of the study species and the scale of the questions being assessed.

A final example, which highlights the importance of applying inter-disciplinary and complimentary tools for understanding fish populations, was a five-year, multi-agency, E.U. funded project investigating the migration and distribution of *S. salar* in the north-east Atlantic (the SALSEA project; NASCO, 2008). The purpose was to understand not just where *S. salar* go, but what they eat, migration routes to feeding grounds and which waters and regions they pass through. The SALSEA project used a combination of genetics (microsatellites), SIA, at-sea trawls, tagging and gut contents analysis to assess the movements and diet of this species across the north-east Atlantic Ocean. As a result of applying these combined approaches, *S. salar* post-smolt movements have been confidently ascertained (Gilbey *et al.*, 2017). Nonetheless, even while this comprehensive study was being finalised, a similarly broad-ranging study was also being undertaken using SNPs (Bourret *et al.*, 2013). Arguably, this method offers both the potential for finer levels of stock discrimination and the ability to better explore patterns among functional loci, which may make microsatellite-based analysis redundant within a short period of time (although see Narum *et al.*, 2008).

Thus, the authors consider the continued development of emerging tools, together with the use of multiple methodologies and inter-disciplinary approaches, to represent the best avenues for further improving our understanding of fish populations. We implore scientists from unrelated fields to collaborate on such projects. The 50th Anniversary Symposium of The Fisheries Society of the British Isles represented one such event, where fish-focused researchers across diverse fields, came together to advance the state of fish biology.

We thank the organizing committee of the 50th Anniversary Symposium of The Fisheries Society of the British Isles for enabling the working group discussion that led to the development

of this review. Thanks also to the University of Exeter for hosting the symposium and to the numerous sponsors for funding its success. J.R.S. acknowledges the support of the SAMARCH project, an E.U. funded France Channel–Manche England Interreg initiative.

References

- Allendorf, F. W., Hohenlohe, P. A. & Luikart, G. (2010). Genomics and the future of conservation genetics. *Nature Reviews Genetics* **11**, 697–709. <https://doi.org/10.1038/nrg2844>
- Amann, R. L., Ludwig, W. & Schleifer, K. H. (1995). Phylogenetic identification and *in situ* detection of individual microbial cells without cultivation. *Microbiological Reviews* **59**, 143–169.
- Austin, B. (2006). The bacterial microflora of fish, revised. *The Scientific World Journal* **6**, 931–945. <https://doi.org/10.1100/tsw.2006.181>
- Austin, C. B., Hua Tan, M., Harrison, K. A., Peng Lee, Y., Croft, L. J., Sunnucks, P., Pavlova, A. & Ming Gan, H. (2017). *De novo* genome assembly and annotation of Australia's largest freshwater fish, the Murray cod (*Maccullochella peelii*) from Illumina and Nanopore sequencing reads. *Gigascience* **6**(8), 1–6. <https://doi.org/10.1093/gigascience/gix063>
- Benestan, L., Moore, J.-S., Sutherland, B. J. G., Le Luyer, J., Maaroufi, H., Rougeux, C., Normandeau, E., RYcroft, N., Atema, J., Harris, L. N., Tallman, R. F., Greenwood, S. J., Clark, F. K. & Bernatchez, L. (2017). Sex matters in massive parallel sequencing: evidence for biases in genetic parameter estimation and investigation of sex determination systems. *Molecular Ecology* **26**(24), 6767–6783. <https://doi.org/10.1111/mec.14217>
- Blanchet, S., Prunier, J. G. & De Kort, H. (2017). Time to go bigger: emerging patterns in macrogenetics. *Trends in Genetics* **33**, 579–580. <https://doi.org/10.1016/j.tig.2017.06.007>
- Boecklen, W. J., Yarnes, C. T., Cook, B. A. & James, A. C. (2011). On the use of stable isotopes in trophic ecology. *Annual Review of Ecology, Evolution and Systematics* **42**, 411–440. <https://doi.org/10.1146/annurev-ecolsys-102209-144726>
- Bograd, S. J., Block, B. A., Costa, D. P. & Godley, B. J. (2010). Biologging technologies: new tools for conservation. Introduction. *Endangered Species Research* **10**, 1–7. <https://doi.org/10.3354/esr00269>
- Bonney, R., Shirk, J. L., Phillips, T. B., Wiggins, A., Ballard, H. L., Miller-Rushing, A. J. & Parrish, J. K. (2014). Next steps for citizen science. *Science* **343**, 1436–1437. <https://doi.org/10.1126/science.1251554>
- Botsford, L. W., Brumbaugh, D. R., Grimes, C., Kellner, J. B., Largier, J., O'Farrell, M. R., Ralston, S., Soulanile, E. & Wespestad, V. (2009). Connectivity, sustainability and yield: bridging the gap between conventional fisheries management and marine protected areas. *Reviews in Fish Biology Fisheries* **19**, 69–95. <https://doi.org/10.1007/s11160-008-9092-z>
- Bourret, V., Kent, M. P., Primmer, C. R., Vasermägi, A., Karlsson, S., Hindar, K., McGinnity, P., Verspoor, E., Bernatchez, L. & Lien, S. (2013). SNP-array reveals genome-wide patterns of geographical and potential adaptive divergence across the natural range of Atlantic salmon (*Salmo salar*). *Molecular Ecology* **22**, 532–551.
- Budding, A. E., Grasman, M. E., Eck, A., Bogaards, J. A., Vandenbroucke-Grauls, C. M., van Bodegraven, A. A. & Savelkoul, P. H. (2014). Rectal swabs for analysis of the intestinal microbiota. *PLoS One* **9**, e101344. <https://doi.org/10.1371/journal.pone.0101344>
- Chang, D. & Shapiro, B. (2016). Using ancient DNA and coalescent-based methods to infer extinction. *Biology Letters* **12**, 20150822.
- Chistiakov, D. A., Helleman, B. & Volckaert, F. A. M. (2006). Microsatellites and their genomic distribution, evolution, function and applications: a review with special reference to fish genetics. *Aquaculture* **255**, 1–29.
- Christensen, V. & Walters, C. J. (2004). Ecopath with Ecosim: methods, capabilities and limitations. *Ecological Modelling* **172**, 109–139.
- Clusa, L., Ardura, A., Fernández, S., Roca, A. A. & García-Vázquez, E. (2017). An extremely sensitive nested PCR-RFLP mitochondrial marker for detection and identification of salmonids in eDNA from water samples. *PeerJ* **5**, e3045.

- Cooke, S. J., Midwood, J. D., Thiem, J. D., Klimley, P., Lucas, M. C., Thorstad, E. B., Eiler, J., Holbrook, C. & Ebner, B. C. (2013). Tracking animals in freshwater with electronic tags: past, present and future. *Animal Biotelemetry* **1**, 5.
- Cooke, S. J., Brownscombe, J. W., Raby, G. D., Broell, F., Hinch, S. G., Clark, T. D. & Semmens, J. M. (2016). Remote bioenergetics measurements in wild fish: opportunities and challenges. *Comparative Biochemistry and Physiology A* **202**, 23–37.
- Dahlgren, C. P., Buch, K., Rechisky, E. & Hixon, M. A. (2016). Multiyear tracking of Nassau grouper spawning migrations. *Marine and Coastal Fisheries: Dynamics, Management and Ecosystem Science* **8**, 522–535. <https://doi.org/10.1080/19425120.2016.1223233>
- Davey, J. W. & Blaxter, M. L. (2011). RADSeq: next-generation population genetics. *Briefings in Functional Genomics* **9**, 416–423.
- DeAngelis, D. L. & Grimm, V. (2014). Individual-based models in ecology after four decades. *F1000Prime Reports* **6**, 39. <https://doi.org/10.12703/P6-39>
- DeAngelis, D. L. & Mooij, W. M. (2005). Individual-based modelling of ecological and evolutionary processes. *Annual Review of Ecology, Evolution and Systematics* **36**, 147–168.
- DeNiro, M. J. & Epstein, S. (1977). Mechanism of carbon isotope fractionation associated with lipid synthesis. *Science* **197**, 261–263.
- Doi, H., Uchii, K., Takahara, T., Matsushashi, S., Yamanaka, H. & Minamoto, T. (2015). Use of droplet digital PCR for estimation of fish abundance and biomass in environmental DNA surveys. *PLoS One* **10**, e0122763.
- E.C (2000). Directive 2000/60/EC, establishing a framework for community action in the field of water policy. *Official Journal of the European Communities* **327**, 1–71 Available at www.eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2000:327:0001:0072:EN:PDF.
- Eklom, R. & Wolf, J. B. W. (2014). A field guide to whole-genome sequencing, assembly and annotation. *Evolutionary Applications* **7**, 1026–1042.
- Ellegren, H. & Galtier, N. (2016). Determinants of genetic diversity. *Nature Reviews Genetics* **17**, 422–433. <https://doi.org/10.1038/nrg.2016.58>
- Ellis, J. S., Gilbey, J., Armstrong, A., Balstad, T., Cauwelier, E., Cherbonnel, C., Consuegra, S., Coughlan, J., Cross, T. F., Crozier, W., Dillane, E., Ensing, D., García de Leániz, C., García-Vázquez, E., Griffiths, A. M., Hindar, K., Hjørleifsdóttir, S., Knox, D., Machado-Schiaffino, G., McGinnity, P., Meldrup, D., Nielsen, E. E., Olafsson, K., Primmer, C. R., Prodohl, P., Stradmeyer, L., Vähä, J. P., Verspoor, E., Wennevik, V. & Stevens, J. R. (2011). Microsatellite standardization and evaluation of genotyping error in a large multi-partner research programme for conservation of Atlantic salmon (*Salmo salar* L.). *Genetica* **139**, 353–367.
- Ellison, A. M. (2004). Bayesian inference in ecology. *Ecology Letters* **7**, 509–520.
- FAO (2014). The State of World Fisheries and Aquaculture Opportunities and Challenges 2014. Rome: FAO Available at <http://www.fao.org/3/a-i3720e.pdf>
- FAO (2015). *Voluntary Guidelines for Securing Sustainable Small-Scale Fisheries in the Context of Food Security and Poverty Eradication*. Rome: FAO Available at <http://www.fao.org/fishery/ssf/guidelines/en>
- Ferguson, M. M. & Danzmann, R. G. (1998). Role of genetic markers in fisheries and aquaculture: useful tools or stamp collecting? *Canadian Journal of Fisheries and Aquatic Sciences* **55**, 1553–1563.
- Freeland, J. R. (2016). The importance of molecular markers and primer design when characterizing biodiversity from environmental DNA. *Genome* **60**, 358–374.
- Gaulke, C. A., Barton, C. L., Proffitt, S., Tanguay, R. L. & Sharpton, T. J. (2016). Triclosan exposure is associated with rapid restructuring of the microbiome in adult zebrafish. *PLoS One* **11**, e0154632.
- Ghanbari, M., Kneifel, W. & Domig, K. J. (2015). A new view of the fish gut microbiome: advances from next-generation sequencing. *Aquaculture* **448**, 464–475. <https://doi.org/10.1016/j.aquaculture.2015.06.033>
- Gilbey, J., Coughlan, J., Wennevik, V., Prodohl, P., Stevens, J. R., Garcia de Leániz, C., Ensing, D., Cauwelier, E., Cherbonnel, C., Consuegra, S., Coulson, M. W., Cross, T. F., Crozier, W., Dillane, E., Ellis, J. S., García-Vázquez, E., Griffiths, A. M., Gudjonsson, S., Hindar, K., Karlsson, S., Knox, D., Machado-Schiaffino, G., Meldrup, D., Nielsen, E. E., Ólafsson, K., Primmer, C. R., Prusov, S., Stradmeyer, L., Vähä, J.-P., Veselov, A. J., Webster,

- L. M. I., McGinnity, P. & Verspoor, E. (2017). A microsatellite baseline for genetic stock identification of European Atlantic salmon (*Salmo salar* L.). *ICES Journal of Marine Science*. <https://doi.org/10.1093/icesjms/fsx184>
- Gobalet, K. W. (2001). A critique of faunal analysis; inconsistency among experts in blind tests. *Journal of Archaeological Science* **28**, 377e386.
- Goodrich, J. K., Di Rienzi, S. C., Poole, A. C., Koren, O., Walters, W. A., Caporaso, J. G., Knight, R. & Ley, R. E. (2014). Conducting a microbiome study. *Cell* **158**, 250–262. <https://doi.org/10.1016/j.cell.2014.06.037>
- Goodwin, J. C. A., King, R. A., Jones, J. I., Ibbotson, A. & Stevens, J. R. (2016). A small number of anadromous females drive reproduction in a brown trout (*Salmo trutta*) population in an English chalk stream. *Freshwater Biology* **61**, 1075–1089.
- Gordon, T. A. C., Harding, H. R., Clever, F. K., Davidson, I. K., Davison, W., Montgomery, D. W., Nedelec, S. L., Weatherhead, R. C., Windsor, F. M., Armstrong, J. D., Bardonnnet, A., Bergman, E., Britton, J. R., Côté, I. M., D'Agostino, D., Greenberg, L. A., Harborne, A. R., Kahilainen, K. K., Metcalfe, N. B., Mills, S. C., Milner, N. J., Mittermayer, F. H., Montorio, L., Prokkola, L. A., Rutterford, L. A., Salvanes, A. G. V., Simpson, S. D., Vainikka, A., Pinnegar, J. K. & Santos, E. M. (2018). Fishes in a changing world: learning from the past to promote sustainability of fish populations. *Journal of Fish Biology* **92**, 804–827. <https://doi.org/10.1111/jfb.13546>
- Gorokhova, E. (2017). Individual growth as a non-dietary determinant of the isotopic niche metrics. *Methods in Ecology and Evolution*. <https://doi.org/10.1111/2041-210X.12887>
- Gruss, A., Robinson, J., Heppell, S. S., Heppell, S. A. & Semmens, B. X. (2014). Conservation and fisheries effects of spawning aggregation marine protected areas: what we know, where we should go and what we need to get there. *ICES Journal of Marine Science* **71**, 1515–1534. <https://doi.org/10.1093/icesjms/fsu038>
- Guillot, G., Leblois, R., Coulon, A. & Frantz, A. C. (2009). Statistical methods in spatial genetics. *Molecular Ecology* **18**, 4734–4756. <https://doi.org/10.1111/j.1365-294X.2009.04410.x>
- Halfyard, E. A., Webber, D., Del Papa, J., Leadley, T., Kessel, S. T., Colborne, S. F. & Fisk, A. T. (2017). Evaluation of an acoustic telemetry transmitter designed to identify predation events. *Methods in Ecology and Evolution* **8**, 1063–1071. <https://doi.org/10.1111/2041-210X.12726>
- Hänfling, B., Lawson Handley, L., Read, D. S., Hahn, C., Li, J., Nichols, P., Blackman, R. C., Oliver, A. & Winfield, I. J. (2016). Environmental DNA metabarcoding of lake fish communities reflects long-term data from established survey methods. *Molecular Ecology* **25**, 3101–3119.
- Harborne, A. R., Mumby, P. J., Cappel, C. V., Dahlgren, C. P., Micheli, F., Holmes, K. E., Sanchirico, J. E., Broad, K., Elliot, I. A. & Brumbaugh, D. R. (2008). Reserve effects and natural variation in coral reef communities. *Journal of Applied Ecology* **45**, 1010–1018.
- Hart, M. L., Meyer, A., Johnson, P. J. & Ericsson, A. C. (2015). Comparative evaluation of DNA extraction methods from feces of multiple host species for downstream next-generation sequencing. *PLoS One* **10**, e0143334. <https://doi.org/10.1371/journal.pone.0143334>
- Hawkins, J. P., O'Leary, B. C., Bassett, N., Peters, H., Rakowski, S., Reeve, G. & Roberts, C. M. (2016). Public awareness and attitudes towards marine protection in the United Kingdom. *Marine Pollution Bulletin* **111**, 231–236.
- Hayden, E. C. (2015). Pint-sized DNA sequencer impresses first users. *Nature* **521**, 15–16.
- Hays, G. C., Bradshaw, C. J. A., James, M. C., Lovell, P. & Sims, D. W. (2007). Why do Argos satellite tags deployed on marine animals stop transmitting? *Journal of Experimental Marine Biology and Ecology* **349**, 52–60.
- Helyar, S. J., Hemmer-Hansen, J., Bekkevold, D., Taylor, M. I., Ogden, R., Limborg, M. T., Cariani, A., Maes, G. E., Dopere, E., Carvalho, G. & Nielsen, E. E. (2011). Application of SNPs for population genetics of nonmodel organisms: new opportunities and challenges: analytical approaches. *Molecular Ecology Resources* **11**, 123–136.
- Heupel, M. R., Semmens, J. M. & Hobday, A. J. (2006). Automated acoustic tracking of aquatic animals: scales, design and deployment of listening station arrays. *Marine and Freshwater Research* **57**, 1–13.
- Hight, B. V. & Lowe, C. G. (2007). Elevated body temperatures of adult female leopard sharks, *Triakis semifasciata*, while aggregating in shallow nearshore embayments: evidence for

- behavioral thermoregulation? *Journal of Experimental Marine Biology and Ecology* **352**, 114–128.
- Hussey, N. E., Kessel, S. T., Aarestrup, K., Cooke, S. J., Cowley, P. D., Fisk, A. T., Harcourt, R. G., Holland, K. N., Iverson, S. J., Kocik, J. F., Flemming, J. E. M. & Whoriskey, F. G. (2015). Aquatic animal telemetry: a panoramic window into the underwater world. *Science* **348**, 1221–1255642. <https://doi.org/10.1126/science.1255642>
- Ingerslev, H. C., von Gersdorff Jørgensen, L., Strube, M. L., Larsen, N., Dalsgaard, I., Boye, M. & Madsen, L. (2014). The development of the gut microbiota in rainbow trout (*Oncorhynchus mykiss*) is affected by first feeding and diet type. *Aquaculture* **424–425**, 24–34. <https://doi.org/10.1016/j.aquaculture.2013.12.032>
- Iwamoto, E. M., Myers, J. M. & Gustafson, R. G. (2012). Resurrecting an extinct salmon evolutionary significant unit: archived scales, historical DNA and implication for restoration. *Molecular Ecology* **21**, 1567–1582.
- Jansen, H. J., Liem, M., Jong-Raadsen, S. A., Dufour, S., Weltzien, F.-A., Swinkels, W., Koelewijn, A., Palstra, A. P., Pelster, B., Spaink, H. P., Van den Thillart, G. E., Dirks, R. P. & Henkel, C. V. (2017). Rapid de novo assembly of the European eel genome from nanopore sequencing reads. *Sci Rep* **7**, 101907. <https://doi.org/10.1101/101907>
- Jentoft, S. & Chuenpagdee, R. (2009). Fisheries and coastal governance as a wicked problem. *Marine Policy* **33**, 553–560. <https://doi.org/10.1016/j.marpol.2008.12.002>
- Jepsen, N., Thorstad, E. B., Havn, T. & Lucas, M. C. (2015). The use of external electronic tags on fish: an evaluation of tag retention and tagging effects. *Animal Biotelemetry* **3**, 1–23. <https://doi.org/10.1186/s40317-015-0086-z>
- Kansa, E. C. (2010). Open context in context: cyberinfrastructure and distributed approaches to publish and preserve archaeological data. *The SAA Archaeological Record* **10**, 12–16.
- Kurle, C. M. & McWhorter, J. (2017). Spatial and temporal variability within marine isoscaes: implications for interpreting stable isotope data from marine systems. *Marine Ecology Progress Series* **568**, 31–45. <https://doi.org/10.3354/meps12045>
- Lacoursière-Roussel, A., Rosabal, M. & Bernatchez, L. (2016). Estimating fish abundance and biomass from eDNA concentrations: variability among capture methods and environmental conditions. *Molecular Ecology Resources* **16**, 1401–1414.
- Lauridsen, R. B., Moore, A., Privitera, L., Gregory, S. D., Beaumont, W. R. C. & Kavanagh, A. J. (2017). Migration behaviour and loss rate of trout smolts in the transitional zone between freshwater and saltwater. In *Proceedings of the 2nd International Sea Trout Symposium Sea Trout: Science & Management* (Harris, G., ed), pp. 292–307. Leicester: Matador.
- Laval, G. & Excoffier, L. (2004). SIMCOAL 2.0: a program to simulate genomic diversity over large recombining regions in a subdivided population with a complex history. *Bioinformatics* **20**, 2485–2487. <https://doi.org/10.1093/bioinformatics/bth264>
- Layman, C. A., Araujo, M. S., Boucek, R., Hammerschlag-Peyer, C. M., Harrison, E., Jud, Z. R., Matich, P., Rosenblatt, A. E., Vaudo, J. J., Yeager, L. A., Post, D. M. & Bearhop, S. (2012). Applying stable isotopes to examine food-web structure: an overview of analytical tools. *Biological Reviews* **87**, 545–562. <https://doi.org/10.1111/j.1469-185X.2011.00208.x>
- Lea, J. S. E., Humphries, N. E., von Brandis, R. G., Clarke, C. R. & Sims, D. W. (2016). Acoustic telemetry and network analysis reveal the space use of multiple reef predators and enhance marine protected area design. *Proceedings of the Royal Society B* **283**(1834), 20160717. <https://doi.org/10.1098/rspb.2016.0717>
- Lien, S., Koop, B. F., Sandve, S. R., Miller, J. R., Matthew, P., Leong, J. S., Minkley, D. R., Zimin, A., Grammes, F., Grove, H., et al. (2016). The Atlantic salmon genome provides insights into rediploidization. *Nature* **533**, 200–205.
- Linden, M. V. & Webley, L. (2012). Introduction: development-led archaeology in north-west Europe. In *Development-led Archaeology in North-West Europe* (Bradley, R., Haselgrove, C., Linden, M. V. & Webley, L., eds). Oxford: Oxbow Books.
- Llewellyn, M. S., Boutin, S., Hoseinifar, S. H. & Derome, N. (2014). Teleost microbiomes: the state of the art in their characterization, manipulation and importance in aquaculture and fisheries. *Frontiers in Microbiology* **5**, 1–17. <https://doi.org/10.3389/fmicb.2014.00207>
- Llewellyn, M. S., McGinnity, P., Dionne, M., Letourneau, J., Thonier, F., Carvalho, G. R., Creer, S. & Derome, N. (2016). The biogeography of the Atlantic salmon (*Salmo salar*) gut microbiome. *The ISME Journal* **10**, 1280–1284. <https://doi.org/10.1038/ismej.2015.189>

- MacDonald, A. J. & Sarre, S. D. (2017). A framework for developing and validating taxon-specific primers for specimen identification from environmental DNA. *Molecular Ecology Resources* **17**, 708–720.
- MacKenzie, K. M., Longmore, C., Preece, C., Lucas, C. H. & Trueman, C. N. (2014). Testing the long-term stability of marine isoscapes in shelf seas using jellyfish tissues. *Biogeochemistry* **121**, 441–454. <https://doi.org/10.1007/s10533-014-0011-1>
- Macqueen, D. J., Primmer, C. R., Houston, R. D., Nowak, B. F., Bernatchez, L., Bergseth, S., Davidson, W. S., Gallardo-Escárate, C., Goldammer, T., Guiguen, Y., Iturra, P., Kijas, J. W., Koop, B. F., Lien, S., Martin, S. A. M., McGinnity, P., Montecino, M., Naish, K. A., Nichols, K. M., Ólafsson, K., Omholt, S. W., Palti, Y., Plastow, G. S., Rexroad, C. E. III, Rise, M. L., Ritchie, R. J., Sandve, S. R., Schulte, P. M., Tello, A., Vidal, R., Vik, J. O., Wargelius, A., Yáñez, J. M. & The FAASG consortium (2017). Functional annotation of all salmonid genomes (FAASG): an international initiative supporting future salmonid research, conservation and aquaculture. *BMC Genomics* **18**, 1–9.
- Magozzi, S., Yool, A., Van der Zanden, H. B., Wunder, M. B. & Trueman, C. N. (2017). Using ocean models to predict spatial and temporal variation in marine carbon isotopes. *Ecosphere* **8**, e01763. <https://doi.org/10.1002/ecs2.1763>
- McClelland, J. W. & Montoya, J. P. (2002). Trophic relationships and the nitrogen isotopic composition of amino acids in plankton. *Ecology* **83**, 2173–2180.
- McClenachan, L. (2009). Documenting loss of large trophy fish from the Florida keys with historical photographs. *Conservation Biology* **23**, 636–643.
- McMahon, K. W. & McCarthy, M. D. (2016). Embracing variability in amino acid $\delta^{15}\text{N}$ fractionation: mechanisms, implications and applications for trophic ecology. *Ecosphere* **7**, e01511.
- Méthé, B. A., Nelson, K. E., Pop, M., Creasy, H. H., Giglio, M. G., Huttenhower, C., Gevers, D., Petrosino, J. F., Abubucker, S., Badger, J. H. & Chinwalla, A. T. (2012). A framework for human microbiome research. *Nature* **486**, 215–221. <https://doi.org/10.1038/nature11209>
- Methling, C., Tudorache, C., Skov, P. V. & Steffensen, J. F. (2011). Pop up satellite tags impair swimming performance and energetics of the European eel (*Anguilla anguilla*). *PLoS One* **6**, e20797. <https://doi.org/10.1371/journal.pone.0020797>
- Mintram, K. S., Brown, A. R., Maynard, S. K., Thorbek, P. & Tyler, C. R. (2017). Capturing ecology in modeling approaches applied to environmental risk assessment of endocrine active chemicals in fish. *Critical Reviews in Toxicology*. <https://doi.org/10.1080/10408444.2017.1367756>
- Miya, M., Sato, Y., Fukunaga, T., Sado, T., Poulsen, J. Y., Sato, K., Minamoto, T., Yamamoto, S., Yamanaka, H., Araki, H. & Kondoh, M. (2015). MiFish, a set of universal PCR primers for metabarcoding environmental DNA from fishes: detection of more than 230 subtropical marine species. *Open Science* **2**, 150088.
- Murchie, K. J., Shultz, A. D., Stein, J. A., Cooke, S. J., Lewis, J., Franklin, J., Vincent, G., Brooks, E. J., Claussen, J. E. & Philipp, D. P. (2015). Defining adult bonefish (*Albula vulpes*) movement corridors around Grand Bahama in the Bahamian Archipelago. *Environmental Biology of Fishes* **98**, 2203–2212. <https://doi.org/10.1007/s10641-015-0422-4>
- Murray, G. G. R., Soares, A. E. R., Novak, B. J., Schaefer, N. K., Cahill, J. A., Baker, A. J. & Shapiro, B. (2017). Natural selection shaped the rise and fall of passenger pigeon genomic diversity. *Science* **358**(6365), 951–954. <https://doi.org/10.1101/154294>
- Musyl, M. K., Domeier, M. L., Nasby-Lucase, N., Brill, R. W., McNaughton, L. M., Swimmer, J. Y., Lutcavage, M. S., Wilson, S. G., Galuardi, B. & Liddle, J. B. (2011). Performance of pop-up satellite archival tags. *Marine Ecology Progress Series* **433**, 1–28.
- Naito, Y. I., Chikaraishi, Y., Drucker, D. G., Ohkouchi, N., Semal, P., Wißing, C. & Bocherens, H. (2016). Ecological niche of Neanderthals from spy cave revealed by nitrogen isotopes of individual amino acids in collagen. *Journal of Human Evolution* **93**, 82–90.
- Narum, S. R., Banks, M., Beacham, T. D., Bellinger, M. R., Campbell, M. R., DeKoning, J., Elz, A., Guthrie, C. M. III, Kozfkay, C., Miller, K. M., Moran, P., Phillips, R., Seeb, L. W., Smith, C. T., Warheit, K., Young, S. F. & Garza, J. C. (2008). Differentiating salmon populations at broad and fine geographic scales with microsatellites and SNPs. *Molecular Ecology* **17**, 3464–3477.

- Nash, K. L., Bijoux, J., Robinson, J., Wilson, S. K. & Graham, N. A. J. (2016). Harnessing fishery-independent indicators to aid management of data-poor fisheries: weighing habitat and fishing effects. *Ecosphere* **7**, e01362.
- Neuenschwander, S. (2006). AQUASPLATCHE: a program to simulate genetic diversity in populations living in linear habitats. *Molecular Ecology Notes* **6**, 583–585.
- Newsome, S. D., del Rio, C. M., Bearhop, S. & Phillips, D. L. (2007). A niche for isotopic ecology. *Frontiers in Ecology and the Environment* **5**, 429–436.
- Nielsen, J. M., Clare, E. L., Hayden, B., Brett, M. T. & Kratina, P. (2017). Diet tracing in ecology: method comparison and selection. *Methods in Ecology and Evolution (Online)*. <https://doi.org/10.1111/2041-210X.12869>
- Ólafsdóttir, G. Á., Westfall, K. M., Edvardsson, R. & Pálsson, S. (2014). Historical DNA reveals the demographic history of Atlantic cod (*Gadus morhua*) in medieval and early modern Iceland. *Proceedings of the Royal Society B* **281**, 20132976.
- Orton, D. C. (2016). Archaeology as a tool for understanding past marine resource use and its impact. In *Perspectives on Oceans Past* (Schwerdtner Máñez, K. & Poulsen, B., eds), pp. 47–69. Dordrecht: Springer.
- Papastamatiou, Y. P., Meyer, C. G., Carvalho, F., Dale, J. J., Hutchinson, M. R. & Holland, K. N. (2013). Telemetry and random-walk models reveal complex patterns of partial migration in a large marine predator. *Ecology* **94**, 2595–2606.
- Paris, J. R., Stevens, J. R. & Catchen, J. M. (2017). Lost in parameter space: a road map for stacks. *Methods in Ecology and Evolution* **8**, 1360–1373. <https://doi.org/10.1111/2041-210X.12775>
- Pauli, J. N., Newsome, S. D., Cook, J. A., Harrod, C., Steffan, S. A., Baker, C. J. O., Ben-David, M., Bloom, D., Bowen, G. J., Cerling, T. E., Cicero, C., Cook, C., Dohm, M., Dharampal, P. S., Graves, G., Gropp, R., Hobson, K. A., Jordan, C., MacFadden, B., Pilaar Birch, S., Poelen, J., Ratnasingham, S., Russell, L., Stricker, C. A., Uhen, M. D., Yarnes, C. T. & Hayden, B. (2017). Opinion: why we need a centralized repository for isotopic data. *Proceedings of the National Academy of Sciences* **114**, 2997–3001. <https://doi.org/10.1073/pnas.1701742114>
- Pauls, S. U., Alp, M., Bálint, M., Bernabò, P., Čiampor, F., Čiamporová-Začovičová, Z., Finn, D. S., Kohout, J., Leese, F., Lencioni, V., Paz-Vinas, I. & Monaghan, M. T. (2014). Integrating molecular tools into freshwater ecology: developments and opportunities. *Freshwater Biology* **59**, 1559–1576. <https://doi.org/10.1111/fwb.12381>
- Peng, B. & Kimmal, M. (2005). SimuPOP: a forward-time population genetics simulation environment. *Bioinformatics* **21**, 3686–3687.
- Pine, W. E., Pollock, K. H., Hightower, J. E., Kwak, T. J. & Rice, J. A. (2003). A review of tagging methods for estimating fish population size and components of mortality. *Fisheries* **28**, 10–13. [https://doi.org/10.1577/1548-8446\(2003\)28](https://doi.org/10.1577/1548-8446(2003)28)
- Richter, K. K., Wilson, J., Jones, A. K. G., Buckley, M., van Doorn, N. & Collins, M. J. (2011). Fish 'n chips: ZooMS peptide mass fingerprinting in a 96 well plate format to identify fish bone fragments. *Journal of Archaeological Science* **38**, 1502–1510.
- Ritchie, M. D., Holzinger, E. R., Li, R., Pendergrass, S. A. & Kim, D. (2015). Methods of integrating data to uncover genotype-phenotype interactions. *Nature Reviews Genetics* **16**, 85–97. <https://doi.org/10.1038/nrg3868>
- Rodríguez-Ezpeleta, N., Bradbury, I. R., Mendibil, I., Álvarez, P., Cotano, U. & Irigoien, X. (2016). Population structure of Atlantic mackerel inferred from RAD-seq-derived SNP markers: effects of sequence clustering parameters and hierarchical SNP selection. *Molecular Ecology Resources* **16**, 991–1001. <https://doi.org/10.1111/1755-0998.12518>
- Rogers, A., Blanchard, J. L. & Mumby, P. J. (2014). Vulnerability of coral reef fisheries to a loss of structural complexity. *Current Biology* **24**, 1000–1005.
- Salipante, S. J., Kawashima, T., Rosenthal, C., Hoogstraat, D. R., Cummings, L. A., Sengupta, D. J., Harkins, T. T., Cookson, B. T. & Hoffman, N. G. (2014). Performance comparison of Illumina and ion torrent next-generation sequencing platforms for 16S rRNA-based bacterial community profiling. *Applied and Environmental Microbiology* **80**, 7583–7591. <https://doi.org/10.1128/AEM.02206-14>

- Schrider, D. R. & Kern, A. D. (2014). Discovering functional DNA elements using population genomic information: a proof of concept using human mtDNA. *Genome Biology and Evolution* **6**, 1542–1548. <https://doi.org/10.1093/gbe/evu116>
- Serafy, J. E., Shideler, G. S., Araújo, R. J. & Nagelkerken, I. (2015). Mangroves enhance reef fish abundance at the Caribbean regional scale. *PLoS One* **10**, e0142022. <https://doi.org/10.1371/journal.pone.0142022>
- Sevellec, M., Pavey, S. A., Boutin, S., Filteau, M., Derome, N. & Bernatchez, L. (2014). Microbiome investigation in the ecological speciation context of lake whitefish (*Coregonus clupeaformis*) using next-generation sequencing. *Journal of Evolutionary Biology* **27**, 1029–1046. <https://doi.org/10.1111/jeb.12374>
- Shin, Y.-J., Shannon, L. J., Bundy, A., Coll, M., Aydin, K., Bez, N., Blanchard, J. L., Borges, M. D. F., Diallo, I., Diaz, E., Heymans, J. J., Hill, L., Johannesen, E., Jouffre, D., Kifani, S., Labrosse, P., Link, J. S., Mackinson, S., Masski, H., Möllmann, C., Neira, S., Ojaveer, H., Abdallahi, K. O. M., Perry, I., Thiao, D., Yemane, D. & Cury, P. M. (2010). Using indicators for evaluating, comparing and communicating the ecological status of exploited marine ecosystems. 2. Setting the scene. *ICES Journal of Marine Science* **67**, 692–716.
- Shogren, A. J., Tank, J. L., Andruszkiewicz, E., Olds, B., Mahon, A. R., Jerde, C. L. & Bolster, D. (2017). Controls on eDNA movement in streams: transport, retention and resuspension. *Scientific Reports* **7**, 5065. <https://doi.org/10.1038/s41598-017-05223-1>
- Sigsgaard, E. E., Nielsen, I. B., Bach, S. S., Lorenzen, E. D., Robinson, D. P., Knudsen, S. W., Pedersen, M. W., Al Jaidah, M., Orlando, L., Willerslev, E. & Møller, P. R. (2016). Population characteristics of a large whale shark aggregation inferred from seawater environmental DNA. *Nature Ecology & Evolution* **1**, 0004.
- Silvertown, J. (2009). A new dawn for citizen science. *Trends in Ecology & Evolution* **24**, 467–471. <https://doi.org/10.1016/j.tree.2009.03.017>
- Smith, N. S. & Zeller, D. K. (2015) Unreported catch and tourist demand on local fisheries of small island states: these case of The Bahamas, 1950–2010. *Fishery Bulletin* **114**, 117–131. <https://doi.org/10.7755/FB.114.1.10>
- Song, W., Li, L., Huang, H., Jiang, K., Zhang, F., Chen, X., Zhao, M. & Ma, L. (2016). The gut microbial community of antarctic fish detected by 16S rRNA gene sequence analysis. *BioMedical Research International*. <https://doi.org/10.1155/2016/3241529>
- Sousa, L. L., Xavier, R., Costa, V., Humphries, N. E., Trueman, C., Rosa, R., Sims, D. W. & Queiroz, N. (2016). DNA barcoding identifies a cosmopolitan diet in the ocean sunfish. *Scientific Reports* **6**, 1–9.
- Stallings, C. D. (2009). Fishery-independent data reveal negative effect of human population density on Caribbean predatory fish communities. *PLoS One* **4**, e5333. <https://doi.org/10.1371/journal.pone.0005333>
- Star, B., Nederbragt, A. J., Jentoft, S., Grimholt, U., Malmstrom, M., Greger, T. F., Rounge, T. B., Paulsen, J., Solbakken, M. H., Sharma, A., Wetten, O. F., Lanzén, A., Winer, R., Knight, J., Vogel, J.-H., Aken, B., Andersen, Ø., Lagesen, K., Tooming-Klunderud, A., Edvardsen, R. B., Tina, K. G., Espelund, M., Nepal, C., Previti, C., Karlsen, B. O., Moum, T., Skage, M., Berg, P. R., Gjølven, T., Kuhl, H., Thorsen, J., Malde, K., Reinhardt, R., Du, L., Johansen, S. D., Searle, S., Lien, S., Nilsen, F., Jonassen, I., Omholt, S. W., Stenseth, N. C. & Jakobsen, K. S. (2011). The genome sequence of Atlantic cod reveals a unique immune system. *Nature* **477**, 207–210.
- Stark, A., Lin, M. F., Kheradpour, P., Pedersen, J. S., Parts, L., Carlson, J. W., Crosby, M. A., Rasmussen, M. D., Roy, S., Deoras, A. N., Ruby, J. G., Brennecke, J., Harvard FlyBase Curators, Berkeley Drosophila Genome Project, Hodges, E., Hinrichs, A. S., Caspi, A., Paten, B., Park, S.-W., Han, M. V., Maeder, M. L., Polansky, B. J., Robson, B. E., Aerts, S., van Helden, J., Hassan, B., Gilbert, D. G., Eastman, D. A., Rice, M., Weir, M., Hahn, M. W., Park, Y., Dewey, C. N., Pachter, L., Kent, W. J., Haussler, D., Lai, E. C., Bartel, D. P., Hannon, G. J., Kaufman, T. C., Eisen, M. B., Clark, A. G., Smith, D., Celniker, S. E., Gelbart, W. & Kellis, M. (2007). Discovery of functional elements in 12 drosophila genomes using evolutionary signatures. *Nature* **450**, 219–232. <https://doi.org/10.1038/nature06340>
- Stuart-Smith, R. D., Bates, A. E., Lefcheck, J. S., Duffy, J. E., Baker, S. C., Thomson, R. J., Stuart-Smith, J. F., Hill, N. A., Kininmonth, S. J., Airoldi, L., Becerro, M. A., Campbell, S. J., Dawson, T. P., Navarrete, S. A., Soler, G. A., Strain, E. M. A., Willis, T. J. & Edgar,

- G. J. (2013). Integrating abundance and functional traits reveals new global hotspots of fish diversity. *Nature* **501**, 539–542.
- Tarnecki, A. M., Burgos, F. A., Ray, C. L. & Arias, C. R. (2017). Fish intestinal microbiome: diversity and symbiosis unravelled by metagenomics. *Journal of Applied Microbiology* **123**, 2–17. <https://doi.org/10.1111/jam.13415>
- Thomsen, P. F., Kielgast, J., Iversen, L. L., Møller, P. R., Rasmussen, M. & Willerslev, E. (2012). Detection of a diverse marine fish fauna using environmental DNA from seawater samples. *PLoS One* **7**, e41732.
- Thurstan, R. H. & Roberts, C. M. (2010). Ecological meltdown in the firth of Clyde, Scotland: two centuries of change in a coastal marine ecosystem. *PLoS One* **5**, e11767.
- Thurstan, R. H., Brockington, S. & Roberts, C. M. (2010). The effects of 118 years of industrial fishing on UK bottom trawl fisheries. *Nature Communications* **1**, 15.
- Trueman, C. N., MacKenzie, K. M. & Palmer, M. R. (2012). Identifying migrations in marine fishes through stable-isotope analysis. *Journal of Fish Biology* **81**, 826–847.
- Turner, C. R., Barnes, M. A., Xu, C. C., Jones, S. E., Jerde, C. L. & Lodge, D. M. (2014). Particle size distribution and optimal capture of aqueous microbial eDNA. *Methods in Ecology and Evolution* **5**, 676–684.
- Uchii, K., Doi, H. & Minamoto, T. (2016). A novel environmental DNA approach to quantify the cryptic invasion of non-native genotypes. *Molecular Ecology Resources* **16**, 415–422.
- Ulibarri, R. M., Bonar, S. A., Rees, C., Amberg, J., Ladell, B. & Jackson, C. (2017). Comparing efficiency of American Fisheries Society standard snorkeling techniques to environmental DNA sampling techniques. *North American Journal of Fisheries Management* **37**, 644–651.
- Valentini, A., Taberlet, P., Miaud, C., Civade, R., Herder, J., Thomsen, P. F., Bellemain, E., Besnard, A., Coissac, E., Boyer, F. & Gaboriaud, C. (2016). Next-generation monitoring of aquatic biodiversity using environmental DNA metabarcoding. *Molecular Ecology* **25**, 929–942.
- Van Neer, W. & Ervynck, A. (2010). Defining “natural” fish communities for fishery management purposes: biological, historical and archaeological approaches. In *Ecologies and Economies in Medieval and Early Modern Europe* (Bruce, S. G., ed), pp. 188–218. Leiden: Brill.
- Vellend, M. (2003). Island biogeography of genes and species. *The American Naturalist* **162**, 358–365.
- Vellend, M., Lajoie, G., Bourret, A., Múrria, C., Kembel, S. W. & Garant, D. (2014). Drawing ecological inferences from coincident patterns of population and community-level biodiversity. *Molecular Ecology* **23**, 2890–2901. <https://doi.org/10.1111/mec.12756>
- Voegeli, F. A., Smale, M. J., Webber, D. M., Andrade, Y. & O’Dor, R. K. (2001). Ultrasonic telemetry, tracking and automated monitoring technology for sharks. *Environmental Biology of Fishes* **60**, 267–281.
- Wang, Z., Gerstein, M., & Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics*, **10**(1), 57–63.
- Weijerman, M., Fulton, E. A., Kaplan, I. C., Gorton, R., Leemans, R., Mooij, W. M. & Brainard, R. E. (2015). An integrated coral reef ecosystem model to support resource management under a changing climate. *PLoS One* **10**, e0144165.
- Wheat, R. E., Wang, Y., Byrnes, J. E. & Ranganathan, J. (2013). Raising money for scientific research through crowdfunding. *Trends in Ecology & Evolution* **28**, 71–72. <https://doi.org/10.1016/j.tree.2012.11.001>
- Wheeler, A. & Jones, A. K. G. (1989). *Fishes. Cambridge Manuals in Archaeology*. Cambridge: Cambridge University Press.
- Whoriskey, F. G. (2015). The ocean tracking network: a global partnership uses electronic tagging technologies to track the movements of aquatic animals, answer science questions, stimulate new technology development and assist with sustainable development of the ocean. *Development and International Law* **47**, 221–232.
- Wilcox, T. M., McKelvey, K. S., Young, M. K., Lowe, W. H. & Schwartz, M. K. (2015). Environmental DNA particle size distribution from brook trout (*Salvelinus fontinalis*). *Conservation Genetics Resources* **7**, 639–641.
- Worm, B., Hilborn, R., Baum, J. K., Branch, T. A., Collie, J. S., Costello, C., Fogarty, M. J., Fulton, E. A., Hutchings, J. A., Jennings, S., Jensen, O. P., Lotze, H. K., Mace, P. M.,

- McClanahan, T. R., Minto, C., Palumbi, S. R., Parma, A. M., Ricard, D., Rosenberg, A. A., Watson, R. & Zeller, D. (2009). Rebuilding global fisheries. *Science* **325**, 578–585. <https://doi.org/10.1126/science.1173146>
- Xia, J. H., Lin, G., Fu, G. H., Wan, Z. Y., Lee, M., Wang, L., Liu, X. J. & Yue, G. H. (2014). The intestinal microbiome of fish under starvation. *BMC Genomics* **15**, 266. <https://doi.org/10.1186/1471-2164-15-266>
- Yamamoto, S., Minami, K., Fukaya, K., Takahashi, K., Sawada, H., Murakami, H., Tsuji, S., Hashizume, H., Kubonaga, S., Horiuchi, T. & Hongo, M. (2016). Environmental DNA as a ‘snapshot’ of fish distribution: a case study of Japanese jack mackerel in Maizuru Bay, sea of Japan. *PLoS One* **11**, e0149786.
- Zeller, D., Harper, S., Zyllich, K. & Pauly, D. (2015). Synthesis of underreported small-scale fisheries catch in Pacific island waters. *Coral Reefs* **34**, 25–39.

Electronic References

- Callou, C. (2009). Inventaires archéozoologiques et archéobotaniques de France (Inventaire National Du Patrimoine Naturel). *Muséum National d'Histoire Naturelle* Available at inpn.mnhn.fr/espece/jeudonnees/3471?lg=en
- Davis, R., Baumgartner, M., Comeau, A., Cunningham, D., Davies, K., Furlong, A., Johnson, H., L’Orsa, S., Ross, T., Taggart, C. & Whoriskey, F. (2016). Tracking whales on the Scotian Shelf using passive acoustic monitoring on ocean gliders. In *Oceans 2016*. Piscataway, NJ: Institute of Electrical and Electronics Engineers Available at www.ieeexplore.ieee.org/document/7761461/authors
- Froese, R. & Pauly, D. (Eds) (2017). FishBase. Available at www.fishbase.org.
- NASCO (2008) Salmon at Sea – SALSEA – An International Cooperative Research Programme on Salmon at Sea. *NASCO Report: SAL 04(5)*. Edinburgh: North Atlantic Salmon Conservation Organization. www.nasco.int/sas/pdf/archive/other_reports/salsea_programme.pdf (accessed 31st October 2017).
- OSPAR Convention (2013). *An Assessment of the Ecological Coherence of the OSPAR Network of Marine Protected Areas in 2012*. London: OSPAR Commission Available at www.ospar.org/about/publications?q=An+assessment+of+the+Ecological+coherence+of+the+OSPAR+Network+of+Marine+Protected+Areas+in+2012&a=&y=2013&s=

Annexe III :

Informations pratiques sur le protocole d'échantillonnage du matériel biologique effectué pour ce travail de thèse :

1. Description des sites de collecte de la zone d'étude :

- Villages :

Le Haut-Maroni recouvre au total 19 points d'habitation Amérindiens répertoriés par la municipalité de Maripasoula, et deux groupements de villages sous la juridiction de Papaïchton. Les cinq sites choisis pour cette étude correspondent toutefois aux cinq principaux bassins de vie Amérindiens et Bushinengués du bassin versant. Ils correspondent également aux sites d'étude suivis par Guillaume Longin lors de la première partie du Programme Pêche, au cours de laquelle il avait débuté la collecte de matériel génétique en vue de cette thèse. La figure III.1 indique, pour chaque village, l'étendue spatiale des principales activités de pêche à la journée (Longin 2016).

La collecte de matériel biologique dans les villages s'est effectuée de manière relativement régulière dans le temps entre le premier trimestre 2014 et le second trimestre 2016. Dans chacun des villages, la (les) personne(s) s'occupant de la collecte étai(en)t accompagnée(s) d'un habitant du village (ou d'une localité voisine) employé par le Parc Amazonien de Guyane pour assurer la communication avec les pêcheurs. Généralement, des prélèvements de tissu épithélial (nageoire) étaient réalisés sur les spécimens débarqués par les pêcheurs - sous réserve d'avoir obtenu leur accord au préalable. Une à deux « tournées » pouvaient ainsi avoir lieu au cours de la journée de manière à visiter chaque pêcheur : à l'aube lors du relevé des filets posés de nuit, et au crépuscule pour le relevé de certains filets posés en journée. Plus sporadiquement, des sorties supplémentaires pouvaient avoir lieu, généralement en compagnie de l'accompagnant employé du Parc Amazonien de Guyane, dans le but de cibler plus spécifiquement certains spécimens et atteindre les quotas de collecte fixés pour certaines espèces plus rarement rencontrées.

Papaïchton :

La dénomination Papaïchton recouvre ici l'ensemble du bassin de vie qui s'étend du bourg de Papaïchton-Cormotibo en lui-même, passant par le bassin de vie Loca-Assici-Agodé, jusqu'au niveau des Abattis Cottica. Ce bassin de vie essentiellement Bushinengué s'étend le long du cours principal plutôt calme de la Lawa, en aval de Maripasoula. Nous n'avons pas fait ici la distinction entre les spécimens capturés à différents points géographiques de cette zone, car trop peu nombreuses pour pouvoir constituer des sous-échantillons. Cependant, l'essentiel des captures réalisées dans cette région proviennent des Abattis Cottica bien que la zone de pêche dans son ensemble soit la plus importante en surface (Fig.III.1).

Wempi (Kayodé) :

Le village de Kayodé se situe sur une portion calme du tributaire du Tampok. A ce niveau, les pêches s'effectuent principalement entre la confluence Tampok et Lawa à l'aval, et le saut Degrad Roche à l'amont, au niveau de la confluence entre Tampok et Waki.

Taluen & Twenké (Opoya) :

Taluen et Twenké sont deux villages attenants qui se situent à quelques kilomètres en amont de la confluence entre Lawa et Tampok. Ils se situent sur une portion très calme de la Lawa et l'essentiel des

captures s'effectue entre les deux sauts importants que sont Domofou Soula en amont et Awara Soula en aval. La zone de pêche aval peut toutefois s'étendre parfois jusqu'à la confluence avec le Tampok.

Antecume-Pata :

Ce village se trouve en plein de cœur de la très importante zone de sauts, de rapides et d'îlets que constitue la confluence entre les tributaires du Litani et du Marouini. Les captures peuvent donc provenir du cours principal de la Lawa - qui prend naissance à cet endroit, ou des quelques kilomètres de tronçon aval des deux tributaires susmentionnés.

Pilima-Pata (Pidima) :

Pilima-Pata n'est distant d'Antecume-Pata que de quelques kilomètres. Ce village, situé sur une portion calme du Litani, est cependant beaucoup plus isolé du reste du bassin versant du Maroni par une succession de sauts très importants et difficilement franchissables, qui limite les déplacements quotidien des habitants et les pousse à une plus forte autonomie. La zone de pêche de Pilima-Pata se limite en aval à l'amorce des premiers sauts qui la séparent de la confluence avec la Lawa, et s'étend un peu plus en amont du Litani.

- **Sites de référence :**

La collecte de matériel biologique en dehors des villages, c'est-à-dire au niveau des sites de référence situés sur les différents tributaires du Haut-Maroni, s'est effectuée au cours de campagnes scientifiques dédiées. Les quatre sites les plus amont (Saut l'Inspecteur sur la Waki, Saut Pierkourou sur le Tampok, Saut Wayo (Wayo Gaan Soula) sur le Marouini et Saut Lavaud (Elëüéletpe icholi) sur le Litani) correspondent à des zones au-delà desquelles l'accessibilité en pirogue avec matériel n'était plus possible. Les trois sites intermédiaires (Saut Tampok sur le Tampok, Langa Soula sur le Marouini et Apsik Icholi sur le Litani) ont été choisis pour leur localisation relativement équidistante des sites amont et de l'embouchure respective de chaque tributaire, et leur capacité à accueillir un bivouac de quelques jours.

La collecte de matériel biologique sur les sites de référence s'est effectuée de manière similaire, entre le premier trimestre 2014 et le début d'année 2016, par du personnel du Parc Amazonien de Guyane, du bureau d'étude Hydreco et de l'INRA mais également par des pêcheurs de différents villages du Haut-Maroni. Les captures ont été effectuées à l'aide de plusieurs engins de pêche, notamment le filet tramail et le filet épervier.

Sur l'ensemble des sites d'étude (villages et sites de référence), chaque spécimen pêché a été identifié, mesuré et pesé (poids frais non-vidé) avant prélèvement de tissu biologique. Chaque prélèvement de tissu a été fixé dans l'éthanol et stocké à température ambiante.



Figure III.1 : Délimitation des zones de pêche à la journée pour chacun des villages suivis durant le premier volet du Programme Pêche, d'après les observations de Guillaume Longin (Issu de Longin 2016).

2. Bilan des données biologiques disponibles :

Table III.1 : Nombre total d'individus et d'espèces collectés sur chaque site (Villages et sites de référence) au cours du programme Pêche entre fin 2013 et début 2016, qu'elles soient ou non intégrées aux espèces focales de cette thèse.

Site de collecte	Coordonnées		Cours d'eau	Nb total d'espèces collectées	Nb total d'individus collectés	
	Lat.	Long.				
Papaïchton	3.8083	-54.1491	Lawa	51	1912	
Kayodé/Wempi	Villages	3.3890	-53.9246	Tampok	45	1546
Opoya/Taluen		3.3813	-54.0504	Lawa	50	1074
Antecume-Pata		3.2980	-54.0713	Lawa	40	1391
Pilima-Pata		3.2875	-54.1085	Litani	45	1274
Saut l'Inspecteur		3.1879	-53.4592	Waki	48	509
Saut Tampok	Reference sites	3.3169	-53.8340	Tampok	52	1246
Saut Pierkourou		2.8323	-53.5455	Tampok	62	1051
Langa Soula		2.8704	-53.9770	Marouini	71	1085
Wayo Gaan Soula		2.6771	-53.9985	Marouini	59	925
Apsik Icholi		2.9358	-54.1748	Litani	68	1063
Eléüéletpe icholi		2.7496	-54.2359	Litani	85	1346

Sur l'ensemble des espèces collectées, qu'elles fassent ou non partie des 18 espèces focales de ce travail de thèse, 47 espèces sont représentées par plus de 50 individus, à l'échelle de la zone d'étude complète du Haut-Maroni (c'est-à-dire villages et sites de référence confondus), pour lesquels l'information de taille et de poids frais sont disponibles. Ces données biométriques pourraient permettre d'évaluer les relations taille-poids pour ces espèces de grands fleuves, à l'instar de ce qui a été effectué sur certaines espèces de petites masses d'eau (Allard et al. 2015).

Annexe IV :

Génotypage : La spectrométrie de masse MALDI-TOF :

1. Principe et déroulement :

a. La spectrométrie de masse MALDI-TOF

La technologie MassARRAY (Agena Bioscience) est une méthode de génotypage de SNP par spectrométrie de masse MALDI-TOF, qui repose sur une PCR classique suivie d'une extension d'amorce. Le produit de cette extension est analysé sur un spectromètre utilisant une source d'ionisation laser assistée par une matrice (MALDI = matrix assisted laser desorption/ionisation) et un analyseur à temps de vol (TOF = time-of-flight) (Gabriel et al. 2011). Les produits d'extension d'amorce sont rendus volatils par la matrice MALDI. La séquence ADN précise des produits d'extension étant connue a priori pour chaque locus et chaque allèle attendu, il est possible d'en déduire leur masse et donc leur temps de vol, c'est-à-dire la durée nécessaire pour qu'ils atteignent un détecteur après ascension dans la colonne verticale du spectromètre. On peut faire l'analogie simple avec le concept de séparation de produits d'amplification de PCR sur un gel d'électrophorèse, qui permet d'identifier les différents allèles en fonction de la distance parcourue de cathode à anode en une unité de temps donnée. Ici le détecteur, qui connaît à l'avance le temps de vol théorique de chaque allèle l (T_l) rapportera à l'utilisateur la présence ou l'absence, ainsi que la qualité de l'observation effectuée à la fin de chaque instant T_l jusqu'à ce que tous les produits d'extension aient théoriquement pu achever leur ascension dans le spectromètre.

b. Protocole opérationnel

Après extraction, les quantités et la qualité de l'ADN extraits ont été dosés par PHERAstar FS et redilués si nécessaires pour une concentration finale de $10 \text{ ng.}\mu\text{L}^{-1}$. Une fois cette standardisation effectuée, les produits d'extraction sont amplifiés par PCR (Fig.IV.1) pour chacun des quatre multiplexes de 34-40 SNP préalablement mis au point à partir du fichier final issu du pipeline informatique présenté en Chapitre 1 (voir également Annexe VIII). Chaque multiplexe est optimisé de manière à ce que l'ensemble des produits d'extension obtenus en fin de manipulation puissent se répartir le long d'un spectre de masse à intervalles réguliers, de manière à être aisément discriminés les uns des autres. Au cours de ce protocole de génotypage, nous avons, pour toutes les espèces étudiées, fixé cet intervalle à 25 daltons.

L'ensemble du protocole MassARRAY s'effectue sur plaque ou puce à 384 puits. Sur chaque plaque ou puce, 4 puits sont dédiés à l'incorporation de contrôles positifs et négatifs. Une seule manipulation permet donc de génotyper, par exemple, 380 individus pour un seul multiplexe, ou encore 190 individus sur deux multiplexes, et ainsi de suite. Après la phase de PCR classique, les produits d'amplification sont traités par phosphatase alcaline (SAP) afin de neutraliser les dNTP non-incorporés lors de la réaction de PCR. Cela permet alors de procéder à la phase d'extension d'amorce dont les produits, après dilution et désalinisation, seront ensuite transférés de plaque 384 à puce qui servira de support à l'analyse du spectromètre de masse. La matrice de désorption et ionisation des produits d'extension pour chacun des puits de la puce intervient donc de manière automatique ainsi que l'analyse de temps de vol. Les données brutes sont transmises à l'utilisateur en fin de processus, et l'étape d'obtention des génotypes peut alors débiter.

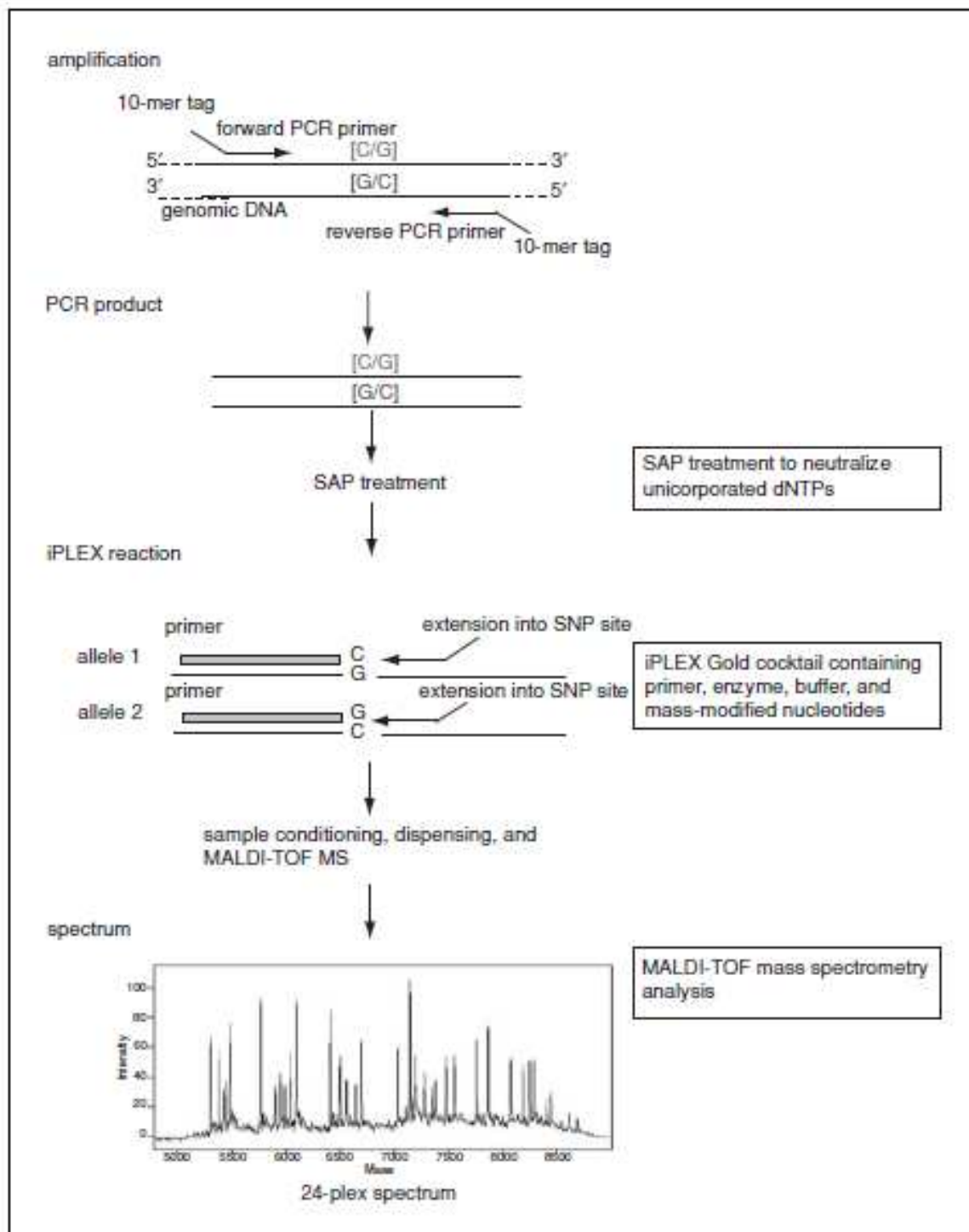


Figure IV.1 : Résumé des principales étapes de génotypage par spectrométrie de masse MassARRAY, depuis la phase d'amplification classique jusqu'au tracé du spectre de répartition des quantités de produits d'extension détectés pour chaque masse attendue. L'exemple pris ici concerne un SNP biallélique avec, comme allèles attendus, une cytosine ou une guanine [C/G] et intégré dans un multiplexe de 24 SNP au total. Adapté de Gabriel et al. (2011).

2. Analyse des données brutes et obtention des génotypes :

Le spectre représenté Fig.IV.1 correspond au spectre global de la densité de produits d'extension observés le long d'un gradient de masses. Afin de faciliter la compréhension de ce spectre, reprenons l'exemple d'un SNP biallélique à deux allèles attendus C et G. La figure IV.2. correspond à un zoom du spectre global jusqu'au niveau du locus considéré. Les deux produits d'extension attendus à ce locus

montrent généralement des masses relativement proches puisqu'ils ne diffèrent, dans leur séquence, que d'un seul nucléotide correspondant au SNP génotypé. Cette figure montre les trois cas optimaux attendus sur un locus à deux allèles chez un individu diploïde. Le panel de gauche et le panel central montrent chacun un cas où l'ensemble des produits d'extension détectés portent un allèle identique. L'individu est donc considéré comme homozygote à ce locus. Le panel de droite montre un cas où les produits d'extension présentent les deux allèles attendus, chacun en proportion un peu moindre par rapport au cas précédent puisque la concentration totale des produits d'extension se répartit entre les deux allèles portés par l'individu, qui est ici considéré comme hétérozygote.

On constate ainsi que la taille du pic est proportionnelle à la concentration de produits d'extension détecté par le spectromètre pour la masse correspondante. L'analyse des génotypes est d'autant plus aisée que la taille des pics est importante.

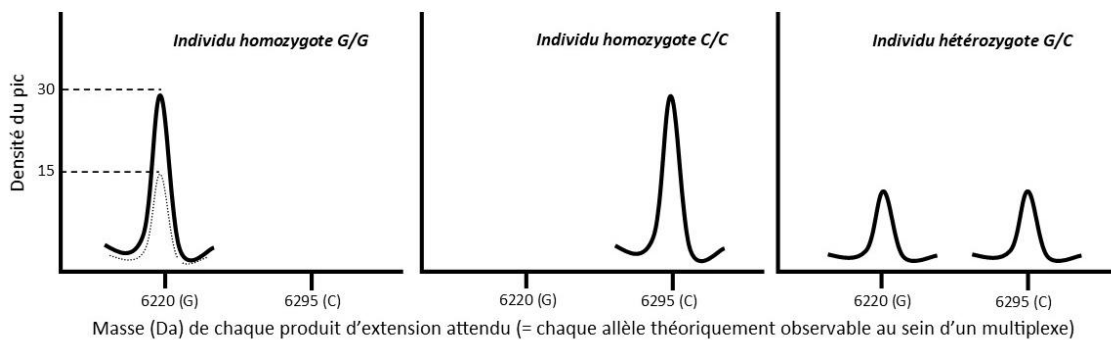


Figure IV.2 : Représentation schématique d'un spectre de masse obtenu pour le génotypage d'un SNP à deux allèles C ou G chez un individu diploïde.

Ainsi, un pic de très petite taille voire l'absence totale de pic correspond à un locus difficile à caractériser, qui ne sera généralement pas pris en compte dans les futures analyses. Le logiciel *TyperViewer* dédié à l'analyse des données brutes de *MassARRAY* convertit, pour chaque locus, l'information contenue par les spectres de tous les individus pour la représenter sous forme d'un biplot synthétique dont la Fig.IV.3 fournit un exemple. Il s'agit, plus précisément, du biplot obtenu pour le génotypage du SNP BFA16483 chez 190 individus *Brycon falcatus*. Le biplot représente en abscisse la densité du pic obtenu pour l'allèle « léger » G, et en ordonnée la densité du pic obtenu pour l'allèle « lourd » C. Le coin inférieur gauche du biplot correspond à des densités de pic nulles ou trop faibles pour fournir un génotype non-ambigu. L'utilisateur peut paramétrer lui-même la taille minimale d'un pic souhaitée, en dessous de laquelle tout génotype sera considéré comme non caractérisé (« no call »). Nous pouvons en outre observer les trois nuages de points homo- et hétérozygote constitués par l'ensemble des individus génotypés. En cliquant sur un individu particulier (comme l'individu entouré sur la Fig.IV.3), il est possible d'analyser visuellement le spectre de masse qui lui est associé. Un individu qui présenterait un spectre ambigu (comme l'individu indiqué par une flèche sur la Fig.IV.3, qui présenterait un pic pour chaque allèle mais avec des tailles inégales) pourrait être assigné comme non valide par l'utilisateur - et donc passé en « no call ».

Ainsi, pour chaque puce *MassARRAY*, chaque locus est vérifié et analysé individuellement à la lumière des spectres de masse présentés par les différents individus génotypés. Au cours de cette thèse, j'ai procédé à l'analyse d'un total de 47 puces et 3029 loci.

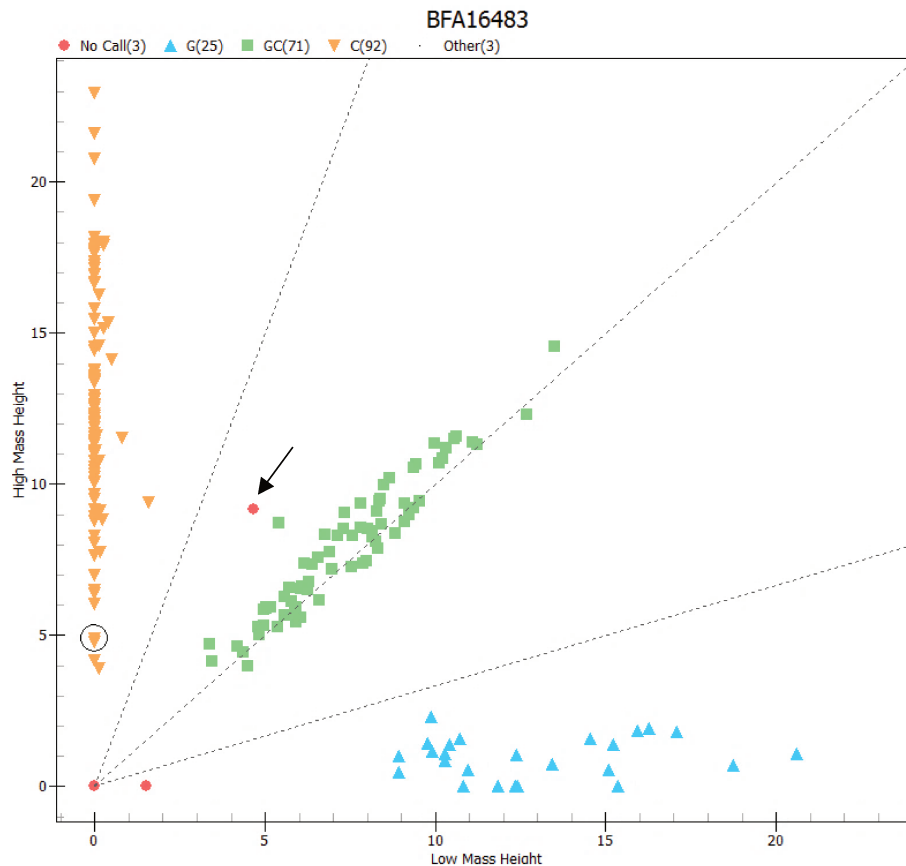


Figure IV.3 : Biplot synthétisant la nature et la taille des pics obtenus pour 190 individus (+4 témoins) de l'espèce *Brycon falcatus* génotypés pour le SNP BFA16483, issu du pipeline informatique présenté dans le Chapitre 1 de cette thèse. Les trois nuages de points correspondent aux trois génotypes possibles que l'on peut rencontrer sur ce locus à deux allèles G ou C. Les individus apparaissant en rouge sont considérés comme invalides et constitueront des données manquantes à ce locus dans la table de génotypes finale.

Après cette phase de contrôle qualité de chacun des loci, l'élimination des loci mal caractérisés, et la réattribution en « no call » des individus présentant un génotype ambigu ; l'ensemble des données de la puce analysée sont exportées et reformatées sous forme d'une table de génotypes classique en format tableur. C'est cette table de génotypes qui sert alors de support à l'élaboration de chaque jeu de données de génétique des populations pour les espèces de poissons du Haut-Maroni.

Référence:

Gabriel, S., Ziaugra, L., & Tabbaa, D. (2009). SNP Genotyping Using the Sequenom MassARRAY iPLEX Platform. In J. L. Haines, B. R. Korf, C. C. Morton, C. E. Seidman, J. G. Seidman, & D. R. Smith (Éd.), *Current Protocols in Human Genetics*. Hoboken, NJ, USA: John Wiley & Sons, Inc. <https://doi.org/10.1002/0471142905.hg0212s60>.

Annexe V :

Détection et gestion des phénomènes de contaminations croisées rencontrés lors de la procédure de séquençage RADSeq pour la mise au point des marqueurs moléculaires utilisés durant la thèse :

Le protocole de développement de ressources moléculaires présenté dans le Chapitre 1 implique des phases de collecte, de transport et de laboratoire (échantillonnage sur le terrain, extraction d'ADN total, constitution de pools équimolaires d'ADN, transport vers la plateforme de génomique Genotoul et préparation de bibliothèques de séquençage RADSeq) qui mobilisent des échantillons biologiques originaires de diverses espèces.

Ces différentes phases ne sont pas sans risques de contaminations cross-spécifiques (c'est-à-dire des transferts d'ADN entre échantillons de deux espèces différentes ou plus). L'échange accidentel d'ADN entre des pools d'ADN appartenant à des espèces différentes résulte, dans la plupart des cas, en la mise au point de marqueurs SNP « faux-positifs » et donc inutilisables dans le cadre d'approches de génétique des populations.

Cette annexe a pour objectif de présenter les différents types de contamination croisée que nous avons rencontrés puis de décrire de quelle manière nous avons pu y faire face.

1. Spécificité des marqueurs SNP

Contrairement au cas des marqueurs microsatellites, les marqueurs SNP sont difficilement transférables entre espèces. Plus précisément, un marqueur polymorphe chez une espèce n'aura qu'une faible probabilité de l'être également chez une autre, quand bien même celle-ci en serait phylogénétiquement proche. Miller et al. (2011), par exemple, ont utilisé la puce SNP Infinium *OvineSNP50* d'Illumina, développée pour le mouton domestique, afin d'en dériver des marqueurs moléculaires utilisables pour les deux espèces sauvages proches que sont le mouflon Canadien et le mouflon de Dall. Sur plus de 54 000 SNP disponibles via la puce Infinium, moins de 1% étaient également polymorphes chez l'une des deux espèces sauvages, et seulement 86 SNPs étaient polymorphes chez chacune des trois espèces.

Cette faible proportion de polymorphisme partagé entre espèces a également été constatée par l'analyse de nos propres données de RADSeq. J'ai ainsi pu observer, en effectuant un assemblage *de novo* de l'ensemble des reads obtenus pour les deux espèces proches *Leporinus friderici* et *Leporinus lebaili*, que seuls 6,2% de l'ensemble des RAD-tags étaient communs aux deux espèces. De ces RAD-tags, seuls 4 marqueurs SNP étaient polymorphes pour chaque espèce et chaque réplicat. Chez les trois espèces de Loricariidae (*Pseudancistrus barbatus*, *Hypostomus gymnorhynchus* et *Harttia guianensis*), seulement 4% des RAD-tags ont été séquencés chez les bibliothèques des trois espèces et la grande majorité des SNP sont, d'après observation qualitative, inter- et non intraspécifiques. Aucun SNP polymorphe dans toutes les bibliothèques, tous Loricariidae confondus, n'a pu être mis en évidence. Ces considérations ont justifié le choix de ne pas consacrer davantage de temps à la recherche de marqueurs SNP cross-spécifiques (chose que l'on aurait pu envisager, essentiellement pour des questions de coûts qui diminueraient d'autant qu'un même marqueur pouvait être remobilisé pour plusieurs espèces), et de se concentrer sur chaque espèce de manière indépendante des autres.

Lors de la phase de génotypage par spectrométrie de masse, j'ai pu ajouter, sur les puces dédiées à certaines espèces, des extraits d'ADN d'une autre espèce plus ou moins proche agissant comme témoins négatifs. Pour ces témoins, je m'attendais à observer un succès de génotypage faible à nul (résultant d'un échec d'amplification) pour une espèce d'un autre genre ou d'une autre famille, ou encore à l'obtention de génotypes quasi exclusivement homozygotes (résultant d'une amplification réussie, mais ciblée sur un locus homologue non-variable) pour une espèce de même genre ou de même famille. Ces attendus se sont vérifiés dans la majorité des cas, corroborant une nouvelle fois le caractère spécifique des marqueurs SNP que nous avons développés. Cette propriété nous a permis de faire face à certains des événements de cross-contamination que nous avons rencontrés.

2. Origine et détection des cross-contaminations :

Nous avons été confrontés à des échanges d'ADN entre pools d'ADN à différents niveaux.

Une contamination précoce peut résulter de l'incorporation, dans le panel de détection de SNP (*SNP discovery-set*), d'un ou plusieurs individus n'appartenant pas à l'espèce ciblée. Cela peut résulter de l'erreur d'identification d'un spécimen sur le terrain, d'une erreur d'étiquetage ou encore d'une erreur de manipulation en laboratoire. Une contamination peut aussi intervenir si deux extraits d'ADN de différentes espèces se mélangent lors du transport des échantillons en plaque 96 puits (si les puits sont mal scellés par exemple).

Nous avons pu, dans certains cas, détecter ces accidents et donc éviter leurs conséquences sur le développement et la validation des ressources moléculaires. Je parlerai donc dans ce cas de « gestion précoce » des cross-contaminations. Dans d'autres cas, ces accidents n'ont pas pu être détectés, ou bien leurs conséquences n'ont pas pu être évitées avant de procéder au génotypage des marqueurs SNP retenus. Je parlerai alors de « gestion tardive » des cross-contaminations. Je présenterai les situations rencontrées dans l'ordre chronologique, car ces expériences successives ont permis d'affiner les processus de développement de ressources génétiques au cours du temps. En effet, la procédure de développement et de génotypage de ressources pour les 18 espèces ne s'est pas faite simultanément mais en trois périodes de temps, centrées sur six espèces à chaque fois.

➤ Une gestion précoce de contamination a pu être effectuée lorsque j'ai remarqué l'incorporation, par erreur, d'un spécimen *Leporinus friderici* dans le panel de détection de SNP de *Serrasalmus eigenmanni* alors que la procédure de RADSeq était déjà aboutie (et donc chaque pool déjà séquencé). Après avoir sélectionné un certain nombre de SNP candidats pour cette espèce, j'ai procédé à l'alignement de chaque séquence consensus des RAD-tags retenus contre l'ensemble des RAD-tags assemblés de *Leporinus friderici*, à l'aide du logiciel BLAST2-n. Les deux espèces étant de familles différentes, j'ai pu éliminer des marqueurs candidats tous ceux, relativement peu nombreux, qui montraient une forte homologie entre les deux espèces et portaient donc le risque d'être faux-positifs pour *Serrasalmus eigenmanni*. Sur les 160 marqueurs testés après ce filtrage, 150 ont été informatifs pour cette espèce.

➤ Dans un deuxième temps, une gestion tardive de contamination a été effectuée lors de l'analyse des données de certaines espèces de Serrasalmidés. Je présenterai ici le cas de l'espèce *Myloplus rubripinnis*, mais j'ai appliqué une procédure identique pour ses voisines *Acnodon oligacanthus* et *Myloplus rhomboidalis* qui présentaient un phénomène analogue. Pour ces trois espèces, l'analyse des données de spectrométrie de masse puis des génotypes obtenus ont révélé la présence d'individus présentant des profils génotypiques particuliers et très différents de tous les

autres. Ces individus présentaient des profils homozygotes pour la majorité des marqueurs mais, en revanche, soutenaient la totalité du polymorphisme pour une proportion plus faible d'autres marqueurs, révélant ainsi des marqueurs moléculaires fixés entre les quelques individus concernés et les autres. En collaboration avec le Museum d'Histoire Naturelle de Genève, ces individus ont été génotypés pour le gène mitochondrial COI afin de procéder à leur réidentification par barcoding. Cette procédure a permis de conclure que, pour le cas de *Myloplus rubripinnis*, deux individus intégrés au panel de détection appartenaient en réalité à l'espèce *Myloplus ternetzi*, morphologiquement difficile à distinguer de *M. rubripinnis* pour l'œil peu exercé. L'ensemble des individus mal identifiés ont été retirés du jeu de données de *Myloplus rubripinnis* et, avec eux, plusieurs marqueurs SNP devenus monomorphes suite à leur retrait (et qui étaient donc des faux-positifs). Cela explique le nombre final un peu plus faible de SNPs utilisés pour cette espèce (mais aussi dans une moindre mesure pour *Acnodon oligacanthus* et *Myloplus rhomboidalis*). Par souci de simplicité, nous n'avons pas détaillé ce phénomène dans le Chapitre 1, dans la mesure où sa conséquence a pu être maîtrisée et ne remettait pas en question la validité du protocole et du pipeline utilisé.

➤ Enfin, une gestion précoce de contamination, consécutive à l'échec partiel d'une première tentative de développement de marqueurs SNP pour l'espèce *Cynodon meionactis*, a permis le succès de la seconde tentative. Un fort déséquilibre du nombre de RAD-tags avait été constaté entre les deux réplicats pour cette espèce, ce qui avait abouti à un relâchement d'un des paramètres du pipeline de développement de marqueurs lors du premier essai. Nous avons en effet autorisé, pour cette espèce, qu'un SNP candidat soit conservé même en apparaissant dans un seul des deux réplicats, suite à la conclusion erronée que le réplicat le plus pauvre en RAD-tags avait possiblement souffert d'un artefact de séquençage. Le génotypage des marqueurs développés durant ce premier essai a mené à un échec important : sur 160 SNPs génotypés, seuls 31 se sont révélés informatifs. Pour déterminer l'origine de cette faible proportion, nous avons procédé à la comparaison, à l'aide de BLAST2-n, de la séquence de chaque RAD-tags des deux réplicats de *Cynodon meionactis* avec les séquences de chaque RAD-tags de toutes les autres espèces. Nous avons ainsi réalisé qu'un des réplicats de cette espèce présentait de très nombreuses homologues avec les RAD-tags obtenus pour l'espèce *Brycon falcatus*. Les séquences complètes d'origine des 31 SNPs ayant été génotypés avec succès étaient retrouvées dans les deux réplicats et étaient, par contre, absentes des RAD-tags de *Brycon falcatus*. A l'inverse, les séquences des SNPs n'ayant pas été amplifiés lors du génotypage étaient présentes uniquement dans l'un des deux réplicats de *Cynodon meionactis* et se retrouvaient également parmi les RAD-tags de *Brycon falcatus*. Un test d'amplification, par PCR, d'échantillons de *Brycon falcatus* par les quatre multiplexes de SNPs de *Cynodon meionactis* a confirmé que la plupart de ces marqueurs devaient bien, en réalité, être spécifiques à *Brycon falcatus* (Fig.V.1).

Cette contamination, par l'ADN de *Brycon falcatus*, d'un seul des deux réplicats constitués pour *Cynodon meionactis* explique ainsi le déséquilibre du nombre de RAD-tags préalablement signalé entre les deux réplicats. Cette contamination a très probablement eu lieu durant la phase de préparation des pools ou lors du transport de l'ensemble des pools d'ADN vers la plateforme Genotoul. Nous recommandons un transport d'ADN en tubes, plutôt qu'en plaque, afin d'éviter au maximum ce type d'accident.

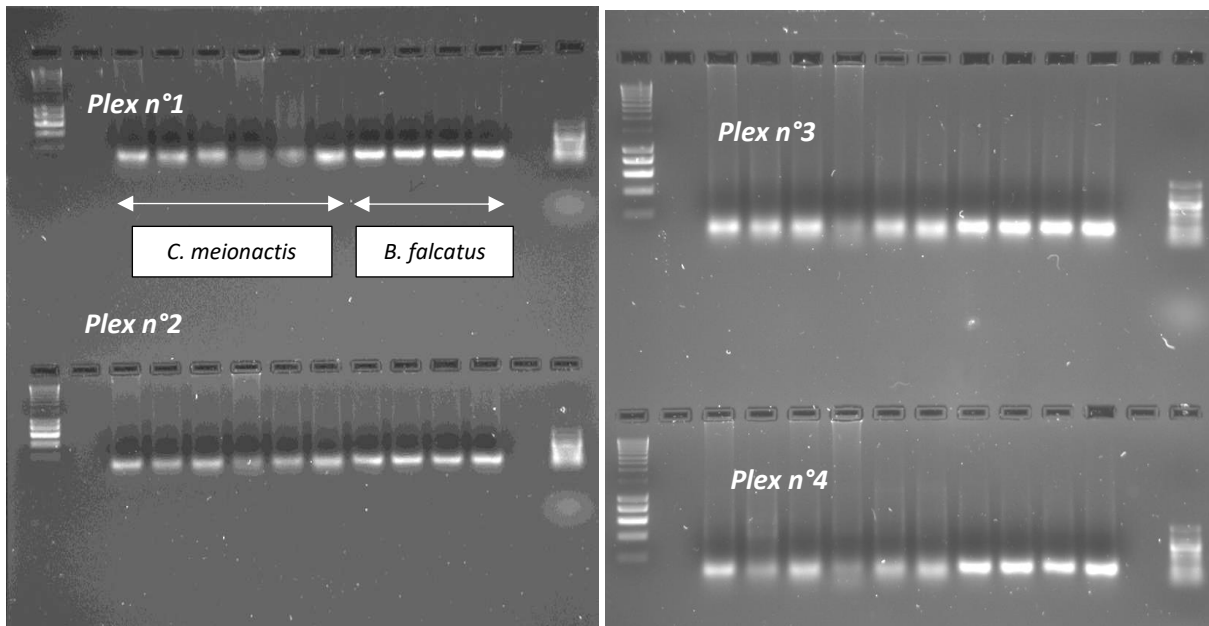


Figure V.1 : Photographies de gels d'agarose obtenus suite à une électrophorèse effectuée afin de vérifier la présence ou l'absence d'amplification, par PCR, de l'ADN issu de 10 spécimens de *Brycon falcatus* et de *Cynodon meionactis* (1 spécimen par puits). Chaque rangée correspond à un des quatre multiplexes créés pour le génotypage par spectrométrie de masse MassARRAY. Les spécimens génotypés pour ces différents multiplexes, d'une rangée à l'autre, sont les mêmes. On constate l'amplification plus importante chez les quatre spécimens *Brycon falcatus*, quel que soit le multiplexe considéré.

➤ Afin de vérifier l'occurrence d'autres cross-contaminations entre pools de différentes espèces, nous avons par la suite comparé deux à deux les séquences consensus des RAD-tags de toutes les paires possibles de pools, toutes espèces confondues, à l'aide de BLAST2-n. Nous nous attendions ainsi à des proportions élevées de RAD-tags homologues entre pools provenant d'une même espèce (c'est-à-dire entre répliquats), tandis que deux espèces phylogénétiquement éloignées devraient présenter de faibles proportions de RAD-tags homologues sauf en cas de cross-contamination majeure. Ceci nous a permis de détecter le cas des espèces *Ageneiosus inermis* et *Ludwigia grandiflora*, évoqué dans le Chapitre 1. Une contamination bi-directionnelle a en effet été constatée chez ces deux espèces, empêchant de pouvoir discriminer de manière certaine les SNPs exclusifs de l'une ou de l'autre. Cette situation a été approchée par une gestion à la fois précoce et tardive des contaminations. En sélectionnant les marqueurs candidats dont la profondeur de couverture était beaucoup plus élevée chez *Ageneiosus inermis* que chez la seconde espèce, nous avons tenté d'augmenter au maximum la probabilité de sélectionner des SNP exclusifs de cette espèce. Cependant, nous avons malgré tout dû faire face à un certain nombre de faux-positifs qui n'ont pu être révélés qu'après la procédure de génotypage. C'est pourquoi un total de 111 SNPs uniquement a pu finalement être conservé pour cette espèce.

3. Conclusion et recommandations

Malgré la survenue de cross-contaminations à diverses étapes de la phase de développement de ressources moléculaires multi-espèces, nous avons été en mesure de faire face à leurs conséquences en éliminant les SNPs faux-positifs soit avant, soit après la procédure de génotypage. Ceci a été permis par les comparaisons de séquences consensus entre différentes espèces, mais aussi par la faible transférabilité des marqueurs SNP même entre espèces proches (et donc, le fait que les marqueurs

testés étaient exclusifs et spécifiques). Cette propriété a en effet permis de détecter des individus mal identifiés, et donc des marqueurs SNP faux-positifs, même dans le cas de deux espèces proches phylogénétiquement et pour lesquelles la procédure de génotypage peut montrer un succès correct par cross-amplification de séquences homologues du génome par un couple d'amorces donné. Dans le cas de *Myloplus rubripinnis* par exemple, si tous les marqueurs testés avaient présenté un polymorphisme partagé avec l'espèce *Myloplus ternetzi* échantillonnée par erreur, nous aurions conservé dans le jeu de données des individus d'une autre espèce sans le savoir.

Afin d'éviter au mieux l'occurrence des cross-contaminations lors du développement de ressources moléculaires pour plusieurs espèces en simultané, nous soulignons deux principales recommandations.

Tout d'abord, il est recommandable de systématiquement ré-identifier, à l'aide par exemple du gène mitochondrial COI ou tout autre marqueur adapté au barcoding, chaque individu avant de l'intégrer la constitution d'un panel de détection de SNPs. Cette procédure permet en effet de détecter d'éventuelles erreurs d'identification sur le terrain, ou même la présence d'ADN contaminant provenant d'instruments mal nettoyés entre deux prélèvements sur différentes espèces par exemple. Ensuite, et comme indiqué dans le Chapitre 1, nous conseillons de procéder de façon systématique à des comparaisons, entre espèces, des séquences consensus de toutes les paires possibles de RAD-tags. Nous envisageons d'ailleurs d'intégrer l'utilisation systématique du logiciel BLAST2-n dans une éventuelle future version du pipeline présenté en Chapitre 1.

Référence :

Miller, J. M., Poissant, J., Kijas, J. W., Coltman, D. W., & the International Sheep Genomics Consortium. (2011). A genome-wide set of SNPs detects population substructure and long range linkage disequilibrium in wild sheep: PERMANENT GENETIC RESOURCES ARTICLE. *Molecular Ecology Resources*, 11(2), 314-322. <https://doi.org/10.1111/j.1755-0998.2010.02918.x>

Annexe VI :

Matériel supplémentaire de l'article n°1 présenté en Chapitre 1.

A cost-and-time effective procedure to develop SNP markers for multiple species: a support for community genetics.

Delord C.^{1,2}, Lassalle G.¹, Oger A.¹, Barloy D.¹, Coutellec M.A.¹, Delcamp A.³, Evanno G.¹, Genthon C.⁴, Guichoux E.³, Le Bail P.Y.⁵, Le Quilliec P.¹, Longin G.⁶, Lorvelec O.¹, Massot M.³, Reveillac E.¹, Rinaldo R.⁶, Roussel J.-M.¹, Vigouroux R.², Launey S.¹, Petit E.J.¹.

¹ESE, Ecology and Ecosystem Health, Agrocampus Ouest, INRA, 35042, Rennes, France

²HYDRECO Guyane SARL, Laboratoire-Environnement de Petit Saut, 97310, Kourou, France

³BIOGECO, INRA, Université de Bordeaux, 33610, Cestas, France

⁴Plateforme génomique INRA GenoToul, 31320, Castanet-Tolosan, France

⁵LPGP, INRA, 35042, Rennes, France

⁶PARC AMAZONIEN DE GUYANE, 97354, Rémire-Montjoly, France

Corresponding author: Chrystelle DELORD

email: chrys.delord@gmail.com

SUPPORTING INFORMATION

This document contains additional information about the workflow used to obtain SNP markers for each of the 40 species, from treatment of sequencing raw data and RAD-tags assembly to pipeline filtering for *in silico* validation (Appendix C1.S1). Appendix C1.S2 gives the geographical origin of each individual constituting the SNP-discovery set for each 18 fish species from French Guiana.

Appendix C1.S1: Detailed description of the bioinformatic workflow used for the development of SNP markers: Stacks catalog of RAD-tags assembly and pipeline intervention

A. Read assembly and construction of Stacks catalog of RAD-tags per species:

Commands listed hereafter are relative to the Stacks version 1.37. Options of reads processing and *de novo* assembly are slightly different from the last update of Stacks (version 2.0b from April, 2018).

We demultiplexed file of raw reads according to library-specific barcodes using the *process_radtags* command as following:

```
process_radtags -1 [file_for_reads1(SE).fastq.gz] -2 [file_for_reads2(PE).fastq.gz]
-b [barcodes_file.csv] -o [output_folder] -e sbfl -i gzfastq -y gzfastq -c -q
```

We eliminated PCR duplicates using the *clone_filter* command as following:

```
clone_filter -1 [file_for_reads1_speciesXX.1.fastq.gz] -2 [file_for_reads_speciesXX.2.fastq.gz]
-o [output_folder] -i gzfastq
```

Loci assembly and catalog construction was performed using the *denovo_map* command as following:

```
denovo_map -s [file_for_reads1_speciesXX_duplicate1_.1.1.fastq.gz]
-s [file_for_reads1_speciesXX_duplicate2_.1.1.fastq.gz] -o [output_folder] -m 4 -M 10 -n 1
-T 6 -B database_speciesXX -b 1 -X "ustacks:--max_locus_stacks 5"
--bound_low 0 --bound_high 0.001 --alpha 0.05
```

Comment: We set a minimum depth coverage (argument “-m”) to four (4) to build a stack (the word “stack” referring here to the assembly of strictly homologous SE reads within a sequencing library). Maximum number of mismatches allowed between two stacks to build a RAD-tag (argument “-M”) was voluntarily set quite high (-M=10) in order to force over-merging (e.g., *in silico* assembly of two homologous stacks even though these are biologically distinct – coming from different loci within the genome). The avoidance of over-merging or over-splitting (*in silico* separation of two slightly divergent stacks that actually come from the same locus within the genome) is a critical issue in typical RAD data analysis (Harvey et al., 2015; Mastretta-Yanes et al., 2015). In our case, however, over-merging is recommendable, as we need to eliminate highly similar loci as much as possible to keep most specific loci for further amplification.

The maximum number of stacks allowed within an assembled RAD-tag (argument “--max_locus_stacks”) was set to 5 in order to take into account the fact we may expect more than only two haplotypes per locus, considering each library was built by mixing several individuals. Maximum number of mismatches between two homologous RAD-tags - from different libraries - to build final Stacks catalog (argument “-n”) was kept as default.

The bounded-maximum likelihood SNP-calling model implemented in Stacks is usually better fitted for individual RAD data (i.e., each library corresponding to a single diploid individual). Consequently, we identified SNPs by relaxing the bounded model to allow low coverage minor alleles in our pool to be considered as variants (lower-bound=0, upper-bound=0.001, alpha=0.05). From the perspective of the model, this equals to allowing a sample to be considered as “heterozygous” (i.e., with biologically

relevant polymorphism) even if the ratio between major and minor allele depth coverage is strongly unbalanced. This phenomenon is indeed expected with pooled samples; as there is a high probability that alleles have been sampled unevenly, grossly reflecting their actual frequency in the natural population combined with sampling and sequencing stochasticity.

Reference:

Mastretta-Yanes, A., Arrigo, N., Alvarez, N., Jorgensen, T. H., Piñero, D., & Emerson, B. C. (2015). Restriction site-associated DNA sequencing, genotyping error estimation and *de novo* assembly optimization for population genetic inference. *Molecular Ecology Resources*, 15(1), 28–41. <https://doi.org/10.1111/1755-0998.12291>

B. Parameters of the pipeline:

Table VI.S1: List of user-defined parameters of the pipeline:

User-defined criteria for SNP development at each filtering step of the pipeline	Default parameter	Brief comment
1st filter		
MySQL request		
<i>depthAlleleMin</i>	5	Minimum depth coverage for each allele inside a library.
<i>depthAlleleMax</i>	100	Maximum depth coverage for each allele inside a library.
<i>pos5prim</i>	30	Minimum position (pb) of SNP from 5'-read extremity.
<i>pos3prim</i>	110	Maximum position (pb) of SNP from 5'-read extremity.
<i>numSamplesHomologous</i>	2	Number of libraries covering a locus from Stacks catalog.
<i>bothSamplesPolymorphic</i>	true	Each library should present the polymorphic state of a SNP from Stacks catalog. Usable only if numSamplesHomologous > 1.
2nd filter		
.sam file and .pileup file treatment on single-end reads		
<i>nbReadsSam</i>	3	Maximum number of « outlier » reads accepted after realignment of SE reads to the candidate sequences exported after first filter.
<i>depthFlankmax</i>	2	Maximum depth coverage for any SNP-flanking polymorphism (unrevealed by Stacks) tolerated after realignment of SE reads to keep the candidate sequence.
3rd filter		
.pileup file treatment on paired-end reads		
<i>depthFlankmax</i>	2	Maximum depth coverage for any SNP-flanking polymorphism (unrevealed by Stacks) tolerated after realignment of PE-reads to keep the candidate sequence.

Additional details:

-depthAlleleMin: The minimum depth of coverage for an allele within a RAD-tag library. This value should be high enough to avoid a risk a false-positive candidates (i.e., if a variant is only resulting from sequencing error), but low enough to increase the probability to capture rare alleles and avoid ascertainment bias (i.e., an allele that presents a low MAF might be at risk to display low coverage).

-depthAlleleMax: The maximum depth of coverage for an allele within a RAD-tag library. This value should be higher than the mean depth of coverage of an allele across the whole catalog of RAD-tags, but low enough to avoid the retention of tags that might reveal experimental artifacts or paralogs. For instance, some authors choose to ignore SNPs with an allele depth of coverage greater than the mean depth + 2 fold the coverage standard deviation.

-pos5prim and -pos3prim: These values only depends on the desired length of amplification primers for subsequent genotyping. The position of a SNP should not fall were one of the flanking sequence would be too short for primers design.

-numSamplesHomologous: This value should be equivalent to the number of libraries used for each species. Indeed, we should be expected a locus to be found within each library from a same species, unless there was some locus ‘dropout’ resulting from PCR artifacts. Lowering this value might increase the number of candidate markers by relaxing stringency, but at risk of getting some false-positives if a locus not found within each library results from sequencing errors.

-bothSamplesPolymorphic: This should be added (i.e., “true”) if libraries come from technical replicates, since libraries built from the same DNA should harbor the same variants. Removing this option might increase the number of selected candidate markers by relaxing stringency, but at risk of getting some false-positives if a SNP not found within each library results from sequencing errors.

-nbReadsSam: After SE reads from a given species were aligned to its previously exported candidate sequences: if a candidate displays alignment with some ambiguous reads (i.e., reads that align just partly, or does align on full-length but with many mismatches), then these reads are considered as “outliers” and if their number is higher than $-nbReadsSam$, then the candidate is eliminated. Lowering this parameter leads to more stringency and might reduce the number of candidates successfully going through the filter.

-depthFlankmax: This value should not be higher than the expected depth of coverage for a sequencing error. If the value is too high, then the flanking variant might be an actual rare polymorphism and hamper primer design to amplify the SNP of interest. Lowering this parameter leads to more stringency and might reduce the number of candidates successfully going through the filter.

C. Main commands performed by the pipeline:

From now on, all commands being presented here are normally performed automatically through the pipeline. This section aims to decompose these features and processes to help readers follow the three different filtering steps and the principles these are based on.

SQL query

The first filter selects candidate SNPs from the Stacks catalog of RAD-tags. An equivalent manual query for candidate SNPs would be the following:

```
select distinct ma.catalog_id, col, seq # Extraction of RAD-tag ID within the catalog, SNP position,
                                and complete 145-bp sequence.
from matches ma
join catalog_snps cs on ma.catalog_id=cs.tag_id
join catalog_tags ct on ma.catalog_id=ct.tag_id
join catalog_index ci on cs.tag_id=ci.tag_id
where # Insertion of first filtering criteria (Edit: to be soon integrated into a paramfile)
ma.catalog_id not in (select catalog_id FROM matches where depth < 5) # depthAlleleMin
and ma.catalog_id not in (select catalog_id FROM matches where depth > 100) # depthAlleleMax
and cs.col > 30 and cs.col < 110 # Pos5prim and Pos3prim
and ci.snps = 1
and ci.alleles = 2
and ci.parents = 2 # numSamplesHomologous
and ci.ratio = 'ab:2(100.0%);' # if bothSamplesPolymorphic=TRUE
order by ma.depth;
```

bwa-mem and samtools pileup

The second filter needs an input file in .sam format to check homologies between candidate sequences and all other SE reads for the focal species. Then, another file in .pileup format is generated, this time to check for the absence of remaining, undetected flanking polymorphisms around the focal SNP.

.sam and .pileup files are generated as following:

```
bwa mem -a -M
[fasta_file_of_candidateSNPmarkers_speciesXX] [file_for_allreads1_speciesXX_1.1.fastq.gz]
> [Output folder]/SpeciesXX_SE.sam
```

The third filter also needs as input a .pileup file with PE reads. This file is built in a same way as previously, after converting the .sam file to a .bam file:

```
samtools mpileup -d 1000 -O --ff 4 -f [fasta_file_of_SEvalidatedSNPmarkers_speciesXX]
[SpeciesXX_PE.bam]
> [Output folder]/SpeciesXX_PE.pileup # pileup build for 3rd filtering (PE reads).
```

D. Details on the use of .sam format to perform second filtering:

This filtering was partly done using the 'NM' and 'CIGAR' flags provided in the .sam file (A nice description of the .sam format is available at the URL <https://samtools.github.io/hts-specs/SAMv1.pdf>). The objective of this filter is to detect "outlier" reads, i.e. reads that do not fully align to the query candidate sequence or displays a higher-than-expected number of mismatches comparing to the query. If the number of "outlier" reads overtakes the parameter *-nbReadsSam* (def=3), the candidate query is discarded to avoid unspecific further PCR amplification. In a .sam file, 'CIGAR' summarizes the information about the alignment between a read and the query sequence. In our case, it is expected to be '145M' since a read must align fully to the 145 bp candidate query sequence. The pipeline considers as an outlier any read with a different CIGAR value. Then, we consider the 'NM' value, which is the number of mismatches between an aligned read and the candidate query sequence. 'NM' should be either null (equal to zero: the read displays the exact same sequence as the query) or unity (the read displays one mismatch corresponding either to the alternate allele on the SNP position, or to a punctual other mismatch), and a value of 2 is tolerated. If $NM > 2$, the read is flagged as an outlier.

E. Details on the use of .pileup format to complete second filtering and perform third filtering:

Still aiming to keep primer design as simple as possible, we finally checked for the absence of flanking polymorphism by scanning each position of the alignment but the expected SNP. This was partly done during second filtering on .sam file as a result of the elimination of a candidate sequence associated with reads displaying 'NM'>2, but without considering the depth coverage of each flanking polymorphism when 'NM'=1 or 2.

In a .pileup file, each line corresponds to the alignment of several reads on a given base pair of the candidate query sequence. The candidate SNP position is known so it is possible to check all other positions to check for the absence of unexpected variants. If another variant is found with depth coverage higher than *-depthFlankMax* (def=2), the corresponding query candidate sequence is discarded. The pipeline applies this treatment on an identical way to complete second filtering on SE reads, and to perform third filtering with PE-reads.

Appendix C1.S2: SNP-discovery set composition and origins for each 18 fish species from French Guiana.

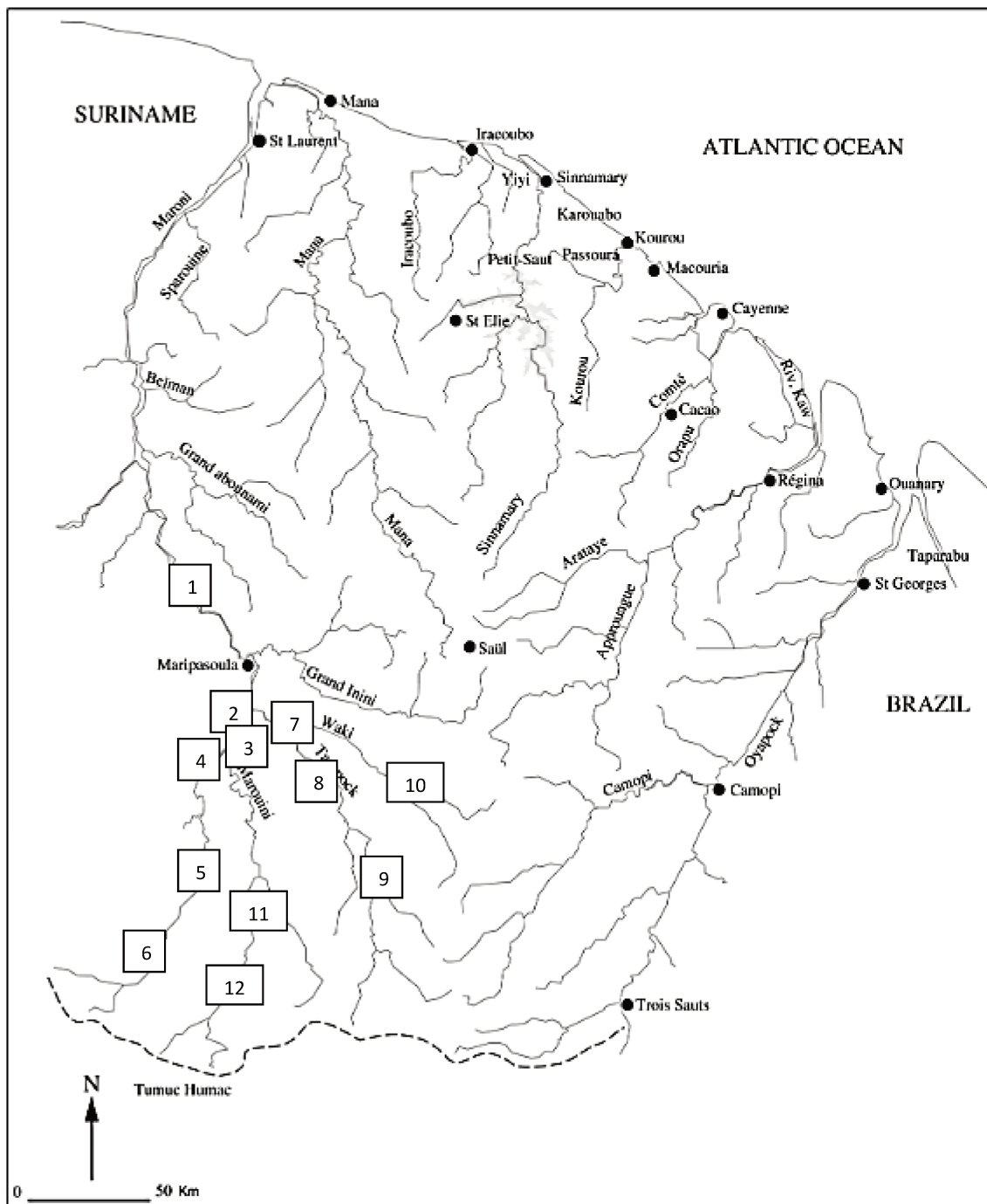


Figure VI.S2 : Map of French Guiana (adapted from Le Bail et al., 2012) and study area where the 18 fish species used for SNP validation were sampled. Numbers indicates the 12 sampling sites.

Table VI.S2.1: Tributary and coordinates (latitude, longitude) of each sampling site from the Maroni:

	Site	Tributary	Latitude	Longitude
1	<i>Papaichton</i>	Lawa	3,808	-54,149
2	<i>Opoya (Taluen)</i>	Lawa	3,381	-54,05
3	<i>Antecume pata</i>	Lawa	3,298	-54,071
4	<i>Pilima pata</i>	Litani	3,287	-54,108
5	<i>Apsik icholi, Yaleime</i>	Litani	2,935	-54,175
6	<i>Elüéletpe icholi</i>	Litani	2,749	-54,23
7	<i>Wempi (Kayode)</i>	Tampok	3,389	-53,924
8	<i>Saut Tampok</i>	Tampok	3,316	-53,833
9	<i>Saut Pierkourou</i>	Tampok	3,187	-53,459
10	<i>Saut l'Inspecteur</i>	Waki	2,832	-53,545
11	<i>Langa Soula</i>	Marouini	2,861	-53,975
12	<i>Wayo Gaan Soula</i>	Marouini	2,677	-53,998

Reference:

Le Bail, P.-Y., Covain, R., Jégu, M. et al. (2012) Updated checklist of the freshwater and estuarine fishes of French Guiana. *Cybium*, 36, 293–319.

Table VI.S2.2: Number of samples used to build the SNP discovery-set, per sampling site and per species:

Tributary	Lawa		Litani				Marouini		Tampok			Waki
	Papaichton	Opoya	Antecume-Pata	Pilima-Pata	Apsik Icholi	Eléüéletpe icholi	Langa Soula	Wayo Gaan Soula	Wempi	Saut Tampok	Saut Pierkourou	Saut l'Inspecteur
1. <i>Brycon falcatus</i>	2	3	-	-	3	1	2	2	2	-	2	1
2. <i>Cynodon meionactis</i>	2	3	1	3	1	2	1	1	2	1	2	3
3. <i>Triportheus brachypomus</i>	2	2	-	-	2	2	2	1	-	2	1	1
4. <i>Hoplias aimara</i>	4	3	-	2	-	1	2	2	1	1	2	1
5. <i>Leporinus friderici</i>	2	4	-	-	2	2	2	2	-	2	2	1
6. <i>Leporinus lebaili</i>	3	-	2	-	1	2	2	3	2	1	2	1
7. <i>Acnodon oligacanthus</i>	1	3	-	1	1	2	2	2	2	1	2	2
8. <i>Myloplus rhomboidalis</i>	2	2	-	-	2	1	2	2	2	-	2	1
9. <i>Myloplus rubripinnis</i>	3	3	-	1	1	2	2	1	2	-	2	1
10. <i>Tometes lebaili</i>	2	1	1	-	1	2	-	4	1	-	2	1
11. <i>Serrasalmus eigenmanni</i>	1	2	-	-	1	3	2	2	1	1	2	2
12. <i>Serrasalmus rhombeus</i>	3	2	-	-	2	2	2	1	1	1	2	1
13. <i>Ageneiosus inermis</i>	4	-	-	4	-	3	2	3	2	1	2	2
14. <i>Doras micropoeus</i>	-	3	1	2	1	-	4	3	3	-	-	-
15. <i>Harttia guianensis</i>	1	2	2	1	-	2	3	3	3	-	1	2
16. <i>Hypostomus gymnorhynchus</i>	3	1	-	2	-	2	2	1	2	1	1	2
17. <i>Pseudancistrus barbatus</i>	2	3	1	2	1	2	1	3	2	-	2	2
18. <i>Geophagus harreri</i>	2	1	1	1	1	1	1	3	4	-	2	2

Annexe VII :

Scripts Python constitutifs du pipeline informatique présenté en Chapitre 1 :

Ces scripts sont ceux que l'on peut actuellement consulter en ligne sur Zenodo :

<https://doi.org/10.5281/zenodo.1252193>, et qui ont été utilisés pour générer les données mobilisées dans l'article n°1.

Ils sont actuellement en cours d'optimisation par Gilles Lassalle, bioinformaticien au sein de l'unité Ecologie et Santé des Ecosystèmes, qui permettra notamment à l'utilisateur de renseigner les paramètres de départ (c.f. Annexe précédente, Table C1.S1) via un fichier *paramfile*.

Le pipeline se compose de cinq scripts Python distincts.

- Le script '**RAD4SNPs_main.py**' est le script principal. Il lance les commandes successives qui génèrent l'ensemble des fichiers input et outputs, intermédiaires et finaux, qui sont nécessaires afin d'atteindre l'objectif du pipeline. Il lance successivement chacun des quatre autres scripts au(x) moment(s) où ceux-ci doivent intervenir :
- Le script '**RAD4SNPs_SQL2fasta.py**' procède à la requête SQL permettant d'appliquer le premier filtre.
- Le script '**RAD4SNPs_SamFilter.py**' débute l'application du second filtre.
- Le script '**RAD4SNPs_PileupFilter.py**' termine l'application du second filtre et procède au troisième filtre.
- Le script '**RAD4SNPs_FinalSQLExtract.py**' génère le fichier final contenant l'ensemble des séquences consensus pour les RAD-tags contenant un SNP susceptible d'être génotypé par amplification classique.

Beginning of RAD4SNPs_main.py

```
#!/usr/bin/python

# RAD4SNPs: A set of Python scripts to select and validate independent SNPs markers from a list of read files #
# MAIN PROGRAM

# Authors: G.LASSALLE (gilles.lassalle@inra.fr) & C.DELORD (chrystelle.delord@inra.fr)
# Last update: AUGUST 2017

#----- PRE-CONDITIONS
#- [-i] Working directory where to store results of the pipeline for the focal species X
#- [-d] exact name of MySQL database where denovo_map.pl Stacks data are available for the focal species X
#- [-i1] single-end reads (reads 1) for focal species X duplicate 1
#- [-i2] single-end reads (reads 1) for focal species X duplicate 2
#- [-i3] paired-end reads (reads 2) for focal species X duplicate 1
#- [-i4] paired-end reads (reads 2) for focal species X duplicate 2

#- BWA and SAMtools available
#- Connexion to the Stacks MySQL database available: databases of Stacks 'denovo_map output' for each species.

#-----
import argparse
import os
import sys
import MySQLdb
#-----
parser = argparse.ArgumentParser()
parser.add_argument('-i', action='store', dest='InputDir', help='Working Directory')
parser.add_argument('-d', action='store', dest='database', help='Stacks database')
parser.add_argument('-c', action='store', dest='CodeSp', help='ID of the species')
parser.add_argument('-i1', action='store', dest='R11', help='First R1 file')
parser.add_argument('-i2', action='store', dest='R12', help='Second R1 file')
parser.add_argument('-i3', action='store', dest='R21', help='First R2 file')
parser.add_argument('-i4', action='store', dest='R22', help='Second R2 file')
parser.add_argument('--version', action='version', version='% (prog)s 0.1')

results = parser.parse_args()
print 'input directory          =', results.InputDir

#-----
# Arguments testing
#-----
```

```

if results.InputDir:
    if os.path.isdir(results.InputDir):
        print "Working directory is valid."
    else :
        print "Caution: working directory is invalid, please ckeck [-i]."
        sys.exit()
else :
    print "Please insert path for working directory [-i]. End of program."
    sys.exit()

#
if results.database:
    db = MySQLdb.connect(host="",          # your host, usually localhost
                        user="",          # your username
                        passwd="",        # your password
                        db=results.database) # name of the Stacks SQL database
    cur1= db.cursor()                    # connexion
    print "Currently working on MySQL database: "+str(results.database)
else:
    print "Incorrect ID for database: database not found, please check [-d]"
    sys.exit()

#
if results.R11:
    if os.path.isfile(results.R11):
        print "First file of single-end reads: found."
    else :
        print "Path to single-end reads data is not a file: please check out [-i1]."
        sys.exit()
else :
    print "Please insert path to single-end read files [-i1]. End of program."
    sys.exit()

#
if results.R12:
    if os.path.isfile(results.R12):
        print "Second file of single-end reads: found."
    else :
        print "Path to single-end reads data is not a file: please check out [-i2]."
        sys.exit()
else :
    print "Please insert path to single-end read files [-2]. End of program."

```



```

sys.exit()

#
if results.R21:
    if os.path.isfile(results.R21):
        print "First file of paired-end reads: found."
    else :
        print "Path to paired-end reads data is not a file: please check out [-i3]."
        sys.exit()
else :
    print "Please insert path to paired-end read files [-i3]. End of program."
    sys.exit()

#
if results.R22:
    if os.path.isfile(results.R22):
        print "Second file of paired-end reads: found."
    else :
        print "Path to paired-end reads data is not a file: please check out [-i4]."
        sys.exit()
else :
    print "Please insert path to paired-end read files [-i4]. End of program."
    sys.exit()

#
if results.CodeSp:
    CodeEspece=str(results.CodeSp)
    if CodeEspece[:1]!="_":
        CodeEspece=str(results.CodeSp)+str("_")
else:
    CodeEspece="std_"

WorkDir=os.path.abspath(results.InputDir) # Current working directory
# Formatting name of candidates fasta file -output of MySQL filtering
FastaCatalog=str(WorkDir)+"/"+str(results.CodeSp)+"Catalog.fasta"

#####
# Main program
#####

if os.path.isfile("/usr/bin/bwa"):
    print "BWA program is found."

```

```

else :
    print "Cannot find BWA: please check out pipeline requirements."
    sys.exit()
if os.path.isfile("/usr/bin/samtools"):
    print "SAMtools program is found."
else :
    print "Cannot find SAMtools: please check out pipeline requirements."
    sys.exit()

# Check if working directory is writable

filepath = results.InputDir+'/file.txt'

try:
    filehandle = open( filepath, 'w' )
except IOError:
    sys.exit( 'Working directory is not accessible' + filepath )

#-----
# PIPELINE COMMANDS:
#-----

#----- FIRST FILTERING -----#
print os.getcwd()
commandeExtractFasta="./RAD4SNPs_SQL2Fasta.py  -o  "+str(FastaCatalog)+"  -d  "+str(results.database)+"  -c
"+str(CodeEspece)

print "Extraction du fichier fasta"
print commandeExtractFasta
os.system(commandeExtractFasta)

## Fusion of single-end reads ##
if results.R11:
    if results.R12:
        commandFusionR1="cat "+str(results.R11)+" "+str(results.R12)+" > "+str(WorkDir)+"/allR1.fq.gz"

    else :
        commandFusionR1="cp "+str(results.R11)+" "+str(WorkDir)+"/allR1.fq.gz"
#----- End of fusion.

## Fusion of paired-end reads ##
if results.R21:

```

```

if results.R22:
    commandFusionR2="cat "+str(results.R21)+" "+str(results.R22)+" > "+str(WorkDir)+"/allR2.fq.gz"
else :
    commandFusionR2="cp "+str(results.R21)+" "+str(WorkDir)+"/allR2.fq.gz"
#----- End of fusion.

#----- SECOND FILTERING (1) -----#

command1="bwa index "+str(FastaCatalog) # Indexing
command2="bwa mem -a -M "+str(FastaCatalog)+" "+str(WorkDir)+"/allR1.fq.gz > "+str(WorkDir)+"/PremierAlign.sam" # SE
reads alignment
command3="samtools view -Sb "+str(WorkDir)+"/PremierAlign.sam | samtools sort - "+str(WorkDir)+"/PremierAlign1Sorted"
# Conversion to bam file
command4="samtools view -F4 "+str(WorkDir)+"/PremierAlign1Sorted.bam > "+str(WorkDir)+"/PremierAlign1Sorted-
F4.sam" # Elimination of unmapped SE reads

print "SE reads merging: "+str(commandFusionR1)
os.system(commandFusionR1)
print "PE reads merging: "+str(commandFusionR2)
os.system(commandFusionR2)

print "BWA indexing: "+str(command1)
os.system(command1)
print "Alignment: "+str(command2)
os.system(command2)
print "Conversion to bam file: "+str(command3)
os.system(command3)
print "Elimination of unmapped SE reads: "+str(command4)
os.system(command4)

print " *****"
print "          Second filtering (1) with default parameters          "
print " *****"

print os.getcwd()
commande5="./RAD4SNPs_SamFilter.py -i "+str(WorkDir)+"/PremierAlign1Sorted-F4.sam"
os.system(commande5)

Candidatfasta1=str(WorkDir)+"/PremierAlign1Sorted-F4R1Filtered.fa" # Obtention of incomplete SE-validated fasta file

if os.path.isfile(Candidatfasta1):
    print "SE-validated fasta file about to be completed. Re-aligning to complete second filtering."

```

else :

```
sys.exit( '****ERROR**** A problem occurred. Please check out alignment outputs.')
```

```
#----- SECOND FILTERING (2) -----#
```

```
command21="bwa index "+str(Candidatfasta1)
```

```
command22="bwa mem -a -M "+str(Candidatfasta1)+" "+str(WorkDir)+"/allR1.fq.gz > "+str(WorkDir)+"/SecondAlign.sam"
```

```
command23="samtools view -Sb "+str(WorkDir)+"/SecondAlign.sam | samtools sort - "+str(WorkDir)+"/SecondAlign1Sorted"
```

```
command25="samtools index "+str(WorkDir)+"/SecondAlign1Sorted.bam"
```

```
command25bis="samtools faidx "+str(Candidatfasta1)
```

```
command26="samtools mpileup -d 1000 -O --ff 4 -f "+str(Candidatfasta1) + " "+ str(WorkDir)+"/SecondAlign1Sorted.bam"+"
> "+str(WorkDir)+"/CandidatsR1.pileup"
```

```
print "BWA indexing: "+str(command21)
```

```
os.system(command21)
```

```
print "Alignment: "+str(command22)
```

```
os.system(command22)
```

```
print "Conversion to bam file: "+str(command23)
```

```
os.system(command23)
```

```
print "Indexing of bam file: "+str(command25)
```

```
os.system(command25)
```

```
print "Indexing for pileup file: "+str(command25bis)
```

```
os.system(command25bis)
```

```
print "Construction of SE pileup file: "+str(command26)
```

```
os.system(command26)
```

```
print " *****"
```

```
print "          Second filtering (2) with default parameters          "
```

```
print " *****"
```

```
print os.getcwd()
```

```
command27="/RAD4SNPs_PileupFilter.py -i "+str(WorkDir)+"/CandidatsR1.pileup"
```

```
print "End of second filtering: elimination of flanking variants: "+str(command27)
```

```
os.system(command27)
```

```
command28="/RAD4SNPs_FinalSQLExtract.py -i"+str(WorkDir)+"/CandidatsR1NoMulti.txt -d "+str(results.database)+" -c
"+str(CodeEspece)+" > "+str(WorkDir)+"/CandidatFin.fasta"
```

```
print "Complete SE-validated fasta file: "+str(command28)
```

```
os.system(command28)
```

```
command28bis="sed -i '1d' "+str(WorkDir)+"/CandidatFin.fasta"
```

```
os.system(command28bis)
```

```

#----- THIRD FILTERING -----#

CandidatFin=str(WorkDir)+"/CandidatFin.fasta"
if os.path.isfile(CandidatFin):
    print "SE-validated fasta file is completed. Re-aligning to perform third filtering."
else :
    sys.exit( '****ERROR**** A problem occurred. Please check out alignment and/or pileup outputs. ')

command29="bwa index "+str(CandidatFin)
command30="bwa mem -a -M "+str(CandidatFin)+" "+str(WorkDir)+"/allR2.fq.gz > "+str(WorkDir)+"/ThirdAlign.sam"
command31="samtools view -Sb "+str(WorkDir)+"/ThirdAlign.sam | samtools sort - "+str(WorkDir)+"/ThirdAlign2Sorted"
command32="samtools index "+str(WorkDir)+"/ThirdAlign2Sorted.bam"
command32bis="samtools faidx "+str(CandidatFin)
command33="samtools mpileup -d 1000 -O --ff 4 -f "+str(CandidatFin)+" "+str(WorkDir)+"/ThirdAlign2Sorted.bam"+" > "+str(WorkDir)+"/Candidats3.pileup"

print "BWA indexing: "+str(command29)
os.system(command29)
print "Alignment: "+str(command30)
os.system(command30)
print "Conversion to bam file: "+str(command31)
os.system(command31)
print "Indexing of bam file: "+str(command32)
os.system(command32)
print "Indexing for pileup file: "+str(command32bis)
os.system(command32bis)
print "Construction of PE pileup file: "+str(command33)
os.system(command33)

print " *****"
print "          Third filtering with default parameters          "
print " *****"

print os.getcwd()
command34="./RAD4SNPs_PileupFilter.py -i "+str(WorkDir)+"/Candidats3.pileup"
print "End of third filtering: elimination of flanking variants: "+str(command34)
os.system(command34)
command35="./RAD4SNPs_FinalSQLExtract.py -i "+str(WorkDir)+"/CandidatsR2NoMulti.txt -d "+str(results.database)+" -c "+str(CodeEspece)+" > "+str(WorkDir)+"/SNPs_out.fasta"
print "Complete PE-validated fasta file: "+str(command35)
os.system(command35)

```

End of RAD4SNPs_main.py

Beginning of RAD4SNPs_SQL2fasta.py

```
#!/usr/bin/python

# RAD4SNPs: A set of Python scripts to select and validate independent SNPs markers from a list of read files #
# SUB-PROGRAM FOR FIRST FILTERING FROM MYSQL DATABASE

# Authors: G.LASSALLE (gilles.lassalle@inra.fr) & C.DELORD (chrystelle.delord@inra.fr)
# Last update: AUGUST 2017

#----- PRE-CONDITIONS
#- [-i] File
#- Connexion to the Stacks MySQL database available: databases of Stacks 'denovo_map output' for each species.

#-----
import argparse
import os
import csv
import sys
import MySQLdb
#-----
parser = argparse.ArgumentParser()
# General arguments
parser.add_argument('-o', action='store', dest='OutputFile', help='Fichier Fasta de sortie') # Name of the candidate
sequences file
parser.add_argument('-d', action='store', dest='Database', help='Stacks database')
parser.add_argument('-c', action='store', dest='CodeSp', help='ID of the species')
# SQL request arguments
parser.add_argument('-dMin', action='store', dest='depthAlleleMin', help='')
parser.add_argument('-dMax', action='store', dest='depthAlleleMax', help='')
parser.add_argument('-p5', action='store', dest='pos5prim', help='')
parser.add_argument('-p3', action='store', dest='pos3prim', help='')
parser.add_argument('-q', action='store', dest='numSamplesHomologous', help='')
parser.add_argument('-p', action='store', dest='bothSamplesPolymorphic', help='Default "TRUE"')
parser.add_argument('--version', action='version', version='%(prog)s 0.1')

results = parser.parse_args()

# Defining default values:
if results.depthAlleleMin:
```

```

        dMin=int(results.depthAlleleMin)
else:
    dMin=5

if results.depthAlleleMax:
    dMax=int(results.depthAlleleMax)
else:
    dMax=100

if results.pos5prim:
    p5=int(results.pos5prim)
else:
    p5=30

if results.pos3prim:
    p3=int(results.pos3prim)
else:
    p3=110

if results.nbMaxAlleles:
    x=int(results.nbMaxAlleles)
else:
    x=2

if results.numSamplesHomologous:
    q=int(results.numSamplesHomologous)
else:
    q=2

if results.bothSamplesPolymorphic:
    p=str("\'aa:1(50.0%);ab:1(50.0%);\'")
else:
    p=str("\'ab:2(100.0%);\'")

# Server connexion
DBHost="127.0.0.1"          # localhost

#-----
# Output file checking (.fasta file in text format for the listing of candidate SNPs)
#-----
if results.OutputFile:
    try:

```

```

    FastaFileOut=open(results.OutputFile,"w")
    print "Creation of candidate SNPs fasta file: done."
except:
    print "Creation of candidate SNPs fasta file has failed."
    print("Unexpected error:", sys.exc_info()[0])
    sys.exit()
else :
    print "Please check out inserted path for Output file [-o] of candidate sequences from MySQL. End of program."
    sys.exit()

#-----
# Database connexion checking
#-----

print "Initialisation of Stacks MySQL database"
if results.Database:
    db = MySQLdb.connect(host="",      # your host, usually localhost
        user="",                    # your username
        passwd="",                  # your password
        db=results.Database)        # name of the Stacks SQL database
    cur1= db.cursor()              # connexion
else:
    print "Database not found, please check out inserted ID for the database [-d]. End of program."
    sys.exit()

if results.CodeSp:
    CodeEspece=str(results.CodeSp)
else:
    CodeEspece=""
print "The species ID that will be used is: "+str(CodeEspece)

#-----
# MySQL request
#-----

print "Retrieving candidate markers from Stacks MySQL database..."

Request="select distinct ma.catalog_id,col,rank_1,rank_2,seq from matches ma join catalog_snps cs on
ma.catalog_id=cs.tag_id join catalog_tags ct on ma.catalog_id=ct.tag_id join catalog_index ci on cs.tag_id=ci.tag_id where
ma.catalog_id not in (select catalog_id FROM matches where depth < "+str(dMin)+") and ma.catalog_id not in (select
catalog_id FROM matches where depth > "+str(dMax)+") and cs.col > "+str(p5)+" and cs.col < "+str(p3)+" and ci.snps = 1 and
ci.alleles = 2 and ci.parents = "+str(q)+" and ci.ratio = "+str(p)+" order by ma.depth;"

```



```

print Request

#-----
# Preparation to write candidate markers information in the output .fasta file
#-----
def AjoutFasta(FastaFile,IdSeq,Seq,code):
    IdSeq=">" +str(CodeEspece)+str(IdSeq)
    FastaFile.write("%s\n" % (IdSeq))
    FastaFile.write("%s\n" % (Seq))

#-----
# Running the sub-program for first filtering
#-----

cur1.execute(Request)

# Print all the first cell of all the rows
for row in cur1.fetchall():
    IdSeq=str(row[0])+"|"+str(row[1])+"|["+str(row[2])+"|"+str(row[3])+"]"
    Seq=str(row[4])
    print "idseq="+str(IdSeq)
    print "seq="+str(Seq)

    AjoutFasta(FastaFileOut,IdSeq,Seq,CodeEspece) # Writing candidate markers information in the output .fasta file
db.close()

```

End of RAD4SNPs_SQL2fasta.py

Beginning of RAD4SNPp_SamFilter.py

```

#!/usr/bin/python

# RAD4SNPs: A set of Python scripts to select and validate independent SNPs markers from a list of read files #
# SUB-PROGRAM FOR SECOND FILTERING PART 1 : Filtering on a .sam file.

# Authors: G.LASSALLE (gilles.lassalle@inra.fr) & C.DELORD (chrystelle.delord@inra.fr)
# Last update: AUGUST 2017

#----- PRE-CONDITIONS
#- [-i] A sorted .sam file of aligned single-end reads

```

```

#-----
import argparse
import os
import csv
import sys
#-----

parser = argparse.ArgumentParser()
parser.add_argument('-i', action='store', dest='InputFile', help='Sorted .sam file from the first alignment of single-end reads')
parser.add_argument('-x', action='store', dest='nbReadSam', help='Maximum number of outlier reads in an alignment against a candidate marker (def=3)')
parser.add_argument('--version', action='version', version='%s 0.1' % (prog))
results = parser.parse_args()

print 'inputfile          =', results.InputFile

#-----
# Input file checking
#-----
if results.InputFile:
    if os.path.isfile(results.InputFile):
        print "Inserted input is a file: OK."
        ReadCsvFile=csv.reader(open(results.InputFile,"r"),delimiter='\t')
        Base, BaseExt = os.path.splitext(results.InputFile)
        if BaseExt == ".sam" :
            print ".sam extension: OK."
            # Output files initialisation
            FastaFileName=str(Base)+"R1Filtered.fa"
            FastaFileOut=open(FastaFileName,"w")
        else:
            print "The input file extension is not .sam. End of program."
            sys.exit()
    else :
        print "Input file is not found. End of program."
        sys.exit()
else :
    print "Please check out inserted path for the .sam input file [-i]. End of program."
    sys.exit()

if results.nbReadSam :
    nbRead=int(results.nbReadSam)

```

```

else:
    nbRead=3
# Constant values
MaxMismatch=2
# MaxRatio=0.45 (depreciated)

#-----
# Sub-programs for the analysis of input .sam file
#-----
#
def testRead(ligneSamLue,MaxMismatch):
    # Input : a line of the .sam file
    # OutPut : The state of a read (outlier or not) + The state of the candidate marker
    # A candidate with more than >nbReads outliers is discarded.
    # An "outlier" read is a read that does not align end-to-end to the candidate sequence, or display >2 mismatches with the
    candidate sequence.

    ReadState=1 # The state of a read is considered as not-outlier by default.
    NbreNMMax=0
    idSeqCurrent=str(ligneSamLue[2])
    Flag=str(ligneSamLue[1])
    PosSnp=(idSeqCurrent.split('|'))[1]
    Seq=str(ligneSamLue[9])
    CIGAR=str(ligneSamLue[5])
    NM=(ligneSamLue[11].split(":"))[2]
    if int(Flag)==0:
        if str(CIGAR)=="145M":
            SNP=Seq[int(PosSnp)]
            if int(NM)> int(MaxMismatch):
                ReadState=0
            if int(NM)==int(MaxMismatch):
                NbreNMMax=1
        else:
            SNP=""
            ReadState=0 # Aligned reads flagged as secondary (i.e. with Flag=="246") are automatically considered as outliers.
    else:
        SNP=""
        ReadState=0
    return idSeqCurrent,ReadState,SNP,NbreNMMax

#
def ajoutAllele(Dico,allele):

```

```

# Input : aligned reads against a candidate sequence
# OutPut : a dictionary containing the number of reads per allelic state at the SNP candidate marker

if str(allele)<>"":
    if str(allele) in Dico.keys():
        Dico[allele]=int(Dico[allele])+1
    else:
        Dico[allele]=1
return Dico

#
def RatioAllele(listeTuple): # unused in the default settings
    SommeEffectif=0
    for allele in SortTabSnp:
        SommeEffectif= SommeEffectif+allele[1]

    ratio=(float(listeTuple[1][1])/float(SommeEffectif))
    return ratio,SommeEffectif

#
def AjoutLineFasta(Fichier, ListSamLine): # Writing the partially SE pre-validated fasta file.
    # Fichier is linked to a writable file.
    for SamLine in ListSamLine:
        CIGAR=str(SamLine[5])
        NM=(SamLine[11].split(":"))[2]
        if str(CIGAR)=="145M":
            if int(NM)==0:
                idSeqRef=str(SamLine[2])
                Seq=str(SamLine[9])
                Fichier.write("%s%s\n" % (">",idSeqRef))
                Fichier.write("%s\n" %      (Seq))
            break

#
def AjoutLineSam(FichierOut, ListSamLine): # Writing a new .sam file containing only candidate sequences (and associated
aligned reads) that passed the second filtering (1).
    for SamLine in ListSamLine:
        NewLine=""
        for elts in SamLine:
            NewLine=NewLine+str(elts)+"\t"
        FichierOut.write("%s\n" % (NewLine))

```

```

#-----
# Intializing variables
#-----

# Variables for alignments on a candidate sequence
TabSnp={}      # Dictionary of alleles of a candidate SNP
Alignement=list() # Buffer table of the alignment correspong to idSeqCurrent (i.e, the current candidate sequence)
NbreBadRead=0  # Counting the number of outlier reads in an alignment
keep=1         # The alignment/current candidate sequence is kept by default
MaxNM=0       # Counting reads with >MaxMismatch in an alignment

# Variables for the whole sub-program
SeqTotal=0    # Counting number of total candidate sequences
SeqGood=0     # Counting number of SE pre-validated candidate sequences
SeqWrong=0    # Counting number of discarded candidate sequences

#-----
# Screening the .sam file for second filtering:
#-----

print "Reading .sam file."

idSeqRef="" # Current read identifiant.

for refligne in ReadCsvFile:
    idSeqCurrent=str(refligne[2])
    # At the beginning of the .sam file screening, idSeqRef is empty.
    if idSeqRef=="":
        idSeqRef=idSeqCurrent
        ResultatsRead=testRead(refligne,MaxMismatch)
        # Counting outliers read in an alignment.
        if int(ResultatsRead[1])==0:NbreBadRead+=1
        if int(ResultatsRead[3])==1:MaxNM+=1
        # Writing a new .sam file with retained alignments.
        Alignement.append(refligne)
        # Filling table of alleles
        TabSnp=ajoutAllele(TabSnp,ResultatsRead[2])
    # Pursuing the .sam file screening.
else:

```

```

if idSeqRef==idSeqCurrent: # While being in the same alignment (i.e., still screening reads of a given candidate
sequence)

```

```

    ResultatsRead=testRead(refligne,MaxMismatch) # Checking the state of the read (outlier or not)

```

```

    # Incrementing the counter of outlier reads inside the alignment

```

```

    if int(ResultatsRead[1])==0:NbreBadRead+=1

```

```

    if int(ResultatsRead[3])==1:MaxNM+=1

```

```

    Alignement.append(refligne)

```

```

    TabSnp=ajoutAllele(TabSnp,ResultatsRead[2])

```

```

else: # idSeqRef is not empty but different from idSeqCurrent, which means we switched to another candidate
sequence alignment.

```

```

    # The former candidate sequence alignment is then analyzed as a whole:

```

```

        # Sorting the allele table as a tuple list

```

```

        SortTabSnp=sorted(TabSnp.iteritems(), key=lambda item: -item[1])

```

```

        # Discard of the candidate if too many outlier reads:

```

```

        if int(NbreBadRead)>int(nbRead):keep=0

```

```

        # Discard of the candidate if supplementary alleles (e.g. tri- quadriallelic marker) is present with depth coverage >1.

```

```

        if len(SortTabSnp)==4:keep=0

```

```

        if len(SortTabSnp)==3:

```

```

            if int(SortTabSnp[int(len(SortTabSnp))-1][1])>1:

```

```

                keep=0

```

```

            else:

```

```

                CalcRatio=RatioAllele(SortTabSnp)

```

```

        # SNP information:

```

```

        if len(SortTabSnp)==2:

```

```

            CalcRatio=RatioAllele(SortTabSnp) # Obtaining the ratio between depth coverage of each two alleles for
information.

```

```

        # It is possible to discard a candidate if the ratio between depth coverage of the two main allele exceed a certain
value. This option has not been retained in Delord et al. due to low depth coverage that may be unrepresentative of actual
allele frequencies in the discovery pool.

```

```

        # With a higher depth coverage (e.g. >50) it is possible to discard an alignment where the ratio approaches 0.50 as it
might result from paralogy.

```

```

        # as follows:

```

```

        # if float(CalcRatio[0])>float(MaxRatio):keep=0

```

```

        # Checking the retained candidates sequences:

```

```

        if keep==1:

```

```

            SeqGood+=1

```

```

            print idSeqRef

```

```

print SortTabSnp
print "Ratio = "+str(CalcRatio[0])
print "Depth coverage of each allele = "+str(CalcRatio[1])
print "Size of the new .sam file = "+str(len(Alignement))
print ""

# Generating the SE pre-validated fasta file:
AjoutLineFasta(FastaFileOut,Alignement)

# Re-Initialisation of variables:
del Alignement[:]
TabSnp={}
NbreBadRead=0
keep=1
MaxNM=0
idSeqRef=idSeqCurrent
ResultatsRead=testRead(refligne,MaxMismatch)
if int(ResultatsRead[1])==0:NbreBadRead+=1
if int(ResultatsRead[3])==1:MaxNM+=1
Alignement.append(refligne)
TabSnp=ajoutAllele(TabSnp,ResultatsRead[2])

print "Number of candidate sequences that passed Second filtering (1) = "+str(SeqGood)

# .sam file screening is complete: analyzing the last candidate.
if int(NbreBadRead)>int(nbRead):keep=0

# Sorting the allele table as a tuple list
SortTabSnp=sorted(TabSnp.iteritems(), key=lambda item: -item[1])

if len(SortTabSnp)==4:keep=0
if len(SortTabSnp)==3:
    if int(SortTabSnp[int(len(SortTabSnp))-1][1])>1:
        keep=0
    else:
        CalcRatio=RatioAllele(SortTabSnp)

# SNP information:
if len(SortTabSnp)==2:
    CalcRatio=RatioAllele(SortTabSnp)

# if float(CalcRatio[0])>float(MaxRatio):keep=0

```

```

if keep==1:
    SeqGood+=1
    print idSeqRef
    print SortTabSnp
    print "ratio = "+str(CalcRatio[0])
    print "Depth coverage of each allele = "+str(CalcRatio[1])
    print "Size of the new .sam file =" +str(len(Alignement))
    print ""

# Completing the SE pre-validated fasta file:
    AjoutLineFasta(FastaFileOut,Alignement)

FastaFileOut.close

```

End of RAD4SNPs_SamFilter.py

Beginning of RAD4SNPs_PileupFilter.py

```

#!/usr/bin/python

# RAD4SNPs: A set of Python scripts to select and validate independent SNPs markers from a list of read files #
# SUB-PROGRAM FOR SECOND FILTERING PART 2, AND THIRD FILTERING : Filtering on a .pileup file for SE reads then for PE
reads.

# Authors: G.LASSALLE (gilles.lassalle@inra.fr) & C.DELORD (chrystelle.delord@inra.fr)
# Last update: AUGUST 2017

#----- PRE-CONDITIONS
#- [-i] File

#-----
import argparse
import os
import csv
import sys
#-----

parser = argparse.ArgumentParser()
parser.add_argument('-i', action='store', dest='InputFile', help=' sorted SAM file')
parser.add_argument('-nv', action='store', dest='depthFlankMax', help='Couverture maximale autorisee variant flanquant
(def=2)')
parser.add_argument('--version', action='version', version='%(prog)s 0.1')

```



```

results = parser.parse_args()
print 'inputfile          =', results.InputFile

#-----
# Input file checking
#-----
if results.InputFile:
    if os.path.isfile(results.InputFile):
        print "Inserted input is a file: OK."
        ReadCsvFile=csv.reader(open(results.InputFile,"r"),delimiter='\t')
        Base, BaseExt = os.path.splitext(results.InputFile)
        # tester acces ecriture
        GoodFileOut=str(Base)+"NoMulti.txt"
        FileIn0=open(GoodFileOut,"w")
        BadFileOut=str(Base)+"WithMulti.txt"
        FileIn1=open(BadFileOut,"w")

        if BaseExt == ".pileup" :
            print ".pileup extension: OK."
        else:
            print "The input file extension is not .pileup. End of program."
            sys.exit()
    else :
        print "Input file is not found. End of program."
        sys.exit()
else :
    print "Please check out inserted path for the .pileup input file [-i]. End of program."
    sys.exit()

if results.depthFlankMax :
    depthFlankMax=int(results.depthFlankMax)
else:
    depthFlankMax=2

#-----
# Sub-programs for the analysis of input .pileup file
#-----
#
def testPosition(lignelue):
    idSeqCurrent=str(lignelue[0]) # Identifiant of the candidate sequence.
    PosSnp= (idSeqCurrent.split('|'))[1]
    Poslue=int(lignelue[1])

```

```

Variant=str(lignelue[4])
CompteurVariant=0
if int(Poslue)<>int(PosSnp)+1: # In Stacks, bp positions start at zero. A SNP given at the 8th position according to Stacks is
actually the 9th base pair.
    for Nuc in Variant:
        if Nuc.lower() in listNuc:
            CompteurVariant+=1

if int(CompteurVariant) >= int(depthFlankMax) :
    print str(idSeqCurrent)+" keep =" +str(keep)
    return 0
# 0 = Discard of candidate sequences where a flanking variant is found with depth coverage higher than
depthFlankMax.

#
def AjoutLineTxt(Fichier, IdSeq):

    Fichier.write("%s%s\n" % (">",IdSeq)) # Sorting candidate sequences between those that passed second filtering (2) or
third filtering / those that did not.

#-----
# Intializing variables
#-----

# Variables for a given read aligned against the candidate sequence
Variant=""          # Read sequence

# Variables for alignments on a candidate sequence
idSeqRef=0         # ID of the analyzed candidate
keep=1             # The alignment/current candidate sequence is kept by default

# Variables for the whole sub-program
SeqTotal=0        # Counting number of total candidate sequences
SeqGood=0         # Counting number of SE validated or PE-validated candidate sequences
SeqWrong=0        # Counting number of discarded candidate sequences

#-----
# Screening the .sam file for second filtering:
#-----

print "Reading .pileup file..."

```

```

idSeqRef=""
listNuc=('a','t','c','g')
CompteurPileup=0

for refligne in ReadCsvFile:
    idSeqCurrent=str(refligne[0])
    if idSeqRef=="":
        # At the beginning of the .pileup file screening, idSeqRef is empty.
        idSeqRef=idSeqCurrent
        if testPosition(refligne)==0:keep=0
    else:
        if idSeqCurrent==idSeqRef:
            if testPosition(refligne)==0:keep=0
        else : # Switching to the next candidate sequence alignment:

            if keep==1:
                AjoutLineTxt(FileIn0, idSeqRef) # Listing the IDs of candidate sequences that passed the filter

            if keep==0:
                AjoutLineTxt(FileIn1, idSeqRef) # Listing the IDs of candidate sequences that did not passed the filter (i.e. discarded
sequences)

# Re-Initialisation of variables:
    idSeqRef=idSeqCurrent
    keep=1
    CompteurPileup+=1

    if testPosition(refligne)==0:keep=0

# .pileup file screening is complete: analyzing the last candidate.
if keep==1:
    AjoutLineTxt(FileIn0, idSeqRef)
else :
    idSeqRef=idSeqCurrent
    AjoutLineTxt(FileIn1, idSeqRef)

CompteurPileup+=1
print "Number of analyzed candidates = "+str(CompteurPileup)

```

End of RAD4SNPs_PileupFilter.py

Beginning of RAD4SNPs_FinalSQExtract.py

```
#!/usr/bin/python

# RAD4SNPs: A set of Python scripts to select and validate independent SNPs markers from a list of read files #
# SUB-PROGRAM FOR EXTRACTING SUCCESSFUL CANDIDATE SEQUENCES (i.e. that passed the pileup filtering) FROM
STACKS DATABASE

# Authors: G.LASSALLE (gilles.lassalle@inra.fr) & C.DELORD (chrystelle.delord@inra.fr)
# Last update: AUGUST 2017

#----- PRE-CONDITIONS
#- [-i] File

#-----import argparse
import os
import csv
import sys
import MySQLdb
#-----

parser = argparse.ArgumentParser()
parser.add_argument('-i', action='store', dest='InputFile', help='.txt of candidate sequences that passed the .pileup filter')
parser.add_argument('-d', action='store', dest='Database', help='Stacks database')
parser.add_argument('-c', action='store', dest='CodeSp', help='ID of the species')
parser.add_argument('--version', action='version', version='%(prog)s 0.1')
results = parser.parse_args()

print 'inputfile          =', results.InputFile

#-----
# Input file checking
#-----

if results.InputFile:
    if os.path.isfile(results.InputFile):
        ReadCsvFile=csv.reader(open(results.InputFile,"r"),delimiter='|')
        Base, BaseExt = os.path.splitext(results.InputFile)
        if BaseExt != ".txt" :
            print "The input file extension is not .txt. End of program."
            sys.exit() # On teste le format du fichier via son extension, serait mieux avec PySam.
```

```

else :
    print "Input file is not found. End of program."
    sys.exit()
else :
    print "Please check out inserted path for the .txt input file of good sequences [-i]. End of program."
    sys.exit()

#-----
# INITIATE Mysql connect
#-----
db = MySQLdb.connect(host="",          # your host, usually localhost
                    user="",          # your username
                    passwd="",        # your password
                    db=results.Database) # name of the Stacks SQL database
cur1= db.cursor()

#-----
# PROCEDURES (to be complemented in order to simultaneously generate final file in fasta and MassARRAY format – with
candidate SNP within brackets.)
#-----

#
def AjoutFasta(FastaFile,IdSeq,Seq):
    FastaFile.write("%s\n" % (IdSeq))
    FastaFile.write("%s\n" % (Seq))

#
def AjoutSequenom():
    a=5

#-----
# Writing final file.s
#-----

for refligne in ReadCsvFile:
    idSeq=refligne[0]
    idSeq1=idSeq[1:].split('_')
    idSeq=idSeq1[1]

    requete2= "select distinct ma.catalog_id,col,rank_1,rank_2,seq from matches ma join catalog_snps cs on
ma.catalog_id=cs.tag_id join catalog_tags ct on ma.catalog_id=ct.tag_id where ma.catalog_id=\""+idSeq+"\";"
    cur1.execute(requete2)

```

```
for row in cur1.fetchall():
    IdSeq=">"+str(results.CodeSp)+str(row[0])+"|"+str(row[1])+"|["+str(row[2])+"/"+str(row[3])+"]"
    Seq=str(row[4])
    # Generating a fasta file of candidate sequences that passed the .pileup filter.
    print IdSeq
    print Seq

    # PreSeq=Seq[0:int(row[1])]
    # PostSeq=Seq[(int(row[1])+1):]

    # Generating the list of successful candidates sequences in a shape of a sequence with the SNP indicated between
brackets [].
    # MassARRAYId=str(results.CodeSp)+str(row[0])+"\t"+PreSeq+"["+str(row[2])+"/"+str(row[3])+"]"+PostSeq
    # print MassARRAYId

db.close()
```

End of RAD4SNPs_FinalSQLExtract.py

Annexe VIII :

Exemple de fichier final issu du pipeline informatique présenté en Chapitre 1 pour l'espèce *Serrasalmus rhombeus*, modifié dans son formatage afin de servir de support à la définition des multiplexes pour la spectrométrie de masse MALDI-TOF :

Dans chaque en-tête de ligne, les trois lettres correspondent à l'identifiant de l'espèce (ici, SRH) et les chiffres correspondent à l'identifiant du RAD-tag correspondant tel que référencé dans la base de données SQL générée par le logiciel Stacks. La suite de la ligne correspond à la séquence consensus du RAD-tag portant le SNP d'intérêt, indiqué avec ses deux états alléliques entre crochets (par exemple, pour le consensus SRH1134, [G/A]).

Serrasalmus rhombeus:IdSeq	Seq
SRH1134	TGCAGGGAAGCCTCTCCAGGAAGAGGAAGGCC[G/A]TGGACAACACTGCACTGAACTCGGCACGGATATTGTTGTCCTTTCTTTCACAGATTTCTTTCATATTCTTAAATGTAACAAGTTTCTGTTCAATTAAGACTGTATCT
SRH2072	TGCAGGCTTTCTGAAAATAGTTGCCGT[G/A]ATGACAGTCTAGGGGTGCTGCCAAATACTCTGATAAGTTTTGAACCTCTCAAACCTGCTTACACTCAAAAAACACAGTCAAATGTTGTACAGCTGCTGACAAATCCTTTCTGCGT
SRH2602	TGCAGGTTTCACCTTACCCCTAAATCTAAAGCAAGGCGCTTCTGTTGATTTAG[C/A]TGTGCTGTTATTTGGTCACCCGCTTTCAGTTCAGTTCACAACTAATTTTTTGAAGCTGCATGCAAAACACGCACACACAT
SRH2604	TGCAGGTTCTATAATATGTGCACTCT[T/G]GTGTTGACGAGTGTATGGGCCTCTTGCAGAAGTGCATAAGCAGTAATAAGAATCAACTTATGGTGATACACAGTCTTCAAATGTTGGTAAACAACACATTATTGGTGGTCAGAAA
SRH2967	TGCAGGAGTAAAAGCATGGATGAGCTGCTCTAGGTTTGGTGGACACCAAAACCTTGAATCTGGCTATGTTTTAGATGATAGTAGAAC[T/C]TCAACTAGACAGAGAGTCCACCATCAGTGTGATGGGGTTCTCAAAGTGTTTT
SRH6495	TGCAGGCCAATTCTGGCCAATTAGGA[T/C]GTTAAATTTGGGCTGCCAGAAAGTACCCTAATCTCACACCCCTCATCCAGACAGAGAACAACATATTTTTATACTATGCACTATTACATTTAACAGTTTGAACAGCCTTATTATAT
[...]	
SRH13153	TGCAGGCCGGTGTGGAGTCTAGAGGGGAAAATAAGAGAGCAGAGCCATATTTCCATCTGAATCAGGCTGCTCC[C/A]ACTTCCCTATCACCTTTAAAAAAGAGTCTTTTGTTCAGAATCAGAGCGCAAGGATGTGCTTTCTCT
SRH13562	TGCAGGGCATCTCCAGTTCTCTGCTCTTGAGAACAGAAAGGCTTTGGAGATGTCTGTATCTCTGGCCTCTATCTGTCTGTCTCTGCGTCTTT[T/G]TCTCTCTCTCTTTCTCGACTGTAACAGTGTCTTAATTGTATGGA
SRH15109	TGCAGGCACACCCACCTGCAAGCACACCCACAGTGC[T/C]GTGTGTTTTCTGCCAGATCTAATGGCAGTAATAATAACATACCTTCATATAGCTGCCTGGACGTGCTGGGTCTCCACCCTTCCCTGGTTTACCCCGTTAGACG
SRH15233	TGCAGGCTTAAC TAGGTTAACTCTCTCACAAATGCAGGGTGTGTCAAGGAGCCACATATGAGAGCTCATAAAT[G/A]GGTTATTCTACTGAGTGCATTCTTCTATCAGGGACTTAAATAGGCTAGCTCAGAACGGTTCAACAGC
SRH23401	TGCAGGCTTCCATTCTCACTTCTATAGGTTAAGAAAAGTC[C/A]GTTTGCTCTGGATTCCACATAGTTACTGAGGAATACTGCTTTTCTAAAGGATCATATGATAACATGAATCATTTCATTGAAAACAAGCAGTGAATTTTGTCC
SRH25367	TGCAGGCTGAAGCACTGCAGCCCTTACGCTTCTCCAGACAGAGTCAACCGAGGGTGAATTATATGAGCATGGCAATGGTGATTATTCTTTATTACCATTATCATCGCACAGAC[G/T]CACTCACAGTGGTGCAGGGTGGTCT
SRH49432	TGCAGGCCAGAAGTGTCACTCTTTGAAAAATGGTTTATTACTTGAATTCAAATCTCCATGTACTCAGTTCTCAGGGCT[A/G]AATGTAGCGTGGAGGCAGCTACATCAAGCATATTCCTGCAGTTAGTTTAGTTTATGTTAATGGA
SRH46458	TGCAGGAAGGCTTGATTGACATGTTCAAGTGTAGCCCGCGCTGCAGTCAACTGGGAGAATGTTTACTAAACA[C/T]GCAGGAAATGACTCAAGTTCTAGACATCAGCAGCTCAAAAAGACAATACAAGAAAACAGGCTCTCAGAATC
SRH30202	TGCAGGACGTGAGGACCACACTGCCAGGGAGCCAGACTCCCTGCCCTGGCCCCACTGCAAGCCACTGCAAAACAGCTTCACTTGTCTT[G/C]TCTCCAAGCCAGAGTTTCTTTTTAAAGTAGAGCACGACCCAGCCACTGCTAGG
SRH48194	TGCAGGAAATCTGTCTGAAACAGCGCTCACTCA[C/T]GCCGGTGTGTGATGGTACTAGCTGTGTAGTTCTCCAGGTGGTGTAGAGCGACTGAGTGGTTTTCAGCACGACAGCCTACTTTTATAGATGCAGAAGTCTTCGACTGAA
SRH12818	TGCAGGAATGATTCAACAAAATTAATTTCTCGT[A/G]TAATTTCAAAAAGAATGTCTACATGGAATTTGTACAGAACTGAAATGTATTCATATGTACAAAATTTGCTTATGAAATGAAAAGCACTCAGTCAATTTCTTTCAAGA
SRH27071	TGCAGGAAAGCTCTTCTGTCCCTGGGATGGACCTTCTGTACTGCTCTTCTCCGGCTGGCCACATGGTTCGGTCTGCCGATCCTGGTCTTTCGCCAATGCCAT[T/C]CAGCTTCTTCTCACTCAAGGTACAGAGTCCCT

Annexe IX :

Matériel supplémentaire de l'article n°2 présenté en Chapitre 2 :

Riverscape patterns of within and between-population genetic diversity in eighteen harvested Neotropical freshwater fish species.

Delord C.^{1,2}, Le Bail P.-Y.³, Fourcy D.¹, Longin G.⁴, Rinaldo R.⁴, Vigouroux R.², Roussel J.-M.¹, Launey S.¹

¹ESE, Ecology and Ecosystem Health, Agrocampus Ouest, INRA, 35042, Rennes, France

²HYDRECO Guyane SARL, Laboratoire-Environnement de Petit Saut, 97310, Kourou, France

³LPGP, INRA, 35042, Rennes, France

⁴PARC AMAZONIEN DE GUYANE, 97354, Rémire-Montjoly, France

SUPPORTING INFORMATION

This document contains additional information about per-sampling site genetic information (i.e., at the putative population level), for each of the eighteen species studied (Appendix C2.S1). Pairwise F_{ST} are also given for each species (Appendix C2.S2) along with Neighbor-Joining tree on Nei's genetic distance between sites (Appendix C2.S3).

Appendix C2.S1: Information about the number of specimen collected for each species within each site (N) and P_{95} , H_s , H_o , standardized and unstandardized ($St.$) F_{ST-p} , F_{IS} and I_D values for each species and each sampling site. Bold, underlined F_{IS} values are significantly different from zero.

Site	ID	N	Species	P_{95}	H_s	H_o	F_{ST-p}	StF_{ST-p}	F_{IS}	I_D
Saut l'Inspecteur	R1	13	Brycon falcatus	0,981	0,396	0,389	0,003	0,787	0,009	0,110
Saut Tampok	R2	24		0,998	0,396	0,393	0,001	0,267	0,015	0,000
Saut Pierkourou	R3	19		0,991	0,401	0,390	0,000	0,000	0,023	-0,110
Langa Soula	R4	22		0,979	0,393	0,387	0,001	0,244	0,011	0,130
Wayo Gaan Soula	R5	21		0,972	0,397	0,408	0,004	1,000	-0,026	0,010
Apsik icholi	R6	20		0,985	0,400	0,397	0,001	0,299	0,014	-0,100
Eléüéletpe icholi	R7	23		0,979	0,395	0,386	0,001	0,334	0,019	0,010
Papaïchton	V1	14		0,981	0,404	0,383	0,004	0,869	0,052	-0,060
Wempi	V2	-		na	na	na	na	na	na	na
Opoya (Taluen)	V3	-		na	na	na	na	na	na	na
Antecume-Pata	V4	18	0,991	0,397	0,389	0,000	0,088	0,020	0,000	
Pilima-Pata	V5	-	na	na	na	na	na	na	na	
Saut l'Inspecteur	R1	22	Cynodon meionactis	0,922	0,339	0,349	0,003	0,644	-0,025	0,370
Saut Tampok	R2	20		0,950	0,341	0,335	0,004	0,761	0,022	0,240
Saut Pierkourou	R3	20		0,943	0,342	0,328	0,005	1,000	0,034	0,190
Langa Soula	R4	21		0,919	0,349	0,349	0,002	0,402	-0,009	-0,050
Wayo Gaan Soula	R5	20		0,960	0,353	0,351	0,001	0,328	0,002	-0,220
Apsik icholi	R6	20		0,971	0,361	0,367	0,004	0,924	-0,020	-0,460
Eléüéletpe icholi	R7	21		0,938	0,351	0,350	0,001	0,305	0,004	-0,160
Papaïchton	V1	-		na	na	na	na	na	na	na
Wempi	V2	25		0,942	0,346	0,337	0,001	0,201	0,018	-0,050
Opoya (Taluen)	V3	15		0,931	0,345	0,341	0,000	0,000	0,006	0,150
Antecume-Pata	V4	-	na	na	na	na	na	na	na	
Pilima-Pata	V5	-	na	na	na	na	na	na	na	
Saut l'Inspecteur	R1	29	Triportheus brachypomus	0,990	0,431	0,421	0,003	0,930	0,020	-0,240
Saut Tampok	R2	19		0,997	0,432	0,439	0,001	0,325	-0,014	-0,070
Saut Pierkourou	R3	19		0,991	0,432	0,414	0,000	0,049	0,029	-0,040
Langa Soula	R4	20		0,999	0,436	0,436	0,003	0,859	-0,004	-0,240
Wayo Gaan Soula	R5	17		0,979	0,426	0,432	0,000	0,027	-0,018	0,090
Apsik icholi	R6	15		0,986	0,428	0,432	0,000	0,027	-0,005	0,070
Eléüéletpe icholi	R7	20		0,997	0,427	0,422	0,000	0,000	0,014	0,060
Papaïchton	V1	-		na	na	na	na	na	na	na
Wempi	V2	23		0,990	0,418	0,407	0,003	1,000	0,027	0,310
Opoya (Taluen)	V3	12		0,980	0,432	0,437	0,000	0,027	-0,012	0,060
Antecume-Pata	V4	-	na	na	na	na	na	na	na	
Pilima-Pata	V5	-	na	na	na	na	na	na	na	
Saut l'Inspecteur	R1	20	Hoplias aimara	0,980	0,391	0,397	0,001	0,077	-0,020	-0,140
Saut Tampok	R2	20		0,979	0,390	0,398	0,001	0,124	-0,012	-0,130
Saut Pierkourou	R3	19		0,989	0,377	0,375	0,004	0,380	0,005	0,210
Langa Soula	R4	20		0,974	0,396	0,398	0,004	0,363	0,001	-0,250

Wayo Gaan Soula	R5	20	0,972	0,389	0,395	0,005	0,440	-0,014	-0,090
Apsik icholi	R6	25	0,960	0,375	0,373	0,007	0,590	0,008	0,380
Eléüéletpe icholi	R7	20	0,965	0,383	0,378	0,003	0,221	0,009	0,070
Papaïchton	V1	33	0,934	0,386	0,383	0,011	1,000	0,007	-0,210
Wempi	V2	11	0,948	0,376	0,361	0,003	0,219	0,033	0,240
Opoya (Taluen)	V3	13	0,978	0,398	0,408	0,000	0,000	-0,027	-0,120
Antecume-Pata	V4	16	0,984	0,382	0,382	0,001	0,094	0,016	0,110
Pilima-Pata	V5	12	0,944	0,399	0,385	0,001	0,082	0,039	-0,080
Saut l'Inspecteur	R1	11	0,914	0,355	0,365	0,001	0,361	-0,020	0,180
Saut Tampok	R2	21	0,957	0,359	0,361	0,004	1,000	-0,004	-0,020
Saut Pierkourou	R3	12	0,914	0,363	0,360	0,001	0,369	0,000	0,030
Langa Soula	R4	25	0,950	0,365	0,376	0,000	0,041	-0,023	-0,330
Wayo Gaan Soula	R5	15	0,935	0,366	0,366	0,002	0,658	0,005	0,010
Apsik icholi	R6	10	0,970	0,364	0,352	0,000	0,000	0,019	0,130
Eléüéletpe icholi	R7	21	0,968	0,360	0,354	0,003	0,769	0,019	-0,040
Papaïchton	V1	20	0,930	0,366	0,363	0,004	0,987	0,010	-0,100
Wempi	V2	19	0,947	0,366	0,375	0,001	0,324	-0,023	-0,050
Opoya (Taluen)	V3	21	0,957	0,361	0,361	0,001	0,339	-0,006	-0,100
Antecume-Pata	V4	20	0,950	0,362	0,359	0,000	0,115	-0,003	-0,090
Pilima-Pata	V5	13	0,880	0,349	0,358	0,004	0,928	-0,016	0,330
Saut l'Inspecteur	R1	20	0,971	0,381	0,387	0,001	0,708	-0,015	0,130
Saut Tampok	R2	12	0,937	0,376	0,344	0,000	0,188	0,072	0,280
Saut Pierkourou	R3	16	0,953	0,380	0,357	0,000	0,188	0,059	0,090
Langa Soula	R4	19	0,978	0,390	0,386	0,002	1,000	0,016	-0,390
Wayo Gaan Soula	R5	21	0,975	0,387	0,379	0,000	0,188	0,016	-0,340
Apsik icholi	R6	10	0,976	0,382	0,383	0,000	0,188	-0,013	0,150
Eléüéletpe icholi	R7	18	0,958	0,386	0,376	0,000	0,188	0,036	-0,200
Papaïchton	V1	13	0,956	0,377	0,375	0,001	0,599	0,021	0,250
Wempi	V2	-	na	na	na	na	na	na	na
Opoya (Taluen)	V3	-	na	na	na	na	na	na	na
Antecume-Pata	V4	16	0,965	0,384	0,372	0,000	0,000	0,031	-0,010
Pilima-Pata	V5	-	na	na	na	na	na	na	na
Saut l'Inspecteur	R1	-	na	na	na	na	na	na	na
Saut Tampok	R2	20	0,874	0,310	0,311	0,003	0,597	-0,007	0,048
Saut Pierkourou	R3	20	0,859	0,311	0,310	0,006	1,000	-0,003	-0,053
Langa Soula	R4	20	0,859	0,310	0,302	0,003	0,633	0,026	0,075
Wayo Gaan Soula	R5	23	0,846	0,312	0,294	0,000	0,051	0,049	-0,054
Apsik icholi	R6	20	0,860	0,313	0,311	0,003	0,498	0,003	-0,033
Eléüéletpe icholi	R7	20	0,864	0,311	0,292	0,000	0,000	0,069	0,053
Papaïchton	V1	21	0,862	0,317	0,311	0,003	0,524	0,019	-0,213
Wempi	V2	-	na	na	na	na	na	na	na
Opoya (Taluen)	V3	22	0,815	0,307	0,294	0,001	0,187	0,036	0,117
Antecume-Pata	V4	27	0,849	0,315	0,315	0,002	0,373	0,003	-0,241
Pilima-Pata	V5	25	0,846	0,302	0,295	0,004	0,685	0,009	0,299

Saut l'Inspecteur	R1	-	na	na	na	na	na	na	na	
Saut Tampok	R2	18	0,932	0,344	0,325	0,011	0,359	0,050	0,160	
Saut Pierkourou	R3	14	0,860	0,339	0,346	0,021	1,000	-0,014	0,180	
Langa Soula	R4	20	0,943	0,350	0,342	0,009	0,170	0,013	-0,110	
Wayo Gaan Soula	R5	19	0,927	0,351	0,347	0,011	0,311	0,004	-0,120	
Apsik icholi	R6	29	0,924	0,345	0,343	0,007	0,062	0,003	-0,040	
Eléüéletpe icholi	R7	17	0,936	0,351	0,354	0,010	0,282	-0,013	-0,030	
Papaïchton	V1	24	0,948	0,346	0,340	0,010	0,254	0,010	0,100	
Wempi	V2	13	0,911	0,343	0,316	0,010	0,278	0,059	0,160	
Opoya (Taluen)	V3	20	0,976	0,355	0,337	0,009	0,158	0,034	-0,210	
Antecume-Pata	V4	26	0,918	0,346	0,340	0,007	0,037	0,015	0,030	
Pilima-Pata	V5	42	0,902	0,345	0,340	0,006	0,000	0,009	-0,140	

Saut l'Inspecteur	R1	-	na	na	na	na	na	na	na	
Saut Tampok	R2	23	0,847	0,350	0,323	0,000	0,221	0,060	-0,080	
Saut Pierkourou	R3	20	0,904	0,351	0,349	0,006	1,000	0,005	0,210	
Langa Soula	R4	20	0,906	0,356	0,349	0,000	0,201	0,032	-0,230	
Wayo Gaan Soula	R5	15	0,890	0,346	0,353	0,003	0,606	-0,018	0,270	
Apsik icholi	R6	19	0,906	0,365	0,367	0,002	0,481	-0,006	-0,280	
Eléüéletpe icholi	R7	24	0,863	0,354	0,347	-0,002	0,000	0,016	-0,180	
Papaïchton	V1	-	na	na	na	na	na	na	na	
Wempi	V2	27	0,888	0,362	0,361	0,002	0,446	0,008	-0,240	
Opoya (Taluen)	V3	19	0,889	0,355	0,342	0,001	0,272	0,028	0,050	
Antecume-Pata	V4	32	0,882	0,351	0,345	0,002	0,522	0,013	0,090	
Pilima-Pata	V5	19	0,898	0,351	0,335	0,000	0,201	0,033	0,380	

Saut l'Inspecteur	R1	25	0,846	0,321	0,314	0,048	1,000	0,016	0,750	
Saut Tampok	R2	20	0,881	0,338	0,345	0,024	0,101	-0,013	-0,010	
Saut Pierkourou	R3	20	0,857	0,332	0,332	0,029	0,295	0,005	0,260	
Langa Soula	R4	23	0,857	0,338	0,344	0,022	0,036	-0,012	-0,040	
Wayo Gaan Soula	R5	19	0,866	0,336	0,331	0,023	0,082	0,013	0,110	
Apsik icholi	R6	20	0,889	0,351	0,347	0,029	0,285	0,014	-0,440	
Eléüéletpe icholi	R7	21	0,839	0,345	0,339	0,025	0,130	0,020	-0,210	
Papaïchton	V1	-	na	na	na	na	na	na	na	
Wempi	V2	19	0,859	0,339	0,332	0,021	0,000	0,020	-0,020	
Opoya (Taluen)	V3	11	0,867	0,366	0,353	0,021	0,007	0,027	-0,350	
Antecume-Pata	V4	-	na	na	na	na	na	na	na	
Pilima-Pata	V5	-	na	na	na	na	na	na	na	

Saut l'Inspecteur	R1	17	0,923	0,363	0,362	0,007	0,836	0,003	0,100	
Saut Tampok	R2	22	0,914	0,358	0,366	0,005	0,508	-0,016	0,130	
Saut Pierkourou	R3	18	0,941	0,363	0,367	0,004	0,344	0,007	0,090	
Langa Soula	R4	18	0,941	0,366	0,343	0,006	0,672	0,062	0,110	
Wayo Gaan Soula	R5	22	0,921	0,373	0,367	0,008	1,000	0,009	-0,070	
Apsik icholi	R6	21	0,933	0,365	0,357	0,005	0,508	0,025	-0,020	
Eléüéletpe icholi	R7	22	0,920	0,364	0,356	0,006	0,672	0,019	-0,040	
Papaïchton	V1	24	0,919	0,358	0,359	0,007	0,836	0,008	0,290	
Wempi	V2	26	0,958	0,370	0,357	0,005	0,508	0,038	-0,190	
Opoya (Taluen)	V3	22	0,953	0,374	0,341	0,004	0,344	0,079	-0,210	

Antecume-Pata	V4	20	0,947	0,371	0,343	0,002	0,000	0,070	-0,160
Pilima-Pata	V5	21	0,914	0,369	0,354	0,007	0,836	0,041	-0,050
Saut l'Inspecteur	R1	-	na	na	na	na	na	na	na
Saut Tampok	R2	16	0,909	0,375	0,367	0,017	0,508	0,019	0,240
Saut Pierkourou	R3	13	0,861	0,372	0,354	0,025	1,000	0,048	0,330
Langa Soula	R4	6	0,885	0,370	0,331	0,008	0,000	0,089	0,340
Wayo Gaan Soula	R5	15	0,907	0,393	0,391	0,023	0,890	0,002	-0,220
Apsik icholi	R6	-	na	na	na	na	na	na	na
Eléüéletpe icholi	R7	15	0,919	0,381	0,384	0,015	0,409	-0,002	0,090
Papaïchton	V1	17	0,913	0,391	0,385	0,009	0,043	0,009	-0,410
Wempi	V2	-	na	na	na	na	na	na	na
Opoya (Taluen)	V3	19	0,896	0,381	0,373	0,010	0,104	0,014	0,140
Antecume-Pata	V4	33	0,908	0,389	0,385	0,010	0,104	0,006	-0,670
Pilima-Pata	V5	37	0,898	0,377	0,377	0,010	0,120	-0,001	0,060
Saut l'Inspecteur	R1	18	0,954	0,377	0,386	0,006	0,722	-0,021	-0,121
Saut Tampok	R2	20	0,948	0,370	0,375	0,003	0,312	-0,008	0,075
Saut Pierkourou	R3	20	0,932	0,371	0,360	0,007	1,000	0,037	-0,009
Langa Soula	R4	20	0,935	0,367	0,383	0,004	0,435	-0,043	0,136
Wayo Gaan Soula	R5	21	0,959	0,374	0,364	0,005	0,602	0,019	-0,053
Apsik icholi	R6	20	0,950	0,370	0,372	0,006	0,843	-0,016	0,001
Eléüéletpe icholi	R7	22	0,920	0,374	0,378	0,002	0,204	-0,005	-0,097
Papaïchton	V1	18	0,942	0,369	0,369	0,001	0,021	0,002	0,073
Wempi	V2	21	0,964	0,370	0,362	0,004	0,460	0,018	0,098
Opoya (Taluen)	V3	-	na	na	na	na	na	na	na
Antecume-Pata	V4	23	0,962	0,373	0,375	0,001	0,000	0,003	-0,066
Pilima-Pata	V5	29	0,952	0,372	0,367	0,003	0,257	0,022	-0,142
Saut l'Inspecteur	R1	-	na	na	na	na	na	na	na
Saut Tampok	R2	12	0,900	0,370	0,358	0,000	0,103	0,023	0,110
Saut Pierkourou	R3	-	na	na	na	na	na	na	na
Langa Soula	R4	15	0,938	0,379	0,369	0,001	0,282	0,020	-0,150
Wayo Gaan Soula	R5	20	0,952	0,372	0,382	0,004	0,788	-0,022	-0,040
Apsik icholi	R6	20	0,938	0,370	0,375	0,002	0,348	-0,019	0,090
Eléüéletpe icholi	R7	16	0,930	0,366	0,364	-0,001	0,000	0,003	0,150
Papaïchton	V1	18	0,935	0,369	0,378	0,000	0,151	-0,020	0,050
Wempi	V2	21	0,942	0,377	0,367	0,002	0,371	0,029	-0,260
Opoya (Taluen)	V3	28	0,943	0,370	0,374	0,001	0,285	-0,004	-0,090
Antecume-Pata	V4	15	0,885	0,366	0,373	0,006	1,000	-0,023	0,160
Pilima-Pata	V5	20	0,931	0,373	0,366	0,001	0,195	0,014	-0,050
Saut l'Inspecteur	R1	21	0,883	0,342	0,327	0,038	1,000	0,030	0,270
Saut Tampok	R2	22	0,940	0,353	0,343	0,009	0,082	0,025	-0,060
Saut Pierkourou	R3	20	0,954	0,350	0,357	0,010	0,121	-0,018	0,010
Langa Soula	R4	18	0,968	0,353	0,358	0,010	0,132	-0,009	0,010
Wayo Gaan Soula	R5	20	0,960	0,352	0,352	0,013	0,214	-0,007	-0,010
Apsik icholi	R6	22	0,928	0,355	0,348	0,010	0,114	0,033	-0,170
Eléüéletpe icholi	R7	21	0,961	0,357	0,351	0,012	0,166	0,010	-0,170
Papaïchton	V1	-	na	na	na	na	na	na	na

Wempi	V2	20	0,961	0,348	0,350	0,012	0,185	-0,006	0,100
Opoya (Taluen)	V3	20	0,947	0,353	0,337	0,008	0,060	0,029	-0,030
Antecume-Pata	V4	28	0,972	0,349	0,334	0,006	0,000	0,039	0,040
Pilima-Pata	V5	-	na	na	na	na	na	na	na
Saut l'Inspecteur	R1	20	0,765	0,266	0,264	0,012	0,831	0,002	0,020
Saut Tampok	R2	23	0,727	0,271	0,266	0,005	0,280	0,005	-0,210
Saut Pierkourou	R3	19	0,779	0,277	0,277	0,015	1,000	-0,005	-0,360
Langa Soula	R4	18	0,715	0,259	0,251	0,009	0,551	0,029	0,190
Wayo Gaan Soula	R5	20	0,750	0,257	0,248	0,009	0,546	0,032	0,370
Apsik icholi	R6	-	na	na	na	na	na	na	na
Eléüéletpe icholi	R7	19	0,764	0,264	0,251	0,014	0,968	0,043	-0,010
Papaïchton	V1	20	0,735	0,265	0,258	0,006	0,322	0,014	0,040
Wempi	V2	25	0,741	0,265	0,258	0,004	0,208	0,019	0,000
Opoya (Taluen)	V3	26	0,687	0,264	0,254	0,003	0,102	0,021	-0,070
Antecume-Pata	V4	17	0,740	0,267	0,267	0,001	0,000	-0,004	0,010
Pilima-Pata	V5	10	0,801	0,271	0,243	0,002	0,077	0,083	0,030
Saut l'Inspecteur	R1	21	0,801	0,293	0,296	0,045	0,941	-0,010	0,500
Saut Tampok	R2	22	0,861	0,300	0,303	0,018	0,185	-0,004	0,330
Saut Pierkourou	R3	20	0,868	0,306	0,304	0,048	1,000	0,007	0,190
Langa Soula	R4	22	0,894	0,327	0,321	0,017	0,144	0,010	-0,360
Wayo Gaan Soula	R5	22	0,881	0,309	0,308	0,030	0,524	0,008	0,090
Apsik icholi	R6	21	0,887	0,311	0,306	0,018	0,178	0,015	0,060
Eléüéletpe icholi	R7	20	0,927	0,323	0,323	0,015	0,112	0,007	-0,220
Papaïchton	V1	30	0,899	0,317	0,311	0,016	0,118	0,018	-0,190
Wempi	V2	31	0,906	0,311	0,305	0,013	0,054	0,020	-0,010
Opoya (Taluen)	V3	21	0,874	0,318	0,311	0,011	0,000	0,018	-0,150
Antecume-Pata	V4	23	0,907	0,320	0,309	0,013	0,056	0,029	-0,210
Pilima-Pata	V5	30	0,919	0,311	0,304	0,013	0,055	0,021	-0,040
Saut l'Inspecteur	R1	17	0,888	0,357	0,354	0,072	1,000	0,010	0,470
Saut Tampok	R2	13	0,915	0,372	0,364	0,044	0,436	0,012	0,130
Saut Pierkourou	R3	19	0,894	0,358	0,355	0,057	0,698	0,000	0,570
Langa Soula	R4	15	0,934	0,378	0,367	0,038	0,321	0,024	-0,010
Wayo Gaan Soula	R5	-	na	na	na	na	na	na	na
Apsik icholi	R6	10	0,967	0,371	0,346	0,039	0,328	0,057	0,230
Eléüéletpe icholi	R7	-	na	na	na	na	na	na	na
Papaïchton	V1	21	0,936	0,372	0,350	0,026	0,078	0,054	0,080
Wempi	V2	10	0,945	0,370	0,352	0,053	0,614	0,042	0,220
Opoya (Taluen)	V3	23	0,932	0,385	0,384	0,027	0,091	-0,004	-0,570
Antecume-Pata	V4	21	0,970	0,391	0,389	0,027	0,088	0,008	-0,690
Pilima-Pata	V5	15	0,948	0,395	0,393	0,022	0,000	0,007	-0,480

Appendix C2.S2: Tables of pairwise F_{ST} per species. Grey, bold values are significant (CIs not including zero).

Table of pairwise F_{ST} values for *Brycon falcatus*:

	V1	V4	R1	R2	R3	R4	R5	R6
V1	-							
V4	0,001	-						
R1	0,009	0,001	-					
R2	0,004	0,000	0,001	-				
R3	0,002	-0,003	0,003	0,001	-			
R4	0,002	0,001	-0,001	0,001	-0,003	-		
R5	0,007	0,004	0,005	0,002	0,000	0,005	-	
R6	0,003	-0,002	0,003	-0,001	0,000	0,000	0,006	-
R7	0,003	-0,001	0,005	0,001	-0,004	0,001	0,006	-0,001

Table of pairwise F_{ST} values for *Cynodon meionactis*:

	V2	V3	R1	R2	R3	R4	R5	R6
V2	-							
V3	-0,004	-						
R1	-0,001	0,001	-					
R2	0,002	0,002	0,003	-				
R3	0,004	-0,003	0,006	0,007	-			
R4	0,004	0,002	0,003	0,003	0,005	-		
R5	-0,001	-0,001	0,003	0,003	0,007	-0,001	-	
R6	0,002	0,002	0,008	0,007	0,009	0,001	0,001	-
R7	0,000	0,001	0,002	0,001	0,003	-0,002	0,001	0,005

Table of pairwise F_{ST} values for *Triportheus brachypomus*:

	V2	V3	R1	R2	R3	R4	R5	R6
V2	-							
V3	0,000	-						
R1	0,006	-0,001	-					
R2	0,006	0,000	0,002	-				
R3	0,003	-0,004	-0,001	-0,001	-			
R4	0,006	0,000	0,008	0,003	0,002	-		
R5	0,004	-0,004	0,002	-0,002	0,000	0,002	-	
R6	-0,001	-0,004	0,004	-0,001	0,001	-0,002	-0,003	-
R7	0,002	-0,001	0,003	0,001	-0,002	0,003	-0,006	-0,002

Table of pairwise F_{ST} values for *Hoplias aimara*:

	V1	V2	V3	V4	V5	R1	R2	R3	R4	R5	R6
V1	-										
V2	0,012	-									
V3	0,004	-0,008	-								
V4	0,010	0,005	-0,006	-							
V5	0,010	0,000	-0,008	-0,003	-						
R1	0,012	-0,003	-0,006	0,001	0,001	-					
R2	0,011	-0,001	-0,011	0,001	0,002	-0,006	-				
R3	0,020	0,002	-0,004	-0,001	0,003	-0,001	-0,004	-			
R4	0,009	0,006	-0,009	0,001	0,001	0,003	0,006	0,009	-		
R5	0,009	0,006	0,001	0,002	0,002	0,000	0,003	0,008	0,006	-	
R6	0,017	0,006	0,001	0,004	0,005	0,003	0,009	0,010	0,009	0,010	-
R7	0,012	0,002	-0,009	-0,001	-0,003	0,006	0,005	0,005	0,005	0,006	0,001

Table of pairwise F_{ST} values for *Leporinus friderici*:

	V1	V2	V3	V4	V5	R1	R2	R3	R4	R5	R6
V1	-										
V2	0,005	-									
V3	0,001	-0,002	-								
V4	0,000	-0,002	-0,001	-							
V5	0,005	0,007	0,005	0,004	-						
R1	0,003	-0,002	0,002	0,000	0,003	-					
R2	0,010	0,000	0,005	0,001	0,012	0,006	-				
R3	0,001	0,004	-0,002	-0,001	0,003	0,004	0,002	-			
R4	0,000	-0,002	0,001	-0,003	-0,001	-0,003	0,002	-0,001	-		
R5	0,002	0,001	0,003	-0,001	-0,001	0,000	0,005	0,003	0,002	-	
R6	0,007	-0,001	-0,004	0,002	-0,002	-0,003	-0,001	0,001	-0,009	0,006	-
R7	0,008	0,002	0,004	0,003	0,003	0,002	0,001	0,000	0,001	0,005	0,001

Table of pairwise F_{ST} values for *Leporinus lebaili*:

	V1	V4	R1	R2	R3	R4	R5	R6
V1	-							
V4	0,001	-						
R1	0,006	0,002	-					
R2	-0,004	-0,008	-0,002	-				
R3	-0,007	-0,004	-0,002	-0,010	-			
R4	0,005	0,005	0,002	0,001	-0,005	-		
R5	0,002	0,000	0,004	-0,009	-0,009	0,000	-	
R6	0,002	-0,002	-0,001	-0,010	-0,007	0,002	-0,003	-
R7	0,003	0,003	0,000	-0,010	-0,008	0,003	-0,001	-0,001

Table of pairwise F_{ST} values for *Acnodon oligacanthus*:

	V1	V3	V4	V5	R2	R3	R4	R5	R6
V1	-								
V3	-0,004	-							
V4	-0,001	-0,003	-						
V5	0,001	0,000	-0,001	-					
R2	0,002	0,001	-0,001	0,006	-				
R3	0,006	0,006	0,004	0,009	0,000	-			
R4	0,002	-0,001	0,000	0,003	0,003	0,006	-		
R5	-0,002	-0,004	-0,003	0,000	-0,001	0,001	-0,001	-	
R6	0,000	0,000	0,001	0,003	0,001	0,005	0,005	-0,005	-
R7	0,001	-0,005	-0,002	0,000	-0,001	0,000	-0,003	-0,004	-0,003

Table of pairwise F_{ST} values for *Myloplus rhomboidalis*:

	V1	V2	V3	V4	V5	R2	R3	R4	R5	R6
V1	-									
V2	0,014	-								
V3	0,006	0,008	-							
V4	0,004	0,010	0,006	-						
V5	0,005	0,006	0,005	-0,002	-					
R2	0,012	0,001	0,009	0,012	0,013	-				
R3	0,020	0,014	0,024	0,017	0,020	0,004	-			
R4	0,008	0,010	0,007	0,006	0,005	0,017	0,027	-		
R5	0,015	0,012	0,010	0,009	0,007	0,018	0,028	-0,003	-	
R6	0,002	0,014	0,003	0,001	-0,001	0,013	0,026	0,005	0,005	-
R7	0,013	0,012	0,008	0,004	0,005	0,017	0,027	0,005	0,007	0,004

Table of pairwise F_{ST} values for *Myloplus rubripinnis*:

	V2	V3	V4	V5	R2	R3	R4	R5	R6
V2	-								
V3	-0,002	-							
V4	0,001	0,002	-						
V5	0,005	-0,005	-0,001	-					
R2	-0,003	-0,004	0,003	-0,003	-				
R3	0,004	0,010	0,008	0,003	0,006	-			
R4	0,001	0,000	0,001	-0,009	-0,005	0,000	-		
R5	0,005	0,000	0,003	-0,001	0,007	0,008	0,002	-	
R6	0,006	0,005	0,005	-0,009	0,004	0,009	-0,002	0,004	-
R7	0,000	-0,001	0,000	-0,010	-0,002	0,006	-0,004	0,001	-0,003

Table of pairwise F_{ST} values for *Tometes leballi*:

	V1	V3	V4	V5	R2	R3	R4	R5
V1	-							
V3	0,003	-						
V4	0,001	0,006	-					
V5	0,000	0,004	0,001	-				
R2	0,015	0,018	0,018	0,012	-			
R3	0,023	0,024	0,028	0,020	0,006	-		
R4	0,008	-0,004	0,006	0,010	0,006	0,019	-	
R5	0,018	0,022	0,018	0,023	0,028	0,037	0,008	-
R7	0,006	0,008	0,005	0,004	0,026	0,039	0,010	0,025

Table of pairwise F_{ST} values for *Serrasalmus eigenmanni*:

	V2	V3	R1	R2	R3	R4	R5	R6
V2	-							
V3	0,024	-						
R1	0,014	0,056	-					
R2	0,001	0,026	0,025	-				
R3	0,002	0,037	0,024	0,002	-			
R4	0,031	0,008	0,059	0,027	0,034	-		
R5	0,025	0,011	0,062	0,030	0,038	0,003	-	
R6	0,039	0,006	0,074	0,043	0,050	0,010	0,011	-
R7	0,032	0,002	0,069	0,038	0,046	0,006	0,007	-0,002

Table of pairwise F_{ST} values for *Serrasalmus rhombeus*:

	V1	V2	V3	V4	V5	R1	R2	R3	R4	R5	R6
V1	-										
V2	0,006	-									
V3	0,006	0,005	-								
V4	0,006	0,002	-0,002	-							
V5	0,007	0,010	0,004	0,002	-						
R1	0,008	0,002	0,004	0,004	0,007	-					
R2	0,008	0,001	0,006	0,001	0,010	0,005	-				
R3	0,002	0,006	0,006	0,000	0,011	0,007	-0,002	-			
R4	0,008	0,007	-0,001	0,003	0,004	0,007	0,006	0,006	-		
R5	0,005	0,008	0,008	0,005	0,012	0,008	0,012	0,001	0,007	-	
R6	0,008	0,007	0,004	-0,003	0,003	0,010	0,004	0,001	0,009	0,011	-
R7	0,009	0,003	0,008	0,004	0,007	0,008	0,008	0,005	0,010	0,007	0,001

Table of pairwise F_{ST} values for *Ageneiosus inermis*:

	V1	V2	V4	V5	R1	R2	R3	R4	R5	R6
V1	-									
V2	0,001	-								
V4	-0,004	0,005	-							
V5	-0,002	0,006	0,003	-						
R1	-0,002	0,005	0,004	0,003	-					
R2	0,001	-0,002	-0,002	0,001	0,008	-				
R3	0,003	0,007	0,002	0,008	0,008	0,005	-			
R4	0,006	0,003	-0,003	0,007	0,007	0,005	0,005	-		
R5	0,003	0,005	0,004	0,002	0,012	0,003	0,012	0,001	-	
R6	0,011	0,004	0,002	0,002	0,007	0,006	0,012	0,006	0,007	-
R7	-0,001	0,003	0,001	0,002	0,003	0,003	0,008	0,002	0,006	0,007

Table of pairwise F_{ST} values for *Doras micropoeus*:

	V1	V2	V3	V4	V5	R2	R4	R5	R6
V1	-								
V2	0,002	-							
V3	-0,002	0,001	-						
V4	0,004	0,006	0,007	-					
V5	0,001	-0,002	0,000	0,004	-				
R2	0,002	-0,002	0,000	0,002	-0,005	-			
R4	-0,002	0,000	0,001	0,008	0,002	-0,003	-		
R5	0,004	0,007	0,005	0,008	0,004	-0,001	0,008	-	
R6	-0,001	0,002	0,001	0,009	0,004	-0,001	-0,001	0,000	-
R7	-0,006	0,002	-0,004	0,004	-0,002	-0,003	-0,003	0,005	0,001

Table of pairwise F_{ST} values for *Harttia guianensis*:

	V2	V3	V4	R1	R2	R3	R4	R5	R6
V2	-								
V3	0,009	-							
V4	0,004	0,000	-						
R1	0,039	0,036	0,038	-					
R2	0,002	0,004	0,003	0,034	-				
R3	0,002	0,005	0,004	0,036	0,000	-			
R4	0,013	0,004	-0,002	0,043	0,008	0,014	-		
R5	0,013	0,007	0,005	0,042	0,010	0,014	0,004	-	
R6	0,011	0,002	0,001	0,040	0,008	0,007	0,004	0,012	-
R7	0,017	0,006	0,005	0,035	0,010	0,010	0,007	0,011	0,004

Table of pairwise F_{ST} values for *Hypostomus gymnorhynchus*:

	V1	V2	V3	V4	V5	R1	R2	R3	R4	R5
V1										
V2	0,002									
V3	-0,004	-0,002								
V4	0,000	-0,006	-0,005							
V5	-0,003	-0,002	-0,008	-0,002						
R1	0,007	0,011	0,006	0,003	0,011					
R2	0,006	0,004	0,003	0,002	0,003	0,008				
R3	0,015	0,006	0,011	0,007	0,015	0,017	0,006			
R4	0,010	0,007	0,007	0,004	0,003	0,021	0,004	0,022		
R5	0,009	0,010	0,006	0,003	0,002	0,018	0,005	0,019	-0,001	
R7	0,015	0,011	0,011	0,008	0,005	0,023	0,011	0,029	0,012	0,016

Table of pairwise F_{ST} values for *Pseudancistrus barbatus*:

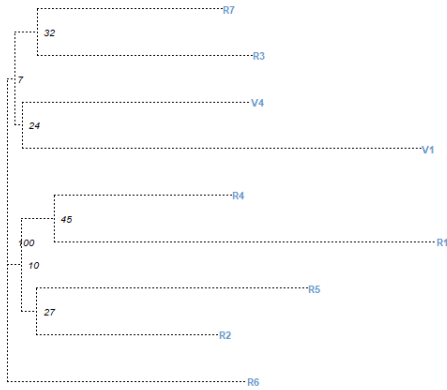
	V1	V2	V3	V4	V5	R1	R2	R3	R4	R5	R6
V1	-										
V2	0,006	-									
V3	0,004	0,001	-								
V4	0,002	0,006	0,000	-							
V5	0,006	0,001	-0,003	0,002	-						
R1	0,046	0,036	0,038	0,042	0,043	-					
R2	0,014	0,008	0,006	0,012	0,012	0,036	-				
R3	0,047	0,036	0,043	0,048	0,051	0,047	0,031	-			
R4	0,011	0,009	0,004	0,008	0,008	0,054	0,017	0,044	-		
R5	0,022	0,025	0,026	0,026	0,020	0,062	0,029	0,065	0,010	-	
R6	0,008	0,007	0,008	0,004	0,005	0,047	0,018	0,058	0,013	0,025	-
R7	0,006	0,012	-0,001	-0,003	0,003	0,050	0,015	0,054	0,005	0,024	0,004

Table of pairwise F_{ST} values for *Geophagus harreri*:

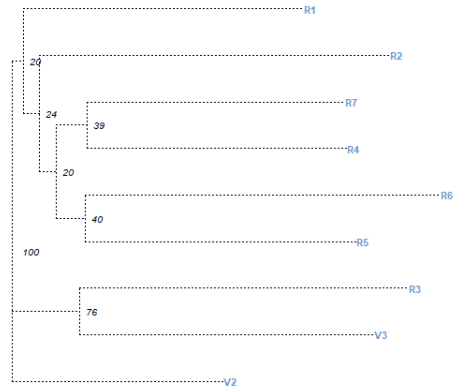
	V1	V2	V3	V4	V5	R1	R2	R3	R4
V1	-								
V2	0,039	-							
V3	0,008	0,041	-						
V4	0,012	0,038	0,008	-					
V5	0,003	0,042	0,006	0,002	-				
R1	0,062	0,068	0,065	0,058	0,059	-			
R2	0,028	0,056	0,030	0,029	0,031	0,059	-		
R3	0,046	0,056	0,048	0,052	0,043	0,099	0,058	-	
R4	0,017	0,069	0,020	0,025	0,009	0,075	0,052	0,054	-
R6	0,021	0,063	0,016	0,015	0,008	0,098	0,050	0,053	0,022

Appendix C2.S3: Neighbor-joining trees of Nei's genetic distance between sampling sites for each species (bootstraps = 999) :

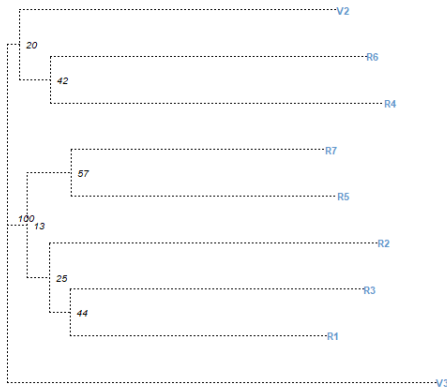
Brycon falcatus



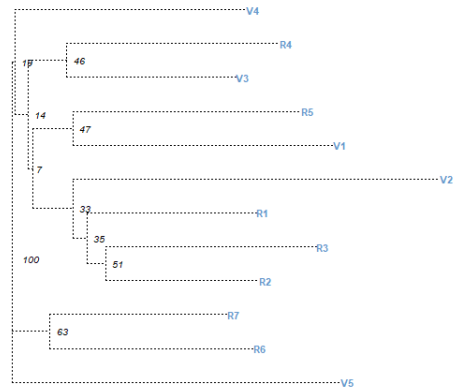
Cynodon meionactis



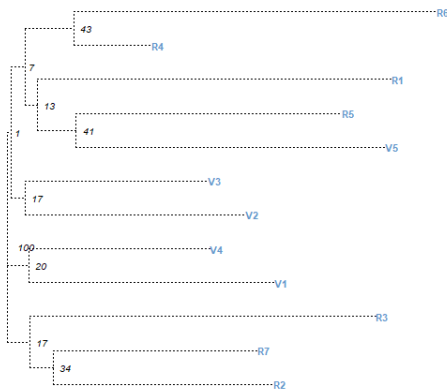
Triportheus brachypomus



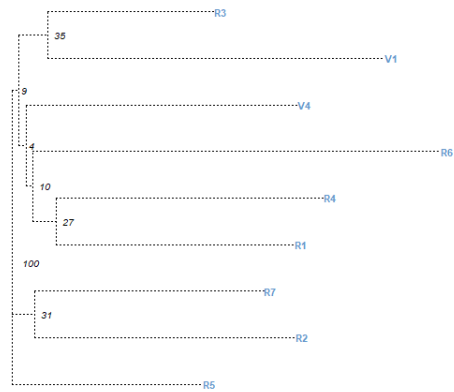
Hoplias aimara



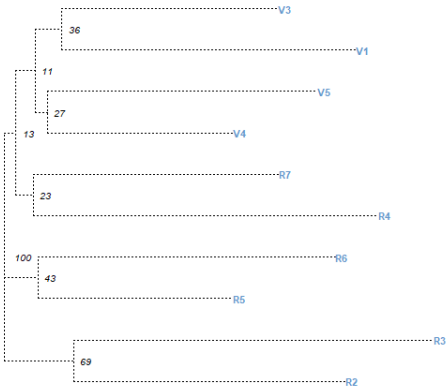
Leporinus friderici



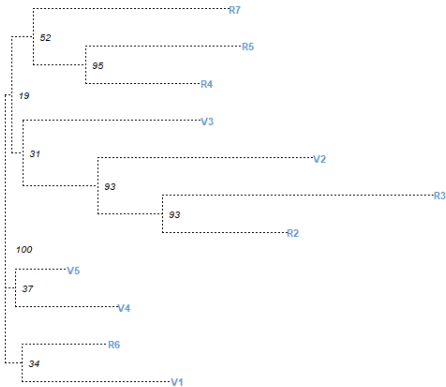
Leporinus lebaili



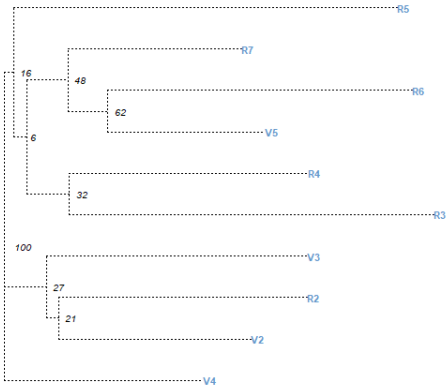
Acnodon oligacanthus



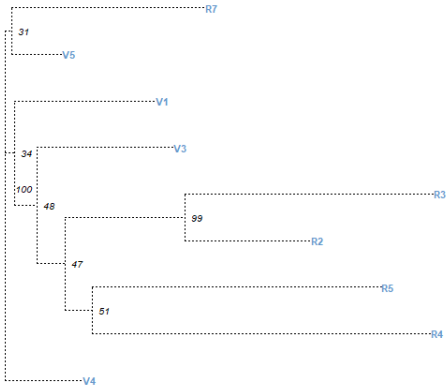
Myloplus rhomboidalis



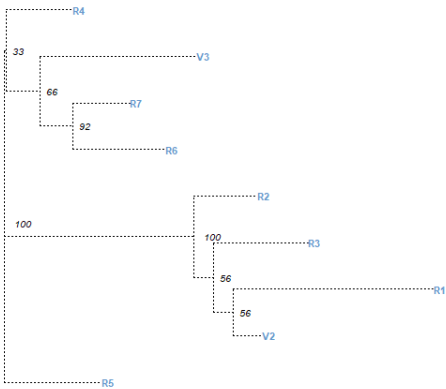
Myloplus rubripinnis



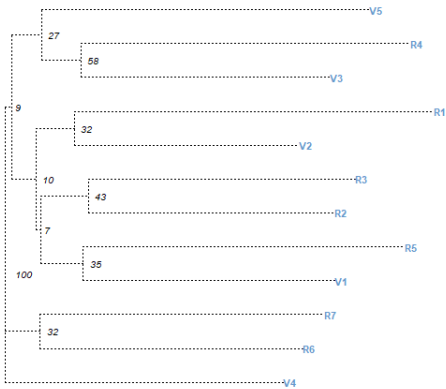
Tometes lebaili



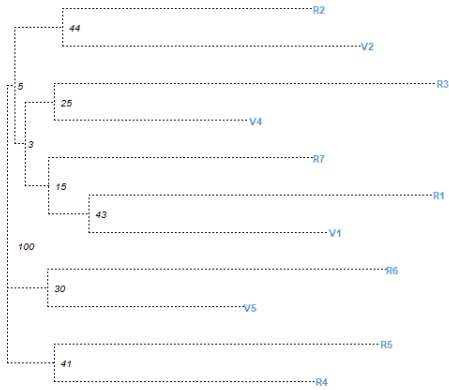
Serrasalmus eigenmanni



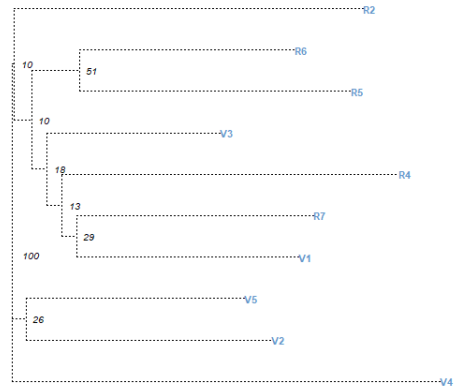
Serrasalmus rhombeus



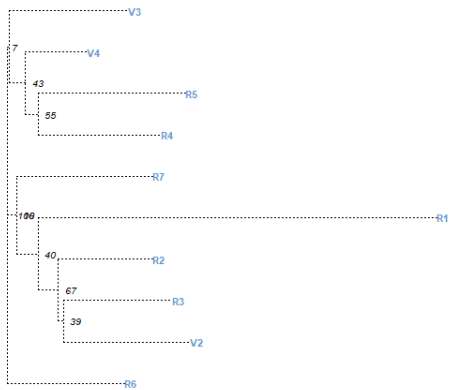
Ageneiosus inermis



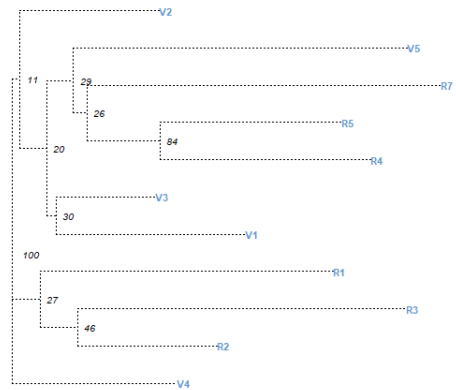
Doras micropoeus



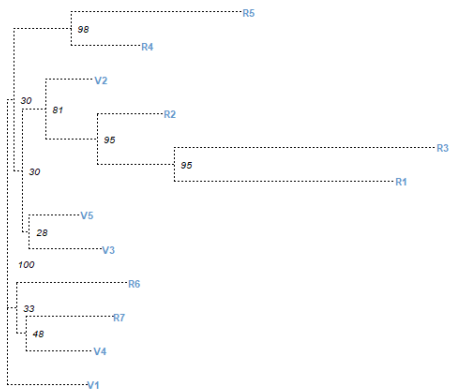
Harttia guianensis



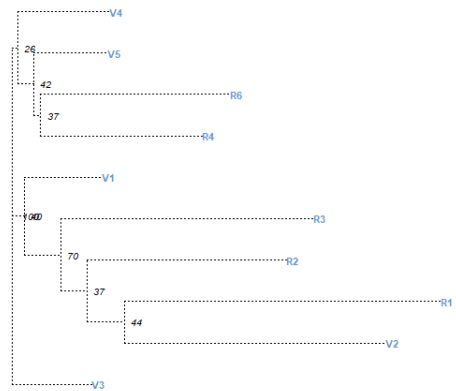
Hypostomus gymnorhynchus



Pseudancistrus barbatus



Geophagus harreri



Annexe X :**Scripts SimuPOP commentés ayant fourni les pseudo-données analysées dans le Chapitre 3 :**

```

#!/usr/bin/python

# Created on Tue Nov 14 08:55:04 2017, updated on Nov 2018
# Author: cdelord
#-----#

# Object: A metapopulation consisting of 16 sub-populations (patches='subPop') arranged in a fractal, river network design.
# Objective:
# - Qualitatively observe the time-lag that is necessary to detect harvesting effect on genetic diversity using SNPs.
# - See how the expected 'dendritic pattern' of genetic variation appears and evolves when simulating downstream population
  size reduction.

# To simulate fishing pressure, individuals are randomly removed from a few selected patches ('subPop') (c.f. Chapter 3).
# CAUTION: IN THIS SCRIPT, each 'subPop' that is about to collapse (i.e. less than 10 individuals left at a given generation) is
  'resized' to a size of 30 whose genotypes are duplicated from the remaining individuals. This was done to avoid SimuPOP to
  crash if there are no individuals left in a 'subPop'. We caution that this is not the same thing as maintaining 'subPop' sizes
  constant.
# Improvements of this script would imply to avoid such an arbitrary resize by finding a way, for instance, of making SimuPOP
  to keep running when a 'subPop' collapses, so the whole metapopulation extinction process is observable. This would
  enable to test for various demographic models (defined within the 'matingSchemes' settings, for example the number of
  offsprings relative to the number of breeders) where the metapopulation is more or less likely to face the simulated fishing.

# Individual removals happen before mating (i.e. during the 'preOps' phase).

#-----#
#----- Importing Python libraries for simulations -----#
#-----#

from simuOpt import setOptions
setOptions(alleleType='binary')
# We are simulating bi-allelic loci only and we are using text-based progress bar.

import simuPOP as sim
# import simuPOP.demography as demo
import numpy
from simuPOP.sampling import drawRandomSample
from simuPOP.utils import importPopulation, export

#-----#
#----- Defining functions to be used during runs -----#
#-----#

# Function to define the number of individuals within each 'subPop' after each mating event:
def model1(pop):
    sim.stat(pop, popSize=True)
    subsize=pop.dvars().subPopSize
    vecsize = numpy.array(subsize)
    vecsize[vecsize < 10] = 30 # A subPop with less than 10 individuals is resized to 30.
    return vecsize

```

```

# Function to define the number of individuals within each 'subPop' after each mating event:
# USED FOR FIGURES OF CHAPTER 3 WITH A CONSTANT SUBPOP SIZE MAINTAINED (Fig.C3.6 to C3.9)
# def model2(pop):
#   sim.stat(pop,popSize=True)
#   subsize=pop.dvars().subPopSize
#   numpop=len(subsize)
#   vecsize=[int(N)]*(numpop) # subPop size is maintained constant and equal to initial N through time.
#   return vecsize

# Function to export samples from all 16 'subPop' at each generation:
def exportSample(pop):
    sample = drawRandomSample(pop, sizes=[20]*16)
    sim.util.Exporter(format='GENEPOP', title="SimuPop output", adjust=1, output='geneSim_%s_%s_%s_%s_%s_gen' %
(str(N), str(m), str(H), str(k), pop.dvars().gen), gui=False).apply(sample)
    return True

#-----#
#----- Defining allele frequencies-----#
#-----#

# Adapted from the methods of Vatsiou et al. 2016, Molecular Ecology (2016) 25, 89–103:
# The following allele frequencies are drawn from empirical data from the Maroni for S. rhombeus (130 randomly drawn
# frequencies from whichever sampling sites). However, the actual allele frequencies that will be used to initialize each run
# will be resampled using a Dirichlet distribution of parameters 'alpha' and 'bhta' obtained from initial frequencies and
# parameters N and m (see below).

snps=[0.184,0.444,0.10,0.10,0.429,0.457,0.206,0.433,0.237,0.421,0.133,0.278,0.10,0.45,0.184,0.385,0.404,0.107,0.361,0.4
5,0.357,0.167,0.133,0.292,0.412,0.3,0.333,0.177,0.10,0.275,0.475,0.139,0.357,0.143,0.281,0.278,0.5,0.182,0.286,0.425,0.3
5,0.444,0.325,0.1,0.5,0.119,0.395,0.175,0.344,0.125,0.5,0.367,0.5,0.25,0.333,0.346,0.18,0.475,0.452,0.375,0.14,0.295,0.12,
0.389,0.357,0.139,0.5,0.188,0.194,0.132,0.10,0.457,0.25,0.281,0.179,0.114,0.429,0.105,0.25,0.471,0.344,0.476,0.481,0.45,
0.5,0.36,0.167,0.10,0.425,0.421,0.119,0.11,0.14,0.156,0.261,0.438,0.2,0.105,0.171,0.265,0.364,0.132,0.25,0.325,0.475,0.45
2,0.083,0.389,0.25,0.135,0.148,0.316,0.312,0.421,0.395,0.381,0.133,0.25,0.312,0.3,0.395,0.5,0.179,0.32,0.5,0.342,0.2,0.40
6,0.368,0.15]

#-----#
#----- Parameters setting within a for loop -----#
#-----#

# Define a loop to run independent simulations.
num_sim=30

# Define parameter space.
Harv_param=[0.000, 0.125]
N=1000 # was also set to 100
m=0.2 # was also set to 0.05

for H in Harv_param:

    # Initializing allele frequencies for the run according to N, m and initial 'snps' object as performed by Vatsiou et al. (2016):
    initFreq=[] # Initial allele frequencies are identical within each subPop.

    for i in range(len(snps)):
        alpha=float(4*int(N)*float(m)*snps[i])
        bhta=float(4*int(N)*float(m)*(1-snps[i]))
        p=numpy.random.beta(alpha,bhta)
        while (p>=0.90 or p<=0.10):
            p=numpy.random.beta(alpha,bhta)

```



```

initFreq.append(p)
print " SNP {snp} with alpha {alpha}, bhta {bhta} and frequency {p}".format(snp=i, alpha=alpha, bhta=bhta, p=p)

# Launching simulation replicates - with identical initialized allele frequencies p[i]
for k in range(1, (num_sim+1)):
    pop = sim.Population(size=[int(N)]*16, ploidy=2, loci=len(snps), infoFields=['migrate_to', 'age', 'ind_id'])

# Defining migration matrix: each individual as an identical probability of moving to an adjacent 'subPop'.
M = [[0.000,float(m),0.000,0.000,0.000,0.000,0.000,0.000,0.000,0.000,0.000,0.000,0.000,0.000,0.000,0.000],
      [float(m),0.000,float(m),0.000,0.000,0.000,0.000,0.000,0.000,0.000,0.000,0.000,0.000,0.000,0.000,0.000],
      [0.000,float(m),0.000,float(m),0.000,0.000,0.000,0.000,0.000,0.000,float(m),0.000,0.000,0.000,0.000,0.000],
      [0.000,0.000,float(m),0.000,float(m),0.000,0.000,0.000,0.000,0.000,0.000,0.000,0.000,0.000,0.000,0.000],
      [0.000,0.000,0.000,float(m),0.000,float(m),0.000,0.000,0.000,0.000,0.000,0.000,float(m),0.000,0.000,0.000],
      [0.000,0.000,0.000,0.000,float(m),0.000,float(m),0.000,0.000,0.000,0.000,0.000,0.000,0.000,0.000,0.000],
      [0.000,0.000,0.000,0.000,0.000,float(m),0.000,float(m),0.000,0.000,0.000,0.000,0.000,0.000,0.000,0.000],
      [0.000,0.000,0.000,0.000,0.000,0.000,float(m),0.000,0.000,0.000,0.000,0.000,0.000,0.000,0.000,0.000],
      [0.000,0.000,0.000,0.000,0.000,0.000,0.000,float(m),0.000,0.000,0.000,0.000,float(m),0.000,0.000,0.000,0.000],
      [0.000,0.000,0.000,0.000,0.000,0.000,0.000,0.000,float(m),0.000,0.000,0.000,0.000,0.000,0.000,0.000,0.000],
      [0.000,0.000,0.000,0.000,0.000,0.000,0.000,0.000,0.000,0.000,float(m),0.000,float(m),0.000,0.000,0.000,0.000],
      [0.000,0.000,0.000,0.000,0.000,0.000,0.000,0.000,0.000,0.000,0.000,float(m),0.000,float(m),0.000,float(m)],
      [0.000,0.000,0.000,0.000,0.000,0.000,0.000,0.000,0.000,0.000,0.000,0.000,float(m),0.000,0.000,0.000,0.000],
      [0.000,0.000,0.000,0.000,0.000,0.000,0.000,0.000,0.000,0.000,0.000,0.000,0.000,0.000,0.000,float(m)],
      [0.000,0.000,0.000,0.000,0.000,0.000,0.000,0.000,0.000,0.000,0.000,0.000,0.000,0.000,0.000,float(m),0.000]]

print("Simulation "+str(k)+" running for N="+str(N)+" , m="+str(m)+" and H="+str(H))

pop.evolve(
    # 'initOps' defines what happens at the very beginning of the simulation at generation 0:
    initOps=[sim.InitSex(maleProp=0.5)] +
    # Sex status is randomly assigned to each individual for a balanced sex-ratio. SimuPOP also assigns genotypes:
    [sim.InitGenotype(freq=[initFreq[i], 1-initFreq[i]], loci=i) for i in range(len(snps))],

    # 'preOps' defines what happens at the beginning of each generation just before mating:
    preOps= [
        sim.Migrator(rate=M, mode=sim.BY_PROBABILITY), # Individuals migrate.
        sim.ResizeSubPops(subPops=[3,4,5,13], proportions=[float(1-float(H)),float(1-float(H)), float(1-float(H)), float(1-
float(H))], begin=150), # A proportion of individuals is randomly removed from 'subPops' 3, 4, 5 and 13 to simulate fishing.
        sim.InitSex(maleProp=0.5) # Sex-ratio is reset to 0.5 at each generation.
    ],

    # 'matingScheme' defines how reproduction happens. Here simulated generations are overlapping:
    matingScheme= sim.HeteroMating([
        sim.CloneMating(weight=2),
        sim.RandomMating(weight=1, sexMode=(sim.PROB_OF_MALES, 0.5), ops=[
            sim.MendelianGenoTransmitter(),
            numOffspring=1)
        ],
        shuffleOffspring=True, subPopSize=model1
    ),

    # 'postOps' defines which information to keep from the simulated generation before going back to 'preOps' for the
    new generation and so on. Here we just export some data in genepop format, each five generations from generation
    140 to the end.
    postOps=[sim.PyOperator(exportSample, begin=140, step=5)],

    # End of run.
    gen=201 )

```

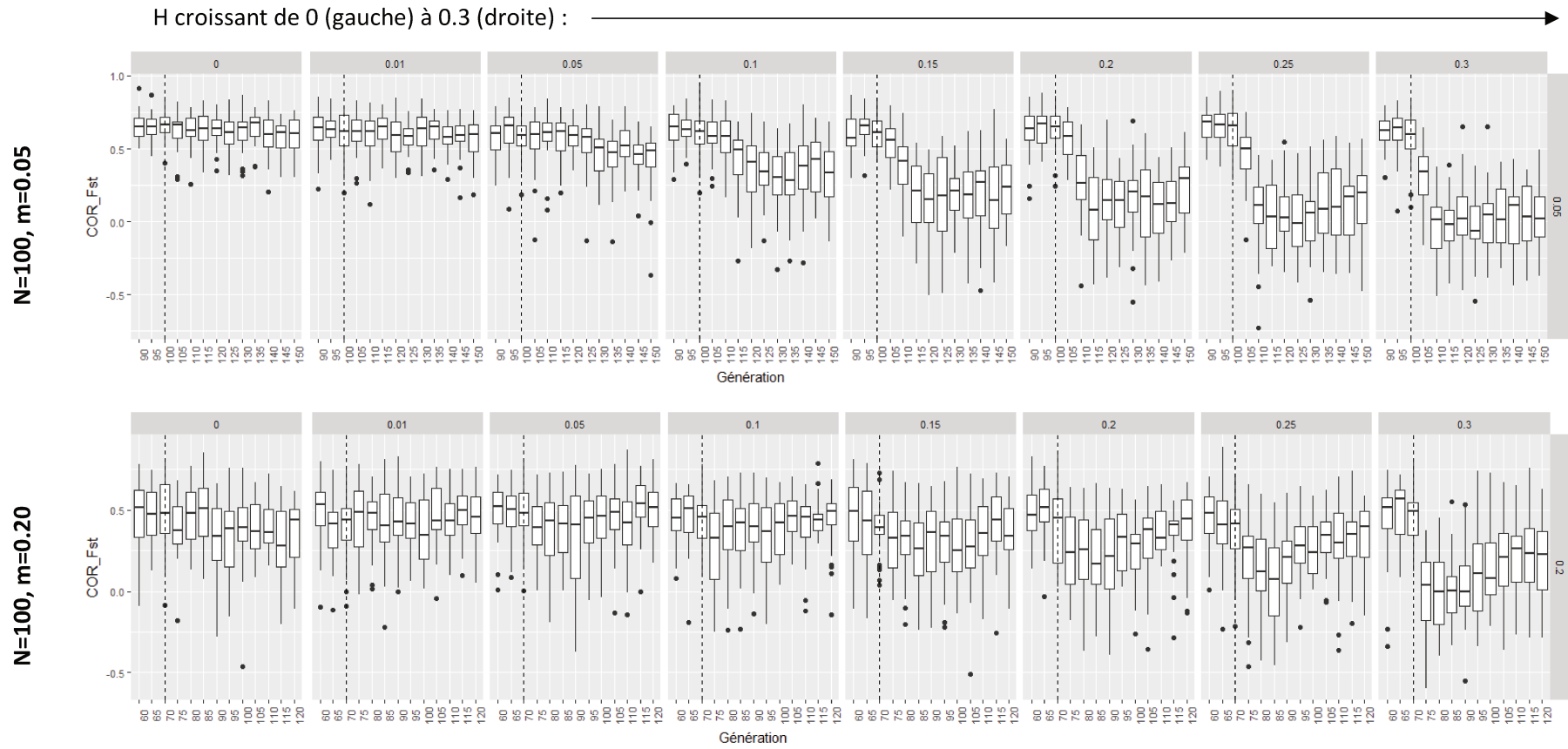


Figure XI.1 : Evolution de COR_{Fst} au cours des générations, avec $N=100$, pour $m=0.05$ (haut) et 0.20 (bas) et pour H variant de 0 à 0.3 (de gauche à droite). Chaque boxplot représente l'ensemble des 30 réplicas simulés pour chaque génération. La ligne verticale en pointillés représente la génération à laquelle la valeur H intervient (H est donc toujours nul pour les générations qui précèdent).

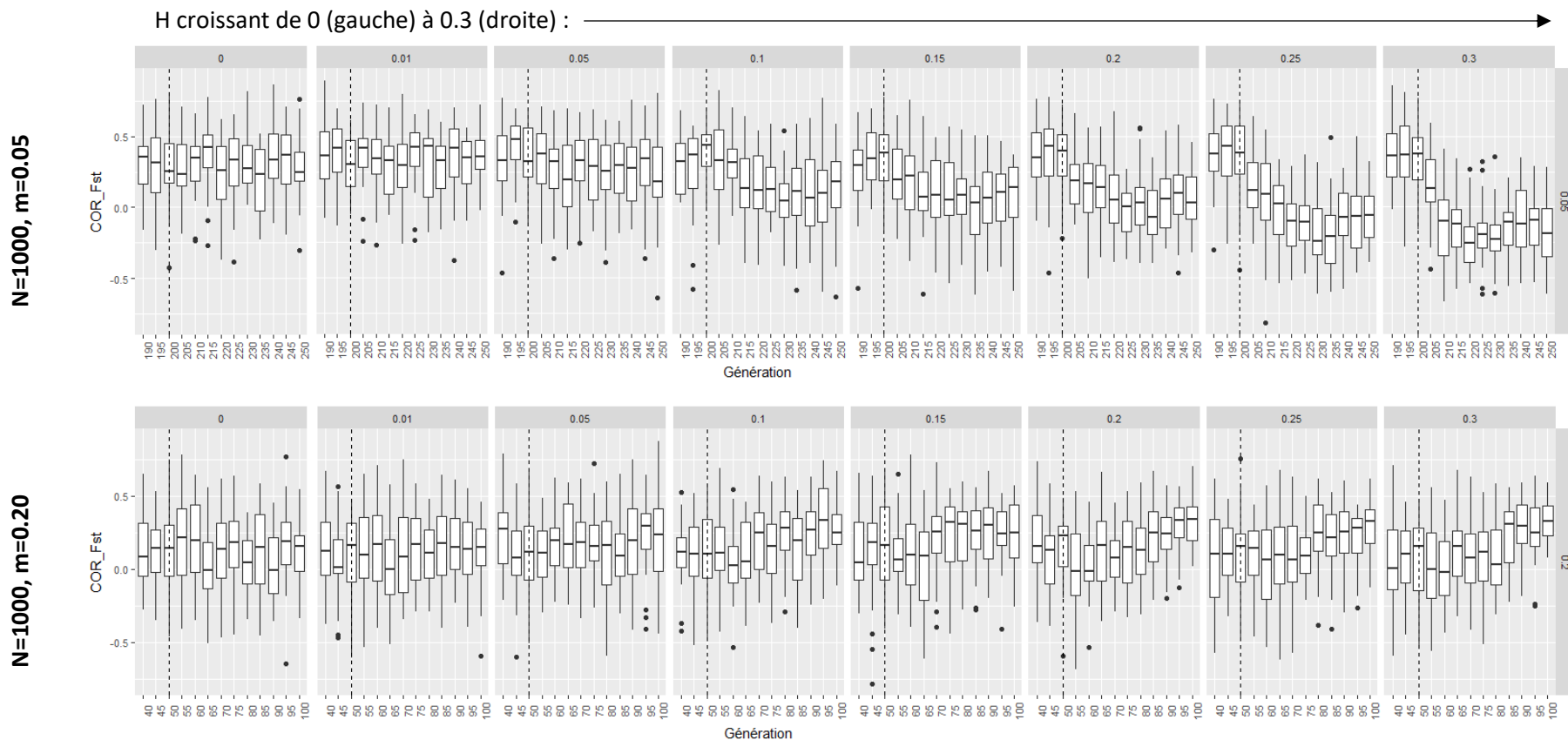


Figure XI.2 : Evolution de COR_{EST} au cours des générations, avec $N=1000$, pour $m=0.05$ (haut) et 0.20 (bas) et pour H variant de 0 à 0.3 (de gauche à droite). Chaque boxplot représente l'ensemble des 30 réplicas simulés pour chaque génération. La ligne verticale en pointillés représente la génération à laquelle la valeur H intervient (H est donc toujours nul pour les générations qui précèdent).

Annexe XII :

Informations supplémentaires relatives aux marqueurs sélectionnés au cours de la phase de mise au point des ressources génétiques pour les espèces étudiées :

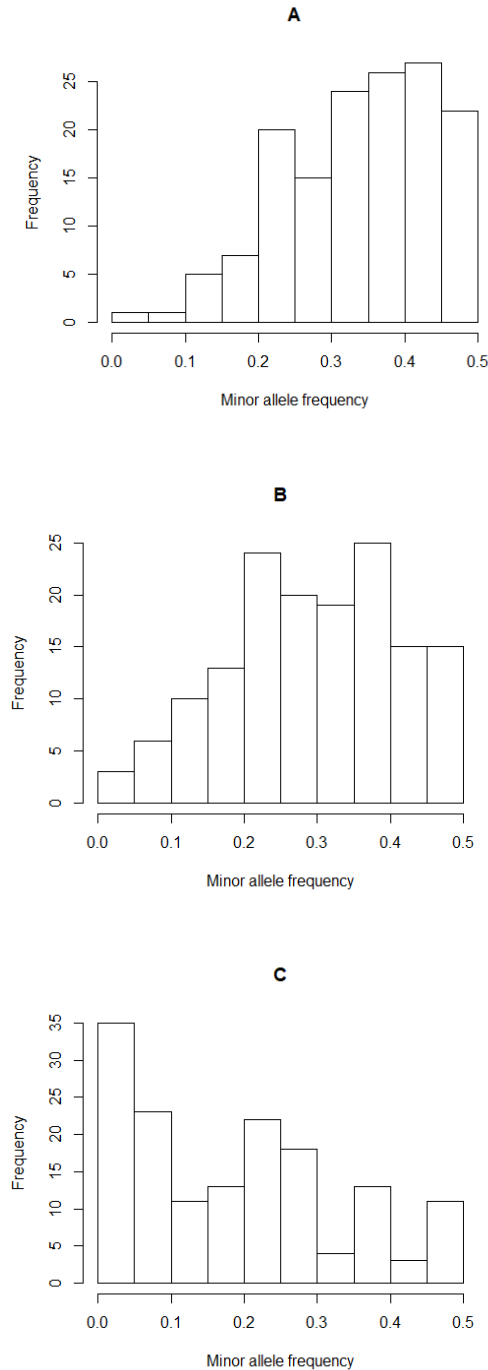


Figure XII.1 : A titre informatif : Exemple de spectre de fréquences alléliques obtenues à l'échelle du Haut-Maroni, pour l'ensemble des marqueurs SNP génotypés et validés, pour les espèces *Triporthus brachypomus* (A), *Geophagus harreri* (B) et *Hypostomus gymnorhynchus*.

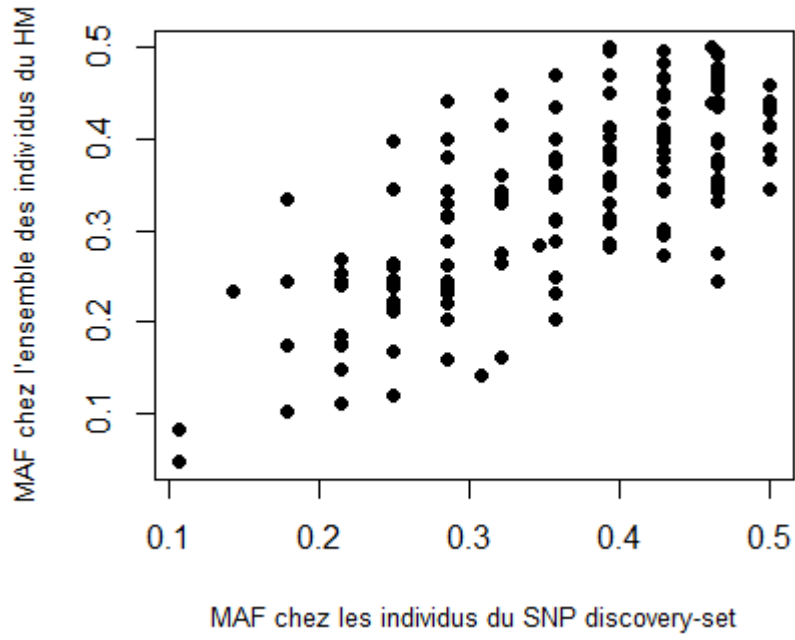
(A) et (C) sont les cas les plus extrêmes observés avec, respectivement, une faible ou forte proportion de loci à fréquences alléliques déséquilibrées (faibles MAF). (B) est un cas intermédiaire et se rapproche de ce qui a été observé chez la majorité des autres espèces.

**Profondeurs de couverture et fréquences alléliques des SNPs
à différentes étapes de leur développement puis de leur génotypage**

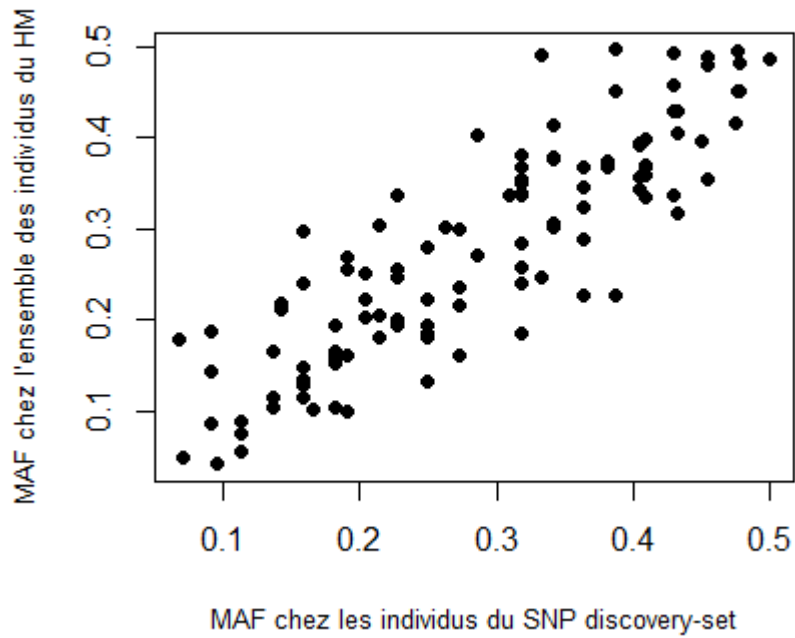
Table XII.1: Coefficient de détermination (R^2 ajusté) (A) entre les fréquences alléliques minoritaires (MAF) des SNPs validés (N=111 à 156*) parmi les individus ayant servi à composer le panel de détection de SNPs, et parmi l'ensemble des spécimens échantillonnés sur le HM. (B) entre les fréquences alléliques minoritaires (MAF) des SNPs validés (N=111 à 156) obtenues via leur profondeur de couverture moyenne dans les RAD-tags fournis par le logiciel Stacks, ou parmi les individus ayant servi à composer le panel de détection de SNPs. (C) entre les fréquences alléliques minoritaires (MAF) des SNPs validés (N=111 à 156) obtenues via leur profondeur de couverture dans les RAD-tags du réplicat 1 et du réplicat 2 (i.e., illustration du degré d'équivalence des fréquences alléliques obtenues à partir des profondeurs de couverture entre deux réplicats). (D) entre les fréquences alléliques minoritaires (MAF) de tous les SNPs détectés dans les RADtags des deux réplicats d'une espèce donnée à l'issue du premier filtrage appliqué par le pipeline présenté en Chapitre 1 (i.e., analogue à (C), mais pas uniquement avec les SNPs validés : avec l'ensemble des SNP validant les critères du premier filtrage du pipeline).

*Les 31 SNP issus d'une procédure parallèle de développement de SNP pour l'espèce *Cynodon meionactis* (voir Chapitre 2 et Annexe VI) ne sont ici pas pris en compte.

	(A) R^2	(B) R^2	(C) R^2	(D) R^2	Remarque
<i>Acnodon oligacanthus</i>	0,819	0,653	0,480	0,274	
<i>Ageneiosus inermis</i>	0,755	0,810	0,610	0,648	
<i>Brycon falcatus</i>	0,663	0,534	0,334	0,237	
<i>Cynodon meionactis</i>	0,790	0,788	0,569	0,472	
<i>Doras micropoeus</i>	0,705	0,649	0,523	0,423	
<i>Geophagus harreri</i>	0,614	0,699	0,146	0,164	Pools différents
<i>Harttia guianensis</i>	0,689	0,712	0,482	0,397	
<i>Hoplias aimara</i>	0,668	0,684	0,100	0,215	Pools différents
<i>Hypostomus gymnorhynchus</i>	0,786	0,509	0,094	-0,014	Pools différents
<i>Leporinus friderici</i>	0,717	0,629	0,400	0,328	
<i>Leporinus lebaili</i>	0,539	0,547	0,377	0,213	
<i>Myloplus rhomboidalis</i>	0,741	0,537	0,378	0,146	
<i>Myloplus rubripinnis</i>	0,793	0,490	0,333	-0,011	
<i>Pseudancistrus barbatus</i>	0,828	0,586	0,182	0,171	Pools différents
<i>Serrasalmus eigenmanni</i>	0,805	0,644	0,190	0,058	
<i>Serrasalmus rhombeus</i>	0,664	0,365	0,252	0,043	
<i>Tometes lebaili</i>	0,655	0,459	0,244	-0,005	
<i>Triportheus brachypomus</i>	0,549	0,292	0,222	0,055	

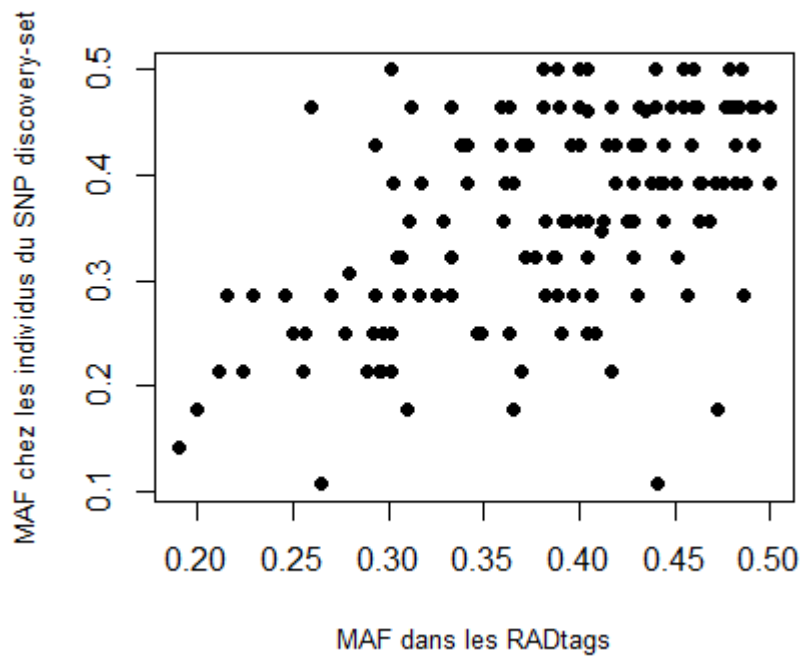


A)

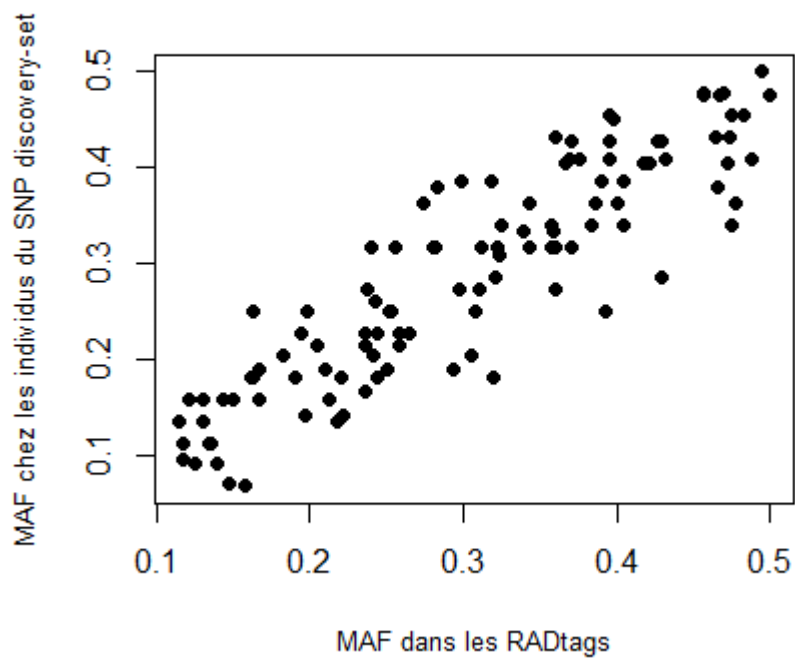


B)

Figure XII.2 : Relation entre les fréquences alléliques minoritaires (MAF) des SNPs validés (N=111 à 156) parmi les individus ayant servi à composer le SNP discovery-panel (abscisse) et parmi l'ensemble des spécimens échantillonnés sur le HM (ordonnée) pour les espèces *Triportheus brachypomus* (A) et *Ageneiosus inermis* (B).

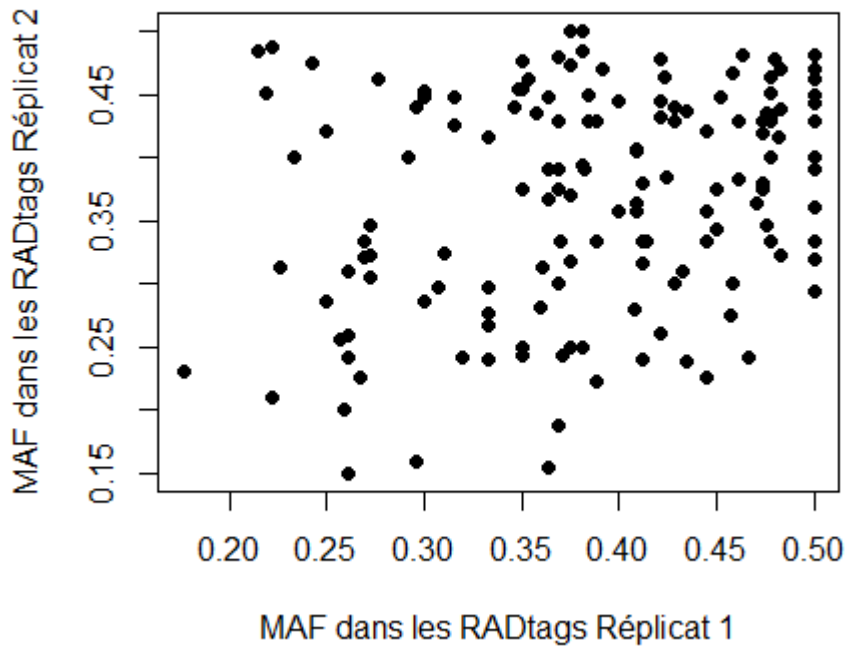


A)

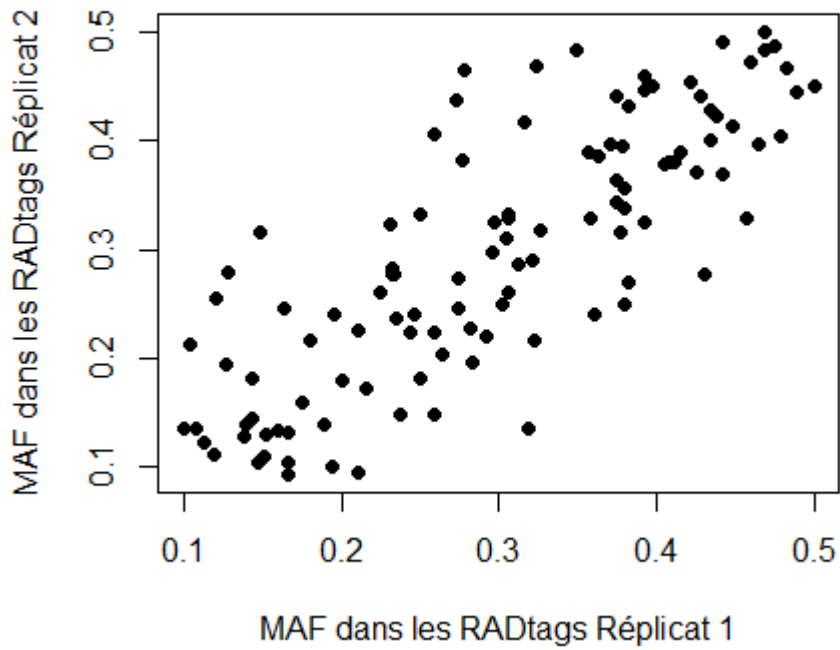


B)

Figure XII.3 : Relation entre les fréquences alléliques minoritaires (MAF) des SNPs validés (N=111 à 156), obtenues via leur profondeur de couverture dans les RAD-tags fournis par le logiciel Stacks (abscisse) ou après génotypage individuel des individus ayant servi à composer le SNP discovery-panel (ordonnée), pour les espèces *Triportheus brachypomus* (A) et *Ageneiosus inermis* (B).



A)



B)

Figure XII.4 : Relation entre les fréquences alléliques minoritaires (MAF) des SNPs validés (N=111 à 156), obtenues via leur profondeur de couverture dans les RAD-tags fournis par le logiciel Stacks entre les deux répliquats, pour les espèces *Triportheus brachypomus* (A) et *Ageneiosus inermis* (B).

Annexe XIII :

Liste des productions et communications scientifiques effectuées au cours de la thèse :

Publications parues :

Delord C., Lassalle G., Oger A., Barloy D., Coutellec M.A., Delcamp A., Evanno G., Genthon C., Guichoux E., Le Bail P.-Y., Le Quilliec P., Longin G., Lorvelec O., Massot M., Reveillac E., Rinaldo R., Roussel J.-M., Vigouroux R., Launey S., Petit E.J. (2018) A cost-and-time effective procedure to develop SNP markers for multiple species: a support for community genetics. *Methods in Ecology and Evolution*, 9(9), 1959-1974. <https://doi.org/10.1111/2041---210X.13034>

Paris, J.R., Sherman, K.D., Bell, E., Boulenger, C., **Delord, C.**, El-Mahdi, M.B.M., Fairfield, E.A., Griffiths, A.M., Gutmann Roberts, C., Hedger, R.D., Holman, L.E., Hooper, L.H., Humphries, N.E., Katsiadaki, I., King, R.A., Lemopoulos, A., Payne, C.J., Peirson, G., Richter, K.K., Taylor, M.I., Trueman, C.N., Hayden, B., & Stevens, J.R. (2018) Understanding & Managing fish populations: keeping the toolbox fit for purpose. *Journal of Fish Biology*, 92(3), 727-751. <https://doi.org/10.1111/jfb.13549>

Publications soumises ou en préparation :

Delord C., Le Bail P.-Y., Fourcy D., Longin G., Rinaldo R., Vigouroux R., Roussel J.-M., Launey S. Riverscape patterns of within and between-population genetic diversity in eighteen harvested Neotropical freshwater fish species. (En préparation pour *Molecular Ecology*)

Interventions en conférences nationales et internationales :

IInd Joint Congress in Ecology and Evolution, du 17 au 23 août 2018, European Society for Evolutionary Biology (ESEB) & Agropolis International, Montpellier (France). **Poster**: *Patterns of within and between-population genetic diversity across a riverine network submitted to anthropogenic pressure: a comparative approach to describe the amerindian multi-species fishery of the Upper-Maroni in French Guiana (South America)*. **Delord, C.**, Lassalle, G., Longin, G., Rinaldo, R., Vigouroux, R., Roussel, J. M., Launey, S., Le Bail, P. Y. *Ce poster a également été présenté au congrès SFE² à Rennes du 22 au 25 octobre 2018.*

50th Anniversary Symposium of the Fisheries Society of the British Isles (FSBI), "Understanding fish populations", du 3 au 7 juillet 2017, Université d'Exeter (Royaume-Uni). **Poster & Speed talk**: *Fishery resources of the upper-Maroni River: Impacts of fishing pressure and life history traits on genetic diversity in several fish species*. Présentation du projet de thèse et résultats préliminaires. **Delord, C.**, Lassalle, G., Longin, G., Rinaldo, R., Vigouroux, R., Roussel, J.-M., Le Bail, P.-Y., Launey, S.

Journées Ouvertes en Biologie, Informatique et Mathématiques (JOBIM), organisées par la Société Française de Bioinformatique, du 27 au 29 juin 2016, Ecole Nationale Supérieure de Lyon (France). **Poster**: *Development of SNPs markers from Pooled Rad-seq data and high-throughput genotyping, Applied to the study of several harvested population genetics in the Upper-Maroni*. **Delord, C.**, Lassalle, G., Launey, S., Le-Bail, P.-Y. (1^{er} prix du meilleur poster présenté par un étudiant).

Enseignement, animation et vulgarisation :

Intervention dans le cadre du cursus ingénieur 'Gestion des ressources halieutiques' 3^{ème} année (Agrocampus Ouest) : Cours magistral portant sur (1) Les apports et limites des technologies de séquençage à haut-débit, (2) Taille efficace : concept et applications. Décembre 2016 (3h), Décembre 2017 (3h).

Salon de l'Agriculture 2017 : participation à la conception et au montage d'un atelier grand public pour illustrer le cycle de vie du saumon et l'importance de la diversité génétique dans les populations naturelles (Paris, les 3 et 4 mars 2018).

Concours « Ma Thèse en 180 secondes » :

Semi-finales régionales (Mars 2017) : <https://www.youtube.com/watch?v=fdLWN3WH5c8>

Journée Portes Ouvertes pour les 70 ans de l'INRA : conception, montage et animation d'un atelier grand public pour illustrer l'importance de la diversité génétique dans les populations naturelles (Agrocampus Ouest de Rennes, les 27 et 28 juin 2016).

Titre : Gestion des ressources halieutiques du Haut-Maroni : Impact de la pression de pêche et des traits d'histoire de vie des espèces sur la diversité génétique intra- et interpopulations

Mots clés : eau douce – poissons – écosystèmes néotropicaux – génétique du paysage – multiespèces – simulations

Résumé : En Guyane française, le bassin versant du Maroni représente un support de vie essentiel pour les populations humaines locales. Au cours des dernières décennies, des bouleversements dans les pratiques de pêche de subsistance posent la question de leur impact sur les populations de poissons. Par une approche de génétique comparative des populations, cette thèse s'intéresse à l'influence de la pression de pêche sur la diversité génétique chez plusieurs espèces plus ou moins exploitées. Nous avons mis au point des ressources moléculaires pour 18 espèces échantillonnées sur le Maroni via l'application d'un protocole original combinant séquençage à haut-débit et un pipeline bioinformatique. En comparant 12 sites soumis ou non à une pression de pêche quotidienne, nous constatons que la structure spatiale dendritique du bassin versant et la qualité de l'habitat cons-

-tituent les principaux facteurs explicatifs de la diversité génétique locale, notamment chez les espèces à capacité de dispersion limitée. Dans la majorité des cas, nous n'observons pas de réduction locale de diversité génétique liée à la pêche, suggérant un impact négligeable ou trop récent pour être détectable, ou encore contrebalancé par des flux de gènes compensatoires à plus grande échelle spatiale. Une approche théorique, basée sur les outils de simulation, a été amorcée pour évaluer l'influence à moyen-terme d'une pression de pêche géographiquement localisée sur les patrons généraux de variabilité génétique au sein des réseaux hydrographiques dendritiques. Néanmoins, l'approche comparative entre sites et entre espèces permet d'en identifier certains nécessitant une vigilance particulière pour une gestion durable des ressources halieutiques.

Title : Management of fishing resources from the Upper-Maroni: Impact of fishing pressure and life history traits on within and between populations genetic diversity

Keywords : freshwater – fish – Neotropical ecosystems – riverscape genetics – multispecies – simulations

Abstract : In French Guiana, the Maroni watershed represents an essential everyday life support for native human communities. During past decades, upheavals in fishing practices have raised concerns about their impact on fish populations. Using a comparative population genetics approach, this thesis focuses on the influence of fishing pressure on genetic diversity of several species with contrasting fishing interest. We developed molecular resources for 18 species sampled across the Maroni through the application of an original protocol combining high throughput sequencing and a bioinformatic pipeline. Comparing 12 sites subject or not to daily fishing pressure, we show that the dendritic spatial structure of the watershed and

habitat quality are the main factors explaining local genetic diversity, especially for species with limited dispersal abilities. In most cases, we observe no local genetic erosion associated with fishing pressure, suggesting its impact to be either negligible, too recent to be detected, or counterbalanced by compensatory gene flow at a higher spatial scale. A theoretical approach, based on computer simulations, was initiated to evaluate medium-term influence of a geographically restrained fishing pressure on global patterns of genetic variability in dendritic river networks. Together, these comparative approaches between sites and species highlight the need for vigilance in order to favor a sustainable use of fishing resources.