Action financée avec le concours de l'Agence de l'Eau Adour-Garonne

Partenariat 2018

LE SILURE GLANE (*Silurus glanis L.*) : CONNAISSANCES ET ESTIMATION DES STOCKS DANS LE BASSIN DE LA GARONNE

RAPPORT D'ÉTUDE

Ivan PAZ-VINAS, Frédéric SANTOUL

Laboratoire Écologie fonctionnelle et environnement ECOLAB ; UMR UPS INP CNRS 5245
Campus UPS ; 118, route de Narbonne – Bâtiment 4R1 31062 Toulouse Cedex 9
France

Décembre 2018

CONTEXTE DE PROGRAMMATION ET DE RÉALISATION

Le silure glane (*Silurus glanis*) est l'un des plus grands poissons d'eau douce au monde et le plus grand en Europe (Cucherousset et al., 2018; Stone, 2007). Les silures adultes peuvent mesurer plus de deux mètres (un individu de 2,74 mètres a été pêché dans le Tarn en 2017). En France, le silure atteint donc plus de deux fois la taille des prédateurs natifs comme le brochet (> 1 m). Ceci a pour conséquence d'augmenter potentiellement la taille refuge des proies, c'est-à-dire la taille maximale à laquelle ces dernières ne sont plus prédatées. Des espèces de poissons comme certaines espèces anadromes (espèces migratrices vivant en mer mais remontant les cours d'eau pour s'y reproduire) qui échappaient, de par leur grande taille, à la prédation dans de tels milieux, sont désormais susceptibles d'être prédatées par le silure.

De précédents travaux ont montré que certaines espèces amphihalines (espèces migrant entre un milieu maritime et un milieu dulçaquicole comme le saumon atlantique ou l'anguille) constituent une part non négligeable du régime alimentaire des silures dans le bassin de la Garonne (Boulêtreau et al., 2018; Guillerault et al., 2017; Syväranta et al., 2009). Néanmoins, les conséquences de la présence de silures dans les grands cours d'eau comme la Garonne restent mal connues. En effet, pour comprendre les effets d'une espèce non-native, comme le silure, sur les espèces autochtones, il est nécessaire de pouvoir estimer les effectifs de silure présents dans de grands cours d'eau comme la Garonne, la Dordogne, le Tarn ou le Lot. Ces estimations démographiques de silure font donc l'objet de cette étude.

Les progrès récents dans le domaine de la génétique des populations permettent désormais de quantifier de façon indirecte des paramètres démographiques d'intérêt comme les flux de gènes entre populations ou les tailles efficaces des populations, et ceci à des échelles assez fines. Ainsi, il est actuellement possible de reconstruire l'histoire démographique d'une espèce grâce aux empreintes que certains évènements (introductions d'individus, expansions ou contractions démographiques) laissent dans le « pool » génique de ses populations. Toutes ces informations, inférées à partir de données génétiques, peuvent être ensuite utilisées pour estimer des abondances et des densités grâce à des traitements statistiques spécifiques.

Au vu de ces avancées, nous avons mené à bien cette étude, dont les objectifs étaient (i) de caractériser génétiquement des silures en utilisant des marqueurs microsatellites déjà décrits chez cette espèce (Krieg et al., 1999; Triantafyllidis et al., 2002) et (ii) d'analyser les données génétiques obtenues en appliquant des modèles avancés de génétique des populations et des méthodes statistiques innovantes pour (iii) **fournir une évaluation du stock de silure dans la Garonne aval et dans ses principaux affluents (Dordogne, Tarn et Lot)**.

Cette étude visait, en particulier, à analyser le profil génétique de plusieurs populations de silures de la partie aval du bassin de la Garonne, zone la plus concernée par la problématique des migrateurs potentiellement prédatés par le silure. Au fur et à mesure de l'avancement de l'étude, de nouvelles analyses non prévues initialement et concernant l'évolution géographique et démographique du silure dans le bassin de la

Garonne ont été ajoutées. Ces nouvelles connaissances nous ont permis de calibrer au mieux les modèles de génétique des populations utilisés pour estimer les abondances et les densités de silure.

LES AUTEURS

Ivan Paz-Vinas
Ingénieur de Recherche (IR) CNRS en CDD
ivanpaz23@gmail.com

Frédéric Santoul
Maître de Conférences – Université Paul Sabatier Toulouse 3 ; UMR 5245 ECOLAB
frederic.santoul@univ-tlse3.fr

LES COLLABORATEURS

Géraldine Loot
Professeure des Universités – Université Paul Sabatier Toulouse 3 ; UMR 5174 EDB

Charlotte Veyssière
Ingénieure d'études – Université Paul Sabatier Toulouse 3 ; UMR 5174 EDB

LES CORRESPONDANTS

Union des Fédérations pour la pêche et la protection du milieu aquatique du Bassin Adour-Garonne (UFBAG) :
Molinié Jean-Louis
Président
jlm.peche47@orange.fr

Partenaire scientifique – Université Paul Sabatier Toulouse 3 ; UMR 5245 ECOLAB :
Frédéric Santoul
Maître de Conférences
frederic.santoul@univ-tlse3.fr

REMERCIEMENTS

Nous tenons à remercier les personnes et les organismes qui, via nos échanges et leurs commentaires, ont contribué à ce que cette étude soit menée à bon terme. Nous remercions les Fédérations de Pêche et de Protection des Milieux Aquatiques de la Haute-Garonne, du Lot, du Lot-et-Garonne, du Tarn et du Tarn-et-Garonne, l'association MIGADO, les pêcheurs professionnels et les pêcheurs à la ligne qui nous ont fourni des échantillons, et tout particulièrement Monsieur Pascal Verdeyroux d'EPIDOR.

Nous remercions l'Agence Française pour la Biodiversité (AFB), en particulier tous les agents de l'AFB qui ont contribué à l'acquisition des données de terrain, à l'élaboration et à la gestion de la base de données « Hydrobiologie » du portail Naïades (http://www.naiades.eaufrance.fr/presentation).

Nous remercions aussi la GENOTOUL de Toulouse pour l'accès au séquenceur, et le groupe CALMIP (Calcul en Midi-Pyrénées ; UMS CNRS 3667) pour l'attribution de ressources computationnelles pour effectuer nos calculs dans le supercalculateur « Eos » (allocation de ressources T18005).

SOMMAIRE

CONTEXTE DE PROGRAMMATION ET DE RÉALISATION 2

LES AUTEURS 4

LES COLLABORATEURS 4

LES CORRESPONDANTS 4

REMERCIEMENTS 5

SOMMAIRE 6

NOTE DE SYNTHÈSE 7

CORPS DU DOCUMENT 11

PARTIE I : ANALYSE DEMOGRAPHIQUE DU SILURE DANS LE BASSIN DE LA GARONNE 11
1. Historique des premiers déversements de silure dans le bassin de la Garonne 11
2. Présentation des données utilisées 13
3. Analyse de l'expansion géographique 15
4. Analyse de l'évolution démographique 19
PARTIE II : DESCRIPTION DE LA DIVERSITE ET DE LA STRUCTURATION GENETIQUE DU SILURE DANS LE BASSIN DE LA GARONNE. ... 25
1. Échantillonnage 25
2. Méthodes moléculaires utilisées 28
Extractions d'ADN et amplifications des marqueurs ADN. 28
Contrôles de la qualité du génotypage. 28
3. Diversité et structure génétique du silure 28
Description de la diversité génétique neutre avec les marqueurs microsatellites. 28
Patrons spatiaux de diversité et de structuration génétique. 32
Description de la structure génétique du silure dans le bassin de la Garonne. 37
PARTIE III : ESTIMATION DE L'ABONDANCE ET DES DENSITES DE SILURE A PARTIR DE DONNEES GENETIQUES. 40
1. Raisonnement soutenant la méthodologie d'estimation d'abondances et de densités proposée 41
2. Estimation des tailles efficaces (Ne) des populations étudiées 41
Principe de la méthode. 43
Définition des modèles utilisés pour simuler les données génétiques. 43
Procédure de simulation. 45
Choix des modèles. 46
Estimation des tailles efficaces des populations de silure. 48
3. Calcul des valeurs de méta-N_e pour les quatre rivières étudiées 48
4. Calcul du rapport Nc / Ne pour le silure 49
5. Estimations de l'abondance méta-N_c et de la densité de silure pour les quatre rivières étudiées 52

CONCLUSIONS GÉNÉRALES 54

BIBLIOGRAPHIE 56

ANNEXES 64

NOTE DE SYNTHÈSE

Cette note résume de façon très synthétique le contexte de l'étude, les méthodes et les résultats obtenus. Des figures et une table sont fournies à titre illustratif. Pour une description plus détaillée et approfondie de tous ces éléments, veuillez vous référer aux sections suivantes.

Le silure glane (*Silurus glanis*) est le plus grand poisson d'eau douce en Europe. En France, où le silure a été largement introduit, il atteint plus de deux fois la taille des prédateurs natifs comme le brochet, ce qui a pour conséquence d'augmenter la taille maximale à partir de laquelle les proies ne sont plus prédatées. Des espèces de poissons (ex. les espèces amphihalines) qui échappaient à la prédation jusqu'alors, sont désormais susceptibles d'être prédatées par le silure.

Il est donc important de comprendre les effets que le silure peut avoir sur les espèces autochtones et les espèces migratrices. Pour cela, il faut pouvoir estimer les effectifs de silure présents dans des grands cours d'eau comme la Garonne, la Dordogne, le Tarn ou le Lot. Ces estimations démographiques ont donc fait l'objet de cette étude, dont l'objectif principal était de **fournir une évaluation du stock de silure dans la Garonne aval et dans ses principaux affluents (Dordogne, Tarn et Lot)**.

Pour répondre à cet objectif, nous avons d'abord conduit une **analyse démographique du silure basée sur les données de suivi hydrobiologique de l'Agence Française pour la Biodiversité** (AFB), afin de comprendre quelle a été l'évolution géographique et démographique du silure dans le bassin de la Garonne depuis son introduction.

Ensuite, nous avons créé et analysé une base de données génétiques concernant 18 populations échantillonnées dans la Garonne, la Dordogne, le Lot et le Tarn, afin de **caractériser la diversité et la structuration génétique du silure dans ces rivières**.

Finalement, nous avons mis en place une approche méthodologique permettant d'**estimer les abondances et les densités de silure dans les quatre rivières considérées** à partir de nos données génétiques, tout en prenant en compte les divers résultats obtenus à partir de nos analyses démographiques et génétiques.

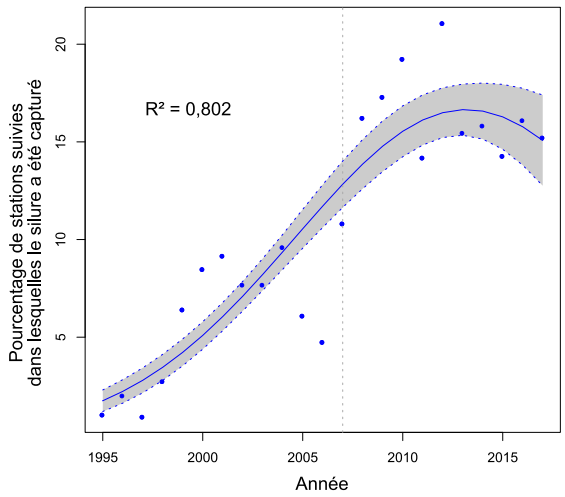
Partie I : Analyse démographique du silure dans le bassin de la Garonne

La première introduction de silures connue dans le bassin de la Garonne date de 1983 (dans le Tarn, à Aiguelèze). Depuis, d'autres introductions de silure ont eu lieu dans le bassin. Divers témoignages recueillis durant cette étude nous ont permis de lister dans ce rapport plusieurs de ces évènements d'introduction. Ils nous ont appris qu'il est très difficile de recueillir avec précision toutes les données concernant les déversements de silure dans le bassin et l'origine géographique des individus déversés.

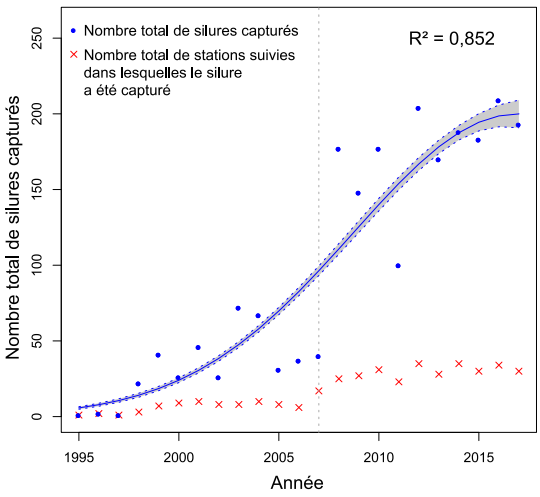
Ainsi, pour caractériser l'évolution de l'expansion démographique et géographique du silure dans le bassin de la Garonne, nous avons analysé les données de suivi de l'AFB

(pour la période s'étalant de 1995 à 2017). L'analyse de l'évolution géographique du silure à l'échelle du bassin nous indique que l'**expansion géographique de cette espèce dans ce bassin semble actuellement se stabiliser**, après avoir connu une phase d'expansion très importante. De même, l'analyse démographique suggère que l**e stock de silure dans le bassin de la Garonne**, après avoir subi une forte croissance lors des deux dernières décennies, **serait peut-être actuellement en train de se stabiliser**.

Pourcentage de stations suivies dans lesquelles le silure a été capturé en fonction des années



Nombre total de silures capturés dans les stations suivies en fonction des années



À l'échelle des rivières, nos analyses indiquent que **les populations de silure seraient actuellement stables dans la Dordogne et le Lot, en expansion démographique dans la Garonne et en décroissance dans le Tarn**.

<u>Partie II : Description de la diversité et de la structuration génétique du silure dans le bassin de la Garonne</u>

Nous avons construit une base de données génétiques concernant 562 individus échantillonnés sur 18 populations différentes réparties dans les quatre cours d'eau principaux du bassin de la Garonne : la Garonne, la Dordogne, le Lot et le Tarn.

L'analyse de cette base de données nous a permis de décrire la diversité génétique de ces populations. Ces analyses indiquent que **les populations situées en amont du Lot présentent des spécificités génétiques qui les distinguent des autres populations** échantillonnées dans le bassin.

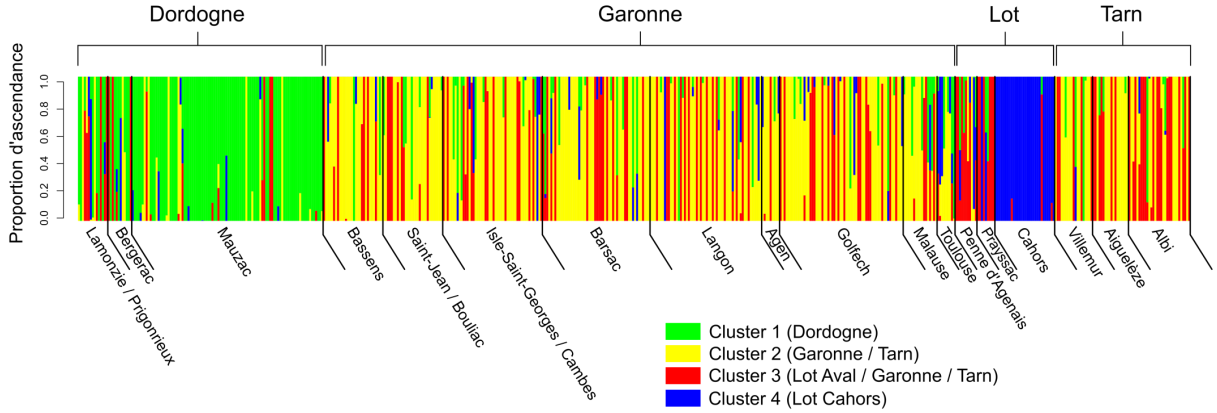
Nous avons ensuite caractérisé les patrons spatiaux de diversité génétique chez les populations de silure du bassin, afin de voir s'ils sont en accord avec des patrons spatiaux généralement prédits pour les organismes aquatiques d'eau douce. Cette analyse nous montre que **deux des trois patrons spatiaux de diversité génétique attendus ne sont pas vérifiés**, bien qu'une tendance vers ce type de patrons semble se profiler. Nous avons confirmé, toutefois, la **présence d'un patron d'isolement par la distance génétique (IBD) significatif** chez les populations de silure dans le bassin de la Garonne.

Au vu de ces résultats, **nous émettons l'hypothèse que le silure n'est pas encore totalement acclimaté au bassin de la Garonne**, étant donné que tous les patrons spatiaux de diversité attendus ne sont pas vérifiés.

Finalement, nous avons mis en évidence qu'il existe une forte structuration génétique des populations de silure dans le bassin de la Garonne. Nos résultats suggèrent plus spécifiquement (i) qu'il existe deux groupes d'individus génétiquement homogènes bien définis spatialement (un groupe occupant la Dordogne ; l'autre occupant le Lot à Cahors), et (ii) qu'il y a présence d'un mélange d'individus issus de deux groupes génétiquement différents dans la Garonne, le Tarn et à l'aval du Lot.

À partir de ces résultats, nous avons formulé l'hypothèse que **les populations de silure du bassin de la Garonne seraient issues *a minima* de quatre introductions importantes, indépendantes et provenant de quatre sources génétiques** (ou « populations sources ») **différentes**.

Diagramme représentant les proportions d'appartenance (ou d'ascendance) de chaque individu aux quatre différents groupes d'individus génétiquement homogènes (ou « clusters »).



Partie III : Estimation de l'abondance et des densités de silure à partir de données génétiques.

Pour pouvoir estimer avec précision l'abondance d'une espèce à partir de données génétiques, il faut :

- Obtenir des estimations de tailles efficaces des populations (N_e) d'intérêt,
- Calculer les valeurs de méta-N_e, c'est-à-dire la taille efficace des métapopulations d'intérêt,
- Connaître le rapport N_c / N_e, c'est-à-dire le rapport existant entre les tailles réelles des populations N_c et leurs tailles efficaces N_e, et
- Appliquer le ratio N_c / N_e aux valeurs de méta-N_e pour obtenir la valeur de méta-N_c, c'est à dire l'abondance de l'espèce dans la métapopulation d'intérêt.

Nous avons donc suivi cette procédure pour estimer l'abondance et les densités de silure. Dans un premier temps nous avons calculé les N_e des populations étudiées en

comparant les données génétiques décrites dans la Partie II avec des données génétiques simulées sous divers modèles démographiques et évolutifs qui prenaient en compte les spécificités des populations de silure dans le bassin de la Garonne mises en évidence dans les Parties I et II. Cette comparaison s'est faite au travers de calculs Bayésiens approchés couplés à des algorithmes d'apprentissage automatique.

Dans un second temps, nous avons calculé à partir des valeurs de N_e obtenues les valeurs méta-N_e de silure pour cinq métapopulations : Garonne aval, Garonne amont, Dordogne, Tarn et Lot. Nous avons utilisé pour cela une méthode adaptée aux environnements de type linéaire comme les rivières.

Ensuite, nous avons estimé un rapport N_c / N_e pour le silure, en calculant pour une même section de la Dordogne une valeur de N_c (à partir de données de Capture-Marquage-Recapture d'une étude effectuée par EPIDOR) et de N_e (à partir de nos données génétiques).

Finalement, nous avons appliqué ce rapport à nos estimations de méta-N_e pour répondre à l'objectif de notre étude. **La Table suivante présente les abondances et les densités moyennes obtenues pour les 5 métapopulations considérées**.

Métapopulation	Valeur de méta-N_c (en nombre d'individus)	Densité (en individus par hectare)
Garonne aval	43 630,176 [26 242,624 – 78 760,130]	11,794 [7,094 – 21,290]
Garonne amont	8 220,524 [4 944,471 – 14 839,490]	7,518 [4,521 – 13,571]
Dordogne	19 990,890 [12 024,095 – 36 087,067]	11,576 [6,962 – 20,896]
Lot	12 587,636 [7 571,195 – 22 722,893]	9,942 [5,980 – 17,948]
Tarn	22 489,434 [13 526,916 – 40 597,377]	19,093 [11,484 – 34,466]

Cette étude fournit donc non seulement une réponse à son objectif principal, à savoir une estimation du stock de silure dans les principales rivières du bassin de la Garonne, mais elle augmente aussi considérablement notre connaissance sur l'histoire de la colonisation du bassin de la Garonne par le silure, et sur la génétique de cette espèce.

Tous ces éléments devraient permettre la mise en place d'études spécifiques concernant les effets du silure sur les espèces autochtones et migratrices dans le bassin de la Garonne.

CORPS DU DOCUMENT

Dans cette section, nous allons détailler les différentes étapes qui nous ont permis de répondre aux objectifs définis dans la section « Contexte de programmation et de réalisation » ici-haut, à savoir :

- Quelle a été l'évolution géographique et démographique du silure dans le bassin de la Garonne depuis son introduction ?
- Quelle est la diversité et la structuration génétique du silure dans la Garonne, la Dordogne, le Lot et le Tarn ?
- Quel est le stock actuel de silure dans les principales rivières du bassin de la Garonne ?

Dans la première partie, nous présenterons une analyse démographique du silure conduite à l'échelle du bassin de la Garonne, basée sur les données de suivi hydrobiologique de l'AFB (Agence Française pour la Biodiversité), disponibles sur le portail Naïades (http://naiades.eaufrance.fr). Nous caractériserons plus spécifiquement l'évolution géographique (spatiale) et démographique (en termes d'effectifs capturés) de cette espèce dans le bassin de la Garonne au cours du temps.

Dans la deuxième partie, nous présenterons le plan détaillé de l'échantillonnage effectué dans le but de construire une base de données génétiques. Nous décrirons aussi les différentes étapes que nous avons suivies pour construire cette base de données. Puis, nous décrirons les diverses analyses que nous avons menées dans le but de caractériser la diversité et la structuration génétique du silure, et de mettre en évidence la présence de patrons spatiaux et génétiques particuliers dans le bassin de la Garonne.

Dans la troisième partie, nous présenterons l'approche statistique que nous avons mise en œuvre pour déterminer les abondances et les densités de silure dans les principales rivières du bassin de la Garonne (Garonne, Dordogne, Tarn et Lot).

À partir de nos résultats, nous avons établi une « note de synthèse » (voir section précédente) ; synthèse qui résume les principaux résultats de cette étude, et qui devrait permettre d'orienter les futures mesures de gestion à mettre en œuvre pour cette espèce.

Partie I : Analyse démographique du silure dans le bassin de la Garonne

Pour pouvoir estimer avec précision et robustesse l'abondance et les densités d'une espèce à partir de données génétiques (voir Partie III), il est important de connaître son histoire démographique récente. Ceci est d'autant plus important dans le cas du silure et dans un bassin comme celui de la Garonne, étant donné que cette espèce a fait l'objet d'une introduction récente et d'une colonisation rapide dans ce bassin.

1. Historique des premiers déversements de silure dans le bassin de la Garonne

La première introduction de silures connue dans le bassin de la Garonne date de 1983. En effet, 2000 à 3000 individus d'environ 20 centimètres auraient été déversés cette année là dans le Tarn, à Aiguelèze (Figure 1). Divers témoignages nous ont indiqué qu'il avait été prévu que ces silures soient déversés dans un plan d'eau pour animer un évènement de pêche au silure dans le cadre d'un grand salon de la pêche à Albi. Pour des raisons inconnues, le déversement ne s'est pas effectué dans un plan d'eau, mais dans le Tarn. Selon ces témoignages, ces silures auraient passé la douane à Colmar et proviendraient d'une pisciculture située à l'étranger (mais non identifiée à ce jour).

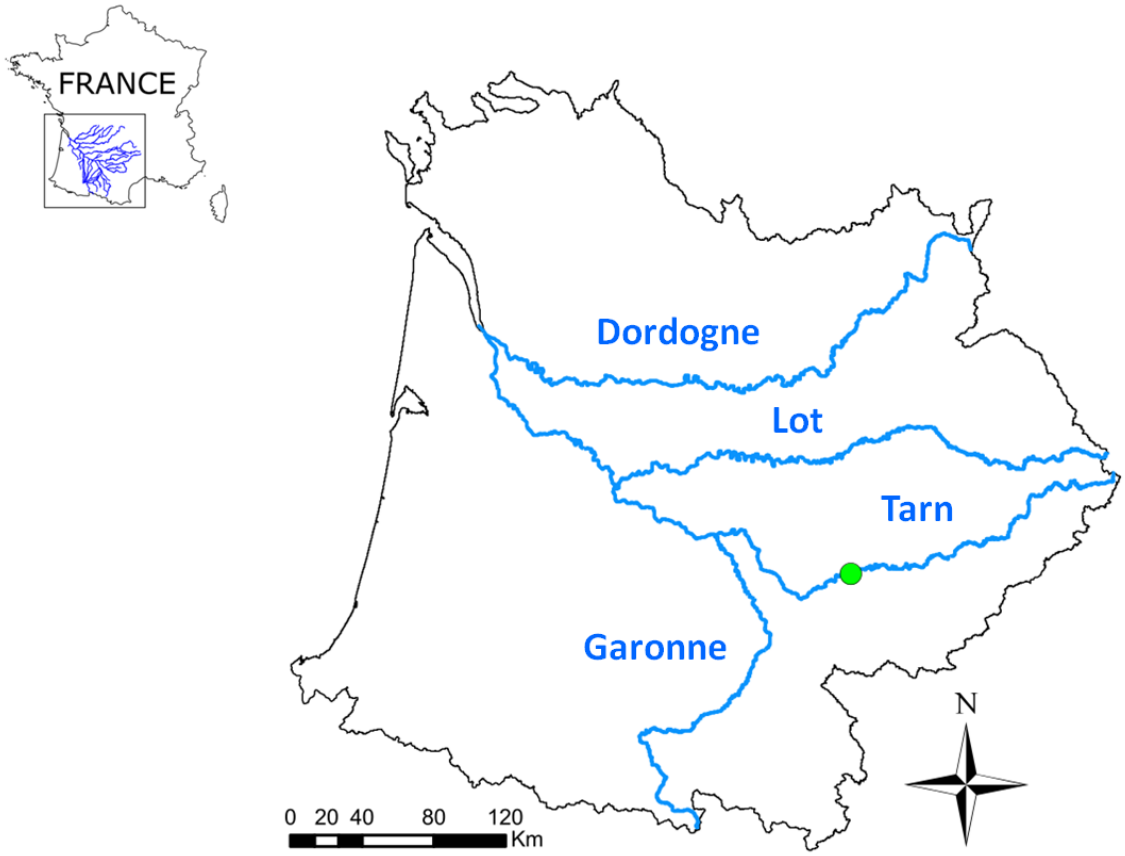


Figure 1 : Carte représentant la zone hydrographique et les quatre rivières principales du bassin de la Garonne (Garonne, Dordogne, Tarn et Lot). Le point vert indique la situation géographique d'Aiguelèze, localité où le premier déversement de silures a eu lieu.

Dès la fin des années 1980, il y a eu un intérêt halieutique et piscicole croissant en France pour le silure (Schlumberger et al., 2001), qui s'est traduit dans la presse (spécialisée ou non) par une communication importante concernant la production de silures en pisciculture et sa commercialisation (dans des buts d'alevinage et/ou gastronomiques). Cet intérêt grandissant, appuyé par le fait que le Conseil Supérieur de la Pêche (CSP) a publié à partir de 1996 des articles dans ses bulletins incitant à l'introduction de silures pour lutter biologiquement contre la présence de poissons-chat, a probablement facilité l'extension du silure dans le bassin de la Garonne dès 1987 (Proteau et al., 2008).

Dans ce contexte, d'autres évènements d'introduction ont eu lieu dans le bassin de la Garonne depuis la première introduction connue en 1983. Divers témoignages fournis par

des membres de diverses Fédérations de pêche rattachées de l'UFBAG nous ont permis de lister un certain nombre de ces évènements :

- En 1987, environ 2000 individus d'une longueur moyenne de 17 cm, provenant d'une pisciculture inconnue mais située probablement à l'étranger, auraient été déversés dans la Dordogne (la localité où le déversement a eu lieu nous est inconnue à ce jour).
- En 1993, malgré le manque d'appui de l'Association Agréée de Pêche et de Protection des Milieux Aquatiques (AAPPMA) de Villeneuve-sur-Lot, une douzaine de pêcheurs, de par leur propre initiative, a organisé une collecte de fonds pour acheter 252 silures de 18 à 20 cm de longueur au Vivier de la Castillonne, pisciculture tenue par les frères André et Gilbert Ribes à Montagnac dans l'Hérault. Ces individus auraient ensuite été introduits dans le Lot à la Meyrade (168 individus), au pont de Romas (42 ind.) et à la cale des pompiers de Villeneuve-sur-Lot (42 ind.).
- En 1994, 42 individus de 50-60 cm de longueur auraient été introduits dans le Lot à Courbiac. Ces poissons proviendraient d'une pisciculture non identifiée, mais située dans le centre de la France (probablement dans la Creuse).
- En 1995, 5 à 6 silures qui se trouvaient dans un plan d'eau privé auraient été capturés et déversés dans le Lot à Puy l'Évêque.
- En 1996 et 1997, quelques déversements supplémentaires d'individus ont eu lieu dans le Lot en Villeneuvois, cette fois-ci à l'initiative de l'AAPPMA de Villeneuve-sur-Lot : 25 individus de 50-60 cm en amont de Madame, 66 individus de 50-60 cm à Tauty, 51 individus de 45-70 cm en amont du barrage de Villeneuve, et 20 kg d'individus de 25-35 cm (dont deux sujets de 50 et 75 cm environ) à La Payssière. Tous ces individus provenaient de la même pisciculture que ceux déversés en 1994 à Courbiac (voir ici-haut).

Ces témoignages sporadiques nous apprennent que les déversements de silure ont pu être encadrés (par des structures de pêche ou de gestion des milieux aquatiques, ex. AAPPMA et CSP), comme réalisés dans le cadre d'initiatives privées. Ces renseignements nous montrent **qu'il est donc très difficile de recueillir avec précision toutes les données concernant les déversements de silure dans le bassin et l'origine géographique des individus déversés**.

Afin de caractériser l'évolution de l'expansion démographique et géographique du silure dans le bassin de la Garonne, nous avons donc analysé les données temporelles de suivi concernant la faune piscicole de l'AFB (Agence Française pour la Biodiversité) disponibles sur le portail Naïades (http://naiades.eaufrance.fr). Plus spécifiquement, nous avons caractérisé l'évolution géographique (en termes d'occupation spatiale) et démographique (en termes d'effectifs capturés) de cette espèce dans le bassin de la Garonne au cours du temps.

2. Présentation des données utilisées

Les données démographiques brutes concernant les occurrences de silure dans le bassin de la Garonne ont été extraites de la base de données « hydrobiologie » du portail **Naïades** (http://naiades.eaufrance.fr). Cette base de données répertorie les résultats des pêches électriques mises en place par le Conseil Supérieur de la Pêche (CSP) en 1990 pour

suivre les communautés piscicoles des rivières françaises. À partir de 2007, ces pêches ont été réalisées par l'Office National des Eaux et des Milieux Aquatiques (ONEMA), devenu aujourd'hui l'« Agence Française pour la Biodiversité » (AFB). Par souci de clarté, nous nous référerons par la suite à ces pêches de suivi temporel des communautés de poissons avec l'expression « pêches AFB ».

Les communautés de poissons ont été suivies chaque année entre les mois de mai et d'octobre dans plusieurs stations réparties sur l'ensemble du réseau hydrographique français en utilisant des protocoles de pêche électrique standardisés (Poulet et al., 2011). Dans les petits cours d'eau, les poissons ont été capturés lors de pêches « à pied », et en utilisant des protocoles de pêche électrique à deux passages (Poulet et al., 2011). Dans les grandes rivières, les prélèvements ont été effectués en bateau, en pêchant à l'électricité des fractions de divers méso-habitats présents dans la station de pêche (Lamouroux et al., 1999).

Les protocoles de pêche utilisés dans le cadre de ce suivi sont tout à fait adaptés pour déterminer la composition en espèces des communautés piscicoles occupant chaque station de suivi (Poulet et al., 2011), mais elles le sont moins pour estimer avec précision l'abondance et la densité des espèces présentes dans ces stations (notamment dans les grands cours d'eau, où il peut y avoir une grande variabilité dans le nombre de captures d'année en année pour une espèce donnée). Ces données restent toutefois très intéressantes et utiles pour caractériser des tendances démographiques sur le long terme pour toutes les espèces de poisson d'eau douce présentes en France, étant donné (i) la grande couverture des bassins hydrographiques français sur le plan spatial (ex. presque 200 stations suivies en 2017 pour le bassin de la Garonne), (ii) la longueur du suivi temporel (plus de 20 ans pour certaines stations) et (iii) le fait d'utiliser des protocoles standardisés d'année en année (permettant des comparaisons interannuelles, car même si les protocoles de pêche présentent des biais, ces mêmes biais auront lieu à chaque année de suivi).

La première capture d'un silure par les pêches AFB dans le bassin de la Garonne date de 1995, année où un individu a été capturé dans le Tarn à Moissac. **Les analyses de l'expansion démographique et géographique du silure dans le bassin de la Garonne que nous avons menées concerneront donc la période s'étalant de 1995** (date de première détection du silure dans le bassin) **à 2017** (dernière année pour laquelle les données d'occurrence sont disponibles dans le portail Naïades au moment de l'écriture de ce rapport).

La Figure 2 représente le nombre de stations suivies dans le bassin de la Garonne entre 1995 et 2017. Ce nombre a augmenté au cours des années, notamment après la dissolution du CSP et la création de l'ONEMA (en 2007), jusqu'à atteindre le nombre d'environ 200 stations suivies ces quatre dernières années.

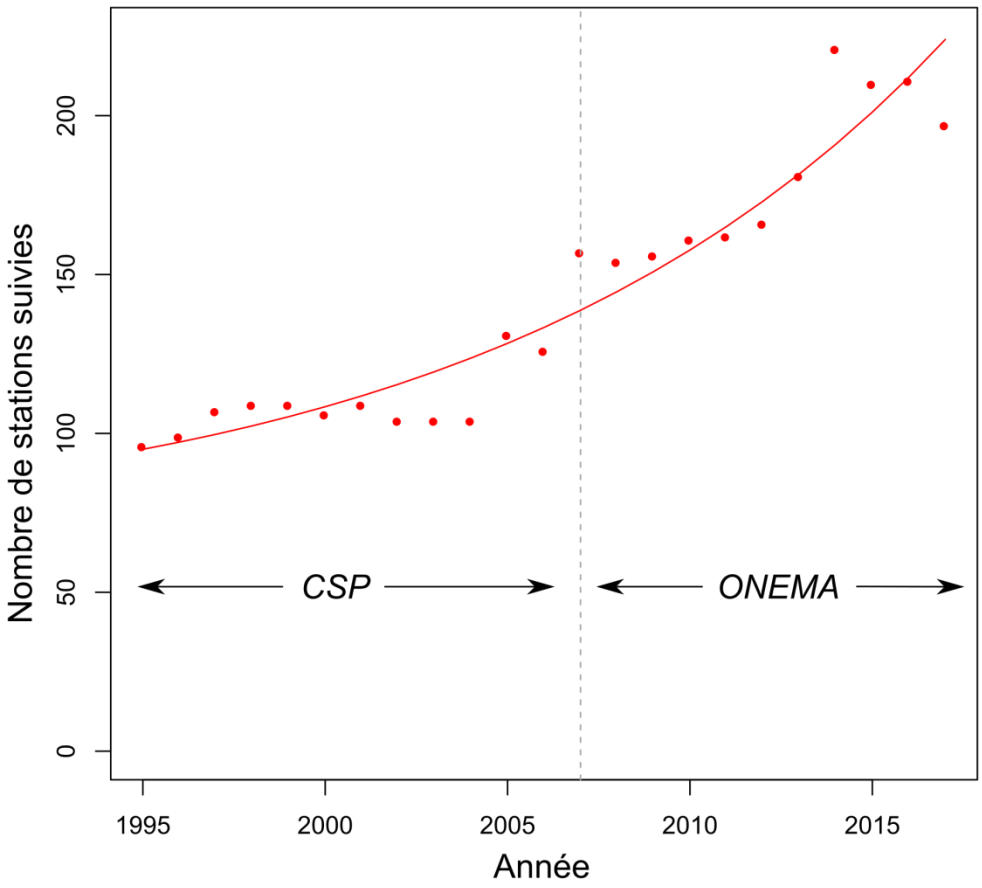


Figure 2 : Graphique représentant le nombre de stations suivies dans le bassin de la Garonne par année (i) entre 1995 et 2006 (pêches électriques organisées par le Conseil Supérieur de la Pêche ou CSP), et (ii) entre 2007 et 2017 (pêches électriques organisées par l'Office National de l'Eau et des Milieux Aquatiques ou ONEMA). La barre en pointillés verticale et grise représente la transition CSP – ONEMA (en 2007), qui s'est traduite par une augmentation conséquente des stations de suivi dans le bassin de la Garonne.

3. Analyse de l'expansion géographique

Nous avons tout d'abord illustré l'évolution géographique du silure dans le bassin de la Garonne sur des cartes de la zone hydrographique correspondante à ce bassin (Figure 3). Ainsi, la localisation spatiale des stations où le silure a été capturé avec succès lors des pêches AFB a été représentée. Ces cartes ont été plus spécifiquement générées pour les années 1995, 1996, 1997, 2000, 2005, 2010 et 2017 (Figure 3 A-G), ceci afin de mettre en relief les premières détections de silures dans les principales rivières d'intérêt de notre étude (à savoir, la Garonne, la Dordogne, le Tarn et le Lot), et de donner un aperçu de la progression de l'occupation spatiale du bassin par le silure au cours du temps. Voici ici-bas les principales informations qui découlent de ces cartes pour chacune des années de suivi considérées :

- **1995 (Figure 3 A)** : il s'agit de l'année où le premier silure a été capturé lors des pêches AFB dans le bassin de la Garonne, plus précisément dans le Tarn à Moissac. Notons que cette première détection lors d'une pêche AFB intervient 12 ans après la première introduction de silures dans le Tarn (1983).

- **1996 (Figure 3 B)** : il s'agit de l'année où les premiers silures ont été détectés dans la Dordogne et dans ses affluents (plus précisément, dans l'Isle à Bénévent) par les pêches AFB. Cette détection intervient 9 ans après le premier déversement d'individus connu dans la Dordogne (1987 ; voir ci-haut).

- **1997 (Figure 3 C)** : il s'agit de l'année où un premier silure a été détecté dans la Garonne (à Nicole) par les pêches AFB. Notons que cette année, le silure n'est ni détecté dans la Dordogne, ni dans le Tarn comme en 1996 et 1995.

- **2000 (Figure 3 D)** : plusieurs silures sont nouvellement capturés dans le Tarn (6 individus en aval de Rabastens et 5 individus à Moissac), dans la Garonne (1 individu à Agen, 4 à Bourret et 4 autres en amont de l'embouchure avec le Lot), et dans certains affluents de ces deux rivières (1 individu dans l'Agout à Saint-Sulpice, 1 individu dans l'Aveyron à Loubéjac et 3 individus dans l'Isle à Bénévent). Au début des années 2000, le silure semble donc s'étendre lentement dans le bassin.

- **2005 (Figure 3 E)** : il s'agit de l'année où les premiers silures sont détectés dans la partie amont du Lot (à Livinhac) par les pêches AFB, bien que des individus aient été détectés précédemment plus à l'aval en 2002 et 2004 (à Clairac, pas loin de la confluence avec la Garonne vers Nicole). En 2005, le silure reste relativement peu représenté dans le bassin, avec des captures dans seulement 8 stations sur les 131 suivies (soit dans 6,11% des stations suivies).

- **2010 (Figure 3 F)** : les captures de silure lors des pêches AFB se généralisent. En 2010, le silure est capturé dans 31 stations sur les 161 stations suivies (soit dans 19,25% des stations suivies). L'expansion géographique du silure dans le bassin semble donc avoir été beaucoup plus intense pendant la période 2005-2010 par rapport à la période 1995-2005.

- **2017 (Figure 3 G)** : les captures de silure dans les pêches AFB restent généralisées dans le bassin de la Garonne, mais il n'y a pas eu de grands changements par rapport à 2010 en termes d'expansion géographique. En 2017, le silure est capturé dans 30 stations sur les 197 stations suivies (soit dans 15,23% des stations suivies, moins que pour l'année 2010).

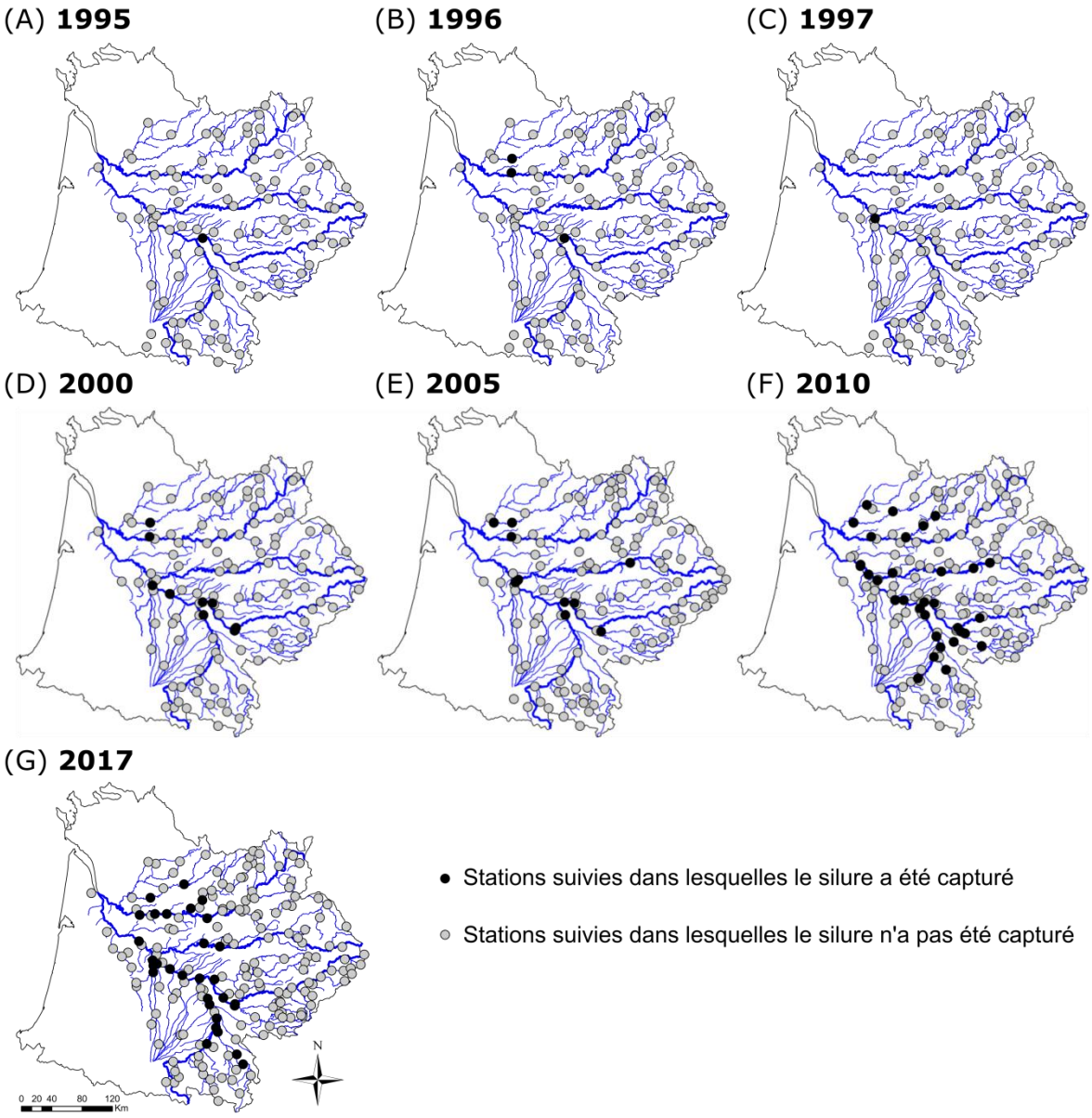


Figure 3 : Cartes représentant la localisation spatiale des stations suivies dans le bassin de la Garonne par le CSP-ONEMA. Les points gris figurent les stations dans lesquelles le silure n'a pas été capturé tandis que les points noirs figurent les stations dans lesquelles le silure a été capturé. Ceci pour les années 1995, 1996, 1997, 2000, 2005, 2010 et 2017 (panels A, B, C, D, E, F et G respectivement).

Afin de mieux caractériser l'évolution géographique du silure dans le bassin de la Garonne, nous avons représenté graphiquement le pourcentage de stations suivies dans lesquelles le silure a été capturé en fonction de l'année de suivi (pour la période 1995-2017 ; Figure 4). En plus de représenter les valeurs de pourcentage brutes, nous avons aussi représenté la courbe de régression de Poisson d'un modèle linéaire généralisé (« glm » ; McCullagh & Nelder, 1999) construit en considérant le pourcentage de stations suivies dans lesquelles le silure a été capturé comme variable à expliquer, et le temps (en années, pour la période 1995-2017) comme variable explicative. Nous avons aussi inclus un terme quadratique dans ce modèle glm. Ce modèle glm, ainsi que tous les autres glms présentés dans ce rapport, ont été construits avec le logiciel de statistiques R v.3.5.1 (R

Development Core Team, 2017). La valeur de R^2 (appelée aussi « coefficient de détermination ») du modèle glm a été calculée grâce au paquet 'rsq' du logiciel R (Zhang, 2017). Le R^2 est un paramètre qui varie entre 0 et 1 et qui indique la qualité de la prédiction d'une régression élaborée à partir d'un modèle, en l'occurrence, un glm dans notre cas. Une valeur proche de 0 indique que la qualité prédictive du modèle est faible, alors qu'une valeur proche de 1 indique que les prédictions du modèle sont bonnes.

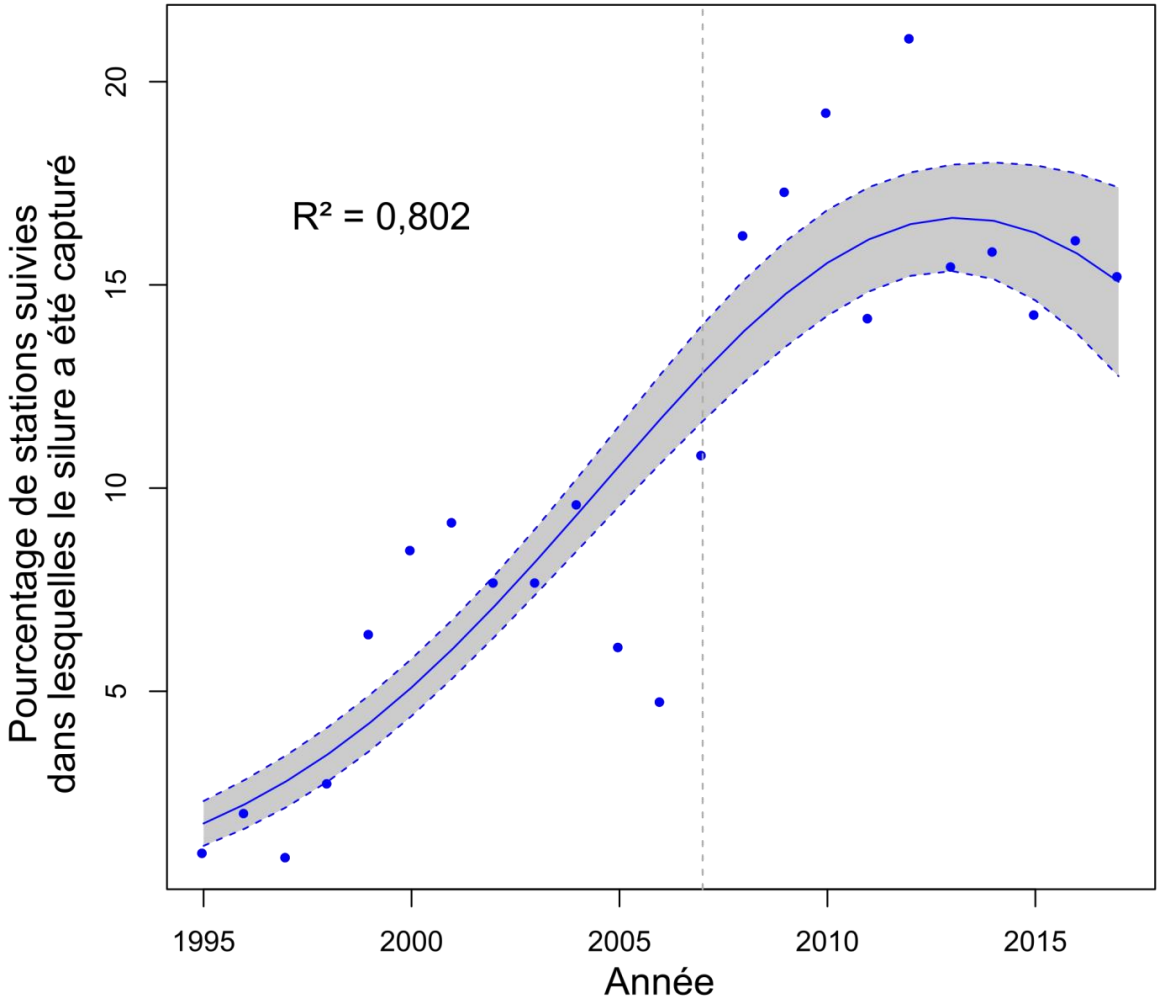


Figure 4 : Graphique représentant, pour la période 1995-2017, le pourcentage par année de stations suivies (dans le bassin de la Garonne) dans lesquelles le silure a été capturé (points bleus). La courbe bleue correspond aux valeurs prédites par un modèle linéaire généralisé (« glm ») visant à expliquer l'évolution du pourcentage de stations suivies dans lesquelles le silure a été capturé (variable à expliquer) en fonction du temps (variable explicative). Les courbes bleues en pointillés représentent l'intervalle de confiance à 95% des prédictions du modèle « glm ». La valeur de R^2 indique la qualité de la prédiction de la régression élaborée par le modèle « glm ». La barre en pointillés verticale et grise représente la transition CSP – ONEMA (en 2007), qui s'est traduite par une augmentation conséquente des stations de suivi dans le bassin de la Garonne.

Ainsi, ce graphique (Figure 4) nous montre qu'après une phase d'expansion géographique lente entre 1995 et 1998, le silure a connu une phase d'expansion très importante entre 1998 et 2012. Depuis 2013, le pourcentage de stations suivies dans lesquelles le silure a été capturé semble se maintenir stable. La courbe de régression que

nous avons obtenue avec le modèle glm a une forme typique de courbe de croissance de type sigmoïde (courbe caractérisé par une phase de croissance lente au début, suivie d'une phase de croissance intermédiaire importante, puis par un ralentissement de sa croissance jusqu'à atteindre un plateau).

Pour résumer, cette analyse nous indique que **l'expansion géographique du silure dans le bassin de la Garonne semble se stabiliser** (voire se réduire), **après avoir connu une phase d'expansion très importante** lors des deux dernières décennies. Il reste toutefois nécessaire d'attendre les résultats des pêches AFB des prochaines années pour vraiment confirmer ou infirmer cette tendance à la stabilisation de l'expansion géographique du silure dans le bassin de la Garonne.

4. Analyse de l'évolution démographique

Nous avons ensuite mis en évidence l'évolution démographique du silure (en termes d'effectifs capturés) dans le bassin de la Garonne et dans les principales rivières du bassin (Garonne, Dordogne, Lot et Tarn).

Tout d'abord, nous avons représenté graphiquement le nombre total de silures capturés parmi toutes les stations suivies dans le bassin de la Garonne en fonction de l'année de suivi (pour la période 1995-2017 ; Figure 5). Nous avons aussi représenté sur le graphique la courbe de régression de Poisson d'un modèle linéaire généralisé construit en considérant le nombre total de silures capturés parmi toutes les stations comme variable à expliquer, et le temps (en années, pour la période 1995-2017) comme variable explicative (en considérant un terme quadratique dans ce modèle glm).

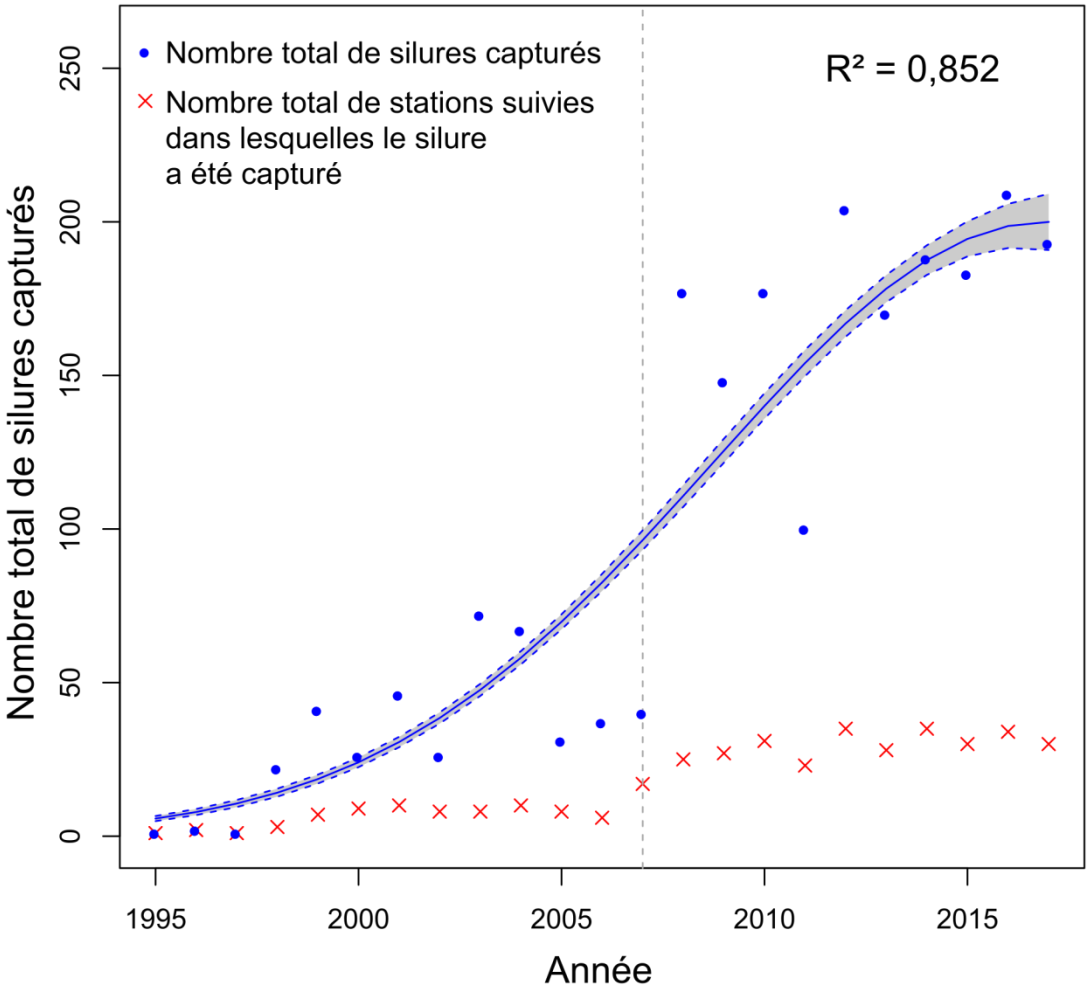


Figure 5 : Graphique représentant, pour la période 1995-2017, le nombre total de silures capturés par année sur l'ensemble des stations suivies dans le bassin de la Garonne (points bleus). Les croix rouges représentent le nombre total de stations suivies dans lesquelles le silure a été capturé par année. La courbe bleue correspond aux valeurs prédites par un modèle linéaire généralisé (« glm ») et visant à expliquer le nombre total de silures capturés (variable à expliquer) en fonction du temps (variable explicative). Les courbes bleues en pointillés représentent l'intervalle de confiance à 95% des prédictions du modèle glm. La valeur de R^2 indique la qualité de la prédiction de la régression élaborée par le modèle glm. La barre en pointillés verticale et grise représente la transition CSP – ONEMA (en 2007), qui s'est traduite par une augmentation conséquente des stations de suivi dans le bassin de la Garonne.

Comme pour le graphique précédent (Figure 4), nous pouvons observer une courbe de régression de type sigmoïde, qui explique l'évolution du nombre total de silures capturés lors des pêches AFB dans le bassin de la Garonne au cours du temps (Figure 5). Ainsi, le nombre total de captures, très faible entre 1995 et 1997, a ensuite augmenté fortement jusqu'en 2012, avant de se stabiliser aux alentours de 200 individus par année de suivi après 2013. Cette analyse suggère que **le stock de silure dans le bassin de la Garonne**, après avoir subi une forte croissance lors des deux dernières décennies, **serait peut-être actuellement en train de se stabiliser**. Comme dans le cas de l'analyse précédente, il faudra attendre les résultats des futures pêches AFB pour confirmer ou infirmer cette tendance. Notons cependant, que la forte augmentation des effectifs

capturés par les pêches AFB à partir de 2007 peut aussi être fortement liée à l'augmentation du nombre de stations suivies dans le bassin suite à la dissolution du CSP et à la création de l'ONEMA (cf. Figure 2 ci-haut et croix rouges sur la Figure 5).

Pour limiter les effets de cette augmentation du nombre de stations suivies dans l'analyse de l'évolution démographique du silure dans le bassin de la Garonne, nous avons effectué une analyse supplémentaire dans laquelle nous avons représenté graphiquement le nombre moyen de silures capturés (parmi toutes les stations suivies dans le bassin de la Garonne où il y a eu au moins une capture de silure) en fonction de l'année de suivi (pour la période 1995-2017 ; cf. Figure 6). Et, comme précédemment, nous avons aussi représenté la courbe de régression de Poisson d'un modèle linéaire généralisé construit en considérant le nombre moyen de silures capturés comme variable à expliquer, et le temps (en années, pour la période 1995-2017) comme variable explicative (en considérant un terme quadratique).

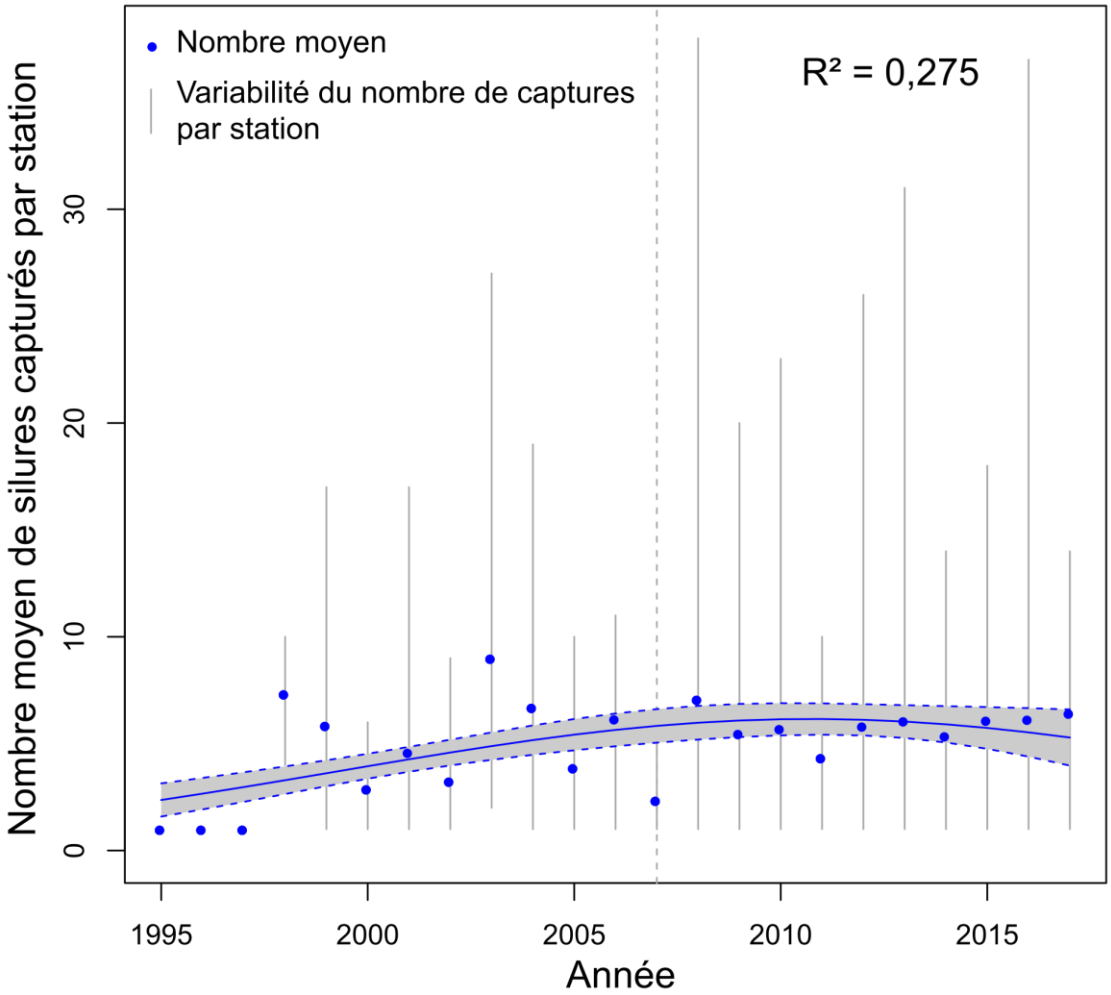


Figure 6 : Graphique représentant, pour la période 1995-2017, le nombre moyen de silures capturés par station et par année dans le bassin de la Garonne (points bleus). Seules les stations où il y a eu au moins un silure capturé ont été prises en compte pour le calcul de la moyenne. Les barres verticales pleines et grises représentent l'étendue du nombre de captures par station. La courbe bleue correspond aux valeurs prédites par un modèle linéaire généralisé (« glm ») et visant à expliquer le nombre moyen de silures capturés par station (variable à expliquer) en fonction du temps (variable explicative). Les

courbes bleues en pointillés représentent l'intervalle de confiance à 95% des prédictions du modèle « glm ». La valeur de R^2 indique la qualité de la prédiction de la régression élaborée par le modèle « glm ». La barre en pointillés verticale et grise représente la transition CSP – ONEMA (en 2007), qui s'est traduite par une augmentation conséquente des stations de suivi dans le bassin de la Garonne.

Ainsi, ce graphique nous montre qu'à partir de 1998 le nombre moyen de silures capturés par station tend à rester stable dans le temps (aux alentours de 5 individus en moyenne par station à partir de 2007), bien qu'il y ait une forte variabilité dans le nombre maximal d'individus capturés par station selon les années (ex. au maximum : 5 individus en 2007, 38 en 2008 et 20 en 2009 ; voir le haut des barres verticales pleines et grises pour ces années dans la Figure 6).

Globalement, l'analyse de l'évolution géographique et démographique du silure à l'échelle du bassin de la Garonne nous suggère que cette espèce, après une phase de latence (de 1983 à 1999 environ), a connu une phase d'expansion très marquée entre les années 2000 et 2012. Cette phase d'expansion semble s'être interrompue, et l'aire de distribution occupée par l'espèce ainsi que ses effectifs semblent se maintenir stables depuis quelques années dans le bassin. Les données des pêches AFB des cinq prochaines années seront décisives pour confirmer ou infirmer cette stabilisation géographique et démographique du silure dans le bassin de la Garonne.

Pour parfaire nos investigations, nous avons effectué une analyse démographique similaire à la précédente, mais cette fois-ci centrée sur les quatre rivières d'intérêt de notre étude (Garonne, Dordogne, Tarn et Lot ; Figure 7 A-D). Nous avons donc, pour la période 1995-2017 et pour chacune de ces quatre rivières, représenté graphiquement le nombre moyen de silures capturés (parmi toutes les stations suivies dans chaque rivière où il y a eu au moins une capture de silure) en fonction de l'année de suivi (Figure 7). Nous avons aussi représenté dans ce graphique (comme précédemment) les courbes de régression de Poisson de modèles glm construits en considérant le nombre moyen de silures capturés comme variable à expliquer, et le temps (en années, pour la période 1995-2017) comme variable explicative (en considérant un terme quadratique).

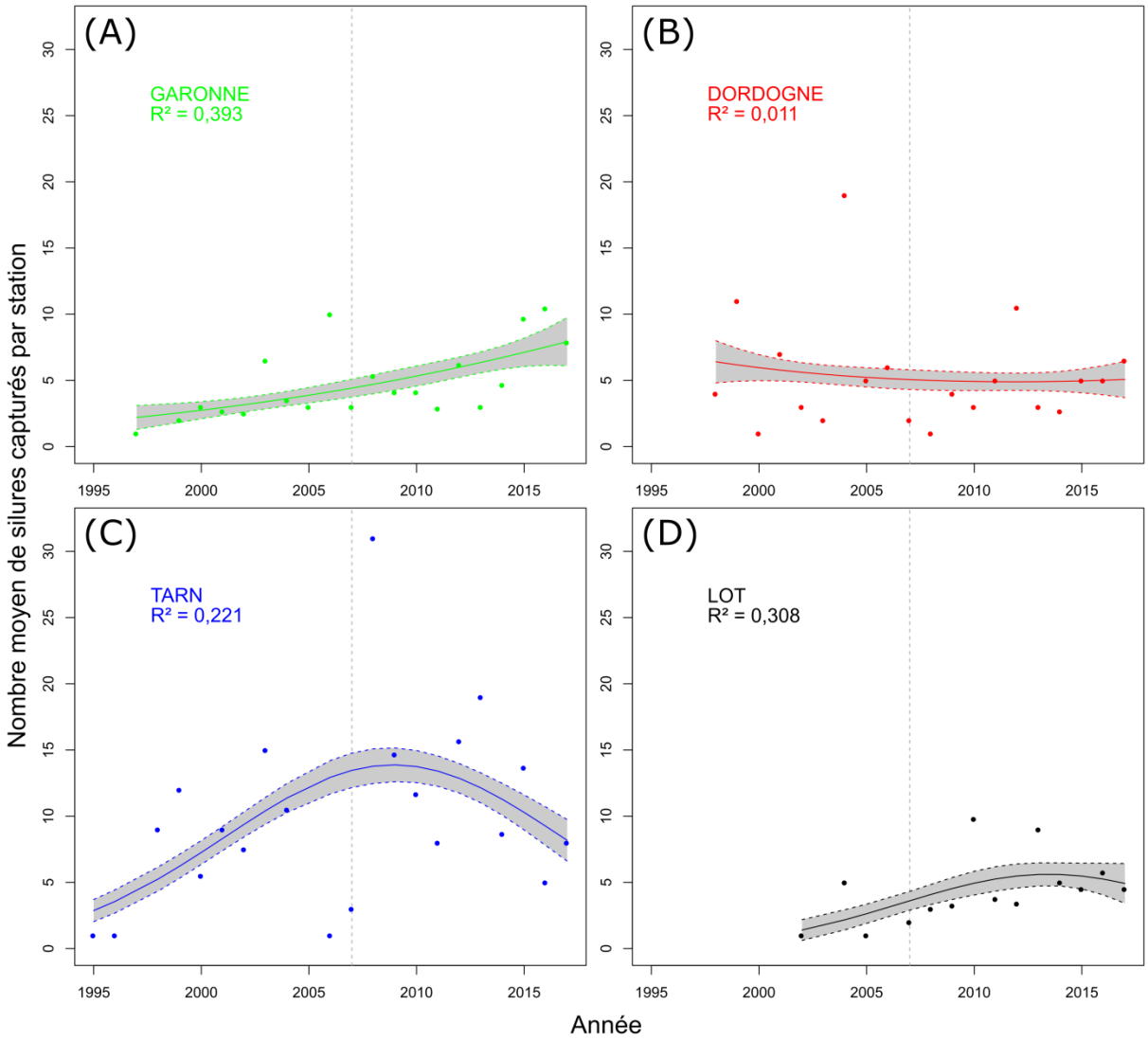


Figure 7 : Graphiques représentant, pour la période 1995-2017, le nombre moyen de silures capturés par station et par année dans (A) la Garonne (points verts), (B) la Dordogne (points rouges), (C) le Tarn (points bleus), et (D) le Lot (points noirs). Seules les stations où il y a eu au moins un silure capturé ont été prises en compte pour le calcul de la moyenne. Les courbes pleines verte, rouge, bleu et noire correspondent aux valeurs prédites par des modèles linéaires généralisés (« glm ») visant à expliquer le nombre moyen de silures capturés (variable à expliquer) en fonction du temps (variable explicative) pour la Garonne, la Dordogne, le Tarn et le Lot respectivement. Les courbes en pointillés verte, rouge, bleue et noire représentent les intervalles de confiance à 95% des prédictions des modèles « glm » construits pour la Garonne, la Dordogne, le Tarn et le Lot respectivement. Les valeurs de R^2 indiquent la qualité de la prédiction des régressions élaborées par les modèles « glm ». Les barres en pointillés verticales et grises représentent la transition CSP – ONEMA (en 2007), qui s'est traduite par une augmentation conséquente des stations de suivi dans le bassin de la Garonne.

Les tendances démographiques mises en évidence par notre analyse ne sont pas les mêmes selon les rivières. En effet, il semblerait que le nombre moyen de silures par station augmente progressivement dans la Garonne depuis 1996, malgré une variabilité interannuelle parfois importante (Figure 7 A), alors que celui-ci reste stable sur la

Dordogne (Figure 7 B). Pour le Tarn, le nombre moyen d'individus a progressé positivement jusqu'en 2008, mais semble en baisse depuis quelques années (Figure 7 C). Finalement, pour le Lot, le nombre moyen de silures capturés a progressé lentement depuis sa première détection en 2002 jusqu'en 2013 (Figure 7 D). Depuis, ce nombre semble demeurer stable.

Globalement, cette analyse par rivières semble indiquer que les populations de silure seraient actuellement stables dans la Dordogne et le Lot, en expansion démographique dans la Garonne et en décroissance dans le Tarn, rivière où l'espèce est présente depuis plus de temps. Comme pour les analyses effectuées à l'échelle du bassin entier, les données des pêches AFB des cinq prochaines années seront décisives pour confirmer ou infirmer ces tendances démographiques observées à l'échelle de ces quatre rivières.

Notons aussi que nous avons effectué une analyse de l'évolution du nombre de silures capturés à une échelle plus locale, dans le cadre d'un concours de pêche au silure organisé chaque année depuis 2007 par la FFPPMA 47 sur le Lot à Penne d'Agenais. Cette analyse et ses résultats sont présentés en détail en Annexe 1 à la fin de ce rapport.

Partie II : Description de la diversité et de la structuration génétique du Silure dans le bassin de la Garonne.

Cette deuxième partie est consacrée à la description de notre échantillonnage, des méthodes moléculaires utilisées pour générer notre jeu de données génétiques, ainsi qu'à la description de la diversité et de la structure génétique observées chez le silure à l'échelle du bassin de la Garonne. **Cette partie a été réalisée en étroite collaboration avec la Professeure Géraldine Loot et l'Ingénieure d'études Charlotte Veyssière du laboratoire « Évolution et Diversité Biologique » de l'Université Paul Sabatier**.

1. Échantillonnage

L'ensemble des échantillons que nous avons récupéré pour construire la base de données génétiques concerne les quatre cours d'eau principaux du bassin de la Garonne : la Garonne, la Dordogne, le Lot et le Tarn (Figure 8 ; Table 1). Plusieurs acteurs ont contribué à cet échantillonnage, qui s'est étalé de 2015 à 2018 : certains échantillons ont été prélevés par des membres d'ECOLAB lors de pêches électriques, tandis que d'autres nous ont été envoyés par Pascal Verdeyroux (EPIDOR), par des pêcheurs professionnels, des pêcheurs à la ligne amateurs, par l'association MIGADO (Migrateurs Garonne Dordogne) ou par les FFPPMA (Fédérations Françaises pour la Pêche et la Protection du Milieu Aquatique) des départements 46, 47 et 81 (cf. Table 1 pour plus de détails concernant l'échantillonnage).

Au total, **562 individus ont été échantillonnés et génotypés avec succès, sur 18 sites d'échantillonnage différents** (9 sites sur la Garonne et 3 sites sur la Dordogne, le Lot et le Tarn respectivement ; cf. Figure 8). Par la suite, nous nous référerons aux 18 sites échantillonnés avec le terme « populations ».

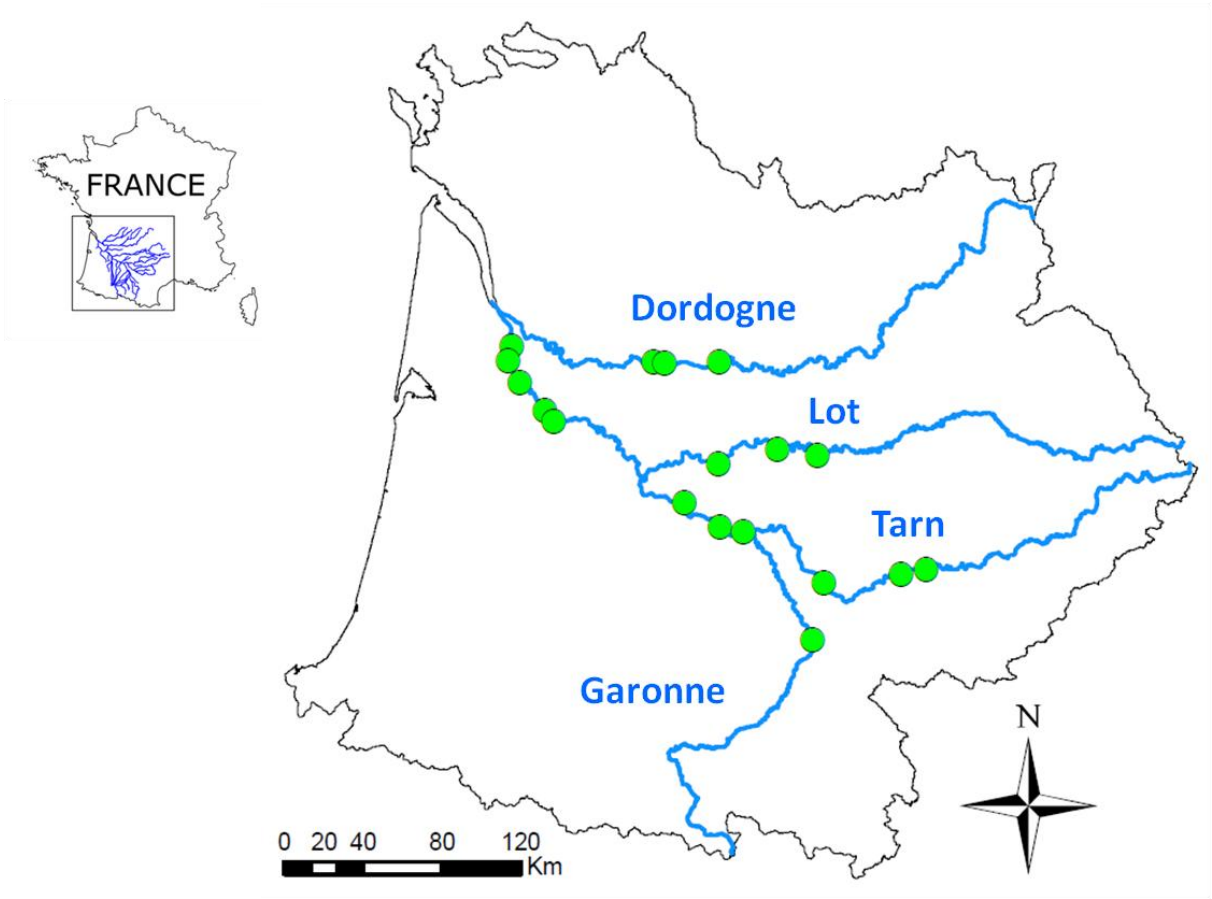


Figure 8 : Carte du Sud-ouest de la France représentant le réseau hydrographique concerné et les différents sites d'échantillonnage du Silure (points verts).

Table 1 : Tableau récapitulatif des différents sites d'échantillonnage avec leur nom de code, leur origine, la date de leur prélèvement, leur localisation, ainsi que le nombre d'individus capturés (N_{ECH}).

Code Site	Origine des échantillons	Année	Rivière	Site	Longitude	Latitude	Distance à l'embouchure (en km)	N_{ECH}
DORLam	Pascal Verdeyroux	2016-2017	Dordogne	Lamonzie / Prigonrieux	N 44°84'96,8"	E 0°38'14,4"	148,125	15
DORBer	Pascal Verdeyroux	2016-2017	Dordogne	Bergerac (Aval)	N 44°84'60,1"	E 0°45'00,3"	155,073	12
DORMau	Pascal Verdeyroux	2016-2017	Dordogne	Mauzac (Aval)	N 44°86'01,6"	E 0°80'41,1"	185,991	96
GARBas	Pêcheurs professionnels	2016	Garonne	Bassens	N 44°89'37,4"	O -0°60'92,9"	9,000	30
GARStj	Pêcheurs professionnels	2016	Garonne	Saint-Jean / Bouliac	N 44°83'09,3"	O -0°55'05,8"	17,942	30
GARIsl	Pêcheurs professionnels	2016-2017	Garonne	Isle-Saint-Georges / Cambes	N 44°73'18,3"	O -0°47'14,4"	32,097	50
GARBar	Pêcheurs professionnels	2016-2017	Garonne	Barsac	N 44°60'82,6"	O -0°30'41,1"	55,287	54
GARLan	Pêcheurs professionnels	2016-2017	Garonne	Langon	N 44°55'78,2"	O -0°24'44,9"	63,004	56
GARAge	Pêcheurs à la ligne	2017	Garonne	Agen	N 44°20'78,3"	E 0°60'55,7"	174,442	9
GARGol	MIGADO	2016-2017	Garonne	Golfech	N 44°10'54,9"	E 0°83'90,8"	201,644	63
GARMal	ECOLAB	2016	Garonne	Malause	N 44°08'26,8"	E 0°98'47,3"	215,556	17
GARTou	ECOLAB	2016	Garonne	Toulouse	N 43°60'06,7"	E 1°43'69,6"	301,476	9
LOTPen	FFPPMA 47 (Silurade)	2018	Lot	Penne d'Agenais	N 44°39'51,1"	E 0°81'77,7"	200,528	11
LOTPra	FFPPMA 47	2018	Lot	Prayssac	N 44°46'74,6"	E 1°19'12,9"	251,963	9
LOTCah	ECOLAB - FFPPMA 46	2016-2018	Lot	Cahors	N 44°48'52,5"	E 1°25'38,4"	300,849	30
TARVil	ECOLAB	2015	Tarn	Villemur	N 43°86'38,3"	E 1°50'37,9"	282,868	19
TARAig	FFPPMA 81	2017	Tarn	Aiguelèze	N 43°90'89,4"	E 1°98'85,7"	342,140	18
TARAlb	ECOLAB	2015	Tarn	Albi	N 43°93'61,6"	E 2°13'14,3"	360,4856	31

2. Méthodes moléculaires utilisées

Extractions d'ADN et amplifications des marqueurs ADN.

Nous avons extrait l'ADN nucléaire des 562 individus échantillonnés à partir de petits fragments de nageoires pelviennes en utilisant un protocole d'extraction dit « au sel » (Aljanabi & Martinez, 1997). L'ADN nucléaire a été amplifié via des techniques de PCR standard (« Polymerase Chain Reaction » ; Mullis et al., 1986) au niveau de 10 loci microsatellites. L'amplification de ces 10 loci (ou marqueurs) microsatellites a été réalisée grâce à la mise en place de 2 kits multiplex différents (voir Annexe 2 pour plus de détails sur les conditions de PCR et de génotypage). Les marqueurs microsatellites sont des séquences d'ADN formées par des répétitions continues de motifs nucléotidiques qui sont très utiles pour mesurer la diversité génétique des espèces, ainsi que pour inférer leur histoire démographique et évolutive récente (David & Samadi, 2011).

Contrôles de la qualité du génotypage.

Nous avons effectué plusieurs contrôles de qualité préliminaires sur nos données génétiques brutes afin d'obtenir une base de données génétique propre et robuste. Nous avons d'abord vérifié si notre jeu de données contenait des erreurs de génotypage et/ou d'amplification (comme par exemple la présence d'allèles nuls). Pour cela, nous avons utilisé le logiciel MICROCHECKER v2.3 (Van Oosterhout et al., 2004). Ensuite, nous avons vérifié si les marqueurs microsatellites étaient en conformité avec l'équilibre d'Hardy-Weinberg grâce au programme GENEPOP v4.0 (Rousset, 2008). Cette analyse permet de détecter si des populations sont fortement affectées par des problèmes liés à de la dépression de consanguinité ou à une forte dérive génétique. Pour finir, nous avons vérifié si les marqueurs étaient ou non en déséquilibre de liaison significatif avec le logiciel FSTAT v2.9.3.2 (Goudet, 1995).

Nous n'avons trouvé des indices de présence d'allèles nuls que pour neuf paires de locus/population sur les 180 paires possibles : trois paires impliquant le marqueur Sgl310 (populations GARGol, GARBar et DORMau ; cf. Table 1 pour retrouver les noms complets de chaque site) ; deux paires impliquant le marqueur Sgl325 (GARGol et GARBar) ; deux paires impliquant le marqueur Sgl5F (GARBas et GARIsl) et une paire impliquant les marqueurs Sgl7e (GARLan) et Sgl7159 (GARTou). Nous avons aussi trouvé deux couples de locus/population non conformes avec l'hypothèse d'équilibre d'Hardy-Weinberg (locus Sgl310 pour DORMau et LOTCah). Nous n'avons pas détecté de signes significatifs de déséquilibre de liaison entre marqueurs, et ceci pour aucune des populations.

Étant donné le faible nombre d'indices de présence d'erreurs de génotypage, ainsi que l'absence de patrons particuliers dans leur occurrence, nous avons pu retenir les 10 marqueurs initiaux pour réaliser toutes les analyses génétiques postérieures.

3. Diversité et structure génétique du silure

Description de la diversité génétique neutre avec les marqueurs microsatellites.

La base de données microsatellites utilisée pour décrire la diversité et la structure génétique du silure (obtenue après avoir effectué tous les contrôles de qualité décrits précédemment) contient des données génotypiques pour 562 individus et 10 marqueurs microsatellites. À partir de cette base de données nous avons d'abord calculé, pour chacune des 18 populations considérées, deux indices qui sont très importants en biologie de la conservation : la richesse allélique (AR ; Petit et al., 1998) et la richesse en allèles privés (PA ; Kalinowski, 2004).

Le premier indice (AR) permet de quantifier à quel point une population est diverse sur le plan génétique, alors que le second indice (PA) permet d'indiquer qu'elle est la proportion moyenne des allèles présents dans une population (tous marqueurs confondus) qui sont exclusifs à cette population. Cette dernière statistique descriptive permet de donner une idée de l'exclusivité de la diversité génétique d'une population. Rappelons qu'un allèle est une version différente d'un même gène (dans notre cas, un marqueur microsatellite). Plus on retrouve de versions des gènes (d'allèles) dans une population, et plus on considère que cette population se caractérise par une forte diversité génétique.

Ces deux indices ont été calculés grâce aux procédures de raréfaction implémentées dans le logiciel ADZE (Szpiech et al., 2008). Ces procédures de raréfaction permettent de standardiser les valeurs obtenues, afin de rendre comparables les valeurs de AR et de PA obtenues pour des populations dont le nombre d'individus échantillonnés est différent. Dans notre cas, nous avons calculé AR et PA en considérant une valeur minimale d'individus échantillonnés de 9, étant donnée l'hétérogénéité du nombre d'individus génotypés avec succès par population (voir Annexe 3). Les Figures 9 A et B illustrent les valeurs de AR et PA calculées pour chacune des 18 populations.

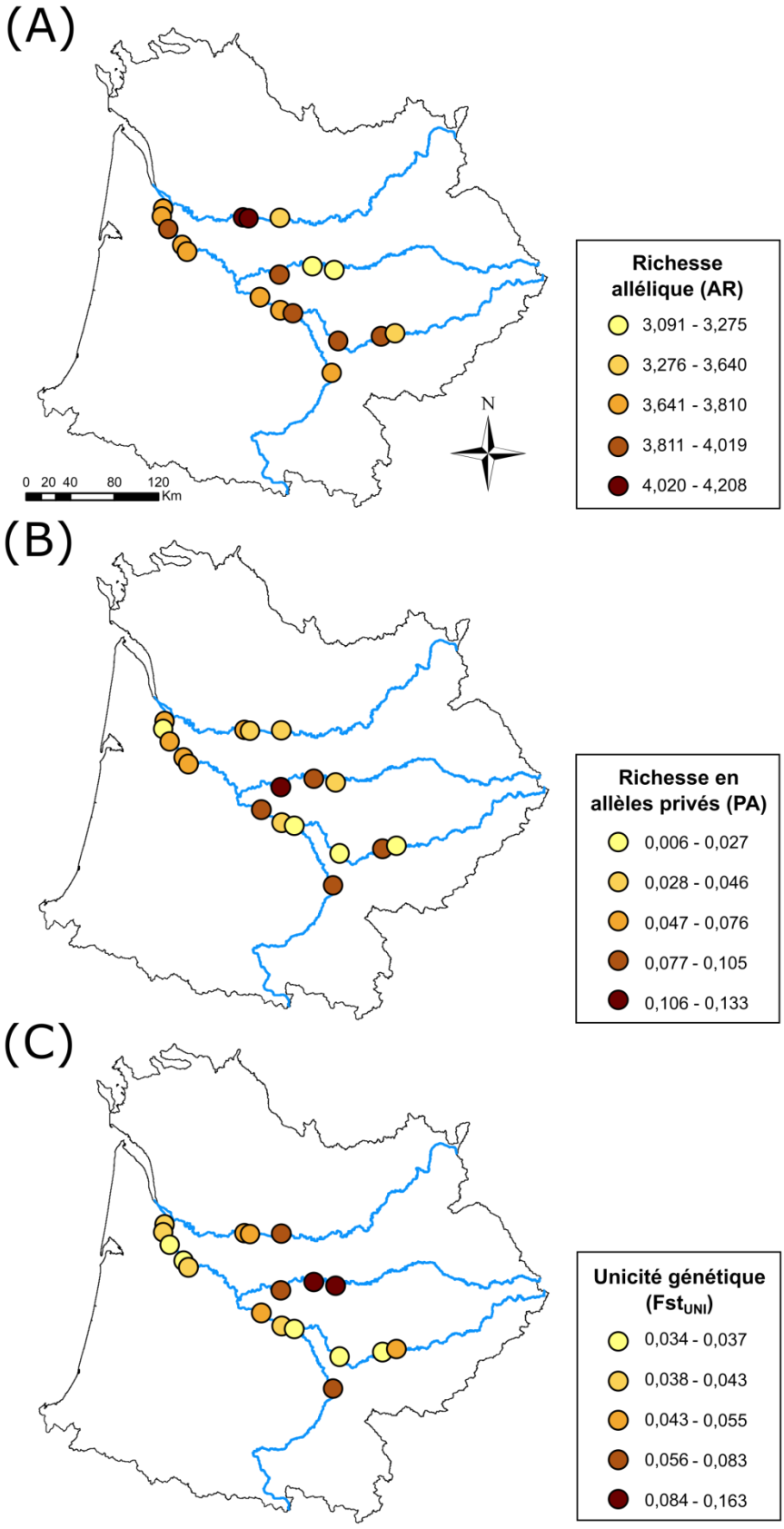


Figure 9 : Cartes représentant les valeurs de richesse allélique AR (A), de richesse en allèles privés PA (B), et d'unicité génétique Fst_{UNI} (C), calculées pour chaque population à partir des 10 marqueurs microsatellites.

Nous pouvons remarquer que les populations qui présentent les plus fortes valeurs de richesse allélique (AR) sont celles de la Dordogne à Lamonzie (4,208) et Bergerac (4,180), ainsi que celle du Lot à Penne d'Agenais (4,020 ; Figure 9 A ; Annexe 3), tandis que les valeurs de AR les plus faibles ont été observées pour les deux autres populations du Lot, situées en amont de Penne d'Agenais (3,276 à Prayssac et 3,091 à Cahors ; Figure 9 A ; Annexe 3).

En ce qui concerne les valeurs de richesse en allèles privés (PA), les populations du Lot à Penne d'Agenais et Prayssac sont celles pour lesquelles nous avons observé les valeurs de PA les plus élevées (0,133 et 0,106 respectivement ; Figure 9 B ; Annexe 3), suivies de celles de la Garonne à Toulouse (0,096) et à Agen (0,086). Les valeurs de PA les plus faibles ont été observées pour les populations du Tarn à Albi (0,007) et de la Garonne à Malause (0,016 ; Figure 9 B ; Annexe 3).

Nous avons ensuite calculé un indice d'unicité génétique (Fst_{UNI}) qui permet de rendre compte à quel point une population est unique sur le plan génétique par rapport aux autres populations. Pour pouvoir estimer une valeur de Fst_{UNI} pour chacune des 18 populations génotypées, nous avons d'abord calculé des valeurs de différenciation génétique (Fst ; Weir & Hill, 2002) entre chaque paire de populations avec le logiciel FSTAT v2.9.3.2 (Goudet, 1995). Ensuite, pour chaque population, nous avons calculé la moyenne des 17 valeurs de Fst estimées entre cette population et les 17 autres. Plus la valeur de Fst_{UNI} d'une population est élevée, et plus cette population sera « unique » ou « différente » sur le plan génétique par rapport à toutes les autres.

Ainsi, les populations pour lesquelles nous avons observé les valeurs de Fst_{UNI} les plus élevées sont celles du Lot à Cahors (Fst_{UNI} = 0,163) et à Prayssac (0,128 ; Figure 9 C ; Annexe 3), alors que les valeurs les plus faibles ont été observées pour les populations de la Garonne à Barsac et à Malause (0,035 ; Figure 9 C ; Annexe 3).

Globalement, **ces analyses nous indiquent que les populations situées en amont du Lot** (Prayssac et Cahors) **présentent des spécificités génétiques qui les distinguent à la fois de la population le plus à l'aval du Lot** (celle échantillonnée à Penne d'Agenais) **et des autres populations échantillonnées dans le bassin** (Figure 9). En effet, les populations à l'amont du Lot (et notamment celle de Cahors) présentent à la fois les richesses alléliques les plus faibles et les valeurs d'unicité génétique les plus fortes observées dans le bassin. Ceci indique que la diversité génétique de ces deux populations semble très particulière par rapport à celle observée dans les autres populations. Les analyses des patrons spatiaux de diversité et de structuration génétique présentés dans les sections suivantes nous permettrons de mieux appréhender les particularités génétiques de chacune des populations échantillonnées.

Finalement, nous avons calculé aussi plusieurs autres statistiques descriptives de la diversité génétique classiques comme l'hétérozygotie attendue (He), l'hétérozygotie observée (Ho), le nombre d'allèles moyen (Na) ou le coefficient de consanguinité (Fis) observés pour chacune des 18 populations considérées. Ces statistiques ont été calculées avec les logiciels GENETIX (Belkhir et al., 1996) et GENEPOP v4.0 (Rousset, 2008). Cependant, par souci de clarté, nous ne décrirons pas en détail les résultats obtenus lors

de ces analyses ; résultats souvent redondants avec ceux présentés jusqu'alors. Nous mettons à la disposition des lecteurs les valeurs brutes de ces statistiques dans l'annexe 3.

Patrons spatiaux de diversité et de structuration génétique.

Afin de mieux déterminer les particularités génétiques des populations de silure échantillonnées dans le bassin de la Garonne, nous nous sommes intéressés à la distribution spatiale des valeurs de AR, PA et Fst_{UNI} calculées dans la section précédente. Nous nous sommes plus particulièrement attachés à caractériser les patrons spatiaux de diversité génétique, afin de voir s'ils sont (ou pas) en accord avec des patrons spatiaux globaux généralement prédits pour les organismes aquatiques d'eau douce.

Nous nous sommes tout d'abord intéressés à la distribution spatiale de la richesse allélique (AR) le long du gradient amont-aval des rivières, afin de déterminer si cette distribution spatiale est en accord avec un patron de type « DIGD » (pour « Downstream Increase in Genetic Diversity (Paz-Vinas et al., 2015; Paz-Vinas, 2014). Le DIGD est un patron spatial de diversité génétique généralisé dans les réseaux hydrographiques selon lequel la diversité génétique augmente le long du gradient amont/aval (Paz-Vinas et al., 2015). Ce patron résulte de l'interaction entre la connectivité de type dendritique d'un réseau hydrographique (Paz-Vinas & Blanchet, 2015) et des processus inhérents aux bassins hydrographiques comme l'asymétrie de flux de gènes biaisée vers l'aval, l'augmentation des tailles efficaces des populations le long du gradient amont-aval ou l'héritage génétique des colonisations postglaciaires (Paz-Vinas et al., 2015). Ce patron DIGD est attendu pour des espèces de poisson d'eau douce non amphihalines et natives. Pour une espèce non-native comme le silure, ce patron DIGD peut être attendu si cette espèce a été présente depuis suffisamment de temps dans un bassin pour s'y être parfaitement acclimatée.

Afin de vérifier la présence ou l'absence de patron DIGD pour le silure dans le bassin de la Garonne, nous avons représenté sur la Figure 10 A les valeurs de AR calculées pour chacune des 18 populations génotypées en fonction de leur distance topologique à l'embouchure (c'est à dire, en suivant le cours d'eau). Comme pour les graphiques de la Partie I, nous avons représenté sur le graphique la courbe de régression de Poisson d'un modèle linéaire généralisé construit en considérant les valeurs de AR comme variable à expliquer, et la distance à l'embouchure comme variable explicative (en considérant un terme quadratique), puis nous avons calculé une valeur de R^2. Les distances topologiques entre chaque population échantillonnée et l'embouchure du bassin de la Garonne ont été calculées grâce au paquet 'riverdist' v. 0.15 du logiciel R v. 2.5.2 (Tyers, 2017).

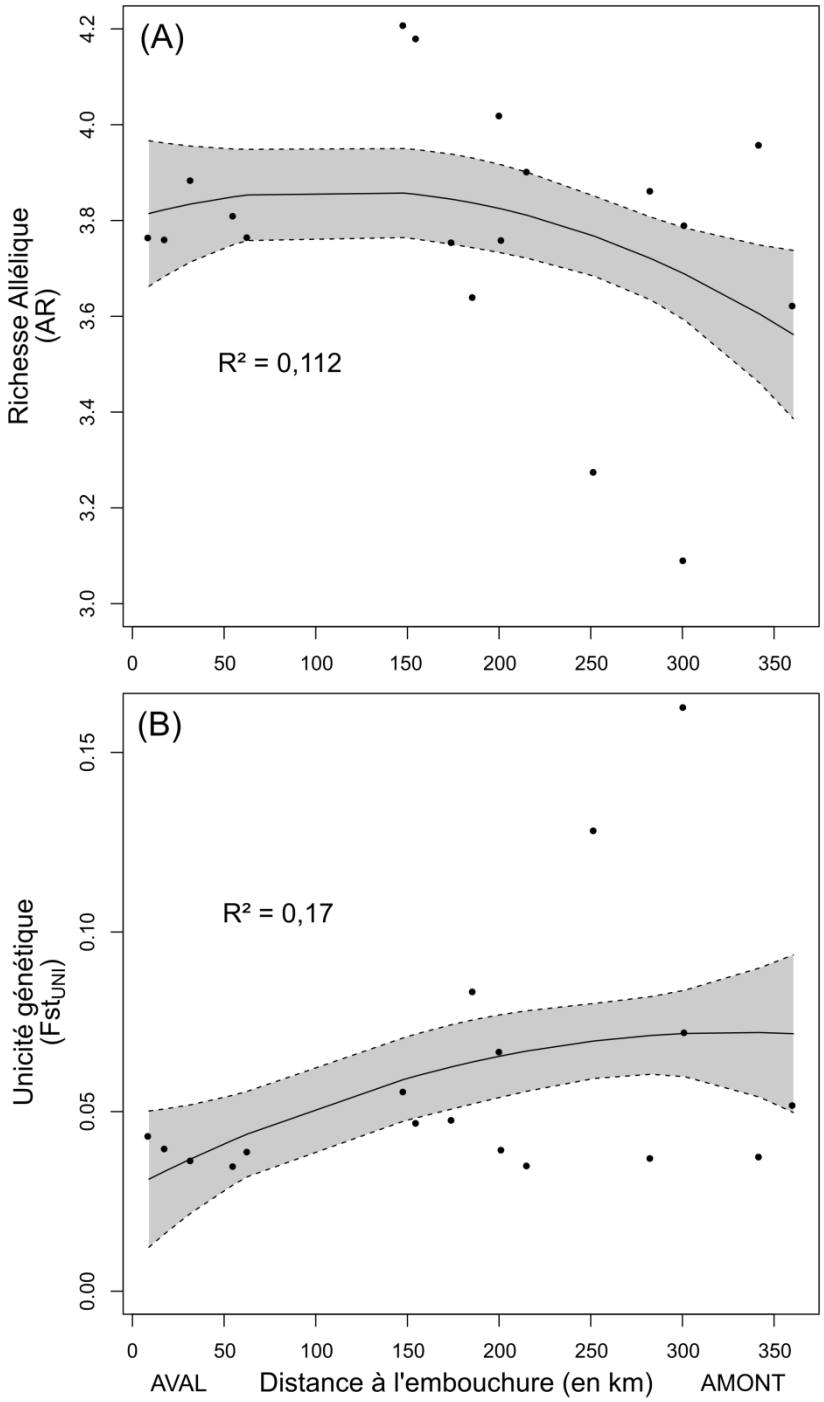


Figure 10 : Graphique représentant (A) la richesse allélique AR et (B) l'unicité génétique Fst_{UNI} calculées pour chacune des 18 populations génotypées dans le bassin de la Garonne en fonction de la distance de chacune de ces populations par rapport à l'embouchure (points noirs). La distance à l'embouchure est une distance topologique (en suivant le cours d'eau). Les courbes noires correspondent aux valeurs prédites par des modèles linéaires généralisés (« glm ») et visant à expliquer les valeurs de AR (A) et de Fst_{UNI} (B) (variables à expliquer) en fonction de la distance à l'embouchure (variable explicative). Les courbes noires en pointillés représentent l'intervalle de confiance à 95% des prédictions du modèle glm. La valeur de R^2 indique la qualité de la prédiction de la régression élaborée par le modèle glm.

Ainsi, cette analyse nous montre qu'**il n'y a pas de patron de DIGD significatif, bien qu'il existe une petite tendance vers l'accumulation de diversité génétique vers l'aval pour le silure dans le bassin de la Garonne** (Figure 10 A). Deux points (vers 250 et 300 km de l'embouchure ; Figure 10 A) se détachent fortement des prédictions : ce sont les points correspondant aux populations les plus à l'amont du Lot (Prayssac et Cahors) qui sont caractérisées par les valeurs de AR les plus faibles (voir section précédente).

Nous nous sommes aussi intéressés à un deuxième patron spatial attendu dans les bassins hydrographiques pour des espèces non amphihalines et natives ou parfaitement acclimatées depuis longtemps : patron selon lequel l'unicité génétique (Fst_{UNI}) augmente le long d'un gradient aval-amont (Finn et al., 2011; Paz-Vinas & Blanchet, 2015). Contrairement au DIGD, qui prédit une augmentation de AR vers l'aval, ce patron prédit une augmentation de Fst_{UNI} vers l'amont. En effet, plus les populations sont situées vers l'amont dans un bassin hydrographique, plus ces populations sont isolées par rapport aux autres populations du bassin. Des processus évolutifs comme la dérive génétique, qui peuvent avoir des effets différentiels selon les populations, vont se mettre en œuvre pour rendre, avec le temps, les populations situées à l'amont de plus en plus uniques sur le plan génétique par rapport aux autres.

Comme dans le cas du DIGD, nous avons représenté graphiquement les valeurs de Fst_{UNI} calculées pour chacune des 18 populations génotypées en fonction de leur distance à l'embouchure (Figure 10 B). De même, nous avons représenté sur le graphique la courbe de régression de Poisson d'un modèle linéaire généralisé construit en considérant les valeurs de Fst_{UNI} comme variable à expliquer, et la distance à l'embouchure comme variable explicative (en considérant un terme quadratique).

Ainsi, cette analyse nous montre qu'**il n'existe pas de patron d'augmentation de l'unicité génétique vers l'amont significatif pour le silure dans le bassin de la Garonne**, **bien qu'une petite tendance vers cette augmentation semble se dégager** (Figure 10 B). Comme dans l'analyse précédente, deux points se détachent fortement des valeurs prédites par le modèle glm : ce sont aussi les points correspondant aux populations les plus à l'amont du Lot (Prayssac et Cahors, à environ 250 et 300 km de l'embouchure ; Figure 10 B) qui sont caractérisées par les valeurs d'unicité génétique les plus fortes (voir section précédente).

Finalement, nous nous sommes intéressés à un troisième patron spatial récurrent et général en génétique des populations : patron selon lequel la différenciation génétique entre deux populations augmente lorsque la distance entre ces deux populations augmente. Ce patron se prénomme « patron d'isolement par la distance », ou « patron IBD » (Sexton et al., 2014; Wright, 1943). Ce patron s'analyse souvent moyennant une régression entre des distances géographiques séparant des paires de populations, et des valeurs de différenciation génétiques calculées entre ces paires de populations (Rousset, 1997; Wright, 1943). Il s'agit d'un patron de diversité génétique qui a été vérifié pour un très grand nombre de taxons et dans plusieurs écosystèmes (aquatiques et terrestres).

Ainsi, nous avons représenté graphiquement la différenciation génétique entre paires de populations échantillonnées en fonction de la distance topologique (en suivant le

cours d'eau) séparant chaque paire de populations (Figure 11). Plutôt que de considérer les valeurs brutes de Fst par paires de populations calculées dans la section précédente comme indice de différenciation génétique entre paires de populations, nous avons appliqué (pour chaque paire de populations) la formule $\frac{\text{Fst}}{(1-\text{Fst})}$, comme préconisé par Rousset (1997), pour des espèces vivant dans des environnements linéaires comme les cours d'eau. Nous avons aussi représenté sur le graphique la courbe de régression d'un modèle linéaire généralisé construit en considérant les valeurs de différenciation génétique comme variable à expliquer, et la distance topologique entre paires de populations comme variable explicative. Nous n'avons pas considéré de terme quadratique comme dans les glm précédents, afin d'obtenir une droite de régression (plutôt qu'une courbe) qui nous permette d'extraire une valeur de pente. La valeur de pente d'une droite de régression d'un patron IBD informe sur les capacités relatives de dispersion d'une espèce (Comte & Olden, 2018). Cette valeur (reportée dans la Figure 11) nous sera très utile dans la Partie III pour estimer les abondances et les densités de silure. Finalement, nous avons aussi testé la significativité du patron d'IBD, en effectuant un test de Mantel entre les valeurs de différenciation génétique entre paires de population et la distance entre ces populations (Figure 11). Les distances topologiques séparant chaque paire de populations ont été calculées grâce au paquet 'riverdist' du logiciel R (Tyers, 2017).

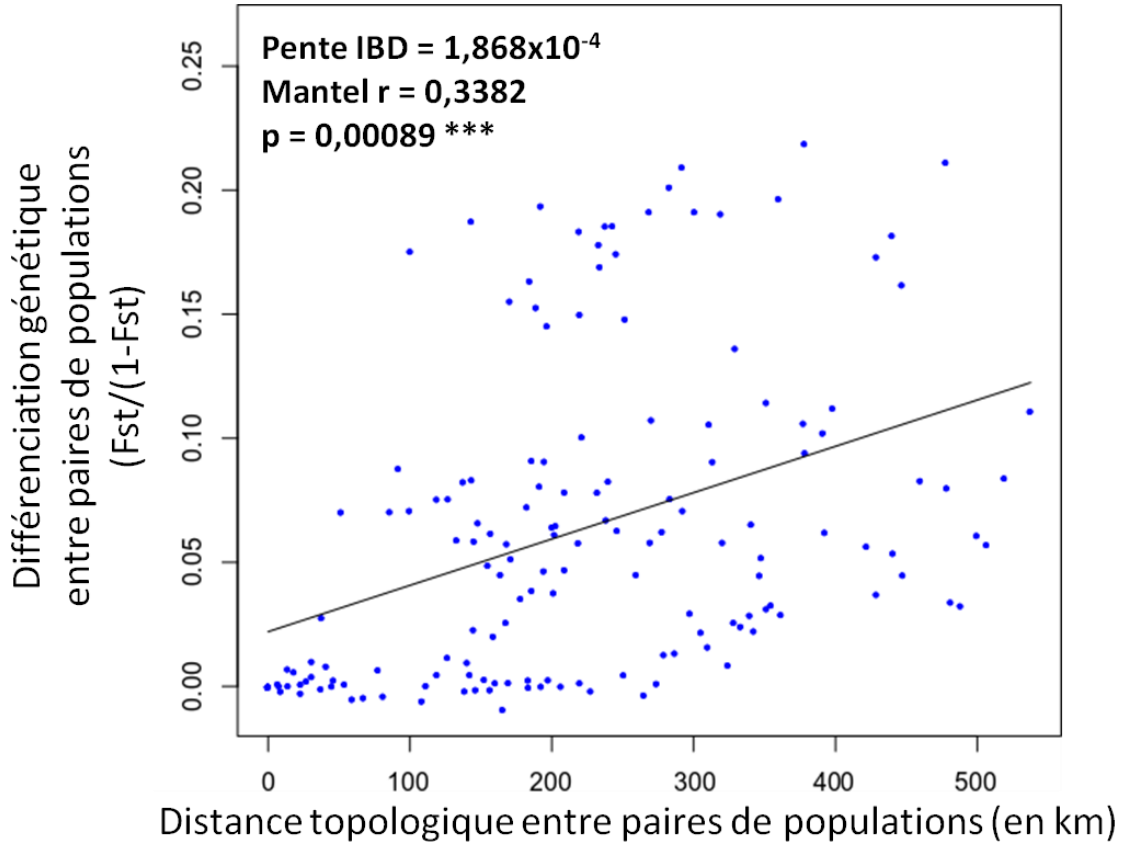


Figure 11 : Graphique représentant la différenciation génétique entre paires de populations échantillonnées ($\frac{\text{Fst}}{(1-\text{Fst})}$) en fonction de la distance topologique (en suivant le cours d'eau) séparant chaque paire de populations (points bleues). La droite noire correspond aux valeurs prédites par un modèle linéaire généralisé (« glm ») visant à

expliquer les valeurs de différenciation génétique entre paires de populations (variable à expliquer) en fonction de la distance entre paires de sites (variable explicative). La valeur de r du test de Mantel mesure la corrélation entre les deux matrices utilisées (matrices de différenciation génétique et matrice de distances topologiques). Elle peut être interprétée comme une valeur de R^2.

Notre analyse confirme la **présence d'un patron d'isolement par la distance génétique (IBD) significatif chez les populations de silure dans le bassin de la Garonne** (r du test de Mantel = 0,338 ; valeur de p = 0,00089 ; Figure 11). La valeur de la pente de la droite de régression du modèle glm (pente IBD = 1,868 x 10^{-4}) pourra donc être utilisée pour renseigner la variance dans les distances de dispersion parcourues par les silures (paramètre σ^2) lors du calcul des tailles efficaces des métapopulations de silures du bassin de la Garonne (voir Partie III ci-après).

Globalement, l'analyse des patrons spatiaux de diversité et de différenciation génétique des populations de silure du bassin de la Garonne nous indiquent :

- que les valeurs de richesse allélique et d'unicité génétique calculées pour ces populations ne forment pas les patrons spatiaux attendus pour des espèces natives ou parfaitement acclimatées depuis longtemps (ex. patron DIGD), bien qu'une tendance vers ce type de patrons semble se profiler,
- que la structuration génétique semble en accord avec un mode de dispersion de type « isolement par la distance »,
- que les populations en amont du Lot (Prayssac et Cahors) semblent particulièrement différentes par rapport à toutes les autres populations du bassin.

Au vu de ces résultats, **nous pouvons émettre l'hypothèse que le silure n'est pas encore totalement acclimaté au bassin de la Garonne, étant donné que tous les patrons spatiaux de diversité attendus ne sont pas vérifiés**. Par contre, cette espèce serait peut-être en voie de correspondre à ce type de patrons. Il faudra probablement attendre quelques années (voire décennies) pour voir si la dynamique des populations de silure du bassin de la Garonne permet de générer tous les patrons spatiaux génétiques attendus.

Le décalage entre les résultats suggérés par les analyses démographiques (qui tendent à montrer une acclimatation presque aboutie de l'espèce dans le bassin), et ceux suggérés par les résultats des analyses génétiques (qui tendent à montrer que cette acclimatation n'est pas encore aboutie) est probablement dû à l'interaction de plusieurs facteurs. Certains facteurs sont intrinsèques à la biologie de l'espèce, comme le temps de génération relativement long du silure (environ 3 ans) par rapport à d'autres espèces de poissons, sa longévité (son espérance de vie est de 60 ans) ou comme le chevauchement des générations de silure (il est fort probable que des individus adultes issus de la première introduction en 1983 dans le Tarn soient encore vivants en 2018). D'autres facteurs sont extrinsèques, comme par exemple la présence de barrières naturelles et/ou artificielles (ex. les barrages ou les chaussées) ou le fait que plusieurs introductions issues de « pools » génétiques différents aient eu lieu dans le bassin (perturbant ainsi les

signaux véhiculés par les analyses génétiques). L'interaction de tous ces facteurs peut donc rallonger le délai naturel qui existe entre les causes écologiques qui affectent la dynamique des populations et leurs conséquences génétiques (Landguth et al., 2010).

Description de la structure génétique du silure dans le bassin de la Garonne.

Nous avons utilisé la méthode de factorisation matricielle non-négative creuse (ou « snmf » pour « sparse non-negative matrix factorization ») implémentée dans le paquet du logiciel R 'LEA' (Frichot & François, 2015) pour identifier des groupes de silure génétiquement homogènes dans le bassin de la Garonne et mettre en évidence la structure génétique fine des populations de silure de ce bassin. Cette méthode permet de déterminer le nombre K le plus probable de groupes génétiquement homogènes (ou « clusters » génétiques) contenus dans un jeu de données génétiques sans utiliser d'information « à priori » sur l'origine des individus. Une fois que le bon nombre de clusters K est déterminé, la méthode assigne chaque individu du jeu de données à ces clusters en fonction de leur proportion d'ascendance génétique envers ces clusters.

Nous avons donc utilisé la fonction « snmf » du paquet 'LEA' pour tester des valeurs de K allant de K=1 à K=20. Pour chaque valeur de K, les analyses ont été répliquées 1000 fois. Nous avons ensuite déterminé quelle valeur de K et quel réplicat (parmi nos 20 000 réplicats) expliquent le mieux nos données génotypiques, au travers du critère d'entropie implémenté dans 'LEA'.

Dans une situation où il n'y aurait eu qu'un seul gros évènement d'introduction de silures dans le bassin de la Garonne (donc que des individus génétiquement similaires tous issus de la même population source), l'analyse devrait détecter un signal homogène, c'est-à-dire une valeur de K égale à 1. D'autant qu'il n'y aurait pas eu assez de générations écoulées depuis l'arrivée de l'espèce dans le bassin (11 générations en considérant un temps de génération de 3 ans) pour générer suffisamment de différenciation génétique menant à la détection de plusieurs clusters.

Or, nous avons trouvé un nombre de clusters égal à 4 (K=4 ; cf. les quatre couleurs dans la Figure 12). Ces quatre clusters représentent quatre groupes génétiquement homogènes. Le premier cluster, en vert sur la Figure 12, est composé d'individus ne se retrouvant quasi exclusivement que dans les populations de la Dordogne (individus avec une proportion d'ascendance au cluster 1 vert supérieure à 0,8 ; Figure 12). Un deuxième cluster (le cluster 4 en bleu sur la Figure 12) n'est composé quasi exclusivement que d'individus présents sur le Lot à Cahors. Les populations du Lot à l'aval de Cahors, du Tarn et de la Garonne sont composées d'un mélange d'individus issus des deux autres clusters (cluster 2 en jaune et cluster 3 en rouge ; Figure 12), avec une majorité d'individus issus du cluster 3 rouge dans les populations du Lot à l'aval de Cahors.

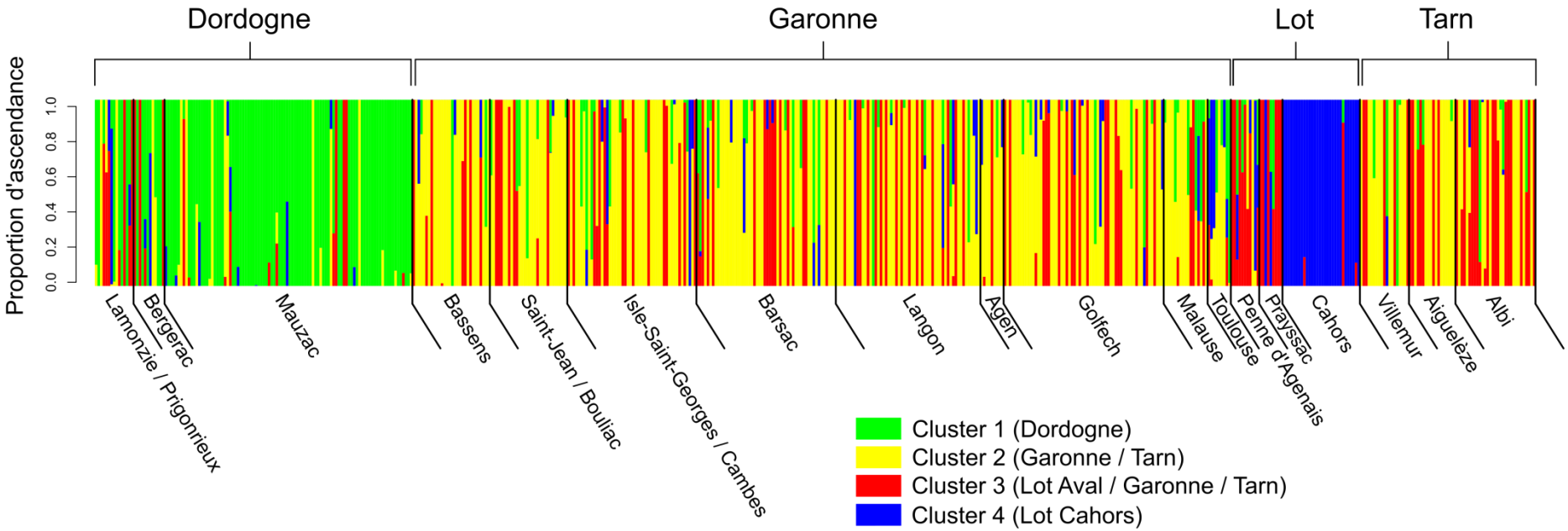


Figure 12 : Diagramme représentant les proportions d'appartenance (ou d'ascendance) de chaque individu aux K=4 différents « clusters » génétiques (ou groupes d'individus génétiquement homogènes) déterminées par les analyses effectuées avec 'LEA'. Chaque barre verticale correspond à un individu, et chaque couleur correspond à un « cluster » différent. Des individus peuvent avoir des ascendances diverses, selon l'appartenance génétique à des clusters différents de leurs géniteurs. Nous considérons ici qu'un individu appartient totalement à un cluster lorsqu'il a une proportion d'ascendance pour ce cluster supérieure ou égale à 0,8. Les individus sont classés par rivières (délimitées par des accolades en haut du diagramme) et par populations (délimitées par des barres verticales noires et nommées par leur localisation en bas du diagramme). Pour chaque rivière, les populations sont classées de la plus à l'aval vers la plus à l'amont.

Globalement, cette analyse nous apprend qu'**il existe une forte structuration génétique des populations de silure dans le bassin de la Garonne**, malgré la colonisation récente du bassin par cette espèce et le peu de générations qui se sont écoulées depuis le premier évènement d'introduction connu dans le bassin. Cette structuration génétique se caractérise par l'**existence de deux groupes d'individus génétiquement homogènes bien définis spatialement (un groupe occupant la Dordogne et l'autre occupant le Lot à Cahors), et par la présence d'un mélange d'individus issus de deux groupes génétiquement différents dans la Garonne, le Tarn et à l'aval du Lot.**

Ces résultats nous permettent de formuler l'hypothèse suivante : **les populations de silure du bassin de la Garonne seraient issues d'au moins quatre introductions importantes de silure, indépendantes et provenant de quatre sources génétiques (ou « populations sources ») différentes.** Néanmoins, il nous est impossible, en l'état, de déterminer l'origine de ces populations sources, étant donné que nous ne possédons pas d'échantillons provenant des populations d'origine potentielles (par exemple, des diverses pisciculture ou des populations natives de l'espèce desquelles proviennent les silures introduits). Ces résultats restent toutefois cohérents avec les informations recueillies en Partie I concernant les premiers déversements connus dans le bassin, à savoir deux grosses introductions dans le Tarn et dans la Dordogne (qui pourraient être issues de deux populations sources génétiquement différentes), et des déversements dans le Lot provenant d'au moins deux pisciculture (dont les individus seraient eux-mêmes issus de deux populations d'origine différentes sur le plan génétique).

Pour finir, notons sur la Figure 12 que des individus appartenant à certains clusters bien définis spatialement (ex. cluster 1 vert pour la Dordogne et cluster 4 bleu pour le Lot à Cahors) ou des descendants d'individus issus de ces clusters, peuvent se retrouver loin de leur zone d'appartenance principale. Ceci peut traduire des mouvements de dispersion de ces individus dans le bassin. Il est cependant difficile de savoir si ces individus sont vraiment des migrants, ou bien s'ils sont issus de déversements ou de translocations d'origine anthropique, étant donné le fort intérêt que cette espèce a suscité ces dernières années (voir Partie I).

Partie III : Estimation de l'abondance et des densités de silure à partir de données génétiques.

« Counting fish is like counting trees, except they are invisible and they keep moving. »

Prof. John Shepherd (dans Hilborn, 2002)

L'estimation de l'abondance et de la densité d'une espèce dans un environnement donné n'est pas une tâche aisée, surtout lorsqu'il s'agit d'une espèce de poisson de rivière. La citation ouvrant cette partie illustre parfaitement la difficulté inhérente à l'estimation des abondances d'espèces de poissons par rapport à d'autres espèces terrestres, plus visibles et moins mobiles, comme les plantes, espèces pour lesquelles des méthodes d'estimation d'abondance directes (ex. comptages) peuvent s'appliquer.

En rivière, lorsque les conditions locales sur un site le permettent, il est possible d'estimer l'abondance et la densité d'une espèce de poisson à partir d'un protocole de pêche électrique standardisée (comme ceux utilisés lors des pêches AFB sur des petits cours d'eau, cf. Partie I). Cependant, sur les grands cours d'eau comme ceux qui nous intéressent dans cette étude (Garonne, Dordogne, Tarn et Lot), les méthodes de pêche électrique ne permettent pas de donner des estimations satisfaisantes, étant donné qu'il est impossible de quadriller parfaitement la station de pêche. De même, ces estimations peuvent varier fortement d'une année à l'autre, et ne sont représentatives que de la présence et de l'abondance relative de l'espèce sur le site. Toute extrapolation à l'ensemble de la rivière des estimations d'abondance et de densité obtenues par pêche électrique pour une ou quelques stations présentes sur celle-ci serait alors hasardeuse, car elle pourrait conduire à de fortes sur- ou sous-évaluations du stock de poisson.

Les outils génétiques peuvent aider à résoudre ce type de problématiques. En effet, la diversité génétique mesurée pour une population donnée en utilisant des marqueurs génétiques neutres (marqueurs n'étant pas soumis aux effets de la sélection naturelle comme par exemple les microsatellites utilisés dans cette étude) est corrélée de façon positive avec sa taille efficace (N_e ; voir description dans la section suivante ; Nei, 1987). Cette taille efficace étant elle-même corrélée avec l'abondance de cette espèce (N_c ; voir section suivante), il est théoriquement envisageable d'estimer N_c à partir de données génétiques neutres (Frankham, 1996; Nei, 1987).

Cependant, d'autres facteurs influencent fortement la diversité génétique observée dans une population, ce qui peut biaiser les estimations de N_e et, par conséquent, les estimations de N_c. En effet, la diversité génétique d'une population est aussi le résultat de diverses forces et processus qui laissent de fortes empreintes génétiques comme par exemple les introductions, les changements démographiques, les patrons de migration particuliers ou l'admixture (c'est-à-dire, le mélange de deux populations différentes sur le plan génétique). Il est donc très important de pouvoir prendre en compte ces facteurs lors de l'estimation des tailles efficaces des populations, notamment pour des populations introduites et ayant subi une forte phase d'expansion comme celles du silure dans le bassin de la Garonne.

1. Raisonnement soutenant la méthodologie d'estimation d'abondances et de densités proposée

Pour pouvoir estimer avec précision l'abondance ou stock d'une espèce à partir de données génétiques, il faut :

- **Obtenir des estimations de tailles efficaces des populations (N_e) d'intérêt** le plus précises possible. La taille efficace d'une population est un paramètre très utilisé en génétique de la conservation. Elle diffère de la taille réelle d'une population (ou N_c pour « census size »). Si nous considérons *X* une population pour laquelle nous voulons estimer la taille efficace, cette taille efficace correspondra au nombre d'individus reproducteurs attendu dans une population idéale pour laquelle la dérive génétique (une force évolutive agissant sur la diversité génétique, comme les mutations ou le flux de gènes) agirait avec la même intensité que pour *X*. La taille efficace d'une population est généralement plus petite que la taille réelle d'une population (Frankham, 2005; Palstra & Fraser, 2012).

- **Calculer le ou les valeurs de méta-N_e**, c'est-à-dire la taille efficace de ou des métapopulations d'intérêt (Gomez-Uchida et al., 2013). Une métapopulation est un ensemble de populations séparées spatialement ou temporellement, et étant interconnectées par la dispersion (Ricklefs & Miller, 2005). En outre, la taille efficace totale d'une métapopulation peut différer de la somme des tailles efficaces de toutes les populations la composant (Gomez-Uchida et al., 2013).

- **Connaître le rapport N_c / N_e,** c'est-à-dire le rapport existant entre les tailles réelles des populations N_c et leurs tailles efficaces N_e (Palstra & Fraser, 2012). Ce rapport est propre à chaque espèce (voire population dans certains cas), et dépend de plusieurs facteurs liés à l'écologie et à la biologie d'une espèce, comme par exemple sa position dans la chaîne alimentaire.

- **Appliquer le ratio N_c / N_e aux valeurs de méta-N_e pour obtenir la valeur de méta-N_c,** c'est à dire l'abondance de l'espèce dans la métapopulation considérée.

Une fois que la valeur de **méta-N_c** d'une espèce à l'échelle d'une rivière est **connue**, **il est possible de calculer la densité moyenne** de cette espèce dans la rivière. Pour cela, il suffit de calculer d'abord la surface totale occupée par l'espèce dans cette rivière, puis de diviser méta-N_c par cette surface pour obtenir la densité d'individus par unité de surface.

Nous allons décrire en détail ces quatre phases dans les sections suivantes.

2. Estimation des tailles efficaces (Ne) des populations étudiées

Précédemment, nous avons mentionné que la taille efficace N_e d'une population est à la fois corrélée de façon positive avec sa diversité génétique (Nei, 1987), mais est aussi fortement liée à d'autres facteurs qui génèrent de fortes empreintes génétiques comme par exemple les effets fondateurs liés aux introductions, les changements démographiques (expansions ou contractions démographiques) ou l'admixture entre

deux populations. Les méthodes d'estimation de N_e devraient pouvoir tenir compte de ces nombreux facteurs dans leurs calculs or, c'est rarement le cas.

Les méthodes les plus couramment utilisées par les généticiens des populations pour estimer les N_e ou les changements de N_e au cours du temps posent des hypothèses de départ très contraignantes (Paz-Vinas et al., 2013). En effet, ces méthodes considèrent souvent que les populations analysées sont conformes au modèle de population idéal de Wright-Fisher (Fisher, 1922; Wright, 1931), c'est-à-dire qu'elles sont à l'équilibre démographique, isolées, n'ont pas de sous-structuration et qu'elles ne reçoivent ni ne fournissent aucun migrant. Ces hypothèses sont rarement vérifiées dans les populations naturelles, et elles le sont encore moins dans les populations introduites comme celles qui nous intéressent dans notre étude.

Ainsi, les méthodes d'estimation classiques, comme par exemple la méthode dite « linkage disequilibrium method » implémentée dans les logiciels NeEstimator (Do et al., 2014) et LDNe (Waples, 2006; Waples & Do, 2008), bien que très populaires (notamment du fait de leur facilité et de leur rapidité d'utilisation), s'avèrent inadaptées à notre étude. De plus, ces méthodes ont tendance à donner des estimations de N_e soit négatives, soit infinies, ce qui empêche tout calcul de N_c postérieur (Paz-Vinas et al., 2013; Waples & Do, 2008).

Afin d'obtenir des estimations de N_e robustes qui limiteraient tous les biais liés à la structure des populations et à leur histoire démographique, nous avons adopté une approche computationnelle basée sur des simulations numériques. Plus concrètement nous avons utilisé des algorithmes de calcul Bayésien approché (« approximate Bayesian computations » ; Beaumont et al., 2002) couplés à des algorithmes d'apprentissage automatique (« random forests » ou forêts aléatoires ; Breiman, 2001; Pudlo et al., 2016) pour simuler des données génétiques sous divers modèles démographiques et évolutifs. Ces modèles prennent en compte toutes les informations que nous avons obtenues grâce aux analyses des Parties I et II et qui concernent les populations de silure du bassin de la Garonne, à savoir : (i) la structure génétique et spatiale des diverses populations de silure étudiées, (ii) l'histoire d'introduction du silure dans le bassin, et (iii) l'expansion démographique subie par le silure depuis son introduction dans le bassin.

Cette approche permet de déterminer le ou lesquels des modèles considérés expliquent le mieux les patrons de diversité et de structuration génétiques observés pour le silure dans le bassin de la Garonne. Par la suite, nous nous réfèrerons à ces méthodes sous le terme de « méthodes ABC ».

Tout d'abord, nous allons donc décrire le principe général des méthodes ABC. Ensuite, nous détaillerons les divers modèles démographiques et évolutifs que nous avons considérés, ainsi que la procédure de simulation utilisée pour simuler les données génétiques. Enfin, nous décrirons la procédure de choix de modèles qui nous a permis de sélectionner le modèle le plus probable, c'est-à-dire celui qui approche le mieux la réalité observée sur le terrain (ou, en d'autres termes, celui qui génère les données génétiques simulées les plus semblables à celles observées dans la réalité). Finalement, nous

décrirons comment nous avons utilisé ce meilleur modèle pour estimer les tailles efficaces des populations de silure du bassin de la Garonne.

Principe de la méthode.

Les algorithmes de type « ABC » (Beaumont et al., 2002) permettent de comparer des données génétiques simulées sous des modèles démographiques et évolutifs complexes, avec des données observées réelles (dans notre cas, les données génétiques utilisées pour décrire la diversité et la structuration génétique présentées en Partie II de ce rapport). Cette comparaison se fait indirectement par le biais du calcul de statistiques descriptives de la diversité génétique, c'est-à-dire des statistiques qui résument l'information génétique contenue dans chaque jeu de données génétique simulé. Dans sa version la plus générique, un algorithme ABC est composé de cinq étapes :

(i) tirage aléatoire des valeurs des paramètres (par exemple, les tailles efficaces des populations ou les taux de migration entre deux populations) d'un modèle démographique donné dans des distributions de valeurs définies « à priori », et couvrant l'ensemble des possibles pour chaque paramètre,
(ii) simulation d'un jeu de données génétiques sous ce modèle en considérant les valeurs de paramètres choisies en (i),
(iii) calcul des statistiques descriptives du jeu de données simulé,
(iv) répétition des trois premières étapes *x* fois, puis
(v) calcul des distances entre les statistiques descriptives calculées pour les données simulées et celles calculées pour les données réelles.

L'application des algorithmes ABC nécessite donc une grande puissance de calcul, car les approximations sont faites au travers de centaines de milliers de jeux de données génétiques simulés selon le cas d'étude. Un grand nombre de simulations permet de faire une exploration approfondie des gammes de valeurs des différents paramètres d'intérêt (ex. les tailles efficaces de nos populations) définis dans le point (i) décrit ci-haut. Pour réaliser ces calculs, nous avons bénéficié des ressources computationnelles offertes par le supercalculateur « Eos » du Mésocentre CALMIP (Calcul en Midi-Pyrénées ; UMS CNRS 3667 ; allocation de ressources T18005). Ainsi, nous avons pu bénéficier de plusieurs milliers d'heures de calcul intensif pour mener à terme nos simulations.

Une fois toutes les simulations générées pour tous les modèles considérés, et toutes les statistiques descriptives calculées, le cadre analytique de l'ABC nous permet :

- de déterminer « à posteriori » quel modèle parmi tous les modèles considérés est le meilleur (c'est-à-dire celui qui explique le mieux les données réelles), et
- d'inférer les valeurs de paramètres démographiques et génétiques d'intérêt, comme les tailles efficaces des populations, en comparant les données réelles et les meilleurs jeux de données simulés sous le meilleur modèle.

Définition des modèles utilisés pour simuler les données génétiques.

En nous basant sur les résultats obtenus dans les Parties I et II, nous avons considéré deux modèles démographiques et évolutifs différents (Figure 13), ainsi que deux variantes supplémentaires de ces deux modèles (en fonction des modalités de dispersion des individus entre les populations, voir description des modèles ici-bas). L'objectif de ces quatre modèles est donc de simuler des jeux de données génétiques pouvant expliquer au mieux la distribution actuelle de la diversité génétique du silure dans le bassin de la Garonne (Figure 13). Ceci nous permet ensuite d'estimer avec précision les N_e de chaque population, tout en limitant au maximum les biais liés à leur structure et à leur histoire démographique.

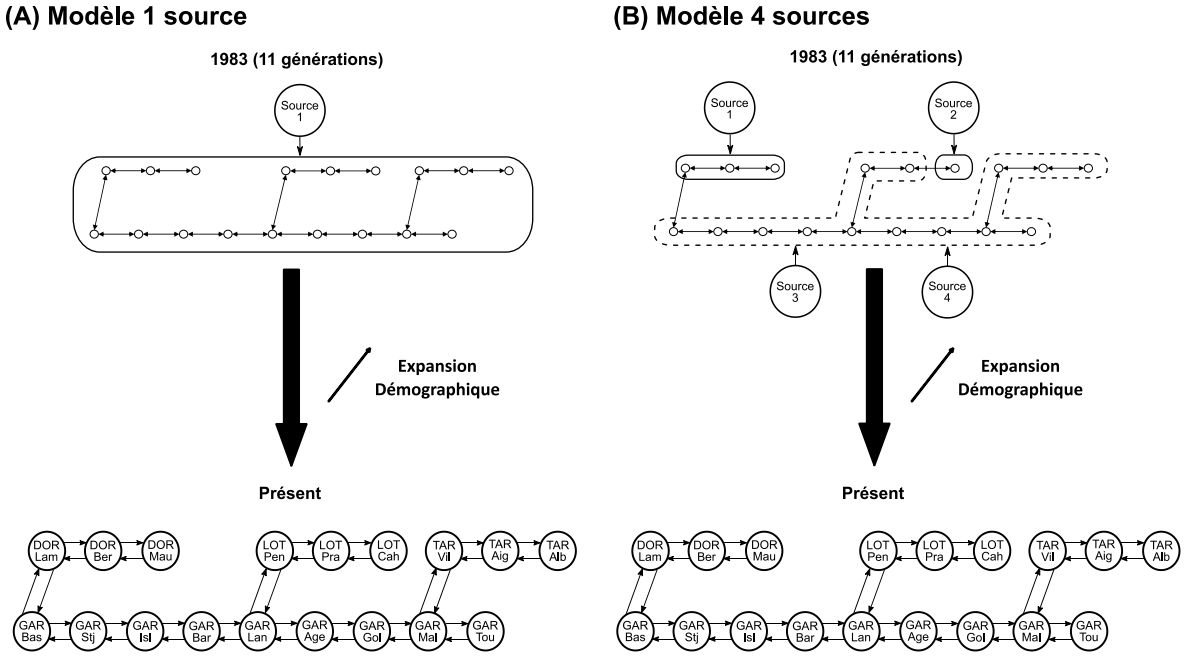


Figure 13 : Représentation schématique des modèles démographiques et évolutifs considérés pour l'analyse ABC. Les grands cercles (un pour le « modèle 1 source » en A et 4 pour le « modèle 4 sources » en B) représentent les populations sources à l'origine des populations de silure introduites dans le bassin de la Garonne. Les 18 petits cercles en haut des figures représentent ces 18 populations après leur introduction. Les 18 cercles moyens représentent ces mêmes populations actuellement, après avoir subi une phase d'expansion démographique. Les petites flèches noires entre paires de populations représentent les taux de migration par génération entre paires de populations adjacentes. Ces taux de migrations peuvent être symétriques ou asymétriques selon la variante du modèle considéré (voir la description des modèles dans le texte principal pour plus de détails).

Pour le premier modèle (Modèle 1 source ; Figure 13 A), nous avons considéré que les 18 populations décrites en Partie II sont toutes originaires d'introductions ayant eu lieu il y a au moins 11 générations (en 1983, si nous considérons un temps de génération de 3 ans pour le silure), et impliquant des individus provenant d'une même population source d'origine. Suite à ces introductions, de petites populations ont été crées (représentées par de petits cercles sur la Figure 13) ; elles vont subir une phase d'expansion similaire à celle décrite en Partie I, jusqu'à atteindre leur taille efficace actuelle (paramètres représentés par les noms de code des populations échantillonnées,

reportés dans les cercles moyens sur la Figure 13). Au cours du temps, ces populations ont aussi pu échanger des migrants suivant un modèle de migration par pas linéaire (ou « stepping stone » linéaire ; Kimura & Weiss, 1964; Paz-Vinas et al., 2015, 2013), modèle selon lequel les échanges (représentés par des flèches à double sens sur la Figure 13) ne peuvent se faire qu'entre paires de populations adjacentes. Ce modèle à 1 source a été décliné en deux versions : une première version qui considère que la migration entre paires de populations se fait de façon symétrique (autant de migrants par génération vers l'aval que vers l'amont), et une seconde version qui considère que la migration entre paires de populations se fait de façon asymétrique et est biaisée vers l'aval (avec plus de migrants par génération vers l'aval que vers l'amont ; Paz-Vinas et al., 2015, 2013). Ces deux modèles, prénommés modèles « ONE SOURCE SYM » et « ONE SOURCE ASYM » respectivement, peuvent être considérés comme des « modèles simplifiés » de colonisation et d'expansion de l'espèce dans le bassin de la Garonne.

Le modèle 4 sources (Figure 13 B) est similaire au modèle 1 source à une différence près. Au lieu de considérer que toutes les populations de silure du bassin de la Garonne sont issues d'une seule et unique population source, ce modèle considère quatre populations sources différentes. La population source 1 (Figure 13 B) serait à l'origine des individus introduits en Dordogne ; la population source 2 serait à l'origine des individus introduits dans le Lot à Cahors. Les populations sources 3 et 4 seraient à l'origine des individus introduits dans la Garonne, le Tarn et à l'aval du Lot (Figure 13 B). Ce modèle a été bâti en accord avec les résultats obtenus dans la Partie II concernant la structuration génétique des populations de silure dans le bassin. Comme pour le modèle 1 source, nous avons décliné ce modèle en deux versions : une version qui considère que la migration entre paires de populations se fait de façon symétrique, et une autre qui considère que la migration entre paires de populations se fait de façon asymétrique et est biaisée vers l'aval. Ces deux modèles seront, prénommés « FOUR SOURCES SYM » et « FOUR SOURCES ASYM » respectivement dans la suite du rapport.

Procédure de simulation.

Nous avons simulé des jeux de données génétiques sous les quatre modèles décrits ci-dessus en suivant quatre étapes :

(i) échantillonnage de valeurs des paramètres pour un modèle donné dans des distributions de valeurs des paramètres définies à priori,
(ii) simulation d'un jeu de données génétique en considérant les valeurs des paramètres tirés en (i),
(iii) calcul d'un vecteur de statistiques descriptives qui résumera l'information contenue dans chaque jeu de données génétiques simulé, et
(iv) répétition des étapes (i) à (iii) 150 000 fois par modèle, afin de créer une grande table de référence.

Nous avons implémenté cette procédure dans un tube computationnel adapté de celui utilisé dans Rey *et al.* (2015), Paz-Vinas *et al.* (2015) et Paz-Vinas et Blanchet (2015) (voir Annexe 3 pour une présentation détaillée de ce tube computationnel).

Pour chaque modèle, nous avons simulé 150 000 jeux de données génétiques ; et pour chacun de ces jeux de données génétiques, nous avons simulé 10 marqueurs microsatellites indépendants évoluant sous un modèle de mutation dit « par pas généralisé » (ou « Generalized Stepwise Mutation model », ou encore GSM ; Kimmel & Chakraborty, 1996). Chaque marqueur microsatellite était caractérisé par un taux de mutation neutre moyen $MSAT_{MUTRATE}$ tiré d'une distribution de type uniforme.

Ensuite, (et pour chaque jeu de données simulé), nous avons échantillonné un nombre d'individus diploïdes identique à celui de notre jeu de données empirique (voir Annexe 3 pour plus de détails sur le nombre d'individus échantillonnés et génotypés avec succès dans les 18 populations), afin que les statistiques descriptives de la diversité génétique calculées pour nos simulations soient comparables à celles de notre jeu de données empirique.

Pour chaque jeu de données simulé, nous avons donc calculé toute une série de statistiques descriptives de la diversité génétique pour créer une table de référence. Au total, nous avons calculé jusqu'à 145 statistiques descriptives pour chaque jeu de données simulé. Par souci de clarté, nous ne détaillerons pas dans ce rapport toutes les statistiques calculées.

Choix des modèles.

Nous avons utilisé la méthode « ABC-RF » (RF pour « Random Forests » ; Pudlo et al., 2016) pour choisir le modèle (parmi les quatre modèles testés) qui explique le mieux les données réelles. Les « Random Forests » ou « Forêts Aléatoires » sont des algorithmes d'apprentissage automatique non-paramétriques qui peuvent avoir des variables de réponse de type catégorique (ou non numériques) comme, dans notre cas, l'identité des quatre modèles ayant servi à simuler les données génétiques (ONE SOURCE SYM, ONE SOURCE ASYM, FOUR SOURCES SYM et FOUR SOURCES ASYM). Dans le cadre d'une procédure de choix de modèles par ABC-RF, les variables explicatives des modèles correspondent à l'ensemble des statistiques descriptives calculées pour chaque simulation. En générant une multitude d'arbres de classification, les Forêts Aléatoires permettent d'obtenir des prédictions moyennes basées sur l'ensemble des arbres composant la forêt (Breiman, 2001; Cutler et al., 2007). L'utilisation des RF en ABC est très récente et présente un grand nombre d'avantages par rapport aux méthodes de choix de modèles ABC classiques (Paz-Vinas et al., 2015; Pudlo et al., 2016; Rougemont et al., 2016). En effet, cette méthode :

- est robuste face au bruit dans les données (par exemple, lorsque des statistiques descriptives sont corrélées entre-elles, ou bien face à l'inclusion de statistiques descriptives non informatives),

- elle s'affranchit de la phase préliminaire de sélection des statistiques descriptives qui prévalait avec les méthodes classiques, ce qui pouvait compromettre la robustesse de la procédure de choix de modèle,

- elle détermine l'importance relative des statistiques descriptives pour discriminer les modèles les uns des autres, et

- elle est moins gourmande en ressources computationnelles que les méthodes traditionnelles, car elle a besoin de moins de simulations dans la table de référence.

Nous avons donc utilisé le paquet du logiciel R 'abcrf' (Pudlo et al., 2016) pour créer une forêt composée de 2000 arbres de classification à partir des 600 000 simulations générées. Nous avons ensuite évalué la robustesse de la procédure de choix de modèles en utilisant les simulations OOB (pour « Out Of Bag », c'est-à-dire des simulations n'ayant pas servi à la construction d'un arbre dans la forêt, mais qui vont nous servir à tester la robustesse du modèle) pour construire une matrice de confusion OOB (Table 2). Une matrice de confusion permet de mesurer la qualité d'un système de classification (ici, notre forêt aléatoire). Pour la construire, nous avons considéré chaque simulation OOB (pour lesquelles nous connaissons les modèles parmi les quatre considérés qui ont permis de les générer) comme des jeux de données observées, et nous avons essayé de retrouver grâce aux prédictions de la forêt aléatoire quel est le bon modèle pour chacune de ces simulations OOB (Table 2).

Table 2 : Matrice de confusion OOB obtenue à partir d'une forêt aléatoire appliquée à l'ABC (ABC-RF) composée de 2000 arbres de classification et de 600 000 simulations (150 000 par modèle). Les nombres d'assignements corrects (simulations générées par un modèle et assignées à ce même modèle par la procédure) sont en gras. Le taux d'erreur de classification OOB de cette forêt est de 28,5%.

Modèle sous lequel les simulations OOB ont été générées :	Pourcentage de simulations OOB assignées aux modèles :			
	ONE SOURCE SYM	ONE SOURCE ASYM	FOUR SOURCES SYM	FOUR SOURCES ASYM
ONE SOURCE SYM	**57,04%**	42,93%	0%	0,03%
ONE SOURCE ASYM	44,43%	**55,53%**	0%	0,04%
FOUR SOURCES SYM	0%	0%	**86,2%**	13,8%
FOUR SOURCES ASYM	0%	0%	12,6%	**87,4%**

Cette analyse montre que plus de la moitié des simulations OOB sont correctement assignées par la forêt aléatoire au modèle qui a permis de les générer (en gras dans la Table 2). Cependant, il y a un nombre non négligeable de simulations mal assignées pour les modèles ONE SOURCE SYM et ONE SOURCE ASYM, ce qui leur confère des taux d'erreur de classification entre eux d'environ 42-44% (contre environ 13% pour les modèles FOUR SOURCES SYM et FOUR SOURCES ASYM entre eux). Ce nombre élevé de mauvais assignements entre ONE SOURCE SYM et ONE SOURCE ASYM est très probablement dû à la grande similarité entre ces deux modèles, qui ne diffèrent que par leurs modalités de dispersion (Figure 13). En effet, ces modalités de dispersion ne doivent pas générer beaucoup de différences génétiques lorsque tous les individus du bassin sont issus d'une même population génétique, ceci à cause de leur fort apparentement. Nonobstant, le nombre global de mauvaises classifications pour les quatre modèles reste bas par rapport aux bonnes classifications ce qui permet d'assurer une prédiction RF robuste. D'ailleurs, la forêt aléatoire discrimine quasiment à la perfection les modèles

FOUR SOURCES et ONE SOURCE (taux d'erreur entre 0 et 0,04%). Nous avons finalement extrait un taux d'erreur de classification OOB de notre RF de 28,5%, ce qui équivaut aux taux d'erreur « à priori » estimés par les procédures classiques de choix de modèles ABC (I. Paz-Vinas et al., 2015; Pudlo et al., 2016).

Finalement, nous avons utilisé notre « Forêt Aléatoire » pour prédire, à partir de nos données observées, **le modèle qui explique le mieux les patrons de diversité et de structuration génétique observés pour le silure dans le bassin de la Garonne.** Cette procédure ABC-RF nous a indiqué que **le modèle FOUR SOURCES SYM est le meilleur des quatre modèles,** avec une probabilité postérieure estimée à 46,8%.

Ainsi, nous considérerons ce modèle démographique et évolutif pour la phase suivante, qui correspond à l'estimation des valeurs des tailles efficaces des populations de silure à partir de nos données observées.

Estimation des tailles efficaces des populations de silure.

Nous avons utilisé une approche de type ABC basée sur des réseaux de neurones pour estimer les tailles efficaces des populations de silure à partir de la comparaison entre les simulations obtenues avec le modèle FOUR SOURCES SYM et nos données réelles (Csilléry et al., 2012). Cette méthode est implémentée dans le paquet 'abc' du logiciel R (Csilléry et al., 2012). Afin d'obtenir les estimations de taille efficaces les plus précises possibles, nous avons généré 850 000 simulations supplémentaires, qui se sont additionnées aux 150 000 déjà utilisées pour le choix du meilleur modèle. Cette analyse sera basée sur ce million de simulations. La Figure 14 récapitule les résultats obtenus pour chaque paramètre. Ces estimations de valeurs de N_e vont être utilisées dans la section suivante pour estimer les valeurs de méta-N_e des quatre rivières d'intérêt.

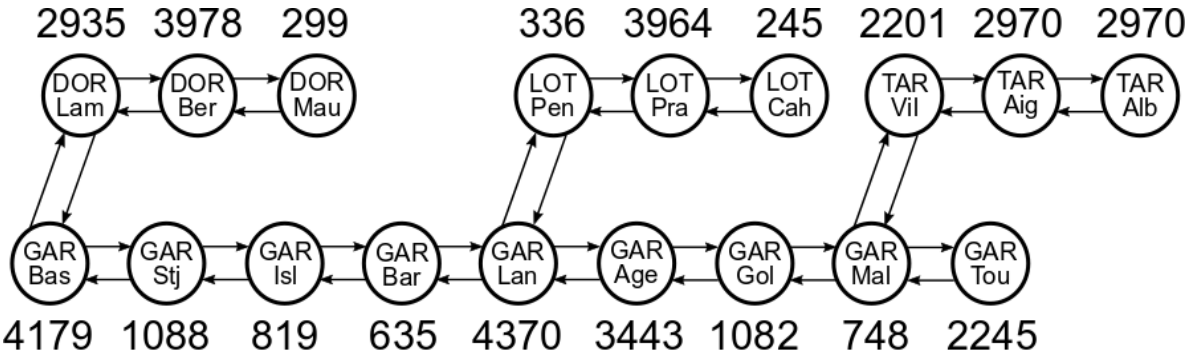


Figure 14 : Estimations des tailles efficaces (N_e) des populations de silure du bassin de la Garonne obtenues aux travers de calculs Bayésiens approchés en utilisant des simulations générées avec le modèle FOUR SOURCES SYM. Chaque cercle correspond à une des 18 populations ; populations auxquelles sont assignées les valeurs de N_e qui ont été estimées.

3. Calcul des valeurs de méta-N_e pour les quatre rivières étudiées

Méta-N_e correspond à la taille efficace d'une métapopulation donnée (Gomez-Uchida et al., 2013). La formule la plus simple pour calculer méta-N_e consiste à additionner

toutes les valeurs de N_e calculées pour toutes les populations composant la métapopulation. Cette formule n'est pas adaptée dans le cas des espèces vivant dans des rivières, étant donné qu'elles vivent dans un milieu avec de fortes contraintes spatiales (ce sont des habitats linéaires ou dendritiques) qui vont affecter la dispersion et la distribution des individus dans l'espace.

Afin de calculer avec précision les valeurs de méta-N_e de silure pour chacune des quatre rivières échantillonnées, nous avons appliqué la méthode dite de « Neighborhood » (Gomez-Uchida et al., 2013; Maruyama, 1971; Wright, 1946). Nous avons considéré une métapopulation par rivière, sauf pour la Garonne, pour laquelle nous en avons considéré deux : une composée de toutes les populations situées à l'aval de Golfech, et l'autre composée de toutes les populations situées en amont de Golfech. Cette méthode est adaptée aux environnements de type linéaire comme les rivières, et prend en compte des paramètres comme la longueur de l'habitat linéaire occupé et les capacités de dispersion de l'espèce. Les valeurs de méta-N_e ont donc été calculées avec la formule suivante :

$$meta - Ne = \sum_{i=1}^{s} Ne_i \times \left(1 + \frac{2L}{\pi^2 \times D\sigma^2}\right)$$

avec :

S = Nombre de populations composant la métapopulation (nombre de populations par rivière dans notre cas)

i = Population *i* faisant partie de la métapopulation

Ne_i = Taille efficace de la population *i*

L = Longueur de l'habitat linéaire (longueur de la rivière étudiée)

D = Somme des valeurs de N_e de toutes les populations divisée par L

σ^2 = Paramètre estimé à partir de la pente de la droite de régression du patron IBD (cf. Partie II) grâce à l'équation suivante :

$$4D\sigma^2 = \left(\frac{1}{Pente\ IBD}\right)$$

La Table 3 présente les résultats obtenus avec cette méthode : ces valeurs devront être multipliées par le rapport N_c / N_e afin de retrouver les valeurs de méta-N_c, c'est-à-dire les estimations d'abondance en silure pour ces métapopulations. La section suivante concerne le calcul de ce rapport.

Table 3 : Valeurs de méta-N_e calculées pour les 5 métapopulations considérées grâce à la méthode dite de « Neighborhood ».

Métapopulation	Valeur de méta-N_e
Garonne aval	16129,455
Garonne amont	3039,011
Dordogne	7390,348
Lot	4653,469
Tarn	8314,023

4. Calcul du rapport Nc / Ne pour le silure

Le rapport N_c / N_e d'une population ou espèce correspond au rapport existant entre les tailles réelles des populations (ou abondances N_c) d'une espèce donnée, et leurs tailles efficaces N_e (Palstra & Fraser, 2012). Ce rapport est propre à chaque espèce (voire à chaque population dans certains cas), et dépend de plusieurs facteurs liés à l'écologie et à la biologie d'une espèce, comme par exemple sa position dans la chaîne alimentaire, sa fécondité, sa longévité ou encore son histoire évolutive. Les rapports N_c / N_e sont souvent utilisés pour estimer indirectement le N_c de populations pour lesquelles des méthodes directes (ex. comptage de visu des individus) sont difficiles à mettre en œuvre. On multiplie alors ce rapport N_c / N_e avec les valeurs de N_e estimées via les analyses génétiques.

De nombreux rapports N_c / N_e ont été calculés pour des espèces de poissons d'eau douce, mais ces estimations concernent surtout des salmonidés (Palstra & Fraser, 2012). Aucune publication scientifique n'a jamais fait écho d'un rapport N_c / N_e concernant le silure, ce qui représentait une difficulté pour notre étude. Cependant, lorsque le rapport N_c / N_e propre à une espèce n'est pas connu, les gestionnaires et les généticiens de la conservation appliquent un rapport N_c / N_e consensus de 10 (c'est-à-dire, ils estiment que la taille réelle de la population considérée est dix fois supérieur à sa taille efficace ; Frankham et al., 2014). Mais appliquer un tel rapport (ou un rapport estimé pour une autre espèce) à nos estimations de N_e sur le silure pour estimer ses N_c reste risqué, surtout si le rapport choisi est très différent en terme d'ordre de grandeur du rapport N_c / N_e réel du silure.

Plutôt que d'utiliser le rapport générique N_c / N_e de 10 ou d'extrapoler au cas du silure un rapport estimé pour une autre espèce de poisson d'eau douce, nous avons donc calculé un rapport spécifique pour le silure.

Pour cela, nous avions besoin de connaître, pour une même section de rivière, le N_c et le N_e de la population de silures vivant dans cette section. Ainsi, nous avons utilisé les résultats d'une étude de Capture-Marquage-Recapture (ou CMR) de silures menée par EPIDOR (Établissement Public Territorial du bassin de la Dordogne) entre avril 2012 et juillet 2016 sur le secteur de la Dordogne allant de Pessac à Mauzac (les rapports de cette vaste étude sont consultables sur le site web d'EPIDOR : https://www.eptb-dordogne.fr/contenu/index/idcontenu/247). En effet, les données de CMR permettent d'estimer le nombre d'individus dans une population, en appliquant des formules concrètes. L'étude CMR menée par EPIDOR n'avait pas pour objectif de calculer le N_c, mais plutôt de caractériser la répartition de l'espèce le long de la Dordogne, la biométrie de l'espèce, son habitat et son régime. Par conséquent, certaines des conditions nécessaires à l'application des méthodes d'estimation de N_c n'ont pas forcément été mises en œuvre ; comme par exemple l'espacement temporel nécessaire entre les diverses campagnes de recaptures (les recaptures et les marquages étaient et continuent d'être réalisées en continu). Ces données restent néanmoins compatibles avec l'application de certaines méthodes d'estimation comme l'estimateur d'abondance ponctuel de Chao (Chao, 1988).

Cette méthode permet d'estimer l'abondance d'individus dans une population à partir de données CMR même lorsque les probabilités de recapture des individus sont faibles (Chao, 1988). L'estimation de la taille de la population (ou abondance N_c) peut donc s'effectuer en appliquant la formule suivante :

$$Nc = S + \frac{{F_1}^2}{2F_2}$$

avec :

S = Nombre d'individus capturés et marqués

F_1 = Nombre d'individus capturés et marqués, mais non recapturés par la suite

F_2 = Nombre d'individus capturés, marqués et recapturés

Cette méthode, contrairement à d'autres méthodes d'estimation, est conservatrice, et tend a sous-estimer les valeurs de N_c plutôt qu'à les surestimer. Nous avons donc calculé un intervalle de confiance à 95%, afin d'obtenir une marge d'erreur autour de l'estimation principale de N_c. Le calcul de l'intervalle de confiance à 95% de l'estimation de N_c s'est fait via les formules suivantes :

$$S + \frac{(Nc-S)}{C} \text{ et } S + (Nc - S) \times C$$

avec :

$$C = \exp\left\{1{,}96 \times \left[\log\left(1 + \frac{\widehat{var}Nc}{(Nc-S)^2}\right)\right]^{1/2}\right\}$$

et

$$\widehat{var}Nc = F_2 \times \left[0{,}25 \times \left(\frac{F_1}{F_2}\right)^4 + \left(\frac{F_1}{F_2}\right)^3 + 0{,}5 \times \left(\frac{F_1}{F_2}\right)^2\right]$$

Ainsi, en considérant S = 902, F1 = 782 et F2 = 120 (valeurs tirées du rapport EPIDOR), nous avons calculé une valeur de N_c de 3450 individus pour la section de la Dordogne qui va de Pessac à Mauzac (intervalle de confiance à 95% = [2926,9 – 4108,3]).

Pour pouvoir ensuite calculer le ratio N_c / N_e, nous avons utilisé un sous-échantillonnage d'individus de nos populations de Dordogne, afin de n'estimer N_e que pour des individus capturés entre Pessac et Mauzac, c'est-à-dire la section de la Dordogne concernée par le protocole de CMR réalisé par EPIDOR (2016). Étant donné que nous avions des données génétiques de silures capturés lors de deux années différentes dans ce secteur (2016 et 2017 ; voir Table 1), nous avons utilisé la méthode temporelle de Jorde et Ryman (1996) implémentée dans le logiciel NeEstimator (Do et al., 2014).

Ainsi, la valeur de N_e obtenue avec cette méthode est de 1275,5 individus (intervalle de confiance à 95% = [841,2 – 1799]).

Grâce à ces estimations de N_c et de N_e (et aux intervalles) calculés pour cette section de la Dordogne, nous en déduisons un rapport N_c / N_e égal à 2,705 [1,627 – 4,883]. Cette valeur est inférieure à la valeur qui est couramment utilisée lorsque le vrai rapport N_c / N_e d'une espèce n'est pas connu (comme dit précédemment, un rapport N_c / N_e = 10 est souvent utilisé dans ces cas là). Cependant notre résultat est en accord avec les rapports

trouvés pour d'autres espèces top-prédatrices (espèces qui, à l'âge adulte, se trouvent au sommet de la chaîne alimentaire d'un écosystème) comme le Brochet (*Esox lucius*, entre 2 et 4; Aguilar et al., 2005; Miller & Kapuscinski, 1997; Palstra & Fraser, 2012).

5. Estimations de l'abondance méta-N_c et de la densité de silure pour les quatre rivières étudiées

Pour finir, nous avons appliqué le rapport N_c / N_e que nous avons trouvé sur la Dordogne pour le silure à toutes les estimations de méta-N_e présentées précédemment, afin d'obtenir l'abondance de silures (méta-N_c) pour les 5 métapopulations considérées (Garonne aval, Garonne amont, Dordogne, Lot et Tarn). De plus, nous avons estimé les densités moyennes de silure dans ces métapopulations (en nombre d'individus par hectare) en divisant ces valeurs de méta-N_c (en nombre d'individus) par la surface totale d'habitat (en hectares) que ces métapopulations occupent. Cette surface a été calculée en multipliant la longueur topologique (c'est-à-dire, la longueur en suivant le cours d'eau) totale de la section de rivière occupée par chaque métapopulation, avec la largeur moyenne de ces sections de rivière. Cette largeur moyenne a été calculée à partir de 30 mesures de largeur effectuées sur des localités distribuées de façon uniforme le long des sections de rivière en question grâce aux outils de mesure du portail national de la connaissance du territoire « Geoportail » (https://www.geoportail.gouv.fr).

La Table 4 répond ainsi à l'objectif de cette étude en présentant les abondances et les densités moyennes obtenues pour les 5 métapopulations considérées.

Table 4 : Valeurs de méta-N_c et de densité calculées pour les 5 métapopulations en appliquant le rapport N_c / N_e égal à 2,705 [1,627 – 4,883] aux valeurs de méta-N_e obtenues grâce à l'approche de type ABC (voir Table 3).

Métapopulation	Valeur de méta-N_c (en nombre d'individus)	Densité (en individus par hectare)
Garonne aval	**43 630,176 [26 242,624 – 78 760,130]**	11,794 [7,094 – 21,290]
Garonne amont	8 220,524 [4 944,471 – 14 839,490]	7,518 [4,521 – 13,571]
Dordogne	19 990,890 [12 024,095 – 36 087,067]	11,576 [6,962 – 20,896]
Lot	12 587,636 [7 571,195 – 22 722,893]	9,942 [5,980 – 17,948]
Tarn	22 489,434 [13 526,916 – 40 597,377]	**19,093 [11,484 – 34,466]**

Ainsi, **le stock de silure le plus grand se trouve dans la partie aval de la Garonne** (environ 43 630 individus ; ou entre 26 000 et 79 000 individus si nous prenons en compte l'intervalle de confiance ; Table 4**). Suivent ensuite par taille de stock de silure décroissante, le Tarn** (entre 13 500 et 40 600 individus), **la Dordogne** (entre 12 000 et 36 000 individus), **le Lot** (entre 7500 et 22 700 individus) **puis la partie amont de la Garonne** (entre 5000 et 14 800 individus ; Table 4). Ces nombres sont difficilement comparables entre eux, étant donné que les surfaces d'habitat linéaire de ces 5 métapopulations sont diverses. **Il convient mieux de comparer les densités moyennes**.

En effet, **le Tarn est la rivière qui présente les densités de silure les plus élevées par hectare** (entre 11 et 34 individus par hectare si nous prenons en compte l'intervalle de confiance ; Table 4). **Il est suivi par la partie aval de la Garonne et par la Dordogne**, avec

entre environ 7 et 21 individus par hectare. **Le Lot** (entre 6 et 18 individus par hectare) **et la Garonne amont** (entre 4 et 14 individus par hectare) **sont les cours d'eau pour lesquels les estimations de densité sont les plus faibles** (Table 4).

Notons bien ici que **ces estimations de densités sont des estimations moyennes** pour les rivières (ou sections de rivière dans le cas de la Garonne) considérées. **Les densités peuvent fortement différer localement**, en fonction, par exemple, de la disponibilité en habitat souhaitable pour le silure à un endroit et à un moment donnés.

CONCLUSIONS GÉNÉRALES

Cette étude nous a permis à la fois de dresser un état des lieux complet concernant (i) les introductions de silure dans le bassin de la Garonne, (ii) l'évolution géographique et démographique du silure dans ce bassin, (iii) la diversité et de la structuration génétique du silure dans les quatre rivières principales du bassin de la Garonne (Garonne, Dordogne, Tarn et Lot), et ceci tout en fournissant (iv) des estimations de l'abondance et des densités moyennes de silure pour ces quatre rivières.

Ainsi, nous avons compilé grâce à plusieurs témoignages des informations concernant les premiers déversements de silures dans le bassin de la Garonne, ce qui devrait aider à mieux comprendre comment s'est effectuée la colonisation du bassin par cette espèce, même si ces informations restent fragmentaires à ce jour.

Nous avons ensuite montré que l'expansion géographique et démographique de cette espèce dans le bassin de la Garonne semble actuellement se stabiliser, après avoir connu une phase d'expansion très importante lors des deux dernières décennies. Outre cette tendance générale, nous montrons aussi que les tendances démographiques observées peuvent différer à l'échelle des rivières, avec des populations qui semblent être stables sur le plan démographique (sur la Dordogne et le Lot), des populations qui semblent en expansion (sur la Garonne) et d'autres qui semblent en déclin (sur le Tarn).

Malgré tous ces résultats, des données de suivi supplémentaires concernant les années à venir sont encore nécessaires pour confirmer ou infirmer ces tendances démographiques.

Notre étude met aussi en lumière les particularités génétiques des populations de silure dans les quatre rivières étudiées du bassin de la Garonne. Nos analyses indiquent que les populations situées en amont du Lot (notamment celle du Lot à Cahors) présentent des spécificités génétiques qui les distinguent des autres populations échantillonnées dans le bassin. L'analyse des patrons spatiaux de diversité et de structuration génétique suggèrent aussi que le silure n'est pas encore totalement acclimaté au bassin de la Garonne, même si nos analyses démographiques tendent à montrer le contraire. Ceci est probablement dû au décalage qui existe souvent entre les dynamiques démographiques des populations (dynamiques rapides) et les dynamiques évolutives associées (dynamiques plus lentes que les démographiques).

Nous avons pu mettre en évidence aussi l'existance de quatre groupes d'individus génétiquement homogènes dans le bassin de la Garonne : (i) deux groupes d'individus génétiquement homogènes bien définis spatialement (un groupe occupant la Dordogne ; l'autre occupant le Lot à Cahors), et (ii) deux groupes génétiquement différents mélangés dans la Garonne, le Tarn et à l'aval du Lot. À partir de ces résultats, nous avons formulé l'hypothèse suivante : les populations de silure du bassin de la Garonne seraient issues *a minima* de quatre introductions importantes, indépendantes et provenant de quatre sources génétiques (ou « populations sources ») différentes.

L'origine de ces sources reste à ce jour inconnue. Des études plus poussées, qui incorporeraient à nos analyses génétiques des individus issus de plusieurs populations (ou piscicultures) étant potentiellement à l'origine des individus introduits dans le bassin, pourraient nous éclairer à ce sujet.

Toutes les informations acquises lors de nos analyses démographiques et génétiques nous ont permis de développer et de calibrer la méthode d'estimation des abondances et des densités de silure. Nos résultats, obtenus pour la Garonne aval, la Garonne amont, la Dordogne, le Tarn et le Lot, sont des estimations conservatrices du stock et de la densité de silure, étant donné que certaines étapes intermédiaires utilisées pour les calculer tendent à sous-évaluer leurs estimations plutôt qu'à les surévaluer. Toutefois, ces estimations, ainsi que les intervalles de confiance associés, donnent une idée assez précise de l'ordre de grandeur de ces abondances et de ces densités moyennes.

Cette étude fournit non seulement une réponse à son objectif principal, à savoir une estimation du stock de silure dans les principales rivières du bassin de la Garonne : elle augmente aussi considérablement notre connaissance sur l'histoire de la colonisation du bassin de la Garonne par le silure, et sur la génétique de cette espèce.

Tous ces éléments devraient permettre la mise en place d'études spécifiques concernant les effets du silure sur les espèces autochtones et migratrices dans le bassin de la Garonne.

BIBLIOGRAPHIE

Aguilar, A., Banks, J. D., Levine, K. F., & Wayne, R. K. (2005). Population genetics of northern pike (*Esox lucius*) introduced into Lake Davis, California. *Canadian Journal of Fisheries and Aquatic Sciences*, *62*(7), 1589–1599.

Aljanabi, S. M., & Martinez, I. (1997). Universal and rapid salt-extraction of high quality genomic DNA for PCR-based techniques. *Nucleic Acids Research*, *25*(22), 4692–4693.

Beaumont, M. A., Zhang, W., & Balding, D. J. (2002). Approximate Bayesian computation in population genetics. *Genetics*, *162*(4), 2025–2035.

Belkhir, K., Borsa, P., Chikhi, L., Raufaste, N., & Bonhomme, F. (1996). GENETIX 4.05, logiciel sous Windows TM pour la génétique des populations. *Laboratoire Génome, Populations, Interactions, CNRS UMR*.

Boulêtreau, S., Gaillagot, A., Carry, L., Tétard, S., De Oliveira, E., & Santoul, F. (2018). Adult Atlantic salmon have a new freshwater predator. *PLOS ONE*, *13*(4), e0196046.

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*(1), 5–32.

Chao, A. (1988). Estimating animal abundance with capture frequency data. *The Journal of Wildlife Management*, *52*(2), 295.

Comte, L., & Olden, J. D. (2018). Fish dispersal in flowing waters: A synthesis of movement- and genetic-based studies. *Fish and Fisheries*, *19*(6), 1063–1077.

Cornuet, J.-M., Ravignié, V., & Estoup, A. (2010). Inference on population history and model checking using DNA sequence and microsatellite data with the software DIYABC (v1.0). *BMC Bioinformatics*, *11*(1), 401.

Csilléry, K., François, O., & Blum, M. G. B. (2012). abc: an R package for approximate Bayesian computation (ABC). *Methods in Ecology and Evolution*, *3*(3), 475–479.

Cucherousset, J., Horky, P., Slavík, O., Ovidio, M., Arlinghaus, R., Boulêtreau, S., Santoul, F. (2018). Ecology, behaviour and management of the European catfish. Reviews in Fish Biology and Fisheries, 28(1), 177–190.

Cutler, D. R., Edwards, T. C., Beard, K. H., Cutler, A., Hess, K. T., Gibson, J., & Lawler, J. J. (2007). Random forests for classification in ecology. *Ecology*, *88*(11), 2783–2792.

David, P., & Samadi, S. (2011). *La théorie de l'évolution: une logique pour la biologie*. [Paris]: Flammarion.

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.

EPIDOR : Verdeyroux, P., Guerri, O. (2016). Étude du silure sur la Dordogne. Synthèse des résultats des 4 premières années (avril 2012 – juillet 2016). https://www.eptb-dordogne.fr/public/content_files/resume_etude_silure-4ans_juillet_2016.pdf

Excoffier, L., & Lischer, H. E. L. (2010). Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, *10*(3), 564–567.

Finn, D. S., Bonada, N., Múrria, C., & Hughes, J. M. (2011). Small but mighty: headwaters are vital to stream network biodiversity at two levels of organization. *Journal of the North American Benthological Society*, *30*(4), 963–980.

Fisher, R. A. (1922). On the dominance ratio. *Proceedings of the Royal Society of Edimbourg*, *42*, 321–341.

Frankham, R. (1996). Relationship of genetic variation to population size in wildlife. *Conservation Biology*, *10*(6), 1500–1508.

Frankham, R. (2005). Genetics and extinction. *Biological Conservation*, *126*(2), 131–140.

Frankham, R., Bradshaw, C. J. A., & Brook, B. W. (2014). Genetics in conservation management: Revised recommendations for the 50/500 rules, Red List criteria and population viability analyses. *Biological Conservation*, *170*, 56–63.

Frichot, E., & François, O. (2015). LEA : An R package for landscape and ecological association studies. *Methods in Ecology and Evolution*, *6*(8), 925–929.

Garza, J. C., & Williamson, E. G. (2001). Detection of reduction in population size using data from microsatellite loci. *Molecular Ecology*, *10*(2), 305–318.

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.

Goudet, J. (1995). FSTAT (Version 1.2): A computer program to calculate F-statistics. *Journal of Heredity*, *86*(6), 485–486.

Guillerault, N., Bouletreau, S., Iribar, A., Valentini, A., & Santoul, F. (2017). Application of DNA metabarcoding on faeces to identify European catfish *Silurus glanis* diet. *Journal of Fish Biology*, *90*(5), 2214–2219.

Hilborn, R. (2002). The dark side of reference points. *Bulletin of Marine Science*, *70*(2), 403–408.

Jorde, P. E., & Ryman, N. (1996). Demographic genetics of brown trout (*Salmo trutta*) and estimation of effective population size from temporal change of allele frequencies. *Genetics*, *143*(3), 1369–1381.

Kalinowski, S. T. (2004). Counting alleles with rarefaction: Private alleles and hierarchical sampling designs. *Conservation Genetics*, *5*(4), 539–543.

Kimmel, M., & Chakraborty, R. (1996). Measures of variation at DNA repeat loci under a general stepwise mutation model. *Theoretical Population Biology*, *50*(3), 345–367.

Kimura, M., & Weiss, G. H. (1964). The stepping stone model of population structure and the decrease of genetic correlation with distance. *Genetics*, (49), 561–576.

Krieg, F., Estoup, A., Triantafyllidis, A., & Guyomard, R. (1999). Isolation of microsatellite loci in European catfish, *Silurus glanis*. *Molecular Ecology*, *8*(11), 1964–1966.

Lamouroux, N., Olivier, J.-M., Persat, H., Pouilly, M., Souchon, Y., & Statzner, B. (1999). Predicting community characteristics from habitat conditions: fluvial fish and hydraulics. *Freshwater Biology*, *42*(2), 275–299.

Landguth, E. L., Cushman, S. A., Schwartz, M. K., McKelvey, K. S., Murphy, M., & Luikart, G. (2010). Quantifying the lag time to detect barriers in landscape genetics. *Molecular Ecology*, *19*(19), 4179–4191.

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*(15), 2485–2487.

Lischer, H. E. L., & Excoffier, L. (2011). PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics*, *28*(2), 298–299.

Maruyama, T. (1971). Analysis of population structure: II. Two-dimensional stepping stone models of finite length and other geographically structured populations. *Annals of Human Genetics*, *35*(2), 179–196.

McCullagh, P., & Nelder, J. A. (1999). *Generalized linear models*. London: Chapman & Hall.

Miller, L. M., & Kapuscinski, A. R. (1997). Historical analysis of genetic variation reveals low effective population size in a Northern Pike *Esox lucius*) population. *Genetics*, *147*(3), 1249.

Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G., & Erlich, H. (1986). Specific enzymatic amplification of DNA in vitro: The polymerase chain reaction. *Cold Spring Harbor Symposia on Quantitative Biology*, *51*(0), 263–273.

Nei, M. (1987). *Molecular evolutionary genetics*. New York: Columbia University Press.

Palstra, F. P., & Fraser, D. J. (2012). Effective/census population size ratio estimation: a compendium and appraisal. *Ecology and Evolution*, *2*(9), 2357–2365.

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.

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.

Paz-Vinas, I., Quéméré, E., Chikhi, L., Loot, G., & Blanchet, S. (2013). The demographic history of populations experiencing asymmetric gene flow: combining simulated and empirical data. *Molecular Ecology*, *22*(12), 3279–3291.

Paz-Vinas, Ivan. (2014). *Distribution de la diversité génétique dans les réseaux dendritiques : patrons, processus et implications pour la conservation de la biodiversité.* Thèse de Doctorat. Université Toulouse 3 Paul Sabatier, Toulouse.

Paz-Vinas, Ivan, Comte, L., Chevalier, M., Dubut, V., Veyssiere, C., Grenouillet, G., Blanchet, S. (2013). Combining genetic and demographic data for prioritizing conservation actions: insights from a threatened fish species. *Ecology and Evolution*, *3*(8), 2696–2710.

Petit, R. J., El Mousadik, A., & Pons, O. (1998). Identifying populations for conservation on the basis of genetic markers. *Conservation Biology*, *12*(4), 844–855.

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.

Proteau, J.-P., Schlumberger, O., & Élie, P. (2008). *Le silure glane (Silurus glanis L.): biologie, écologie, élevage*. Versailles: Éd. Quae.

Pudlo, P., Marin, J.-M., Estoup, A., Cornuet, J.-M., Gautier, M., & Robert, C. P. (2016). Reliable ABC model choice via random forests. *Bioinformatics*, *32*(6), 859–866.

R Development Core Team. (2017). R: A language and environment for statistical computing. *Vienna, Austria, R Foundation for Statistical Computing.*

Rey, O., Fourtune, L., Paz-Vinas, I., Loot, G., Veyssière, C., Roche, B., & Blanchet, S. (2015). Elucidating the spatio-temporal dynamics of an emerging wildlife pathogen using approximate Bayesian computation. *Molecular Ecology*, *24*(21), 5348–5363.

Ricklefs, R. E., & Miller, G. L. (2005). *Écologie*. Bruxelles: De Boeck.

Rougemont, Q., Roux, C., Neuenschwander, S., Goudet, J., Launey, S., & Evanno, G. (2016). Reconstructing the demographic history of divergence between European river and brook lampreys using approximate Bayesian computations. *PeerJ*, *4*, e1910.

Rousset, F. (1997). Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. *GENETICS*, *145*(4), 1219–1228.

Rousset, François. (2008). genepop'007: a complete re-implementation of the genepop software for Windows and Linux. *Molecular Ecology Resources*, *8*(1), 103–106.

Schlumberger, O., Sagliocco, M., & Proteau, J. P. (2001). Biogéographie du silure glane (*Silurus glanis*) : causes hydrographiques, climatiques et anthropiques. *Bulletin Français de La Pêche et de La Pisciculture*, (357–360), 533–547.

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.

Stone, R. (2007). The last of the leviathans. *Science*, *316*(5832), 1684–1688.

Syväranta, J., Cucherousset, J., Kopp, D., Martino, A., Céréghino, R., & Santoul, F. (2009). Contribution of anadromous fish to the diet of European catfish in a large river system. *Naturwissenschaften*, *96*(5), 631–635.

Szpiech, Z. A., Jakobsson, M., & Rosenberg, N. A. (2008). ADZE: a rarefaction approach for counting alleles private to combinations of populations. *Bioinformatics*, *24*(21), 2498–2504.

Triantafyllidis, A., Krieg, F., Cottin, C., Abatzopoulos, T. J., Triantaphyllidis, C., & Guyomard, R. (2002). Genetic structure and phylogeography of European catfish (*Silurus glanis*) populations. *Molecular Ecology*, *11*(6), 1039–1055.

Tyers, M. B. (2017). riverdist: River network distance computation and applications. R Package version 0.15.0.

Van Oosterhout, C., Hutchinson, W. F., Wills, D. P. M., & Shipley, P. (2004). micro-checker: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes*, *4*(3), 535–538.

Waples, R. S. (2006). A bias correction for estimates of effective population size based on linkage disequilibrium at unlinked gene loci. *Conservation Genetics*, *7*(2), 167–184.

Waples, R. S., & Do, C. (2008). LDNE : a program for estimating effective population size from data on linkage disequilibrium. *Molecular Ecology Resources*, *8*(4), 753–756.

Wegmann, D., Leuenberger, C., Neuenschwander, S., & Excoffier, L. (2010). ABCtoolbox: a versatile toolkit for approximate Bayesian computations. *BMC Bioinformatics*, *11*(1), 116.

Weir, B. S., & Hill, W. G. (2002). Estimating F-statistics. *Annual Review of Genetics*, *36*, 721–750.

Wright, S. (1931). Evolution in Mendelian Populations. *Genetics*, *16*(2), 97–159.

Wright, S. (1943). Isolation by Distance. *Genetics*, *28*(2), 114–138.

Wright, Sewall. (1946). Isolation by distance under diverse systems of mating. *Genetics*, *31*(1), 39.

Zhang, D. (2017). A coefficient of determination for generalized linear models. *The American Statistician*, *71*(4), 310–316.

ANNEXES

Annexe 1 : Analyse démographique des données de capture issues d'un concours de pêche au silure : « la silurade ».

Nous avons effectué une analyse de l'évolution du nombre de silures capturés à une échelle plus locale, dans le cadre d'un concours de pêche au silure organisé chaque année depuis 2007 par la FFPPMA 47 sur le Lot à Penne d'Agenais (« la silurade »).

Nous avons représenté graphiquement le nombre total de silures capturés par année depuis 2007, date du premier concours (Figure Annexe 1 ici-bas). Nous avons aussi représenté la courbe de régression de Poisson d'un modèle linéaire généralisé construit en considérant le nombre total de silures capturés comme variable à expliquer, et le temps (en années, pour la période 2007-2018) comme variable explicative (en considérant un terme quadratique dans ce modèle glm). Nous avons exclu de ce modèle les données de 2012. En effet, cette année-là, beaucoup de petits silures avaient été pêchés. Une maille silure a été mise en place pour les années suivantes.

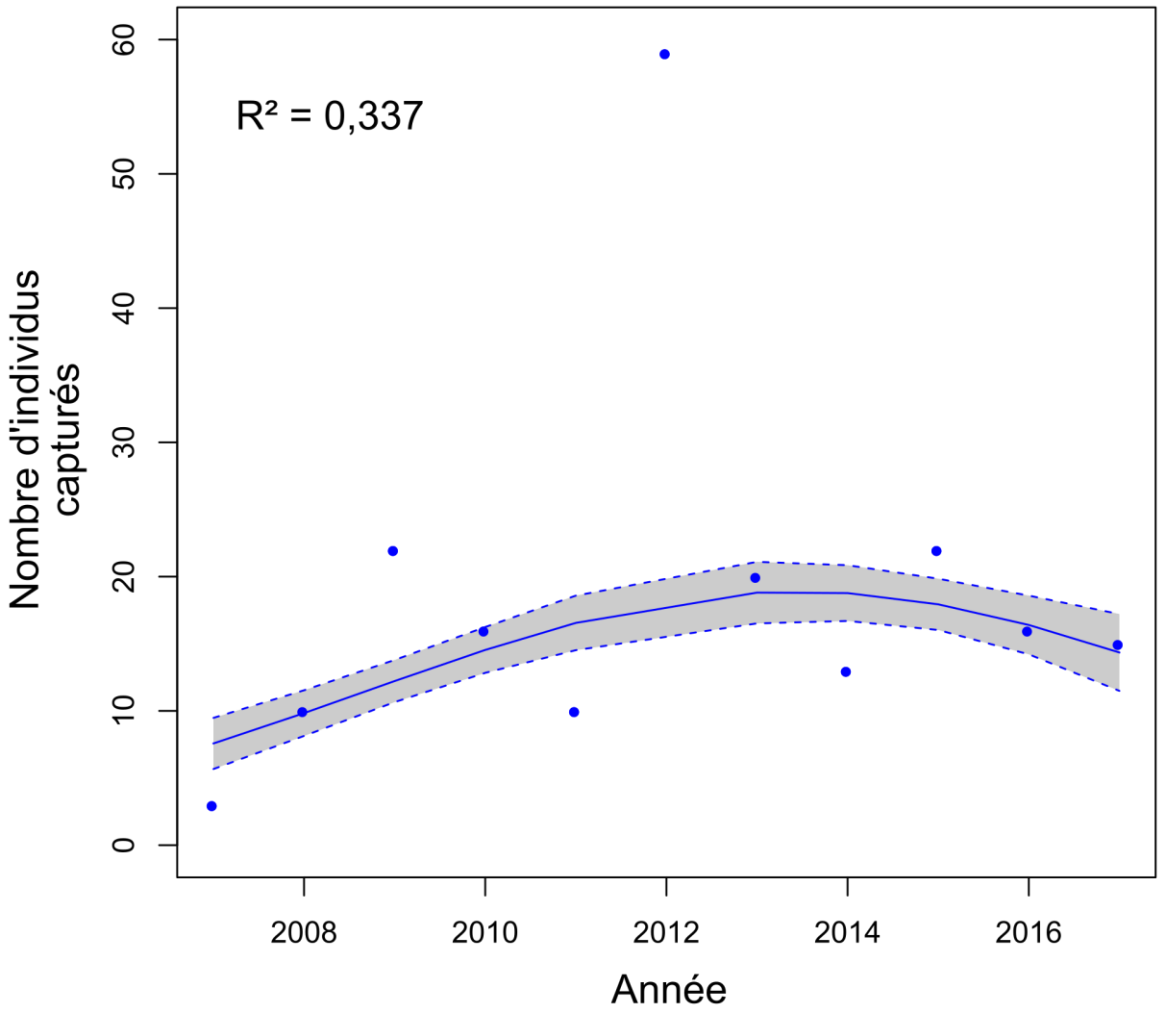


Figure Annexe 1 : Graphique représentant, pour la période 2007-2018, le nombre total de silures capturés au bord du Lot à Penne d'Agenais par année (points bleus), dans le cadre du concours de pêche au silure « la silurade ». La courbe bleue correspond aux valeurs

prédites par un modèle linéaire généralisé (« glm ») et visant à expliquer le nombre total de silures capturés (variable à expliquer) en fonction du temps (variable explicative). Les courbes bleues en pointillés représentent l'intervalle de confiance à 95% des prédictions du modèle glm. La valeur de R^2 indique la qualité de la prédiction de la régression élaborée par le modèle glm.

Nous pouvons observer que le nombre de captures reste relativement stable au cours du temps (malgré l'année exceptionnelle de 2012 où il y a eu beaucoup de petits individus capturés), avec un nombre de captures moyen de 19 silures par an (pour un nombre moyen de participants de 35). Bien que ces données ne permettent pas de caractériser correctement des tendances démographiques temporelles comme c'est le cas avec les données des pêches AFB (les protocoles de pêche, le nombre de pêcheurs et l'effort de pêche ne sont pas standardisés), les résultats de l'analyse de ces données semblent aller dans le même sens que ceux décrits dans la partie I, à savoir une certaine tendance à la stagnation des effectifs au cours du temps.

Annexe 2 : Tableau récapitulant les informations concernant les loci microsatellites utilisés dans cette étude et les kits multiplex utilisés pour les amplifier par PCR.

Locus	GenBank Accession ID	Kit Multiplex
Sgl310INRA	AF146414	SilA
Sgl325INRA	AF146415	SilA
Sgl33INRA	AF146416	SilA
Sgl5fINRA	AF146418	SilA
Sgl7eINRA	AF146423	SilA
Sgl1154aINRA	AF146410	SilB
Sgl140aINRA	AF146412	SilB
Sgl695INRA	AF146419	SilB
Sgl7159INRA	AF146420	SilB
Sgl7fINRA	AF146425	SilB

Référence :

Krieg F, Estoup A, Triantafyllidis A and Guyomard R (1999) Isolation of microsatellite loci in European catfish, *Silurus glanis*. *Molecular Ecology,* **8**, 1957-1969.

Annexe 3 : Tableau récapitulant les valeurs des différentes statistiques descriptives de la diversité génétique calculées pour chaque population. N_{Ech} correspond au nombre d'individus génotypés avec succès, He correspond à l'hétérozygotie attendue, N_{All} correspond au nombre moyen d'allèles observés (parmi les 10 loci microsatellites considérés), AR correspond à la richesse allélique moyenne, PA correspond à la richesse moyenne en allèles privés, Fis correspond au coefficient de consanguinité de la population, et Fst_{UNI} correspond à l'unicité génétique de la population.

Code Site	Rivière	Site	N_{ECH}	He	N_{ALL}	AR	PA	Fis	Fst_{UNI}
DORLam	Dordogne	Lamonzie / Prigonrieux	15	0,715	5,9	4,208	0,068	0,040	0,056
DORBer	Dordogne	Bergerac (Aval)	12	0,698	5,6	4,181	0,037	0,077	0,047
DORMau	Dordogne	Mauzac (Aval)	96	0,666	7	3,641	0,043	-0,015	0,084
GARBas	Garonne	Bassens	30	0,643	6,1	3,765	0,071	0,048	0,043
GARStj	Garonne	Saint-Jean / Bouliac	30	0,637	5,8	3,761	0,024	-0,049	0,040
GARIsl	Garonne	Isle-Saint-Georges / Cambes	50	0,671	7,3	3,885	0,077	0,045	0,036
GARBar	Garonne	Barsac	54	0,664	6,5	3,811	0,060	-0,021	0,035
GARLan	Garonne	Langon	56	0,661	6,8	3,766	0,066	0,049	0,039
GARAge	Garonne	Agen	9	0,596	5	3,755	0,086	0,070	0,048
GARGol	Garonne	Golfech	63	0,660	6,5	3,760	0,046	0,009	0,039
GARMal	Garonne	Malause	17	0,669	5,3	3,903	0,016	-0,030	0,035
GARTou	Garonne	Toulouse	9	0,638	4,8	3,791	0,096	0,083	0,072
LOTPen	Lot	Penne d'Agenais	11	0,689	5,3	4,020	0,133	-0,061	0,067
LOTPra	Lot	Prayssac	9	0,638	3,6	3,276	0,106	0,014	0,128
LOTCah	Lot	Cahors	30	0,604	4,1	3,091	0,042	-0,077	0,163
TARVil	Tarn	Villemur	19	0,652	5,6	3,863	0,028	0,010	0,037
TARAig	Tarn	Aiguelèze	18	0,683	6	3,959	0,083	-0,021	0,038
TARAlb	Tarn	Albi	31	0,672	5,3	3,623	0,007	-0,006	0,052

Annexe 4 : Le tube computationnel que nous avons utilisé dans cette étude pour estimer les tailles efficaces des populations de silure est une adaptation de celui utilisé dans Rey *et al.* (2015), Paz-Vinas *et al.* (2015) et Paz-Vinas et Blanchet (2015). Il combine et interconnecte un grand nombre de programmes informatiques de génétique des populations et de statistiques, qui travaillent en synergie afin d'appliquer la procédure décrite dans la Partie III et de générer en sortie une table de référence contenant un ensemble de statistiques descriptives calculées pour les 1 450 000 jeux de données génétiques simulés (150 000 simulations/modèle pour la partie « choix du meilleur modèle, auxquelles se sont additionnées 850 000 simulations supplémentaires pour le meilleur modèle dans la partie « inférence des tailles efficaces »).

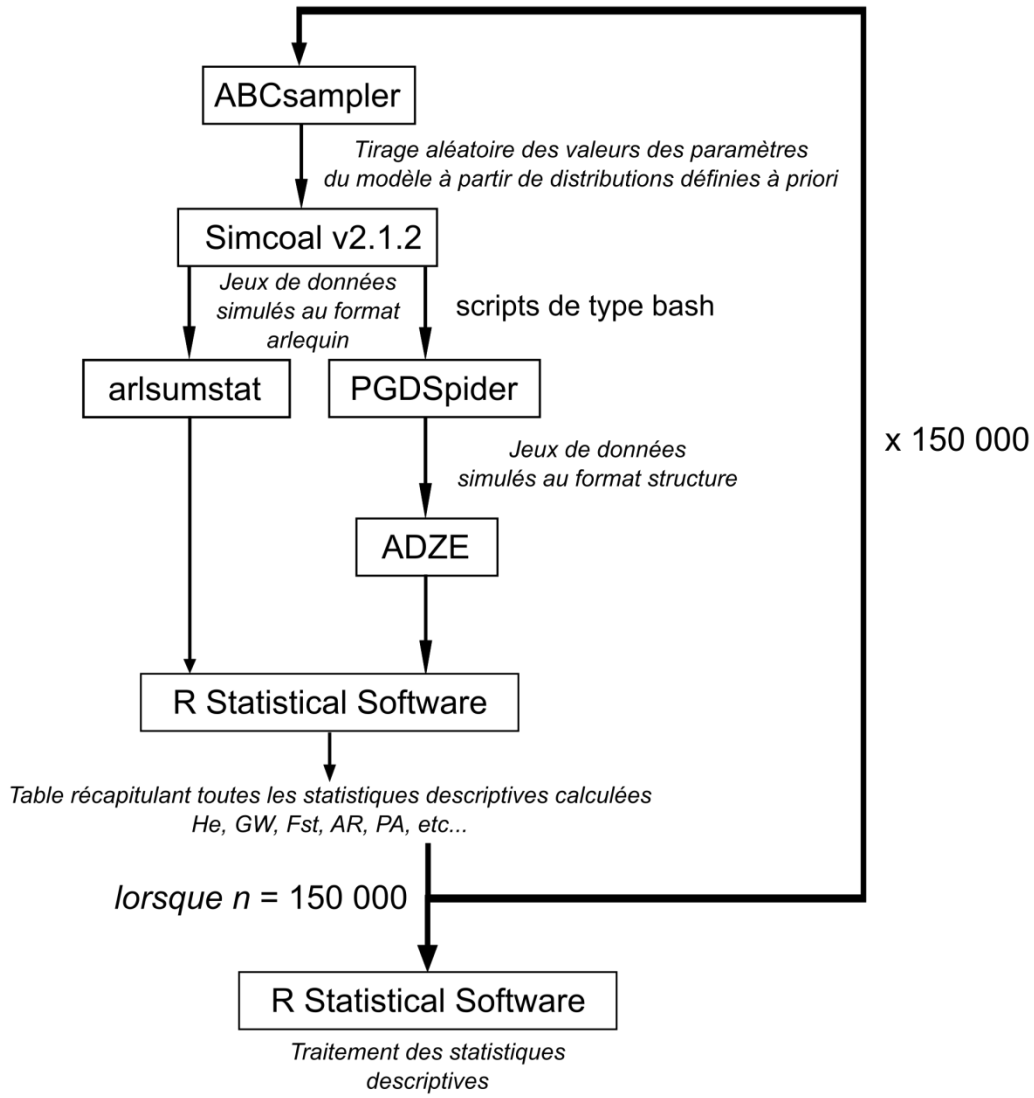


Figure Annexe 4 : Schéma représentant le tube computationnel de logiciels (« pipeline »), utilisé pour l'implémentation des algorithmes ABC et pour la simulation en série de jeux de données génétiques.

La mise en place et l'exécution de ce tube computationnel est plus difficile et plus lourde que celle de programmes dits « stand-alone » ou « user-friendly » couramment utilisés pour appliquer des algorithmes de type ABC, comme par exemple 'DIYABC' (Cornuet et al., 2010). Par contre, la grande flexibilité de ce « pipeline » nous permet de

calculer des statistiques descriptives très importantes comme la richesse allélique (AR ; Petit et al., 1998), la richesse en allèles privés (PA ; Kalinowski, 2004) et la statistique de Garza et Williamson (2001), appelée aussi GW ou M-ratio, une statistique qui est intimement liée à l'histoire démographique d'une population (elle permet de détecter efficacement des goulots d'étranglement génétiques liés à des crashes démographiques récents ou bien à des effets fondateurs via des introductions/colonisations). Un autre avantage de ce tube computationnel est qu'il permet d'utiliser des logiciels de simulation de données génétiques qui sont capables de modéliser les échanges génétiques inter-populationnels qui ont lieu à chaque génération, une caractéristique très importante qui n'est pas implémentée avec des logiciels comme DIYABC (Rey et al., 2015), et qui semble nécessaire dans le contexte d'une espèce qui a colonisé en quelques générations les principales rivières du bassin de la Garonne.

Au cœur du tube computationnel se trouve le programme ABCsampler (Wegmann et al., 2010). Il gère la communication entre les différents programmes du tube computationnel, en s'occupant de tirer aléatoirement les valeurs des paramètres des distributions « à priori » pour les envoyer au simulateur de données génétiques. Ensuite, nous utilisons le programme SIMCOAL v2.1.2 (Laval & Excoffier, 2004) pour simuler les données génétiques. Ce programme est un simulateur basé sur la théorie de la coalescence qui permet de générer divers types de données génétiques en adoptant une stratégie de type « backwards-in-time », c'est-à-dire en reconstruisant des généalogies de gènes partant du temps présent jusqu'au passé. Il permet donc de simuler divers événements historiques comme des colonisations, des changements de tailles efficaces de populations ou bien des modifications de flux de gènes entre sous-populations.

Dans notre cas, nous simulons des données de type microsatellite similaires à celles obtenues pour les données empiriques décrites en Partie II de ce rapport (par exemple, en simulant le même nombre de loci microsatellites et le même nombre d'individus qui ont été considérés pour générer la base de données génétique du silure). Puis, pour chaque jeu de données simulé, nous calculons toute une série de statistiques descriptives de la diversité génétique pour créer une table de référence. L'ensemble de ces statistiques descriptives permet de déceler au mieux les traces génétiques laissées par l'histoire démographique sur les populations étudiées et par les diverses forces évolutives agissant sur leur diversité génétique neutre (flux de gènes, mutation et dérive génétique). Le programme « arlsumstat » (Excoffier & Lischer, 2010) nous permet de calculer la plupart des statistiques, comme par exemple le nombre moyen d'allèles ou les hétérozygoties attendues. Nous avons aussi intégré dans le tube computationnel le logiciel PGDSpider (Lischer & Excoffier, 2011), qui permet de convertir les fichiers simulés par SIMCOAL en format « structure », afin que ceux-ci puissent être analysés par la suite par le programme « ADZE » (Szpiech et al., 2008). Ceci nous a permis de calculer la richesse allélique standardisée pour chaque population, ainsi que la richesse moyenne en allèles privées. Finalement, le logiciel R v3.5.1 (R Development Core Team, 2017) nous a permis de récupérer toutes les statistiques descriptives simulées par les différents programmes pour chaque simulation, afin de les stocker dans un seul et même fichier (la « table de référence »).

Références :

Cornuet J-M, Ravignié V, Estoup A (2010) Inference on population history and model checking using DNA sequence and microsatellite data with the software DIYABC (v1.0). *BMC Bioinformatics*, **11**, 401.

Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular ecology resources*, **10**, 564–567.

Garza JC, Williamson EG (2001) Detection of reduction in population size using data from microsatellite loci. *Molecular Ecology*, **10**, 305–318.

Kalinowski ST (2004) Counting Alleles with Rarefaction: Private Alleles and Hierarchical Sampling Designs. *Conservation Genetics*, **5**, 539–543.

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.

Lischer HEL, Excoffier L (2011) PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics*, **28**, 298–299.

Paz-Vinas I, Loot G, Stevens VM, Blanchet S (2015) Evolutionary processes driving spatial patterns of intraspecific genetic diversity in river ecosystems. *Molecular Ecology*, **24**, 4586–4604.

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.

Petit RJ, El Mousadik A, Pons O (1998) Identifying Populations for Conservation on the Basis of Genetic Markers. *Conservation Biology*, **12**, 844–855.

R Development Core Team (2011) R: A language and environment for statistical computing. *Vienna, Austria, R Foundation for Statistical Computing.*

Rey O, Fourtune L, Paz-Vinas I *et al.* (2015) Elucidating the spatio-temporal dynamics of an emerging wildlife pathogen using approximate Bayesian computation. *Molecular Ecology*, **24**, 5348–5363.

Szpiech ZA, Jakobsson M, Rosenberg NA (2008) ADZE: a rarefaction approach for counting alleles private to combinations of populations. *Bioinformatics (Oxford, England)*, **24**, 2498–2504.

Wegmann D, Leuenberger C, Neuenschwander S, Excoffier L (2010) ABCtoolbox: a versatile toolkit for approximate Bayesian computations. *BMC Bioinformatics*, **11**, 116.