

AGROCAMPUS OUEST

CFR Angers **CFR Rennes**

Année universitaire : 2019 - 2020
Spécialité : Ingénieur agronome
Spécialisation (et option éventuelle) :
Sciences halieutiques et aquacoles
(Ressources et écosystèmes aquatiques)

Mémoire de fin d'études

d'ingénieur de l'École nationale supérieure des sciences agronomiques, agroalimentaires, horticoles et du paysage (AGROCAMPUS OUEST), école interne de l'institut national d'enseignement supérieur pour l'agriculture, l'alimentation et l'environnement

de master de l'École nationale supérieure des sciences agronomiques, agroalimentaires, horticoles et du paysage (AGROCAMPUS OUEST), école interne de l'institut national d'enseignement supérieur pour l'agriculture, l'alimentation et l'environnement

d'un autre établissement (étudiant arrivé en M2)

Structure et diversité génétique des populations de raie bouclée en Méditerranée

Par : Romain LÉCUYER



Soutenu à Rennes, le 17 septembre 2020

Devant le jury composé de :

Président : Étienne Rivot

Maître de stage : Pascal Lorange

Autres membres du jury

Didier Gascuel - enseignant chercheur Agrocampus Ouest

Sophie Launey - chargée de recherche INRAE

Les analyses et les conclusions de ce travail d'étudiant n'engagent que la responsabilité de son auteur et non celle d'AGROCAMPUS OUEST

Structure et diversité génétique des populations de raie bouclée en Méditerranée

Romain Lécuyer - Mémoire de fin d'études

Encadré par Pascal Lorange, Verena Trenkel et David Eme

17 février 2020 - 14 août 2020

Résumé

Les chondrichthyens forment un taxon menacé par diverses perturbations et notamment le développement important de la pêche au cours du siècle passé. La Méditerranée est une région fortement exploitée depuis cette période et qui continue de l'être de manière intense par des nations cherchant avant tout à assurer une rentabilité économique au secteur de la pêche ou une sécurité alimentaire à leur population. Ce contexte ne facilite pas les acquisitions et partages de données pour la recherche. Cette étude a pour objectif principal de mieux appréhender la structure génétique et l'état de la raie bouclée afin d'éviter son déclin dû à la pression imposée par la pêche au chalut démersal, difficilement réversible même en cas de soudaine bonne gestion. Dans ce but, un large panel de marqueurs polymorphes issus du génome nucléaire a été développé et analysé afin d'étudier la structuration sous-jacente ainsi que la différenciation et la diversité génétique intrinsèque aux populations de Méditerranée. Une structuration génétique en quatre populations a ainsi été relevée, avec notamment une population semblant panmictique dans le bassin occidental de la Méditerranée. La différenciation génétique précise que les individus de mer Égée sont les plus distants des autres populations, les populations des mers Adriatique et Ionienne et du bassin occidental formant deux sous-groupes au sein desquels la différenciation est plus ténue. Les indicateurs de diversité génétique estimés révèlent une diminution graduelle de la diversité génétique d'ouest en est tandis que la taille de population efficace, indicateur clé en terme d'évolution de la diversité génétique au fil des générations, semble orienter vers une plus grande pérennité de la diversité génétique de la population de mer Égée par rapport aux autres populations plus à l'ouest.

Mots-clés : Méditerranée, structure de populations, différenciation, diversité génétique, SNP, *Raja clavata*

Table des matières

1	Introduction	3
2	Matériel et méthodes	7
2.1	Plan d'échantillonnage et séquençage des marqueurs	7
2.2	Tri du jeu de données brut	9
2.2.1	Suppression des SNPs sans génotypage ou monomorphes	9
2.2.2	Traitement des SNPs présents en plusieurs exemplaires	9
2.2.3	Filtres sur les marqueurs du polymorphisme	9
2.2.4	Filtre sur les individus	10
2.2.5	Suppression des marqueurs sous sélection	11
2.2.6	Restriction du jeu de données à la Méditerranée	12
2.3	Analyses statistiques	12
2.3.1	Structure de population	12
2.3.2	Différenciation entre populations	13
2.3.3	Indicateurs de diversité génétique	14
3	Résultats	15
3.1	Sensibilité des résultats aux différents seuils des filtres sur la MAF et sur le pourcentage de données manquantes et identification des marqueurs sous sélection	15
3.2	Structuration inférée des populations	16
3.3	Différenciation entre populations inférées et zones d'échantillonnage	18
3.4	Indicateurs comparatifs de diversité génétique	19
4	Discussion	23
4.1	Structure et hiérarchisation des populations	24
4.2	Caractérisation de la diversité génétique des populations	26
4.3	Implications pour la gestion et la conservation de la raie bouclée en Méditerranée	27
5	Conclusion	28
	Bibliographie	29
	Annexes	35
A	Méthode implémentée dans BayeScan v2.1 (Foll, 2012 - Foll & Gaggiotti, 2008)	35
B	Méthode implémentée dans PCAdapt v4.3.3 (Luu <i>et al.</i>, 2017 - Duforet-Frebourg <i>et al.</i>, 2014)	36
C	Liste des marqueurs potentiellement sous sélection	37
D	Compléments sur l'analyse de la structure de population	38
E	Compléments sur l'analyse de la différenciation des populations	40
F	Compléments sur l'analyse de la diversité génétique des populations	41
	Bibliographie annexe	44

1 Introduction

Au cours du 20^{ème} siècle, le développement rapide de l'activité de pêche a lourdement impacté les stocks halieutiques. Cet effet est observable sur l'ensemble de la classe des élastombranches, poissons cartilagineux correspondant aux raies et requins et souvent capturés en tant que prises accessoires (Dulvy *et al.*, 2014). Les captures de requins et raies en Méditerranée connaissent une diminution constante depuis le début des années 1980, passant de 26000 tonnes en 1984 à 7000 tonnes en 2007 (FAO, 2012), alors que l'activité de pêche des chalutiers croît entre 4 et 10 jours par an (Russo *et al.*, 2019). La part des captures attribuée à la classe des élastombranches en Méditerranée était d'environ 2% au début du 21^{ème} siècle (Serena, 2005). Ainsi, bien que les captures d'élastombranches soient sous-représentées (Cashion *et al.*, 2019), elles représentent tout de même une part non négligeable des captures totales (Sea Around Us, 2016). Ce constat s'étend à de nombreuses régions autour du monde, et est dû à diverses négligences dont les principales sont l'erreur ou la faible résolution d'identification des espèces, la non déclaration de capture ou la pêche illégale (Giovos *et al.*, 2020). La surveillance de ces espèces est devenue un objectif en terme de conservation afin d'éviter leur déclin, difficilement réversible même en cas de soudaine bonne gestion (Field *et al.*, 2009). Il reste très compliqué de surveiller de telles espèces car les moyens d'y parvenir sont limités et se concentrent sur les espèces à large intérêt commercial. Cette limitation de suivi des espèces exploitées conduit à la nécessité de développement de nouvelles méthodes d'évaluation de l'état des populations et de définir un point de référence actuellement manquant afin d'être en capacité d'instaurer par la suite une surveillance et une gestion adéquates.

Récemment, les analyses génétiques permettent de mener des études approfondies sur tous types de populations. L'avènement de méthodes de séquençage de nouvelle génération¹ rend accessible la constitution de larges banques de données génétiques conduisant à des analyses d'une résolution plus fine qu'en utilisant la méthode de séquençage traditionnelle de Sanger développée en 1977 (Leitwein *et al.*, 2020). Malgré ce constat, le nombre d'analyses génétiques réalisées sur des populations marines reste faible relativement aux recherches réalisées sur des espèces terrestres, représentant environ 96% des publications scientifiques sur le sujet de la génomique (Liggins *et al.*, 2019). Les analyses génétiques se basent sur l'étude de la diversité génétique et des relations entre individus d'une population. L'estimation de nombreux indicateurs de diversité génétique apporte une compréhension fondamentale sur la structuration spatiale des populations (Yan *et al.*, 2020 - Reiss *et al.*, 2009) le mode de reproduction, le degré de consanguinité, la localisation des sites de nourricerie, de frayère ou de ponte (Bernaś *et al.*, 2020) ainsi que le taux de migration (Ramos *et al.*, 2018). La diversité génétique est un indicateur parallèle à l'effectif d'une population. Un effondrement en terme de diversité génétique implique des effets dramatiques sur l'état d'une population, et est tout autant à éviter qu'un effondrement de son effectif (Frankham, 2005). La taille de population efficace est un indicateur central mesurant l'effectif théorique de reproducteurs d'une population idéale se reproduisant selon le modèle de Wright-Fisher dont le polymorphisme est équivalent à la population étudiée (Excoffier, 2006). Une plus forte taille de population efficace traduit une diversité génétique plus importante, un nombre de reproducteurs plus élevé, un taux de consanguinité plus faible mais aussi un processus de sélection naturelle plus efficace vis-à-vis de la dérive génétique, processus stochastique (Charlesworth, 2009). De plus, une taille de population efficace et une diversité génétique élevées offrent un plus grand potentiel adaptatif des populations et une meilleure résilience face aux perturbations environnementales (Mittell *et al.*, 2015 - Willi *et al.*, 2006) par la présence de nombreux variants génétiques au sein des populations, augmentant par

¹Next-generation sequencing, NGS

exemple les résistances aux parasites et maladies (Luzzatto, 2012). La taille de population efficace n'est pas nécessairement liée à la taille réelle d'une population comme le montre l'étude de la famille des grands singes ou *Hominidae* présentant une taille de population efficace autour de 200.000 pour les orangs-outans, espèce dont l'effectif réel est estimé inférieur à 120.000 individus, et une taille de population efficace de 10.000 à 25.000 pour l'être humain dont l'effectif réel a dépassé 7.500.000.000 d'individus (Schrago, 2013). Ce constat s'explique par l'existence de goulets d'étranglement dans l'histoire de la population d'être humain ayant réduit drastiquement sa diversité génétique (Amos & Hoffman, 2010). La taille de population efficace est davantage à usage relatif qu'absolu (Marandel *et al.*, 2019). L'étude de la diversité génétique d'une espèce est préconisée dans le sens de la Directive Cadre Stratégie pour le Milieu Marin (DCSMM) (Bourlat *et al.*, 2013) car elle quantifie son potentiel adaptatif et sa survie à court et long terme par la prise en compte de deux aspects majeurs en terme de conservation : la préservation d'une quantité non-négligeable de variabilité génétique héritable et la minimisation de la fixation d'allèles délétères (Lonsinger *et al.*, 2018). Conjointement, ces deux aspects contribuent au maintien d'un succès reproductif et d'une adaptabilité évolutive nécessaires pour la survie de l'espèce au sein d'un monde en perpétuel changement (O'Brien, 2008).

Les élasmobranches représentent un taxon encore peu étudié, il n'existe que de rares publications scientifiques portant sur l'étude des possibilités de conservation des élasmobranches à l'échelle régionale (Das & Afonso, 2017). Les élasmobranches sont caractérisés par une croissance lente et une maturité sexuelle qui intervient relativement tard dans la vie de l'animal (Goldman & Cailliet, 2004). La classe des élasmobranches contient de nombreux prédateurs supérieurs, et notamment parmi les requins modernes, du super-ordre *Selachii*. Cela rend les espèces concernées d'importance écologique majeure, de par la place qu'ils occupent au sein de leurs réseaux trophiques propres (Heithaus *et al.*, 2008). Tous ces traits en font un taxon qu'il est important de mieux connaître afin de protéger des espèces aux rôles écosystémiques clés. Mais leur cycle de vie lent les rend d'autant plus vulnérables à l'exploitation anthropique (Dulvy *et al.*, 2014).

Le sujet de cette étude est non seulement orienté sur un taxon peu étudié, mais également sur une région dont les populations font l'objet de trop peu d'analyses (Cardinale *et al.*, 2017). En effet, la Méditerranée est une région géopolitiquement complexe, encadrée par une multitude de pays répartis sur des continents différents ne facilitant pas les acquisitions et partages de données pour la recherche dans la mesure où certains pays du bassin méditerranéen cherchent avant tout à assurer une rentabilité économique au secteur de la pêche et une sécurité alimentaire à leur population (Abis, 2018). Il est possible de trouver divers travaux de recherche sur certaines espèces d'élasmobranches comme l'émissole lisse (Ozcan & Basusta, 2018) ou encore l'ange de mer (Fortibuoni *et al.*, 2016), mais ces travaux sont souvent réalisés sur des zones restreintes, respectivement dans le golfe d'Alexandrette et dans le nord de la mer Adriatique. D'autres études considèrent davantage d'espèces d'élasmobranches mais restent spatialement restreintes comme une étude multispécifique du golfe de Gabès (Enajjar *et al.*, 2015).

Cette étude porte sur la raie bouclée² (Figure 1 - (Couch, 1862)), espèce quasi-menacée sur liste rouge de l'Union Internationale pour la Conservation de la Nature (UICN) (Ellis, 2016). Cet état est considéré comme étant à mettre à jour rapidement avec l'utilisation de données plus récentes car son aire de répartition s'étend sur des zones peu profondes, fortement impactées par la pêche. C'est également une espèce dont la pêche est réglementée au niveau international et notamment européen. La réglementation est étayée par les travaux et les rapports de la

²*Chordata* > *Elasmobranchii* > *Rajiformes* > *Raja clavata* Linnaeus, 1758

Commission OSPAR (OSPAR, 2010). Le choix de cette espèce permet de tendre vers un bon compromis entre son appartenant à un taxon menacé et la capture d'un nombre d'échantillons suffisant pour assurer la significativité des résultats (Follesa *et al.*, 2019). Le corps de l'animal est rhomboïdal et possède un rostre court. La présence d'épines principalement dorsales explique son nom de raie *bouclée* (McEachran & Dunn, 1998). Sa longueur à maturité est d'environ 65cm et intervient autour de 7 ans, alors que l'animal peut vivre jusqu'à 15 ans (Kadri *et al.*, 2014). Cette maturité relativement tardive impacte le renouvellement du stock, bien plus lent que chez des espèces matures au bout de quelques années. La raie bouclée est surtout démersale nocturne et se retrouve sur sédiments fins, ce qui en fait une espèce aisément capturable avec des engins traînants comme le chalut de fond ou des engins dormants comme le filet maillant. Elle vit à des profondeurs allant de 10 à 800 mètres dans les latitudes subtropicales bien qu'elle soit plus commune au-dessus des 100 mètres (Last *et al.*, 2016). La raie bouclée est capable d'effectuer des migrations moyennes entre 50 et 120 km par mois (Hunter *et al.*, 2005). Ces migrations sont verticales, des eaux plus profondes en automne et hiver aux eaux côtières au printemps. Elles sont également horizontales dans un objectif d'accouplement vers et au départ des zones de frayère côtières, lors de la saison de reproduction de février à septembre, avec un pic en juin (Hunter *et al.*, 2005). Un comportement de regroupement par sexe a été observé, et il semble que les femelles migrent ensemble vers les zones de frayère environ un mois avant les mâles pour se reproduire (Holden, 1975). Une étude phylogéographique de la raie bouclée en Atlantique et en Méditerranée a conclu que l'expansion de cette espèce a débuté lors de l'interglaciaire Mindel-Riss (~ -400.000 ans) et que les glaciations quaternaires successives n'ont fait que modifier son aire de répartition sans fortement impacter l'effectif des populations du fait de sa mobilité importante (Chevolot *et al.*, 2006).

Les objectifs de cette étude sont de définir la structuration de la population de raie bouclée en mer Méditerranée ainsi que d'en analyser la diversité génétique. Ces objectifs posent des enjeux multiples parmi lesquels figurent l'étude d'une population d'élastranchés en mer Méditerranée. Également, un enjeu majeur de cette étude est l'utilisation de données génétiques étendues à l'ensemble du génome de la raie bouclée pour évaluer l'état de santé et de renouvellement des populations. Malgré l'intérêt que porte la DCSMM envers les analyses génétiques des populations, encore très peu d'études empiriques se basent sur de larges jeu de données génétiques pour analyser structure et diversité génétique des populations, notamment concernant les élastranchés en Méditerranée. Un précédent article fait état de la phylogéographie et de la structure de population de la raie bouclée dans une vaste zone couvrant l'Atlantique nord-est ainsi que la Méditerranée (Chevolot *et al.*, 2006). Cet article se base sur le génotypage de 5 séquences microsatellites pour 385 individus issus de 20 sites d'échantillonnage et date de presque quinze ans. L'étude réalisée ici est fondée sur un panel de plus de 7700 marqueurs de polymorphisme nucléotidique (SNPs³) identifiés préalablement à l'aide d'un protocole RADSeq (Baird *et al.*, 2008) pour plus de 650 individus issus de 11 sites d'échantillonnage (Le Cam *et al.*, 2019). Ce panel a été utilisé au sein du projet GenoPopTaille financé par l'Agence Nationale de la Recherche⁴ (Marandel *et al.*, 2020) et permet l'emploi de données récentes pour caractériser génétiquement les population d'une espèce de la classe des élastranchés à l'échelle de la Méditerranée. Fortes de ce jeu de données, les analyses réalisées ont permis d'aborder trois questions majeures. (1) Quelle est la distribution géographique de la structure génétique des populations de raie bouclée en Méditerranée ? (2) Quels sont les degrés de différenciation génétique entre ces différentes populations ? (3) Comment se distribue la diversité génétique de ces populations et quelle en est la taille de population efficace ? Pour répondre

³Single Nucleotide Polymorphism

⁴ANR-14-CE02-0006-01

à ces trois interrogations une approche multivariée a été employée dans un premier temps pour étudier la structuration génétique des populations, suivie de l'étude d'un ensemble d'indicateurs de différenciation et de diversité génétiques permettant de quantifier les écarts entre populations ainsi que leur diversité génétique intrinsèque.



Figure 1: Représentation de la raie bouclée - vue dorsale (Couch, 1862)

2 Matériel et méthodes

2.1 Plan d'échantillonnage et séquençage des marqueurs

Les individus échantillonnés ont été capturés durant les campagnes de chalutage MEDITS (Bertrand *et al.*, 2002) pour la part Méditerranée et durant les campagnes EVHOE et STURAT ainsi qu'en criée et à bord de navires de pêche pour la part Atlantique. Chaque individu est défini par un code relatif à son lieu de capture (Table 1).

Parmi les 655 individus du jeu de données, 243 individus ont été prélevés en Méditerranée au sein de six zones géographiques (GSA⁵). Le système de GSA a été mis en place par la Commission Générale des Pêches pour la Méditerranée (CGPM) dépendante de l'Organisation des Nations Unies pour l'agriculture et l'alimentation⁶ dans une optique de gestion des pêches (Figure 2) (FAO). Ce découpage de l'écosystème Méditerranéen est d'ordre géopolitique. Toutefois, il s'avère que ce découpage sépare des écosystèmes de manière pertinente dans certains cas.

Lors de l'échantillonnage, un prélèvement de tissu au niveau de la base proximale de la première nageoire dorsale a été placé dans un tube d'éthanol à 90° pour sa conservation. Le sexe et la longueur totale de chaque individu a également été relevé.

La bibliothèque RADseq a initialement été créée en suivant le protocole original (Baird *et al.*, 2008) sur un ensemble de 159 individus échantillonnés dans le golfe de Gascogne (Marandel *et al.*, 2020). De cette façon, 389 483 SNPs inférés ont été identifiés et filtrés sur critères de fréquence allélique du mineur ainsi que de données manquantes pour ne retenir que 43 088 SNPs. Ce panel a ensuite été génotypé par hybridation de sondes sur 7726 individus du golfe de Gascogne, du Portugal et de Méditerranée sur une puce à ADN. Seuls les 9120 SNPs les plus polymorphes dans le golfe de Gascogne ont été conservés (Le Cam *et al.*, 2019).

Table 1: Zones géographiques échantillonnées et nombre d'individus associés

Zone	Abréviation	Nombre d'échantillons
Brest/Douarnenez	Brest	117
Golfe de Gascogne	Gasc	50
Mer Cantabrique	Cant	28
Gironde	Gir	172
Ouest Portugal	Por	45
Golfe du Lion	Lion	45
Mer Tyrrhénienne nord	Tyrr	49
Sardaigne	Sard	35
Mer Adriatique	Adri	37
Mer Ionienne	Ion	41
Mer Égée	Égée	36

⁵Geographical SubArea

⁶Food and Agriculture Organization of the United Nations, FAO

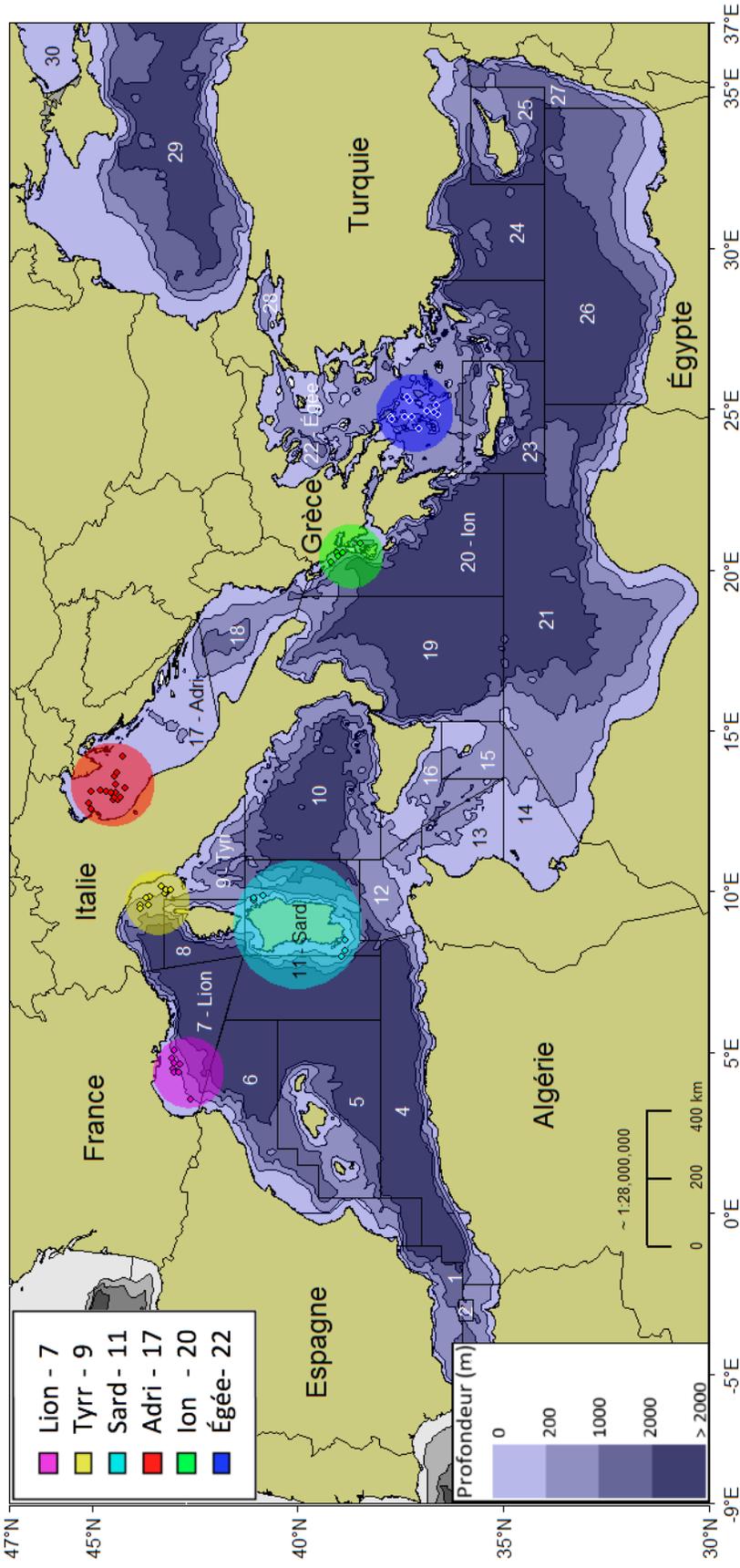


Figure 2: Sites et points d'échantillonnage des 243 individus de raie bouclée du jeu de données superposés au découpage en sous-zones de la Commission Générale des Pêches pour la Méditerranée

2.2 Tri du jeu de données brut

Différents filtres ont dû être appliqués au jeu de données brut constitué de 9120 SNPs sur 655 individus utilisé au long de cette étude afin de supprimer les données non pertinentes. Cette nécessité de tri s'explique notamment par la présence d'une grande quantité de données manquantes, d'une faible variabilité allélique ou encore d'erreurs de génotypage (Reed *et al.*, 2015).

Le tri du jeu de données a été entrepris en cinq étapes. Dans un premier temps, l'ensemble des SNPs monomorphes dont le génotypage a échoué ou dont l'allèle mineur n'a pas pu être identifié ont été supprimés. Dans un deuxième temps, une seule version de SNP faisant doublon ont été conservé pour éviter les problèmes de pseudo-réplication. Dans un troisième temps, nous avons appliqué deux filtres : un premier pour supprimer les SNPs dont l'allèle mineur est trop rare puis un second pour supprimer les SNPs ayant été séquencés sur un nombre trop restreint d'individus. Dans un quatrième temps, nous avons éliminé les individus possédant un trop grand nombre de SNPs manquants. Finalement, les SNPs potentiellement sous influence d'un mécanisme de sélection sont supprimés pour ne conserver que des SNPs neutres.

2.2.1 Suppression des SNPs sans génotypage ou monomorphes

Certains SNPs ($n = 1623$) n'ont été génotypés sur aucun individu et ont été supprimés du jeu de données. Également, aucun allèle mineur n'a pu être génotypé pour d'autres SNPs ($n = 1535$) qui se sont révélés monomorphes pour l'ensemble des individus et sont également supprimés du jeu de données.

2.2.2 Traitement des SNPs présents en plusieurs exemplaires

Pour une partie des SNPs ($n = 1597$), un second SNP se trouve dans la séquence flanquante de 50 bases adjacentes au SNP cible. Deux séquences, qualifiées de MAJ et de MIN par convention selon que le SNP inclus soit sous forme d'allèle majeur ou mineur (Le Cam *et al.*, 2019), ont été utilisées en parallèle pour génotyper chaque SNP cible dans ce cas. Les filtres appliqués précédemment ont conduit à la suppression de séquences MAJ ou MIN, laissant 41 séquences sans doublons. Les 41 SNPs cibles concernés ont été conservé d'office dans le jeu de données. Parmi les séquences restantes ($n = 2 * 778$), cinq SNPs cibles possédant un taux de similitudes inférieur à 95% ont été supprimés du jeu de données car une telle divergence a été considérée indicatrice d'un problème potentiel d'hybridation de la sonde lié à la présence d'un autre SNP en amont du SNP cible (Figure 3 - A). Pour les SNPs cibles restants ($n = 773$), la séquence MAJ a été conservée alors que la séquence MIN a été supprimée pour éviter la pseudo-réplication due à la non-indépendance des SNPs cibles issus de la même séquence. Il est notable que sur un échantillon représentatif de 778 SNPs, soit 15% du jeu de données, on observe un taux inférieur à 1% d'hybridation douteuse ($n = 5$) de la sonde utilisée. Le jeu de données contient désormais 5179 SNPs provenant d'une séquence flanquante unique.

2.2.3 Filtres sur les marqueurs du polymorphisme

La suite du tri a consisté à filtrer les SNPs selon deux critères importants : la fréquence d'occurrence de l'allèle mineur (MAF⁷) et la quantité de données manquantes pour chaque SNP (Reed *et al.*, 2015). La suppression des SNPs possédant une MAF trop faible permet d'éviter le bruit apporté par les allèles les plus rares ne pouvant pas être distingués des erreurs de séquençage

⁷Minor Allele Frequency

et nuisant à la détection de la structure de population (Linck & Battey, 2019). Les SNPs ont été filtrés sur des seuils de MAF de 1% ou 5% (Figure 3 - B), signifiant que les SNPs dont l'allèle mineur intervient sur moins de 1% ou 5% du génotypage ont été supprimés. La suppression des SNPs ayant une quantité de données manquante trop élevée permet d'éviter une diminution du pouvoir d'assignement des méthodes d'analyse génétique (Chattopadhyay *et al.*, 2014). Les SNPs ont été filtrés sur des seuils de données manquantes de 10%, 50% ou 75% (Figure 3 - C), signifiant que les SNPs possédant plus de 10%, 50% ou 75% de données manquantes ont été supprimés. Ainsi, une analyse de sensibilité a été menée vis-à-vis des deux critères de filtre sur la MAF et données manquantes par la génération de six jeux de données pour chaque combinaison des seuils de ces filtres (Table 2).

2.2.4 Filtre sur les individus

Un filtre sur données manquantes a ensuite été appliqué aux individus (Reed *et al.*, 2015) sur une gamme de seuils allant de 1% à 20% (Figure 3 - D), permettant d'identifier le seuil optimal à utiliser lors de cette étude. Les seuils de 2%, 8% et 10% permettent de conserver un nombre d'individus relativement plus important que le seuil plus restrictif qui leur est adjacent. Le seuil de 10% a été sélectionné pour le reste de l'étude, permettant d'éliminer 61 individu dont 28 capturés en Méditerranée.

Table 2: Nombre de SNPs écartés selon le filtre et contenu final du jeu de données

Fréquence du mineur	Données manquantes	SNPs supprimés	SNPs restants
1%	10%	174	5005
1%	50%	34	5145
1%	75%	32	5147
5%	10%	264	4915
5%	50%	126	5053
5%	75%	125	5054

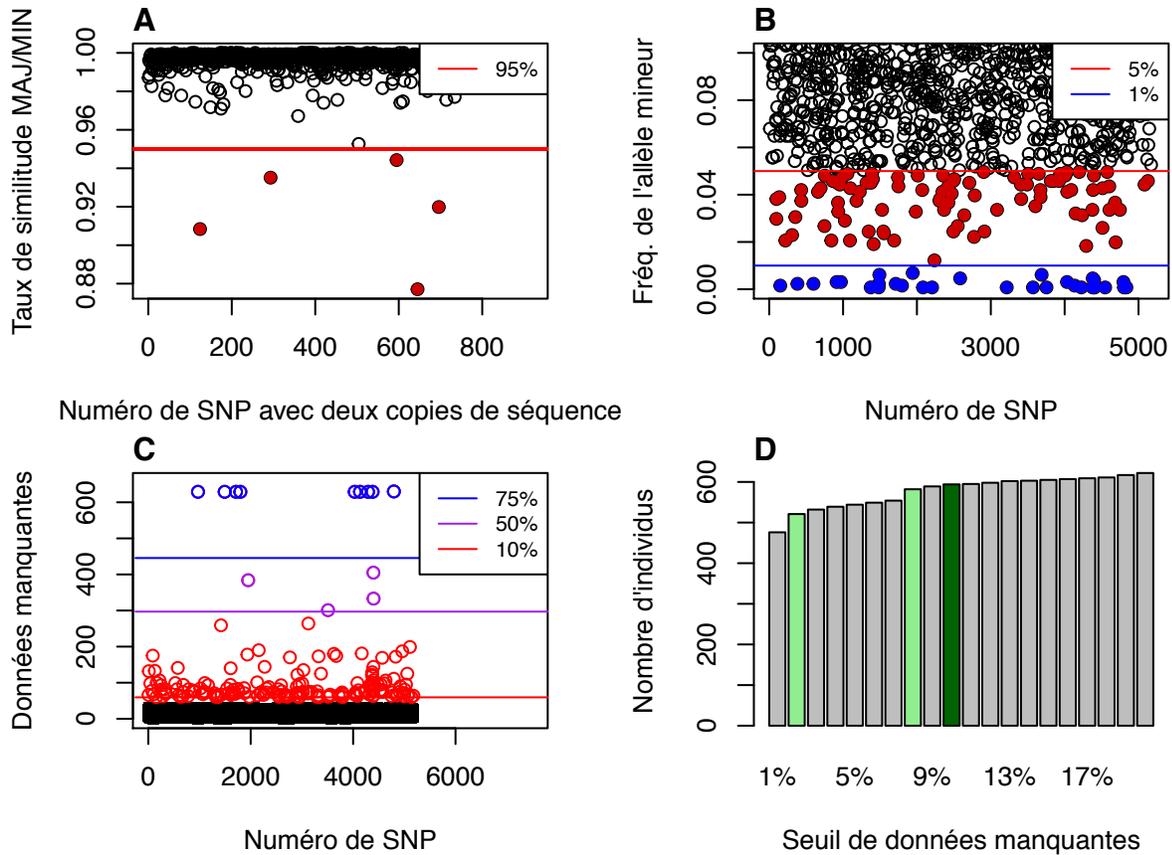


Figure 3: Graphiques issus des étapes de tri des données génétiques de 655 individus de raie bouclée échantillonnés en Atlantique et en Méditerranée - A : Similitudes entre copies majeure (MAJ) et mineure (MIN) d'un SNP possédant ces deux séquences, B : Filtre des SNPs selon le seuil de fréquence allélique du mineur (MAF), C : Filtre des SNPs selon le seuil du nombre d'individus manquants, D : Filtre des individus selon le seuil du nombre de SNPs manquants - les barres vertes représentent les valeurs de seuils avec les plus grands changements, la plus foncée étant le seuil retenu

2.2.5 Suppression des marqueurs sous sélection

Lors de l'analyse de la structure et la diversité génétique d'une population, il est important de s'assurer de n'étudier que la variation génétique des marqueurs dits *neutres*, uniquement sous l'influence de la dérive génétique, et non de marqueurs adaptatifs, sensibles aux effets de la sélection naturelle (Kirk & Freeland, 2011). Les tests de détection de ces marqueurs sous sélection sont d'autant plus puissants que le nombre d'individus étudié par population est élevé. Ainsi, l'ensemble des individus du jeu de données a été conservé, qu'ils soient issus de la région Atlantique ou Méditerranéenne.

La combinaison de plusieurs approches est nécessaire pour détecter les marqueurs sous sélection en réduisant fortement le nombre de faux positifs détectés (Lotterhos & Whitlock, 2015 - Lotterhos & Whitlock, 2014). Dans cette étude, deux logiciels ont été employés, se basant sur une approche bayésienne pour BayeScan (Foll, 2012 - Foll & Gaggiotti, 2008) et sur une approche multivariée pour PCAdapt (Luu *et al.*, 2017 - Duforet-Frebourg *et al.*, 2014).

Le logiciel BayeScan v2.1 (Foll, 2012 - Foll & Gaggiotti, 2008) utilise une approche bayésienne pour estimer la probabilité que chaque SNP soit sous influence de la sélection naturelle en scindant l'indice de fixation de Wright (F_{ST}) (Wright, 1965) en deux sous-composantes spécifiques aux populations d'une part et aux SNPs d'autre part (la méthode est détaillée en **Annexe A**). La méthode pose l'hypothèse que les fréquences alléliques suivent une distribution de Dirichlet (Balding, 2003). Deux modèles, incluant ou excluant les effets de sélection naturelle, sont ensuite comparés pour émettre la probabilité postérieure que chaque SNP soit sous sélection. La détermination d'un seuil sur cette probabilité permet d'identifier les SNPs potentiellement sous influence de la sélection naturelle. La paramétrisation par défaut du logiciel a été conservée à l'exception de la chance *a priori* en faveur du modèle neutre égale à 10 vers une valeur égale à 10.000 (Lotterhos & Whitlock, 2014) dans le but de réduire la détection de faux positifs. Cinq chaînes de Markov ont été simulées afin de s'assurer de la bonne convergence des distributions postérieures.

Le logiciel PCAdapt (Duforet-Frebourg *et al.*, 2014), utilisé via le package R (R Dev. Core Team, 2020) éponyme dans sa version 4.3.3 (Luu *et al.*, 2017) infère la structure de population par analyse en composantes principales (ACP) (la méthode est détaillée en **Annexe B**). Une approche basée sur la distance de Mahalanobis permet d'identifier les SNPs se différenciant de la structuration inférée, ainsi considérés comme étant sous influence de la sélection naturelle.

Cette étude n'a considéré que les 19 SNPs communs aux listes de candidats issues de BayeScan et de PCAdapt (Table S1). Ces marqueurs sous sélection ont été écartés du jeu de données. Tous les autres marqueurs, considérés neutres, ont été conservés pour l'analyse de la structure de la population de raie bouclée.

2.2.6 Restriction du jeu de données à la Méditerranée

Le tri du jeu de données s'est achevé par la suppression de l'ensemble des individus issus de la région Atlantique. Seuls les 213 individus échantillonnés en Méditerranée ont été conservés dans l'analyse. Avec la suppression de ces individus, certains SNPs ($n \in [20, 45]$ selon les filtres appliqués au jeu de données) ont perdu leur caractère polymorphe car fixés en Méditerranée.

2.3 Analyses statistiques

2.3.1 Structure de population

Afin d'étudier la structuration génétique des populations de raie bouclée en Méditerranée, une approche de délimitation de groupes génétiquement semblables (*clustering*) a été utilisée sur la base d'une approche multivariée reposant sur les différents jeux de données, contenant tous environ 5000 SNPs. La DAPC⁸ (Jombart *et al.*, 2010) a été développée afin d'enchaîner une analyse en composantes principales (ACP) et une analyse discriminante (AD). Cet outil est utilisable *via* la fonction `dapc` implémentée dans le package `adegenet` (Jombart & Collins, 2015) utilisé dans sa version 2.1.3 sur le logiciel R (R Dev. Core Team, 2020).

Le principe même de l'ACP consiste à condenser l'information des SNPs, jouant le rôle de variables, sous forme de composantes principales. Les individus sont ensuite séparés selon ces composantes. Dans le cadre de l'analyse de la structure génétique d'une population, la rapidité et la gestion aisée d'un grand nombre de données sont les points forts de l'ACP. Toutefois, ce

⁸Discriminant Analysis of Principal Components

type d’approche multivariée ne s’intéresse qu’à la variabilité globale entre individus là où il est nécessaire de distinguer variabilité entre (*inter*) et au sein (*intra*) des clusters. Cette distinction au sein de la variabilité totale est réalisable par le biais d’une AD. En effet, l’application d’une telle approche multivariée à des données génétiques permet de maximiser la variabilité inter-clusters tout en minimisant la variabilité intra-clusters. L’incorporation d’une information sur les zones d’origine des individus permet de visualiser la coïncidence des clusters avec les découpages utilisés *a priori* lors de l’échantillonnage.

La fonction `dapc` requiert la paramétrisation du nombre de composantes principales (la quantité d’information) ainsi que le nombre de fonctions discriminantes (permettant au mieux de maximiser la variabilité inter) à conserver. Une trop faible quantité d’information conduirait à de nombreuses mésattributions individuelles au sein des clusters, mais une quantité d’information trop importante mènerait à une perte de capacité prédictive par surprédiction des probabilités d’appartenance. La validation croisée, utilisable avec la fonction `xvalDapc` également implémentée dans le package `adegenet`, permet d’estimer le nombre de composantes optimal à conserver pour obtenir un compromis correct entre ces deux situations (la méthode est détaillée en **Annexe D**).

2.3.2 Différenciation entre populations

Des nombres d’individus différents ont été échantillonnés selon le site, or plus le nombre d’échantillons issus d’un site est élevé, plus la probabilité que leur diversité soit large augmente. C’est pourquoi un ré-échantillonnage a été effectué afin d’équilibrer les effectifs au niveau de l’effectif minimal obtenu parmi les populations échantillonnées, soit trente individus en mer Adriatique et en Sardaigne. Pour chaque population, 30 individus ont été sélectionnés aléatoirement sans remise une centaine de fois, et les différents indicateurs obtenus ont été moyennés sur les 100 tirages. Du fait de la présence de 30 individus en mer Adriatique et en Sardaigne, aucun ré-échantillonnage n’a été effectué dans ces deux cas et aucune incertitude n’a été associée aux estimations obtenues.

L’indice de fixation (F_{ST}) de Wright (Wright, 1965) sera utilisé comme indice de différenciation entre paires de populations car il est le plus recommandé pour l’étude de SNPs bialléliques (Whitlock, 2011) :

$$F_{ST} = \frac{\bar{p}(1 - \bar{p}) - \overline{p(1 - p)}}{\bar{p}(1 - \bar{p})}$$

où p est la fréquence d’un allèle dans une population et \bar{p} est la fréquence moyenne d’un allèle parmi l’ensemble des populations. L’estimateur de F_{ST} de Weir et Cockerham (Weir & Cockerham, 1984) a été utilisé dans cette étude. Il considère que la population totale correspond à la plus proche population ancestrale commune aux deux populations comparées. Un intervalle de confiance autour des estimations de F_{ST} a été obtenu par 100 ré-échantillonnages “bootstrap” sur les SNPs.

Une analyse hiérarchique a été appliquée aux indices de différenciation entre paires de populations obtenues afin de construire des dendrogrammes permettant de représenter la structuration hiérarchique des populations. Ces approches sont particulièrement sensibles au choix de l’algorithme de hiérarchisation (Mérigot *et al.*, 2010). La méthode de hiérarchisation utilisée est implémentée dans la fonction `hclust` du package `stats` utilisé dans sa version 3.6.2 sur le logiciel R (R Dev. Core Team, 2020). L’ensemble des algorithmes proposés par la fonction a été testé et la méthode non-pondérée de groupement en paires par moyenne

arithmétique (UPGMA⁹) a été retenue parce qu'elle a renvoyé la corrélation maximale entre la matrice des distances originale et les distances cophénétiques du dendrogramme (Holgersson, 1978).

Enfin, une AMOVA¹⁰ (Excoffier *et al.*, 1992) à deux échelons hiérarchiques a été effectuée pour mieux comprendre la façon dont se répartit la variance génétique entre les composantes inter-population et intra-population. Il s'agit d'un procédé d'analyse similaire à une ANOVA¹¹ hiérarchique prenant en compte les différences de fréquences alléliques et la distance évolutive entre SNPs dans le but de déterminer la proportion de variance attribuée aux différents échelons hiérarchiques.

2.3.3 Indicateurs de diversité génétique

Six indicateurs ont été sélectionnés pour caractériser la diversité génétique des populations de raie bouclée en Méditerranée et ont été calculés comme moyenne des cent valeurs obtenues sur les cent jeux de données ré-échantillonnés. L'hétérozygoté moyenne attendue sous l'hypothèse d'équilibre de Hardy-Weinberg (H_E) donne accès à la richesse et l'uniformité des allèles au sein d'une population. L'hétérozygoté moyenne observée (H_O) permet de quantifier le taux d'homozygotes dans la population par comparaison avec H_E . L'indice de consanguinité (F_{IS}) permet d'évaluer le taux de reproduction entre parents conduisant à un excès d'homozygotes, dangereux face aux maladies exprimées par un allèle récessif. Le pourcentage de marqueurs monomorphes permet de déduire l'adaptabilité de la population à son milieu, un monomorphisme important étant associé à une moindre adaptabilité et une plus grande consanguinité. Le nombre de SNPs privés, fixés dans une population, renseigne sur le taux d'immigration et l'isolement reproducteur d'une population. La taille de population efficace (N_e) permet notamment d'estimer le nombre de reproducteurs mais aussi l'impact des mécanismes de sélection naturelle par rapport à la dérive génétique.

L'hétérozygoté observée est estimée de la façon suivante

$$H_O = 1 - \sum_k P_k/n$$

où P_k représente la proportion d'homozygotes dans la population k et n est le nombre de populations. L'hétérozygoté attendue, ou diversité génétique intrinsèque à une population est estimée par la formule suivante

$$H_E = \tilde{n}/(\tilde{n} - 1)[1 - \overline{p^2} - H_O/2\tilde{n}]$$

où $\tilde{n} = n/\sum_k \frac{1}{a_k}$ avec a_k le nombre d'occurrences d'un allèle dans la population k et $\overline{p^2} = \sum_k p_k^2/n$. L'indice de consanguinité est calculé comme suit

$$F_{IS} = 1 - \frac{H_O}{H_E}$$

Ces trois indicateurs sont calculés par population en tant que moyenne sur l'ensemble des marqueurs et des individus. Ils sont obtenus à l'aide de la fonction `basic.stats` du package `hierfstat` v 0.5.7 (Goudet, 2005). La significativité des différences de moyennes entre toutes les populations de ces trois indicateurs a été testée par un test de Student apparié par SNP au seuil de risque α de 5%. La correction de Holm (Holm, 1979) a été appliquée aux p-values obtenues afin de prendre en

⁹Unweighted Pair Group Method with Arithmetic mean

¹⁰Analysis of Molecular Variance

¹¹Analysis of Variance

compte les comparaisons multiples.

La taille de population efficace est estimée à partir du Δ composite de Burrows corrigé (Weir, 1979)

$$\Delta = \frac{n}{n-1}(D_{gam} + D_{ngam})$$

avec n la taille de population, D_{gam} le déséquilibre de liaison pour un locus gamétique et D_{ngam} le déséquilibre de liaison pour un locus non-gamétique, qui estime le déséquilibre de liaison entre deux loci en tenant compte du fait que les populations ne possèdent qu'un nombre fini d'individus. L'estimateur du coefficient de corrélation entre l'allèle majeur du locus i et l'allèle majeur du locus j est

$$\hat{r}_{\Delta} = \frac{\hat{\Delta}}{\sqrt{[\hat{p}(1-\hat{p}) + (h_i - \hat{p}^2)][\hat{q}(1-\hat{q}) + (h_j - \hat{q}^2)']}}$$

où h_i et h_j sont les fréquences homozygotes majeurs observées aux loci i et j , et \hat{p} et \hat{q} sont les fréquences des allèles majeurs observées. Ce coefficient de corrélation est l'ajustement de Δ sur les fréquences alléliques permettant d'estimer N_e . La taille de population efficace est obtenue à l'aide du logiciel NeEstimator v2.1 (Do *et al.*, 2014) par application de la méthode LDNe, basée sur le déséquilibre de liaison¹² (Waples & Do, 2008). La paramétrisation par défaut a été conservée avec une valeur critique de 0,01. La méthode du déséquilibre de liaison basée sur un accouplement aléatoire a été utilisée car il a été observé une multipaternité de la raie bouclée réfutant l'idée de monogamie (Chevolot *et al.*, 2007). Dans cette situation, NeEstimator calcule un coefficient de corrélation empirique selon la formule $\hat{r}^{2'} = \hat{r}_{\Delta}^2 - E(\hat{r}_{samp}^2)$ où $E(\hat{r}_{samp}^2)$ est la contribution attendue de la taille d'échantillon. S'en suit l'estimation de la taille de population efficace

$$\hat{N}_e = \frac{\frac{1}{3} + \sqrt{\frac{1}{9} - 2,76\hat{r}^{2'}}}{2\hat{r}^{2'}}$$

Le logiciel a reporté des tailles de population efficaces négatives ou extrêmement grandes. Les valeurs négatives et supérieures à un seuil de 50.000 ont été considérées comme indiquant une taille de population efficace infinie et ont été supprimées. Le nombre de valeurs de taille de population efficace conservé par rapport au nombre de jeux de données ré-échantillonnés a été relevée sous la dénomination de "Quantité de données".

3 Résultats

3.1 Sensibilité des résultats aux différents seuils des filtres sur la MAF et sur le pourcentage de données manquantes et identification des marqueurs sous sélection

Les seuils du filtre sur la MAF de 1% et 5% et les seuils du filtre sur données manquantes de 10%, 50% et 75% n'ont jamais retiré plus de 5% de la quantité totale de SNPs (Table 2). L'impact de ces filtres sur la quantité de données étant faible, seuls les résultats issus du jeu de données filtré sur les seuils les plus conservatifs de 1% pour la MAF et de 75% pour les données manquantes contenant 5147 SNPs et 213 individus sont présentés dans la suite.

Le logiciel BayeScan et la fonction `pcadapt` ont respectivement détecté 250 et 51 SNPs sous sélection. Les SNPs communs à ces deux listes ($n = 19$) ont été supprimés (Table S1).

¹²Linkage Disequilibrium, LD

3.2 Structuration inférée des populations

Afin d'évaluer la structure génétique des populations de raie bouclée en Méditerranée par DAPC, la procédure automatique de sélection du nombre de composantes principales par validation croisée a estimé que le meilleur compromis en terme de quantité d'information était atteint en conservant les 80 premières composantes principales pour l'application de l'ACP. Deux fonctions discriminantes, résumant 92,6% de la variabilité génétique totale, ont été conservées pour l'application de l'analyse discriminante. Cette paramétrisation de la DAPC permet d'inférer quatre populations distinctes, les individus issus du bassin occidental de Méditerranée (golfe du Lion, mer Tyrrhénienne nord et Sardaigne) étant associés à la même population inférée (Figure 4). En effet, les trois nuages de points associés aux individus de ces zones se chevauchent largement dans le premier plan discriminant. La première fonction discriminante sépare ces quatre clusters selon un axe qui semble être associé à la longitude de capture des individus. La seconde fonction discriminante sépare les individus des zones les plus extrêmes en terme de longitude d'une part et les individus des zones centrales d'autre part. Aucun SNP ne semble contribuer particulièrement à la discrimination, que ce soit selon la première comme la deuxième fonction discriminante (Figure S1).

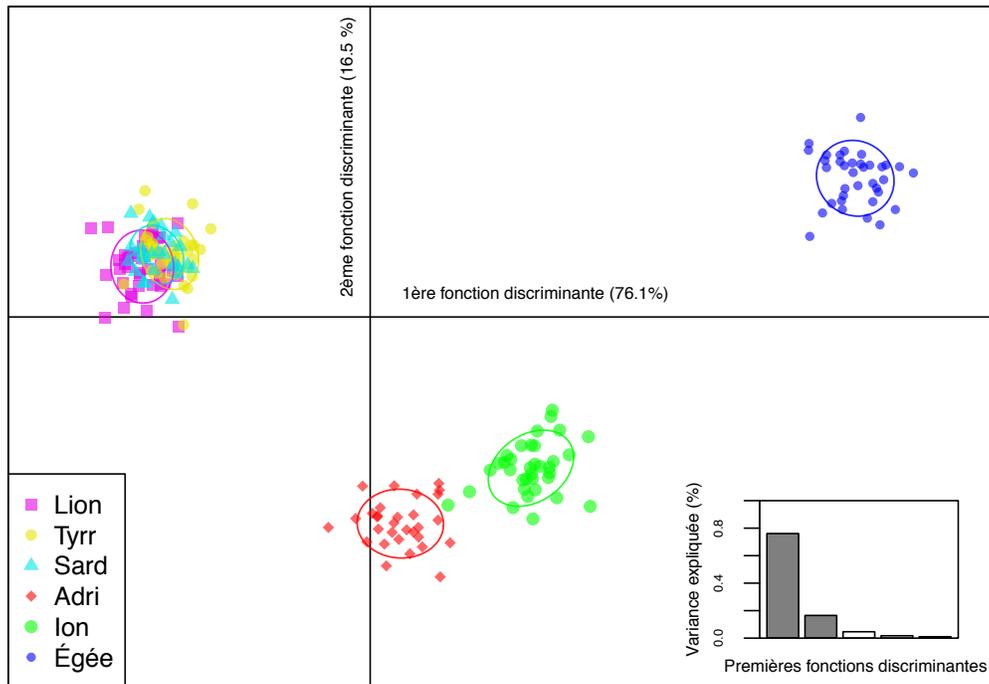


Figure 4: Regroupement des individus de la raie bouclée en Méditerranée selon les deux premières fonctions discriminantes d'une analyse discriminante en composantes principales basée sur 5147 SNPs et 213 individus issus de 6 zones d'échantillonnage - sous-graphique : variance expliquée par chacune des fonctions discriminantes, en gris les fonctions retenues pour l'analyse

Les graphiques de mélange montrent que les populations occidentales ne sont pas différenciées génétiquement car les individus possèdent une probabilité d'appartenance équivalente à chacune des trois zones (Figures 5 & S2). En revanche, les individus capturés dans les mers Adriatique, Ionienne et Égée sont attribués à l'une ou l'autre de ces zones sans mélange notable.

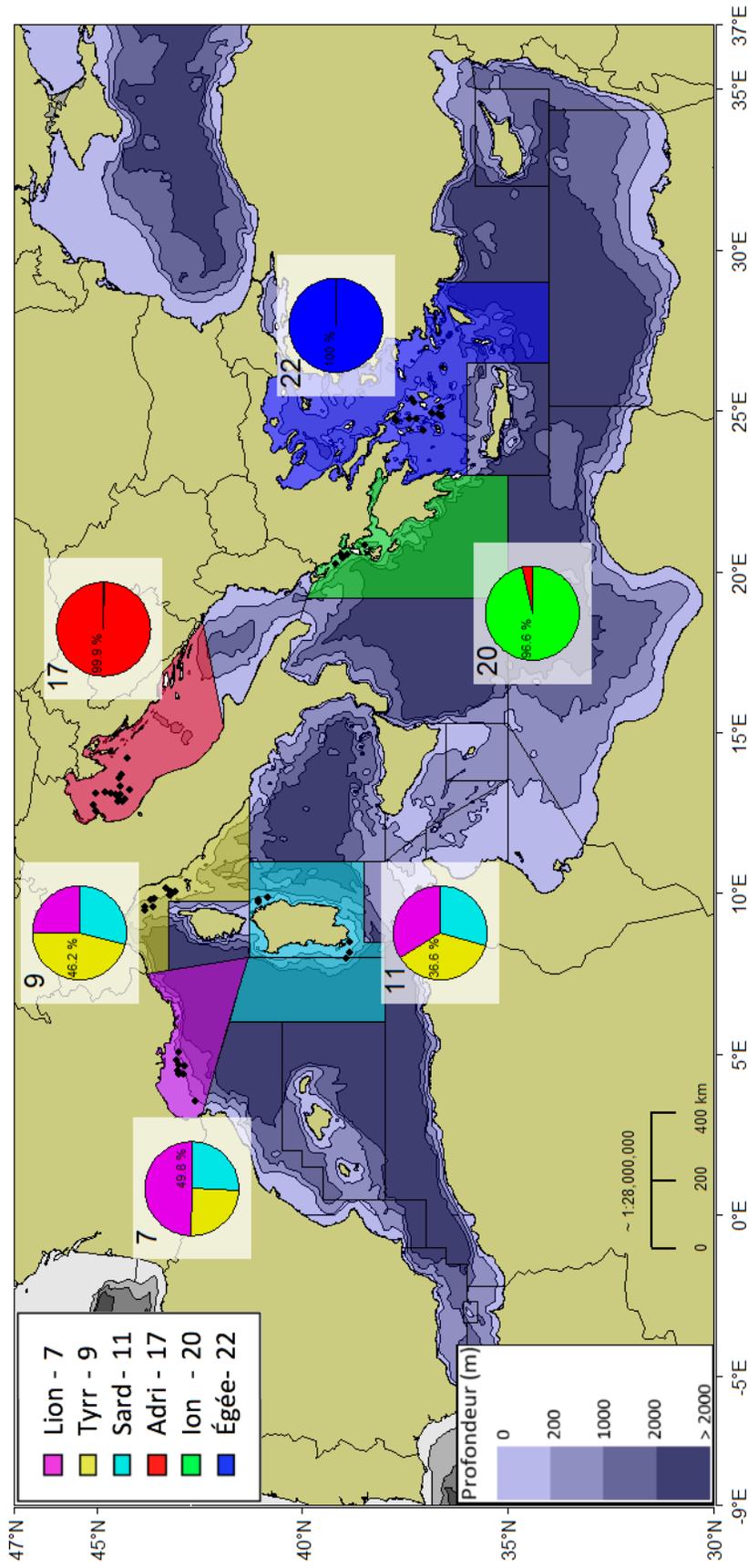


Figure 5: Graphiques des probabilités d'appartenance moyennes aux six zones échantillonnées en Méditerranée pour la raie bouclée selon l'analyse discriminante en composantes principales basée sur 5147 SNPs et 213 individus issus de 6 zones d'échantillonnage - l'échantillonnage des individus est rappelé par les points noirs et le pourcentage de chaque graphique en secteurs correspond à la probabilité d'appartenance maximale moyenne des individus de la zone

Ce premier résultat aboutit à quatre populations inférées pour six zones échantillonnées. Pour la suite de cette étude deux groupements ont été considérés : le groupement initial selon les six zones d’échantillonnage et le groupement inféré par DAPC en quatre populations dans lequel les individus échantillonnés dans le golfe du Lion, en mer Tyrrhénienne nord et autour de la Sardaigne ont été regroupés en une seule population occidentale abrégée “Ouest”. L’analyse parallèle de ces deux groupements a permis de mieux appréhender la façon dont ces découpages capturent la différenciation et la diversité génétique des populations.

3.3 Différenciation entre populations inférées et zones d’échantillonnage

Selon la différenciation visible sur les dendrogrammes (Figure 6) un seuil de F_{ST} de 0,008 conduit à une hiérarchisation structurant les populations de façon similaire aux résultats de la DAPC. La population de mer Égée est la plus distincte des autres populations (gamme de F_{ST} : 0,027 - 0,057). Ensuite, deux groupes de populations se différencient avec d’une part les zones occidentales et d’autre part la mer Adriatique et la mer Ionienne (gamme de F_{ST} : 0,018 - 0,024). Ces dernières sont davantage différentes entre elles (F_{ST} : 0,009) que les populations occidentales entre elles (gamme de F_{ST} : 0,004 - 0,007). Les intervalles de confiance autour des F_{ST} dans le cas de cette hiérarchisation en quatre populations ne se recoupent pas et n’incluent pas 0 attestant de la robustesse du dendrogramme associé (Table S2). Les intervalles de confiance autour des F_{ST} notamment entre le golfe du Lion, la mer Tyrrhénienne nord et la Sadaigne se recoupent dans le cas d’une hiérarchisation en six zones d’échantillonnage (Table S3).

L’analyse moléculaire de la variabilité génétique des individus a dévoilé une plus grande part de variabilité au niveau intra-population qu’au niveau inter-population (Table 3). La variabilité intra-population est très élevée malgré la claire différenciation génétique entre populations relevée sur les dendrogrammes.

Table 3: Décomposition de la variabilité génétique des 213 individus du jeu de données sur les différents niveaux hiérarchiques des populations de raie bouclée en Méditerranée - la variabilité est exprimée entre 0 et 1 en tant que moyenne et écart-type entre parenthèses des valeurs obtenues pour 100 jeux de données ré-échantillonnés à hauteur de 30 individus par population

	Niveau inter-population	Niveau intra-population
Quatre populations (inférées)	0.056 (± 0.0007)	0.944 (± 0.0007)
Six populations (échantillonnées)	0.051 (± 0.0004)	0.949 (± 0.0004)

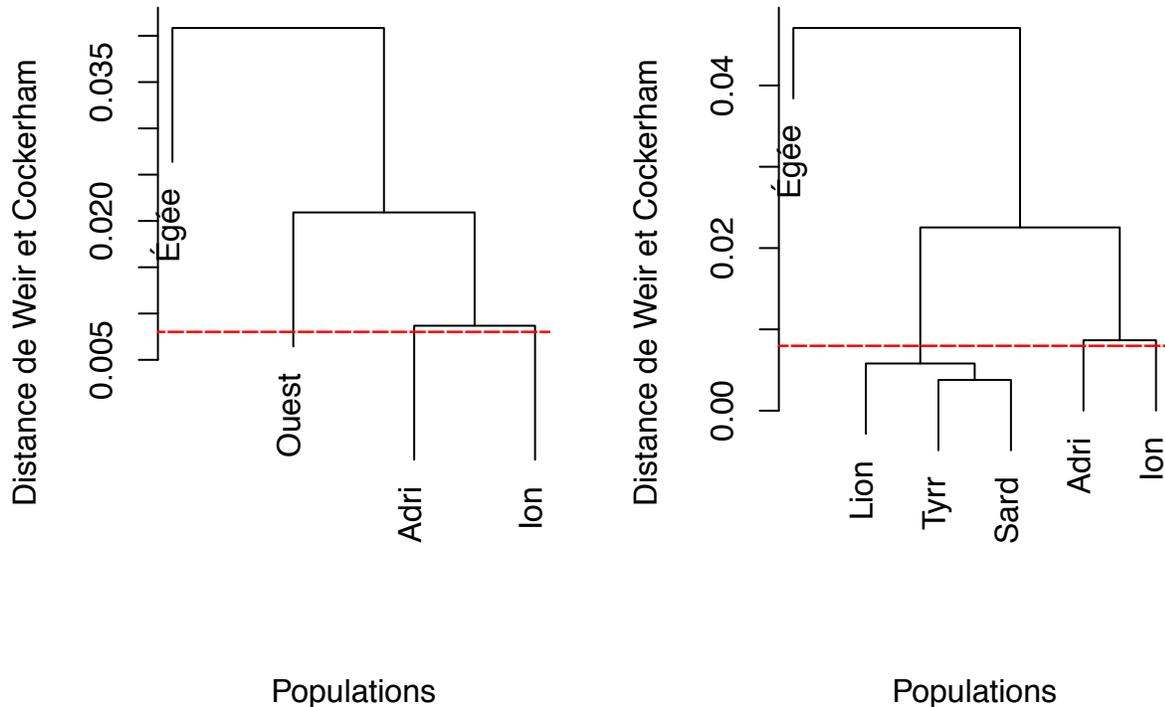


Figure 6: Dendrogramme des F_{ST} de Weir et Cockerham entre zones d'échantillonnage de raie bouclée en Méditerranée - à gauche : quatre populations inférées, à droite : six zones d'échantillonnage - la ligne tiretée rouge représente le seuil de différenciation (0.008) pour un découpage en quatre populations

3.4 Indicateurs comparatifs de diversité génétique

L'analyse de diversité génétique réalisée entre individus d'une même population pour les quatre populations inférées a révélé un changement graduel des indicateurs d'ouest en est. En effet, les indicateurs d'hétérozygoté H_O , H_E et F_{IS} (gammes respectives : 0,269-0,309, 0,273-0,312 et 0,008-0,015) sont plus élevés dans la partie occidentale de la Méditerranée et moindre en mer Égée (Table S6). Les tests de Student appariés réalisés sur ces trois indicateurs dévoilent que les variations de H_O et H_E sont toutes significatives tandis que les variations de F_{IS} ne sont pas significatives (Tables S4 & S5). Au contraire, le pourcentage de marqueurs monomorphes et le nombre d'allèles privés (gammes respectives : 3,10-12,73 et 14,09-256,75) de chaque population inférée sont moindres dans la population occidentale que dans la mer Égée (Figure 7). En revenant aux six zones échantillonnées, les individus de mer Tyrrhénienne nord ont un indice de consanguinité F_{IS} plus faible contrairement à ceux des zones occidentales adjacentes (Figure 8) mais cet écart n'est pas significatif en regard de la Sardaigne. Le reste des indicateurs de diversité génétique ainsi que les tests de significativité de Student sont similaires dans ces trois zones occidentales.

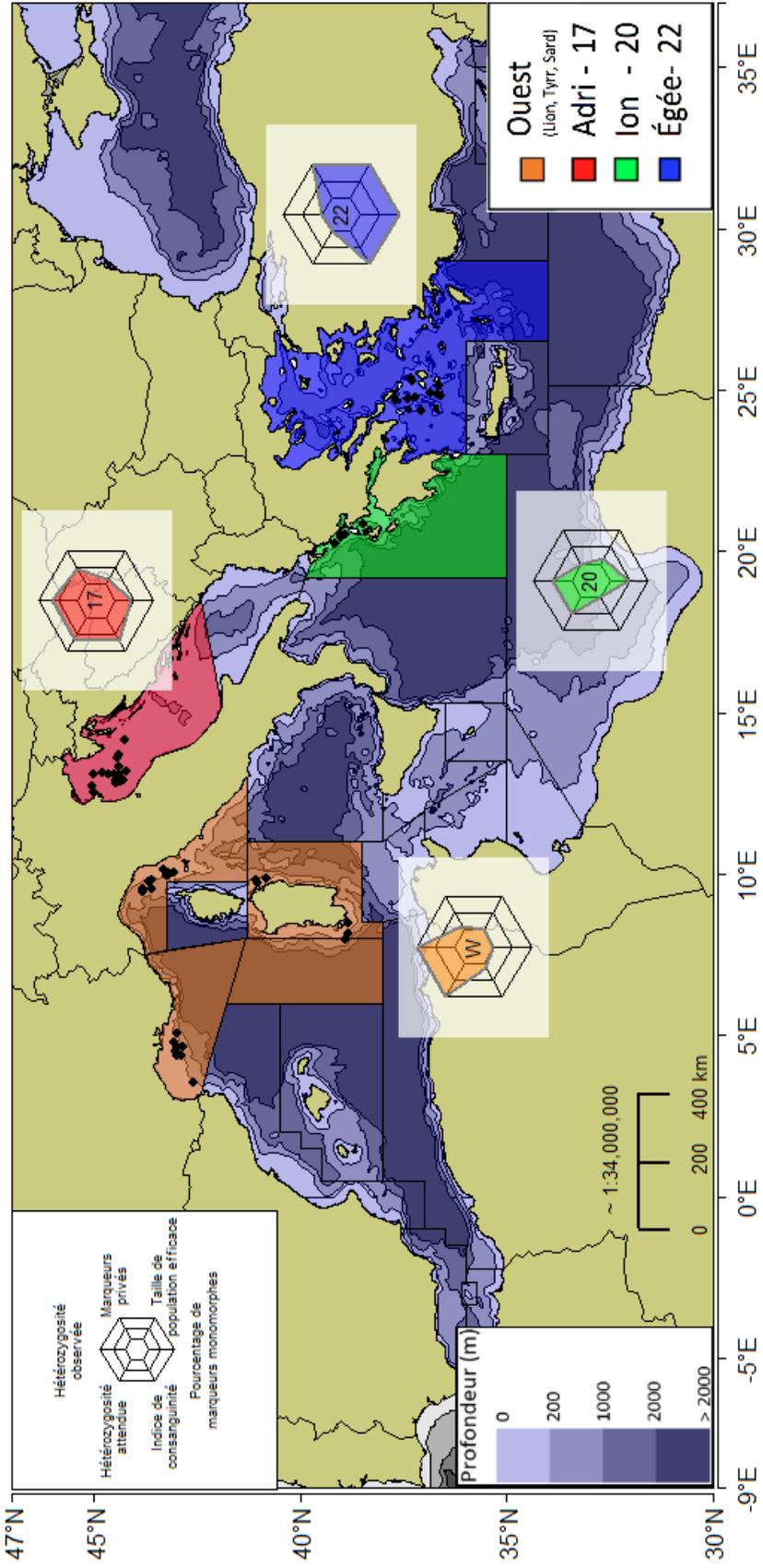


Figure 7: Graphiques comparatifs des indicateurs de diversité génétique étudiés pour les quatre populations inférées de raies bouclées échantillonnées en Méditerranée (échantillonnage rappelé par les points noirs)

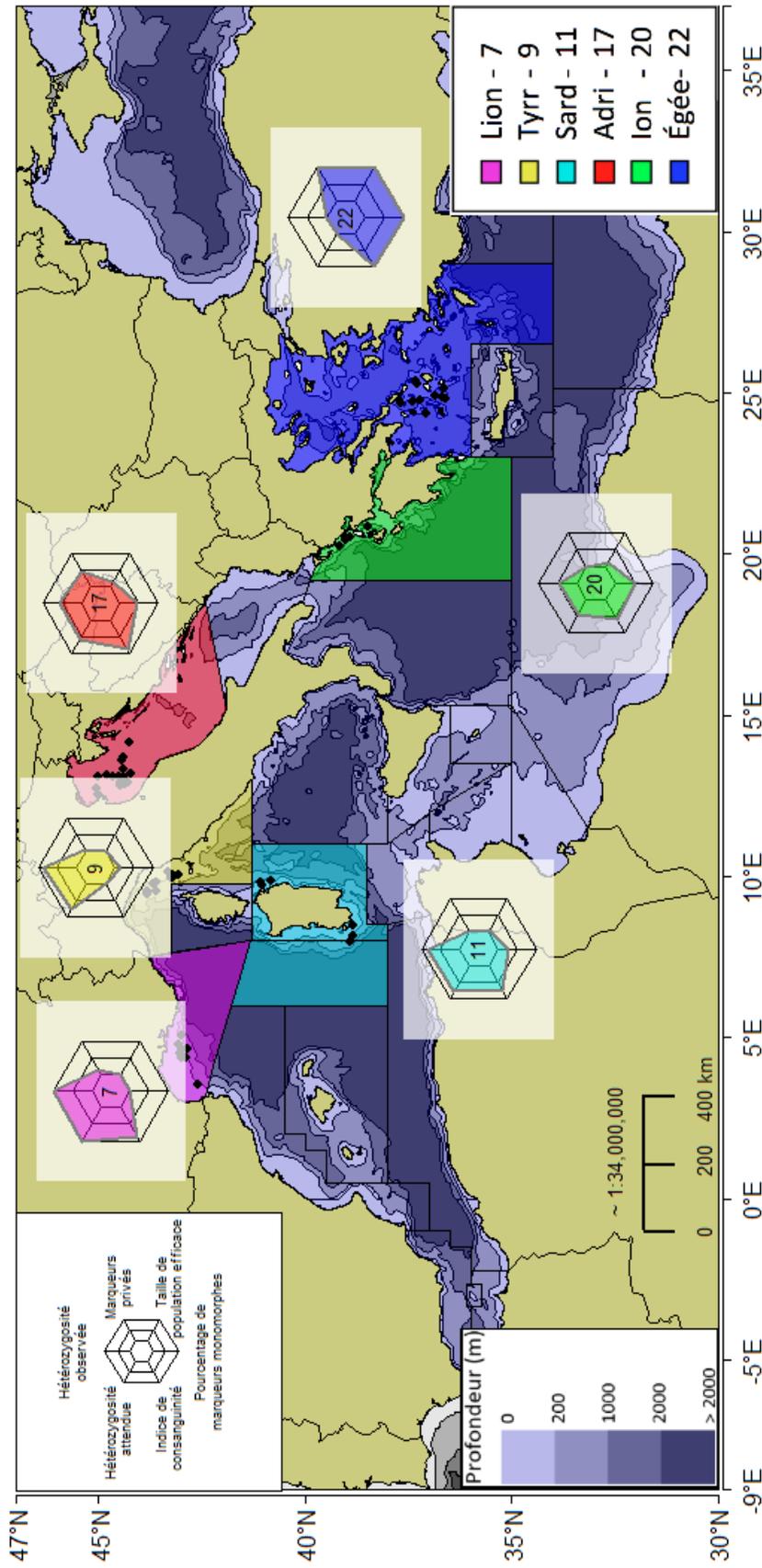


Figure 8: Graphiques comparatifs des indicateurs de diversité génétique étudiés pour les six zones d'échantillonnage de raies bouclées en Méditerranée (échantillonnage rappelé par les points noirs)

Les estimations des tailles de population efficaces ne montrent pas de variation est-ouest monotone comme les autres indicateurs. La plus grande taille de population efficace a été estimée pour la population de mer Égée, suivie de la population de mer Adriatique, la population occidentale et finalement celle de mer Ionienne (Table S7). La densité de répartition des tailles de population efficaces au travers des cent jeux de données ré-échantillonnés dans le cas de quatre populations inférées a permis d’observer que le mode de la densité de taille de population efficace de mer Égée est supérieur aux autres bien que la distribution montre une forte variabilité (Figure 9). Un constat similaire a été observé dans le cas de six populations suivant les zones d’échantillonnage, avec une bimodalité prononcée pour les populations du golfe du Lion ainsi que de la mer Tyrrhénienne nord tandis que la population de Sardaigne possède une distribution unimodale (Figure 10).

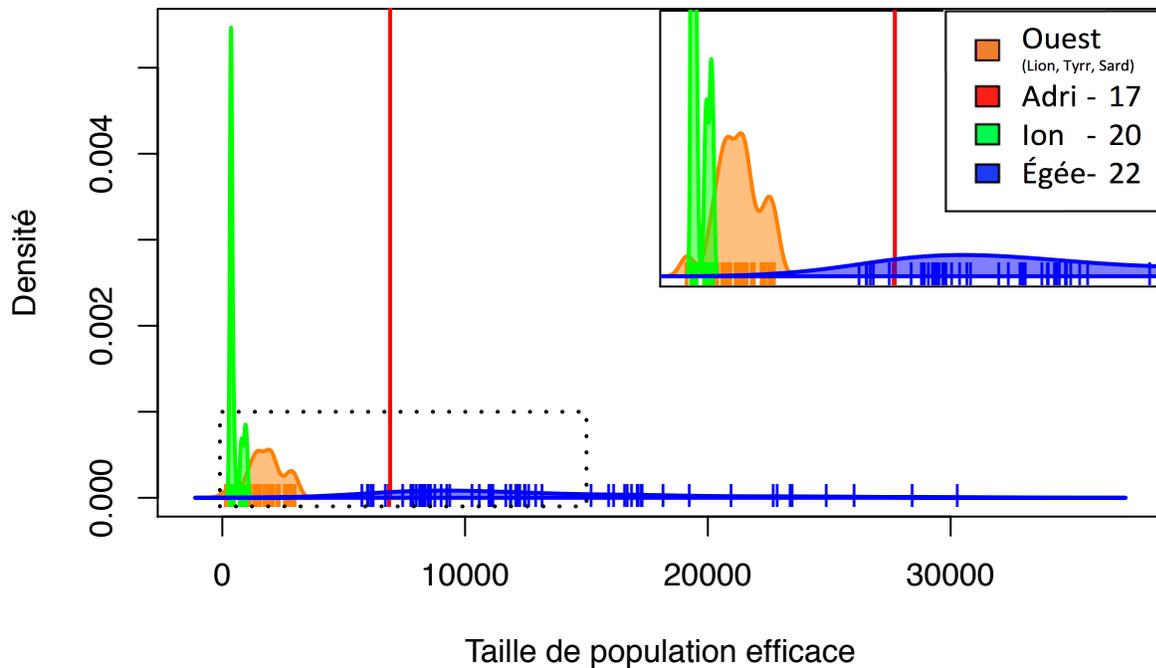


Figure 9: Densité des tailles de population efficaces (N_e) pour les quatre populations inférées de la raie bouclée en Méditerranée. Les distributions correspondent à l’estimation de N_e pour 100 jeux de données de 5147 SNPs ré-échantillonnés à hauteur de 30 individus par population sauf pour l’Adriatique qui comporte initialement 30 individus et pour laquelle une barre verticale a été reportée. Les marques verticales à la base des courbes représentent les estimations de N_e pour chaque jeu de données ré-échantillonné et le sous-graphique est un zoom sur l’encadré en pointillés noir excluant les valeurs extrêmes

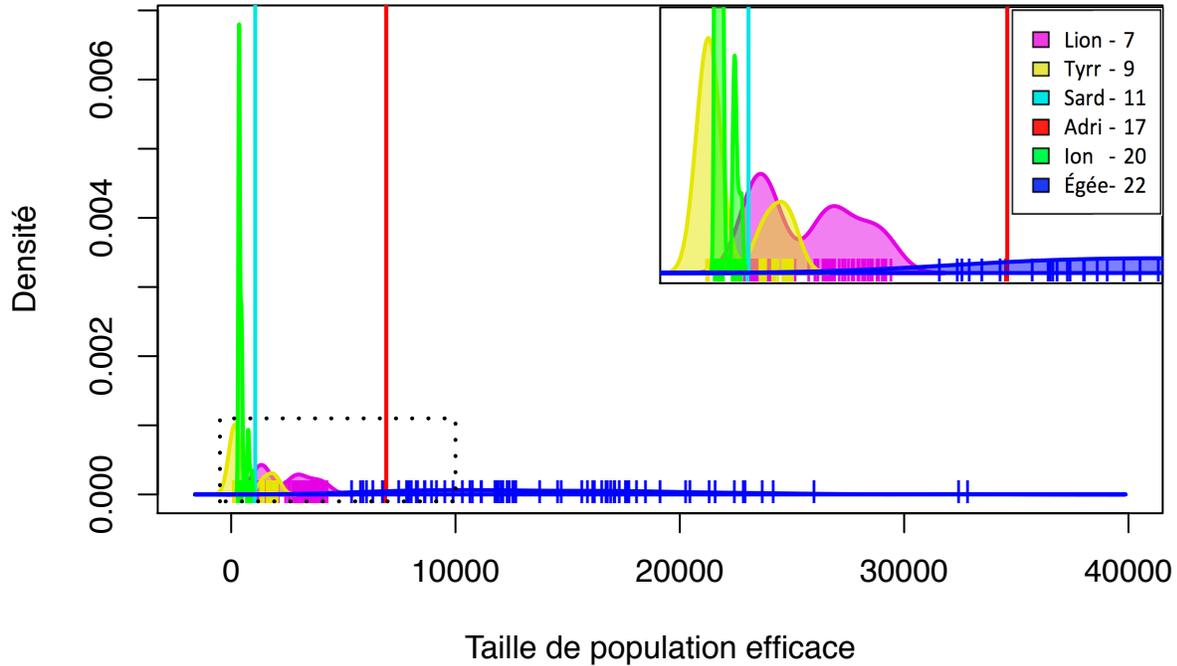


Figure 10: Densité des tailles de population efficaces (N_e) pour les six zones d'échantillonnage de la raie bouclée en Méditerranée. Les distributions correspondent à l'estimation de N_e pour 100 jeux de données de 5147 SNPs ré-échantillonnés à hauteur de 30 individus par population sauf pour l'Adriatique et la Sardaigne qui comportent initialement 30 individus et pour lesquelles une barre verticale a été reportée. Les marques verticales à la base des courbes représentent les estimations de N_e pour chaque jeu de données ré-échantillonné et le sous-graphique est un zoom sur l'encadré en pointillés noir excluant les valeurs extrêmes

4 Discussion

Cette étude de génétique des populations de la raie bouclée en Méditerranée portant sur plus de 5000 SNPs du génome nucléaire de plus de 200 individus a mis en évidence plusieurs résultats majeurs. En premier lieu, les individus se sont révélés très structurés génétiquement entre quatre zones distinctes de Méditerranée. Ce résultat permet d'affirmer que les six zones échantillonnées ne correspondent pas à autant de populations génétiquement distinctes. L'analyse de la structure génétique de population par DAPC a montré que les trois zones d'échantillonnage voisines du golfe du Lion, de mer Tyrrhénienne nord et de Sardaigne ne sont pas discernables tandis que les zones des mers Adriatique, Ionienne et Égée hébergent des populations génétiquement distinctes.

Les analyses de différenciation ont mis en évidence une distinction forte de la population de mer Égée suivie de la population occidentale puis des populations de mer Adriatique et Ionienne. Il est difficile de distinguer les populations de Méditerranée occidentale. L'analyse moléculaire de la variance montre que près de 95% de la variabilité génétique est présente au sein des populations et 5%

se répartit entre les populations. Cette faible part de variabilité inter-population semble suffisante pour différencier correctement les populations car les estimations de F_{ST} obtenues se situent dans des gammes de différenciation intermédiaire ($\sim 0.01 - 0.05$) à forte ($\sim 0.05 - 0.10$) (Haffray *et al.*, 2012).

Les différents indicateurs de diversité génétique mettent en évidence un gradient géographique longitudinal. Les indices d'hétérozygoté diminuent d'ouest en est alors que la tendance inverse apparaît pour la proportion de marqueurs monomorphes et privés. La plus importante taille de population efficace est estimée pour la population de mer Égée tandis que la population adjacente de mer Ionienne possède la plus faible taille de population efficace estimée, séparée d'un facteur 25 avec la population de mer Égée. Toutefois, il est attendu que les estimations de taille de population efficace soient fortement biaisées par l'utilisation de la méthode LDNe basée sur le déséquilibre de liaison ainsi que sur le fait qu'il est recommandé d'échantillonner environ 1% d'une population pour obtenir des estimations fiables (Marandel *et al.*, 2020) sans quoi il est attendu d'obtenir une majorité d'estimations négatives ou infinies, toutes deux exprimant une très grande taille de population efficace.

4.1 Structure et hiérarchisation des populations

Les individus de raie bouclée en Méditerranée semblent organisés en quatre populations avec une panmixie prononcée dans la partie occidentale, contrairement aux populations des mers Adriatique, Ionienne et Égée (Figure 5). Cette structuration possède une différenciation significative au seuil établi (0.01 ± 0.002) sur un jeu de données constitué de 5 marqueurs microsatellites et 483 raies bouclées échantillonnées sur 14 zones en Atlantique nord-est (Chevolot *et al.*, 2006) car un F_{ST} de 0.008 est suffisant pour obtenir une hiérarchie en quatre populations (Figure 6). Toutefois, il convient d'être prudent quant à la comparaison de F_{ST} entre des études basées sur des marqueurs génétiques différents (Lemopoulos *et al.*, 2019) car les marqueurs microsatellites semblent surestimer les indicateurs utilisés du fait d'une moindre corrélation à la diversité génétique que les SNPs (Athrey *et al.*, 2018). Le fait que les intervalles de confiance autour des F_{ST} se recoupent uniquement en considérant six populations et non quatre populations conforte l'idée que les quatre populations inférées reflètent davantage la réalité.

La géographie de la mer Méditerranée pourrait expliquer en partie la forte séparation génétique entre la population de mer Égée et notamment sa plus proche voisine de mer Ionienne. Les échantillonnages n'ont pas été réalisés sur l'ensemble de l'aire couverte par chacune de ces zones mais uniquement de part et d'autre de la presqu'île du Péloponnèse (Figure 2) et le fort gradient bathymétrique des côtes grecques pourrait former une entrave à la dispersion des raies bouclées. La péninsule italienne séparant les bassins oriental et occidental de la Méditerranée pourrait expliquer la différenciation marquée entre la population occidentale et l'ensemble des autres populations par un flux migratoire très réduit créant une absence de connexion entre les deux bassins pour la raie bouclée. Ce faible taux de migration peut être dû à une combinaison complexe de facteurs environnementaux tels que la bathymétrie, la présence de courants, l'absence de ressources trophiques ou encore le manque d'habitat. Les golfes de Squillace et de Taranto ont été identifiés comme point de contact biologique entre les bassins oriental et occidental de Méditerranée (Sá-Pinto *et al.*, 2010) et pourraient former un puits d'hybridation dans lequel les individus hybrides issus de parents provenant de chaque bassin sont moins adaptés (Barton, 1980). Ces hypothèses suggèrent que les péninsules hellénique et italienne pourraient former des barrières géographiques mais la puissance statistique de cette étude est bien trop faible pour estimer le degré de résistance à la dispersion en raison d'un nombre de zones d'échantillonnage

trop faible. La présence de cordons côtiers peu profonds et accessibles à la raie bouclée reliant les zones du bassin occidental de la Méditerranée pourrait favoriser la dispersion et permettre une reproduction panmictique des individus occupant ce bassin. Depuis l'expansion de la raie bouclée (\sim -400.000 ans), les glaciations quaternaires de Riss puis de Würm et les interglaciaires correspondants ont provoqué une série de régressions et transgressions marines au niveau de la mer Adriatique peu profonde. La plus faible différenciation génétique entre les populations de l'Adriatique et de la mer Ionienne porte certainement la signature phylogéographique d'une recolonisation de l'Adriatique par des individus issus de zones adjacentes comme la mer Ionienne. Malgré l'absence de barrière évidente entre ces deux mers, les migrations entre ces populations ne semblent pas compenser les effets de la dérive génétique. Une étude des populations de crénilabre paon (*Symphodus tinca*) basée sur un panel de SNPs neutres et sous sélection a relevé une structuration génétique entre les individus de l'Adriatique et de la mer Ionienne plus forte par l'analyse des SNPs sous sélection que par l'analyse des SNPs neutres (Carreras *et al.*, 2017). Ce constat pourrait être expliqué par des processus d'adaptation locale des populations à des environnements aux conditions différentes peu favorables aux individus non-adaptés (Gaggiotti *et al.*, 2009). Une situation similaire serait peut-être à l'origine de la structuration relevée dans cette étude dans le cas de la raie bouclée et l'analyse de marqueurs sous sélection pourrait renvoyer une différenciation plus forte des populations de l'Adriatique et de la mer Égée.

Un échantillonnage plus dense de la Méditerranée ne suivant pas systématiquement le découpage en GSA proposé par la Commission Générale des Pêches en Méditerranée mais plutôt réparti de manière homogène le long des côtes serait utile afin d'obtenir une vision plus précise de la réelle structuration de la raie bouclée en Méditerranée et de mieux identifier les entraves à la dispersion entre populations. Échantillonner les côtes des îles Baléares, d'Andalousie et d'Algérie permettrait de tester l'existence d'une population panmictique occupant l'ensemble du bassin occidental de la Méditerranée. L'hypothèse de la présence d'une séparation nette au sud de l'Italie pourrait également être testée par l'échantillonnage d'individus dans les golfes de Squillace et de Taranto. Le faible taux de migration entre les populations de l'Adriatique et de la mer Ionienne pourrait être étayé par l'échantillonnage de la partie sud de l'Adriatique. Le jeu de données pourrait être complété par des individus échantillonnés à Chypre, le long des côtes du Levant mais aussi en mer Noire. Il serait également utile d'analyser les marqueurs identifiés comme étant sous sélection naturelle (Table S1) pour observer une structuration de la population sur des critères d'adaptation au milieu de vie comme il est potentiellement le cas entre l'Adriatique et la mer Ionienne.

La variabilité génétique est répartie pour près de 95% au niveau intra-population ce qui indique une forte diversité génétique entre les individus d'une même population (Table 3) bien que la différenciation inter-population marquée des dendrogrammes (Figure 6) traduise intuitivement une forte variabilité génétique inter-population. Ce résultat, *a priori* contre-intuitif, vient du fait que la composante inter-population issue de la décomposition additive de la variabilité génétique totale n'est mathématiquement pas un bon indicateur de la différenciation génétique et réciproquement (Jost, 2008).

L'une des prochaines étapes à entreprendre dans la continuité de cette étude serait de quantifier le taux de migration entre populations et leur direction en utilisant par exemple une approche basée sur le coalescent structuré (Vaughan *et al.*, 2014).

4.2 Caractérisation de la diversité génétique des populations

Les indicateurs de diversité génétique des populations de raie bouclée en Méditerranée présentent des gammes qui sont retrouvées chez de nombreux organismes terrestres et marins (De Lorenzis *et al.*, 2019 - Edea *et al.*, 2017) : les variations d'hétérozygotie attendue H_E sont significatives et elle atteint sa valeur maximale (0,309) au sein de la population occidentale indiquant une plus grande richesse allélique au sein des individus de cette population tandis que les individus de mer Égée présentent la valeur minimale (0,269). Les estimations des indices de consanguinité F_{IS} renvoient des valeurs faibles (< 0.003) sur l'ensemble des populations (Zhang *et al.*, 2020) et leur variation entre populations n'est pas significative indiquant une variabilité élevée entre loci au travers des populations.

Le pourcentage de marqueurs monomorphes élevé au sein de la population de mer Égée (12,74 contre 3,02-7,71 pour les autres populations) indique que sa variabilité allélique est plus réduite. Le nombre maximal de SNPs privés est également retrouvé dans la population de mer Égée (258,15) et étaye encore davantage un isolement reproducteur de ses individus et un faible taux d'immigration. En revanche, la population occidentale présente le moins de SNPs privés (13,3) ce qui conforte l'hypothèse d'une immigration plus importante. L'ensemble des cinq indicateurs utilisés présente un gradient traduisant une plus grande diversité génétique dans la population occidentale conjointe à un plus fort taux d'immigration tandis que les populations plus centrales de l'Adriatique et de mer Ionienne présentent une diversité génétique moindre, qui se réduit encore davantage en mer Égée.

Les variations écologiques induites par les périodes de glaciation quaternaires ont hypothétiquement créé un refuge pour la raie bouclée en Méditerranée car les individus y présentent un haplotype mitochondrial unique et considéré comme la version ancestrale principale (Chevolot *et al.*, 2006). La raie bouclée serait donc installée en Méditerranée depuis son expansion (~ -400.000 ans) ce qui a pu favoriser l'apparition d'écarts en terme de diversité génétique sans occurrence de front de recolonisation post-glaciation.

L'apport de nouveau matériel génétique par des individus ayant migré depuis l'Atlantique est une hypothèse pouvant expliquer en partie les gradients longitudinaux observés sur les différents indicateurs de diversité génétique (Figure 7) (Rolland *et al.*, 2007). Cette hypothèse expliquerait également ce pourquoi les populations se structurent selon leur longitude d'après la première fonction discriminante de la DAPC (Figure 4). L'analyse des individus échantillonnés en Atlantique conduirait à une meilleure compréhension de l'expansion de la raie bouclée à une échelle temporelle plus vaste mais aussi du degré d'isolation génétique entre Méditerranée et Atlantique supposé important par la présence du front Almeria-Oran (Rohling *et al.*, 2009).

La taille de population efficace permet d'obtenir des informations sur les événements que les populations ont subi par le passé comme des goulots d'étranglement plus ou moins intenses (Thuillet *et al.*, 2005) mais aussi sur la façon dont la diversité génétique évoluera au fil des générations futures (Frankham, 2005). De façon générale, la diversité génétique est corrélée positivement à la taille de population efficace. Plus précisément, la corrélation entre perte de diversité génétique et taille de population efficace n'est pas linéaire mais approximativement exponentielle (Reed & Frankham, 2003). La perte de diversité génétique est d'autant plus rapide que la taille de population efficace est faible car la dérive génétique et le taux de fixation d'allèles sont plus forts. Ainsi, la taille de population efficace de la population de mer Égée (9299) est supposée indiquer une diminution de sa diversité génétique plus lente que dans les autres populations et notamment celle de mer

Ionienne possédant la taille de population efficace la plus basse (370). Ce résultat est surprenant dans la mesure où la population de mer Égée cumule la plus faible diversité génétique et la plus grande taille de population efficace. La forte diversité d'habitats offert par l'archipel Égéen pourrait favoriser le développement d'une sous-structuration génétique plus marquée que pour les autres zones d'échantillonnage ne disposant pas de la même complexité d'habitat comme la mer Adriatique. Échantillonner une zone possédant une telle sous-structuration pourrait entraîner une estimation de taille de population efficace élevée ainsi qu'une moindre diversité si des phénomènes de consanguinité apparaissent localement. Une forte sous-estimation des tailles de population efficaces, pouvant aller jusqu'à 99% (Waples *et al.*, 2016), est toutefois attendue malgré la grande majorité d'estimations positives dans des gammes plausibles (Marandel *et al.*, 2019).

4.3 Implications pour la gestion et la conservation de la raie bouclée en Méditerranée

Les populations génétiques de raie bouclée de mer Méditerranée présentent un découpage différent du système géographique de gestion des stocks de cette région. Le bassin occidental de Méditerranée semble contenir une vaste population panmictique sur toute la zone côtière s'étendant de la France à la Sardaigne en longeant la côte italienne (Figure 5). Le bassin oriental présente des populations suivant davantage le système de gestion des stocks car ce dernier semble mieux séparer la diversité d'habitat existante au niveau des barrières physiques et phylogénétiques identifiées et discutées précédemment. L'unité de gestion de la raie bouclée doit être calibrée sur le découpage des populations de cette espèce en Méditerranée afin d'éviter de considérer des individus qui appartiennent à une même population de manière indépendante et inversement. Cette étude propose un découpage génétique nécessitant d'être complété par d'autres analyses afin de constituer une proposition robuste de découpage démographique des populations de la raie bouclée en Méditerranée.

La population présente dans le bassin occidental de Méditerranée semble disposer d'une diversité génétique importante, ce qui conduit à la considérer comme mieux adaptée à réagir face à des perturbations comme l'activité de pêche contrairement à la population de mer Égée dont la moindre diversité génétique la rend davantage vulnérable à des perturbations. Toutefois, ces résultats sont à considérer sous réserve que les estimations de taille de population efficace, même analysées relativement entre populations, soient indicatrices d'un réel signal génétique. Sous cette considération, les estimations de taille de population efficace indiquent que la population occidentale ainsi que la population de mer Ionienne disposent d'une plus faible quantité de reproducteurs et auront plus de mal à stabiliser leur diversité génétique sur le long terme (Figure 9). La population de mer Égée semble être en capacité de subir une perte de diversité génétique plus lente, pouvant lui assurer une meilleure persistance sur le long terme si sa faible diversité génétique actuelle ne la condamne pas rapidement dans un contexte difficile notamment lié à la migration lessepsienne aux impacts écologiques avérés sur le bassin oriental de Méditerranée (Zakaria, 2015). En outre, une mise en parallèle des tendances temporelle des estimations d'abondance de raie bouclée par unité de gestion et des indicateurs de diversité génétique et tailles de population efficaces permettrait de tester de potentielles corrélations. Il serait alors possible d'émettre des hypothèses quant à l'impact de l'activité de pêche sur la diversité génétique de la raie bouclée en Méditerranée. Les traits d'histoire de vie d'une espèce, notamment sa longévité et sa fécondité, semblent également prédire la diversité génétique (Romiguier *et al.*, 2014). Prendre en compte les traits d'histoire de vie de la raie bouclée en Méditerranée serait donc crucial pour mieux évaluer les mesures de protection les plus à même de maintenir la diversité génétique des populations sur le long terme.

5 Conclusion

La raie bouclée est une espèce capturée le plus souvent en tant que prise accessoire en mer Méditerranée notamment par les chalutiers de fond (Dulvy *et al.*, 2014). Son faible intérêt commercial en fait une espèce peu considérée bien qu'elle possède une maturité relativement tardive rendant difficile le renouvellement de sa démographie, à l'instar de nombreux éla-smobran-ches (Goldman & Cailliet, 2004). Les captures de raie bouclée n'ont pas cessé de diminuer bien que la pression de pêche des chalutiers côtiers continue de croître dans les eaux méditerranéennes (Russo *et al.*, 2019 - FAO, 2012). Cette situation signifie généralement que le stock considéré est en déclin et menacé par une activité de pêche intense (Field *et al.*, 2009). La présente étude génétique de la raie bouclée en Méditerranée a permis d'identifier un découpage en quatre populations de cette espèce au sein des six zones échantillonnées. La hiérarchisation de ces populations a été étudiée et l'origine de leur différenciation remontant possiblement aux dernières glaciations quaternaires a été discutée. La diversité génétique relevée au sein de ces populations est hétérogène et l'ensemble des indicateurs utilisés semblent suivre un gradient longitudinal avec une moindre diversité génétique dans le bassin oriental de la Méditerranée. Les tailles de population efficaces estimées ajoutent une composante temporelle en terme d'évolution de la diversité génétique des populations au fil des générations. Dans un objectif de conservation, il serait nécessaire d'approfondir ces premiers résultats en étendant et en densifiant l'échantillonnage, en tenant compte des traits d'histoire de vie de la raie bouclée ainsi qu'en mettant en lumière les liens potentiels entre indicateurs de diversité génétique et tendances temporelles des estimations d'abondance des populations. En l'absence de ces éléments de compréhension, il est difficile d'évaluer l'état de santé des populations de raie bouclée et l'impact de l'activité de pêche sur celles-ci en Méditerranée.

Bibliographie

1. Abis, S. (2018). *Food Security and Conflicts in the Mediterranean Region*. Strateg. Sect. Econ. Territ. 274–277.
2. Amos, W. & Hoffman, J. I. (2010). *Evidence that two main bottleneck events shaped modern human genetic diversity*. Proc. R. Soc. B Biol. Sci. **277** : 131–137.
3. Athrey, G., Faust, N., Charlotte Hieke, A. S. & Brisbin, I. L. (2018). *Effective population sizes and adaptive genetic variation in a captive bird population*. PeerJ. **2018** : 1–21.
4. Baird, N. A., Etter, P. D., Atwood, T. S., Currey, M. C., Shiver, A. L., Lewis, Z. A., Selker, E. U., Cresko, W. A. & Johnson, E. A. (2008). *Rapid SNP Discovery and Genetic Mapping Using Sequenced RAD Markers*. PLoS One. **3** : e3376.
5. Balding, D. J. (2003). *Likelihood-based inference for genetic correlation coefficients*. Theor. Popul. Biol. **63** : 221–230.
6. Barton, N. H. (1980). *The Hybrid Sink Effect*. Comp. Educ. **44** : 277–278.
7. Bernas, R., Poćwierz-Kotus, A., Árnayasi, M., Kent, M. P., Lien, S. & Wenne, R. (2020). *Genetic Differentiation in Hatchery and Stocked Populations of Sea Trout in the Southern Baltic: Selection Evidence at SNP Loci*. Genes. **11** : 184.
8. Bertrand, J., Leonori, I., Dremière, P. Y. & Cosimi, G. (2002). *Depth trajectory and performance of a trawl used for an international bottom trawl survey in the Mediterranean*. Sci. Mar. **66** : 169–182.
9. Bourlat, S. J., Borja, A., Gilbert, J., Taylor, M. I., Davies, N., Weisberg, S. B., Griffith, J. F., Lettieri, T., Field, D., Benzie, J., Glöckner, F. O., Rodríguez-Ezpeleta, N., Faith, D. P., Bean, T. P. & Obst, M. (2013). *Genomics in marine monitoring: New opportunities for assessing marine health status*. Mar. Pollut. Bull. **74** : 19–31.
10. Cardinale, M., Osio, G. C. & Scarcella, G. (2017). *Mediterranean Sea: A Failure of the European Fisheries Management System*. Front. Mar. Sci. **4** : 18.
11. Carreras, C., Ordóñez, V., Zane, L., Kruschel, C., Nasto, I., Macpherson, E. & Pascual, M. (2017). *Population genomics of an endemic Mediterranean fish: differentiation by fine scale dispersal and adaptation*. Sci. Rep. **7** : 43417.
12. Cashion, M. S., Bailly, N. & Pauly, D. (2019). *Official catch data underrepresent shark and ray taxa caught in Mediterranean and Black Sea fisheries*. Mar. Policy. **105** : 1–9.
13. Charlesworth, B. (2009). *Effective population size and patterns of molecular evolution and variation*. Nat. Rev. Genet. **10** : 195–205.
14. Chattopadhyay, B., Garg, K. M. & Ramakrishnan, U. (2014). *Effect of diversity and missing data on genetic assignment with RAD-Seq markers*. BMC Res. Notes. **7** : 841.
15. Chevolut, M., Ellis, J. R., Hoarau, G., Rijnsdorp, A. D., Stam, W. T. & Olsen, J. L. (2006). *Population structure of the thornback ray (*Raja clavata* L.) in British waters*. J. Sea Res. **56** : 305–316.

16. Chevolot, M., Ellis, J. R., Rijnsdorp, A. D., Stam, W. T. & Olsen, J. L. (2007). *Multiple Paternity Analysis in the Thornback Ray Raja clavata L.* J. Hered. **98** : 712–715.
17. Chevolot, M., Hoarau, G., Rijnsdorp, A. D., Stam, W. T. & Olsen, J. L. (2006). *Phylogeography and population structure of thornback rays (Raja clavata L., Rajidae).* Mol. Ecol. **15** : 3693–3705.
18. Couch, J. (1862). *A history of the fishes of the British Islands.* J. Couch. **1** : 26.
19. Das, D. & Afonso, P. (2017). *Review of the Diversity, Ecology, and Conservation of Elasmobranchs in the Azores Region, Mid-North Atlantic.* Front. Mar. Sci. **4** : 1–19.
20. De Lorenzis, G., Mercati, F., Bergamini, C., Cardone, M. F., Lupini, A., Mauceri, A., Caputo, A. R., Abbate, L., Barbagallo, M. G., Antonacci, D., Sunseri, F. & Brancadoro, L. (2019). *SNP genotyping elucidates the genetic diversity of Magna Graecia grapevine germplasm and its historical origin and dissemination.* BMC Plant Biol. **19** : 1–15.
21. Do, C., Waples, R. S., Peel, D., Macbeth, G. M., Tillett, B. J. & Ovenden, J. R. (2014). *NeEstimator v2: Re-implementation of software for the estimation of contemporary effective population size (Ne) from genetic data.* Mol. Ecol. Resour. **14** : 209–214.
22. Duforet-Frebourg, N., Bazin, E. & Blum, M. G. B. (2014). *Genome Scans for Detecting Footprints of Local Adaptation Using a Bayesian Factor Model.* Mol. Biol. Evol. **31** : 2483–2495.
23. Dulvy, N. K., Fowler, S. L., Musick, J. A., Cavanagh, R. D., Kyne, P. M., Harrison, L. R., Carlson, J. K., Davidson, L. N., Fordham, S. V. & Francis, M. P. *et al.* (2014). *Extinction risk and conservation of the world's sharks and rays.* Elife. **3** : e00590.
24. Edea, Z., Dessie, T., Dadi, H., Do, K. T. & Kim, K. S. (2017). *Genetic diversity and population structure of Ethiopian sheep populations revealed by high-density SNP markers.* Front. Genet. **8** : 218.
25. Ellis, J. (2016). *Raja clavata assessment.* IUCN Red List Threat. Species.
26. Enajjar, S., Saidi, B. & Bradai, M. (2015). *The Gulf of Gabes (Central Mediterranean Sea): a nursery area for sharks and batoids (Chondrichthyes: Elasmobranchii).* Cah. Biol. Mar. **56** : 143–150.
27. Excoffier, L. (2006). *Génétique des populations appliquée aux données moléculaires.* Suivre : http://genet.univ-tours.fr/EXCOFFIER/Laurent/GMDP_drift.htm.
28. Excoffier, L., Smouse, P. E. & Quattro, J. M. (1992). *Analysis of Molecular Variance Inferred From Metric Distances Among DNA Haplotypes: Application to Human Mitochondrial DNA Restriction Data.* Genet. Soc. Am. **131** : 479–491.
29. FAO. (2012). *Elasmobranchs of the Mediterranean and Black Sea : Status, Ecology and Biology.* Stud. Rev.
30. FAO. *Raja clavata Fact Sheet.* Suivre : <http://www.fao.org/fishery/species/2851/en>.
31. Field, I. C., Meekan, M. G., Buckworth, R. C. & Bradshaw, C. J. A. (2009). *Susceptibility of Sharks, Rays and Chimaeras to Global Extinction.* Adv. Mar. Biol. **56** : 275–363.
32. Foll, M. (2012). *BayeScan v2.1 User Manual.* 1–10.
33. Foll, M. & Gaggiotti, O. (2008). *A Genome-Scan Method to Identify Selected Loci Appropriate*

- for Both Dominant and Codominant Markers: A Bayesian Perspective. *Genetics*. **180** : 977–993.
34. Follesa, M. C., Marongiu, M. F., Zupa, W., Bellodi, A., Cau, A., Cannas, R., Colloca, F., Djurovic, M., Isajlovic, I. & Jadaud, A. *et al.* (2019). *Spatial variability of Chondrichthyes in the northern Mediterranean*. *Sci. Mar.* **83** : 1–21.
35. Fortibuoni, T., Borme, D., Franceschini, G., Giovanardi, O. & Raicevich, S. (2016). *Common, rare or extirpated? Shifting baselines for common angelshark, Squatina squatina (Elasmobranchii: Squatinidae), in the Northern Adriatic Sea (Mediterranean Sea)*. *Hydrobiologia*. **772** : 247–259.
36. Frankham, R. (2005). *Genetics and extinction*. *Biol. Conserv.* **126** : 131–140.
37. Gaggiotti, O. E., Bekkevold, D., Jørgensen, H. B. H., Foll, M., Carvalho, G. R., Andre, C. & Ruzzante, D. E. (2009). *Disentangling the effects of evolutionary, demographic, and environmental factors influencing genetic structure of natural populations: Atlantic Herring as a case study*. *Evolution*. **63** : 2939–2951.
38. Giovos, I., Arculeo, M., Doumpas, N., Katsada, D., Maximiadi, M., Mitsou, E., Paravas, V., Naasan Aga-Spyridopoulou, R., Stoilas, V.-O., Tiralongo, F., Tsamadias, I. E., Vecchioni, L. & Moutopoulos, D. K. (2020). *Assessing multiple sources of data to detect illegal fishing, trade and mislabelling of elasmobranchs in Greek markets*. *Mar. Policy*. **112** : 103730.
39. Goldman, K. & Cailliet, G. (2004). *Age Determination and Validation in Chondrichthyan Fishes*. *Biol. Sharks their relat.* 399–447.
40. Goudet, J. (2005). *hierfstat , a package for R to compute and test hierarchical F -statistics*. **2** : 184–186. Suivre : <https://cran.r-project.org/web/packages/hierfstat/hierfstat.pdf>.
41. Haffray, P., Malha, R., Ould Taleb Sidi, M., Prista, N., Hassan, M., Castelnaud, G., Karahan-Nomm, B., Gamsiz, K., Sadek, S., Bruant, J. S., Balma, P. & Bonhomme, F. (2012). *Very high genetic fragmentation in a large marine fish, the meagre Argyrosomus regius (Sciaenidae, Perciformes): Impact of reproductive migration, oceanographic barriers and ecological factors*. *Aquat. Living Resour.* **25** : 173–183.
42. Heithaus, M. R., Frid, A., Wirsing, A. J. & Worm, B. (2008). *Predicting ecological consequences of marine top predator declines*. *Trends Ecol. Evol.* **23** : 202–210.
43. Holden, M. J. (1975). *The fecundity of Raja clavata in British waters*. *ICES J. Mar. Sci.* **36** : 110–118.
44. Holgersson, M. (1978). *The limited value of cophenetic correlation as a clustering criterion*. *Pattern Recognit.* **10** : 287–295.
45. Holm, S. (1979). *A Simple Sequentially Rejective Multiple Test Procedure*. *Scand J Stat.* **6** : 65–70.
46. Hunter, E., Buckley, A. A., Stewart, C. & Metcalfe, J. D. (2005). *Migratory behaviour of the thornback ray, raja clavata , in the southern north sea*. *J. Mar. Biol. Assoc. United Kingdom*. **85** : 1095–1105.
47. Jombart, T. & Collins, C. (2015). *A tutorial for Discriminant Analysis of Principal Components (DAPC) using adegenet 1.3-4*. Suivre : <https://cran.r-project.org/web/packages/adegenet/adegenet.pdf>.

48. Jombart, T., Devillard, S. & Balloux, F. (2010). *Discriminant analysis of principal components: a new method for the analysis of genetically structured populations*. *BMC Genet.* **11** : 94.
49. Jost, L. (2008). *G ST and its relatives do not measure differentiation*. *Mol. Ecol.* **17** : 4015–4026.
50. Kadri, H., Marouani, S., Saïdi, B., Bradai, M. N., Bouaïn, A. & Morize, E. (2014). *Age, growth, sexual maturity and reproduction of the thornback ray, Raja clavata (L.), of the Gulf of Gabès (south-central Mediterranean Sea)*. *Mar. Biol. Res.* **10** : 416–425.
51. Kirk, H. & Freeland, J. R. (2011). *Applications and Implications of Neutral versus Non-neutral Markers in Molecular Ecology*. *Int. J. Mol. Sci.* **12** : 3966–3988.
52. Last, P., White, W., De Carvalho, M., Séret, B., Stehmann, M. & Naylor, G. (2016). *Rays of the World*. P. Last & G. Yearsley. **1** : 1–10.
53. Le Cam, S., Bidault, A., Charrier, G., Cornette, F., Lamy, J.-B., Lapègue, S., Lorance, P., Marandel, F. & Trenkel, V. (2019). *SNPs for thornback ray Raja clavata, validated from genotyped individuals*. SEANOE. Suivre : <https://www.seanoe.org/data/00593/70546/>.
54. Leitwein, M., Duranton, M., Rougemont, Q., Gagnaire, P.-A. & Bernatchez, L. (2020). *Using Haplotype Information for Conservation Genomics*. *Trends Ecol. Evol.* **35** : 245–258.
55. Lemopoulos, A., Prokkola, J. M., Uusi-Heikkilä, S., Vasemägi, A., Huusko, A., Hyvärinen, P., Koljonen, M., Koskiniemi, J. & Vainikka, A. (2019). *Comparing RADseq and microsatellites for estimating genetic diversity and relatedness — Implications for brown trout conservation*. *Ecol. Evol.* **9** : ece3.4905.
56. Liggins, L., Treml, E. A. & Riginos, C. (2019). *Seascape Genomics: Contextualizing Adaptive and Neutral Genomic Variation in the Ocean Environment*. Dans : *Population Genomics: Marine Organisms*. M. Oleksiak & O. Rajora. 171–218.
57. Linck, E. & Battey, C. J. (2019). *Minor allele frequency thresholds strongly affect population structure inference with genomic data sets*. *Mol. Ecol. Resour.* **19** : 639–647.
58. Lonsinger, R. C., Adams, J. R. & Waits, L. P. (2018). *Evaluating effective population size and genetic diversity of a declining kit fox population using contemporary and historical specimens*. *Ecol. Evol.* **8** : 12011–12021.
59. Lotterhos, K. E. & Whitlock, M. C. (2015). *The relative power of genome scans to detect local adaptation depends on sampling design and statistical method*. *Mol. Ecol.* **24** : 1031–1046.
60. Lotterhos, K. E. & Whitlock, M. C. (2014). *Evaluation of demographic history and neutral parameterization on the performance of F ST outlier tests*. *Mol. Ecol.* **23** : 2178–2192.
61. Luu, K., Bazin, E. & Blum, M. G. B. (2017). *pcadapt : an R package to perform genome scans for selection based on principal component analysis*. *Mol. Ecol. Resour.* **17** : 67–77.
62. Luzzatto, L. (2012). *Sickle Cell Anaemia and Malaria*. *Mediterr. J. Hematol. Infect. Dis.* **4** : e2012065.
63. Marandel, F., Charrier, G., Lamy, J., Le Cam, S., Lorance, P. & Trenkel, V. M. (2020). *Estimating effective population size using RADseq: Effects of SNP selection and sample size*. *Ecol. Evol.* **10** : 1929–1937.

64. Marandel, F., Lorance, P., Berthel , O., Trenkel, V. M., Waples, R. S. & Lamy, J.-B. (2019). *Estimating effective population size of large marine populations, is it feasible?* Fish Fish. **20** : 189–198.
65. McEachran, J. D. & Dunn, K. A. (1998). *Phylogenetic Analysis of Skates, a Morphologically Conservative Clade of Elasmobranchs (Chondrichthyes: Rajidae)*. Copeia. **1998** : 271–290.
66. M rigot, B., Durbec, J.-P. & Gaertner, J.-C. (2010). *On goodness-of-fit measure for dendrogram-based analyses*. Ecology. **91** : 1850–1859.
67. Mittell, E. A., Nakagawa, S. & Hadfield, J. D. (2015). *Are molecular markers useful predictors of adaptive potential?* Ecol. Lett. **18** : 772–778.
68. O’Brien, S. J. (2008). *Genetic Diversity – Understanding Conservation at Genetic Levels. Dans : Conservation Biology*. F. Van Dyke. 153–184.
69. OSPAR. (2010). *Background Document for Thornback ray Raja clavata*. Biodiversity Series. **1** : 18.
70. Ozcan, E. & Basusta, N. (2018). *Preliminary study on age, growth and reproduction of Mustelus mustelus (Elasmobranchii: Carcharhiniformes: Triakidae) inhabiting the Gulf of Iskenderun, north-eastern Mediterranean Sea*. Acta Ichthyol. Piscat. **48** : 27–36.
71. Ramos, J. E., Pecl, G. T., Moltschaniwskyj, N. A., Semmens, J. M., Souza, C. A. & Strugnell, J. M. (2018). *Population genetic signatures of a climate change driven marine range extension*. Sci. Rep. **8** : 9558.
72. R Dev. Core Team. (2020). *R: A language and environment for statistical computing*. Suivre : <https://www.r-project.org/>.
73. Reed, D. H. & Frankham, R. (2003). *Correlation between fitness and genetic diversity*. Conserv. Biol. **17** : 230–237.
74. Reed, E., Nunez, S., Kulp, D., Qian, J., Reilly, M. P. & Foulkes, A. S. (2015). *A guide to genome-wide association analysis and post-analytic interrogation*. Stat. Med. **34** : 3769–3792.
75. Reiss, H., Hoarau, G., Dickey-Collas, M. & Wolff, W. J. (2009). *Genetic population structure of marine fish: mismatch between biological and fisheries management units*. Fish Fish. **10** : 361–395.
76. Rohling, E. G., Abu-Zied, R. H., Casford, J. S. L., Hayes, A., Hoogakker, B. A. A. & Woodward, J. C. (2009). *The Marine Environment: Present and Past. Dans : The physical Geography of the Mediterranean*. Phys. Geogr. Mediterr. **33**–67.
77. Rolland, J. L., Bonhomme, F., Lagard re, F., Hassan, M. & Guinand, B. (2007). *Population structure of the common sole (Solea solea) in the Northeastern Atlantic and the Mediterranean Sea: revisiting the divide with EPIC markers*. Mar. Biol. **151** : 327–341.
78. Romiguier, J., Gayral, P., Ballenghien, M., Bernard, A., Cahais, V., Chenuil, A., Chiari, Y., Dernet, R., Duret, L. & Faivre, N. et al. (2014). *Comparative population genomics in animals uncovers the determinants of genetic diversity*. Nature. **515** : 261–263.
79. Russo, T., Carpentieri, P., D’Andrea, L., De Angelis, P., Fiorentino, F., Franceschini, S., Garofalo, G., Labanchi, L., Parisi, A., Scardi, M. & Cataudella, S. (2019). *Trends in Effort and*

- Yield of Trawl Fisheries: A Case Study From the Mediterranean Sea.* Front. Mar. Sci. **6** : 1–19.
80. Sá-Pinto, A., Baird, S. J. E., Pinho, C., Alexandrino, P. & Branco, M. (2010). *A three-way contact zone between forms of Patella rustica (Mollusca: Patellidae) in the central Mediterranean Sea.* Biol. J. Linn. Soc. **100** : 154–169.
81. Schrago, C. G. (2013). *The Effective Population Sizes of the Anthropoid Ancestors of the Human-Chimpanzee Lineage Provide Insights on the Historical Biogeography of the Great Apes.* Mol. Biol. Evol. **31** : 37–47.
82. Sea Around Us. (2016). *Catches by Commercial groups in the waters of Mediterranean Sea.* Suivre : <http://www.seararoundus.org/data/#/lme/26?chart=catch-chart&dimension=commercialgroup&measure=tonnage&limit=10>.
83. Serena, F. (2005). *Field identification guide to the sharks and rays of the Mediterranean and Black sea.* FAO Spec. Ident. Guide. 97.
84. Thuillet, A.-C., Bataillon, T., Poirier, S., Santoni, S. & David, J. L. (2005). *Estimation of Long-Term Effective Population Sizes Through the History of Durum Wheat Using Microsatellite Data.* Genetics. **169** : 1589–1599.
85. Vaughan, T. G., Kühnert, D., Popinga, A., Welch, D. & Drummond, A. J. (2014). *Efficient Bayesian inference under the structured coalescent.* Bioinformatics. **30** : 2272–2279.
86. Waples, R. K., Larson, W. A. & Waples, R. S. (2016). *Estimating contemporary effective population size in non-model species using linkage disequilibrium across thousands of loci.* Heredity. **117** : 233–240.
87. Waples, R. S. & Do, C. (2008). *LDNe : a program for estimating effective population size from data on linkage disequilibrium.* Mol. Ecol. Resour. **8** : 753–756.
88. Weir, B. S. (1979). *Inferences about Linkage Disequilibrium.* Int. Biometric Soc. **35** : 235–254.
89. Weir, B. S. & Cockerham, C. C. (1984). *Estimating F-Statistics for the Analysis of Population Structure.* Evolution. **38** : 1358–1370.
90. Whitlock, M. C. (2011). *G'st and D do not replace Fst.* Mol. Ecol. **20** : 1083–1091.
91. Willi, Y., Van Buskirk, J. & Hoffmann, A. A. (2006). *Limits to the Adaptive Potential of Small Populations.* Annu. Rev. Ecol. Evol. Syst. **37** : 433–458.
92. Wright, S. (1965). *The interpretation of population structure by F-statistics with special regard to systems of mating.* Evolution. **19** : 395–420.
93. Yan, R. J., Schnabel, K. E., Rowden, A. A., Guo, X. Z. & Gardner, J. P. A. (2020). *Population Structure and Genetic Connectivity of Squat Lobsters (Munida Leach, 1820) Associated With Vulnerable Marine Ecosystems in the Southwest Pacific Ocean.* Front. Mar. Sci. **6** : 1–18.
94. Zakaria, H. Y. (2015). *Article Review: Lessepsian migration of zooplankton through Suez Canal and its impact on ecological system.* Egypt. J. Aquat. Res. **41** : 129–144.
95. Zhang, Y., Luan, P., Ren, G., Hu, G. & Yin, J. (2020). *Estimating the inbreeding level and genetic relatedness in an isolated population of critically endangered Sichuan taimen (Hucho Bleekeri) using genome-wide SNP markers.* Ecol. Evol. **10** : 1390–1400.

Annexes

A Méthode implémentée dans BayeScan v2.1 (Foll, 2012 - Foll & Gaggiotti, 2008)

BayeScan v2.1 (Foll, 2012 - Foll & Gaggiotti, 2008) permet d'identifier les marqueurs potentiellement sous influence de mécanismes de sélection naturelle par approche bayésienne sur la base d'une comparaison des fréquences alléliques entre populations via l'indice de fixation (F_{ST}) de Wright :

$$F_{ST} = \frac{\bar{p}(1 - \bar{p}) - \overline{p(1 - p)}}{\bar{p}(1 - \bar{p})}$$

où p est la fréquence d'un allèle dans la population considérée. Cet indice permet de quantifier la différenciation entre populations (Wright, 1965). Le logiciel scinde l'indice de fixation en deux coefficients : un coefficient spécifique à chaque population, noté β , et un coefficient propre à chaque locus, noté α . Un coefficient α significativement différent de 0 indique l'action d'une forme de sélection. Pour chaque locus, une probabilité (*posterior*) est calculée dans le cas d'un modèle avec ou sans sélection naturelle (*i.e.* avec ou sans prise en compte de α). Il reste finalement à sélectionner le modèle le plus plausible entre ces deux cas pour chaque locus. Cela passe par le calcul du facteur de Bayes (BF^{13}), un rapport entre probabilités pour chaque modèle permettant de classer les modèles en question selon un gradient de plausibilité. Il reste toutefois nécessaire d'indiquer au logiciel la probabilité qu'il est cohérent d'émettre quant à la chance que chaque locus soit sous sélection. Pour cela, il est nécessaire d'entrer un paramètre nommé "chance *a priori* en faveur du modèle neutre" (prior odds for neutral model) permettant d'obtenir une chance *a posteriori*¹⁴ calculée à partir de BF et de la chance *a priori*. En complément, BayeScan calcule des q-value, équivalent aux p-values pour la prise de décision sur différents tests statistiques. Ces q-values ajoutent une précision supplémentaire à la prise de décision quant à la pression de sélection appliquée sur les loci en incorporant la notion de faux positifs. En effet, sur l'ensemble des marqueurs sous sélection identifiés par BayeScan, une fraction plus ou moins large n'est en réalité pas sous sélection. Ils sont qualifiés de faux positifs. Il est possible de pallier ce souci en filtrant les SNPs identifiés par un taux de faux positifs (FDR^{15}). La q-value associée à un locus est le FDR minimal au-delà duquel il devient significatif de le conserver comme étant sous sélection.

La paramétrisation par défaut du logiciel a été conservée concernant la chaîne de Markov. Dans le but de réduire le FDR sans impacter le pouvoir de détection, il est recommandé d'augmenter la chance *a priori* en faveur du modèle neutre égale à 10 vers une valeur égale à 10000 (Lotterhos & Whitlock, 2014). Il a été décidé d'appliquer un filtre sur FDR de 5% pour tenir compte de la présence de faux positifs. Les SNPs dont la q-value est inférieure à 0,05 sont donc écartés de la liste de sortie. Dans le but de s'assurer de la bonne convergence de la chaîne de Markov, 5 simulations ont été réalisées en parallèle. Ces cinq chaînes ont toutes renvoyées la même liste de marqueurs sous sélection à un SNP près. Ce résultat atteste de la bonne convergence des chaînes de Markov avec le jeu de données.

¹³Bayes factor

¹⁴Posterior Odds, PO

¹⁵False Discovery Rate

B Méthode implémentée dans PCAdapt v4.3.3 (Luu *et al.*, 2017 - Duforet-Frebourg *et al.*, 2014)

La fonction d'analyse du package PCAdapt v4.3.3 (Luu *et al.*, 2017 - Duforet-Frebourg *et al.*, 2014) débute par la réalisation d'une ACP de l'ensemble des SNPs sur la totalité des individus permettant de faire apparaître la structure de la population échantillonnée à partir des données génétiques. Il est nécessaire à ce point de choisir le nombre de composantes principales à conserver par observation de la variance expliquée par chacune de ces composantes. La règle de Cattell est à appliquer : les composantes principales dont la variance expliquée précède l'apparition d'un plateau sont conservées (Cattell, 1966). Les 5 premières composantes sont à conserver selon la règle de Cattell, car la sixième fait partie du plateau. Dès lors que le nombre de composantes principales à conserver est clair, une succession de régressions linéaires des j SNPs sur les k composantes principales est exécutée :

$$G_j = \sum_k \beta_{jk} X_k + \varepsilon_j$$

. Il en résulte une liste de cotes Z (*z-score*), chaque cote étant la mesure centrée réduite de G_j affectée aux j SNPs après régression sur l'ensemble des composantes principales. Cette liste de cotes Z est utilisée pour identifier les marqueurs potentiellement sous sélection. La statistique utilisée pour cette détection est la distance de Mahalanobis robuste notée \mathbf{D} (Luu *et al.*, 2017). Cette distance diffère d'une distance euclidienne classique par le fait qu'elle considère la variance et la corrélation de la série de données. Ainsi, la pondération affectée à chaque donnée diminue plus la dispersion augmente :

$$D_j^2 = (z_j - \bar{z})^\top \Sigma^{-1} (z_j - \bar{z})$$

où Σ est la matrice de covariance des cotes Z et \bar{z} le vecteur des j moyennes des cotes Z sur les k composantes principales.

Afin de limiter la présence de faux positifs, un facteur d'inflation génomique¹⁶ (Devlin & Roeder, 1999) est calculé et permet, par division de \mathbf{D} sur ce facteur, d'obtenir une statistique corrigée suivant la distribution χ^2 attendue du fait de la distribution Gaussienne multivariée des cotes Z . C'est après obtention de ces statistiques correctement distribuées que les p-values sont calculées. Il est à présent préconisé d'appliquer un filtre sur le taux de faux positifs (FDR¹⁷).

L'application d'un filtre sur FDR se fait par la procédure de Benjamini-Hochberg (Benjamini & Hochberg, 1995) qui ajuste les p-values dans le but de les comparer au seuil de FDR sélectionné. Cette procédure est dérivée de la correction de Bonferroni (Dunn, 1959). Après décision d'un seuil de faux positifs, la liste des SNPs dont la p-value corrigée est supérieure à ce seuil est renvoyée en tant que liste de candidats sous sélection.

¹⁶Genomic Inflation Factor, GIF

¹⁷False Discovery Rate

C Liste des marqueurs potentiellement sous sélection

Table S1: Liste des 19 SNPs sous sélection détectés par BayeScan et PCAdapt

Marqueurs sous sélection
RCL_P_1427
RCL_P_1544
RCL_P_1795
RCL_P_2529
RCL_P_2531
RCL_P_2656
RCL_P_2920
RCL_P_3522
RCL_P_36
RCL_P_401
RCL_P_4206
RCL_P_4638
RCL_P_4842
RCL_P_5060
RCL_P_528
RCL_P_581
RCL_P_7983_MAJ_G
RCL_P_9074_MAJ_A
RCL_P_942

D Compléments sur l'analyse de la structure de population

La fonction `xvalDapc` du package `adegenet` (Jombart *et al.*, 2010) utilisable sous R (R Dev. Core Team, 2020) clive les données en deux sous-jeux de données : un jeu initial contenant 90% des individus et un jeu de validation contenant les 10% restants. La validation croisée consiste à quantifier le nombre de réattributions correctes du jeu de validation dans le jeu initial par l'utilisation de DAPC. La fonction répète cette expérience 30 fois pour chaque quantité de composantes principales conservée. Elle estime finalement quelle quantité d'information permet le meilleur taux de réussite dans les réattributions. La sortie de cette fonction est le nombre optimal de composantes à conserver, en l'occurrence 80.

L'analyse de la contribution que chaque SNP apporte à la discrimination des individus est répartie de manière relativement uniforme sur l'ensemble des SNPs (Figure S1). Aucun SNP ne s'est vu attribué une importance supérieure à 1% pour la première fonction discriminante comme pour la deuxième.

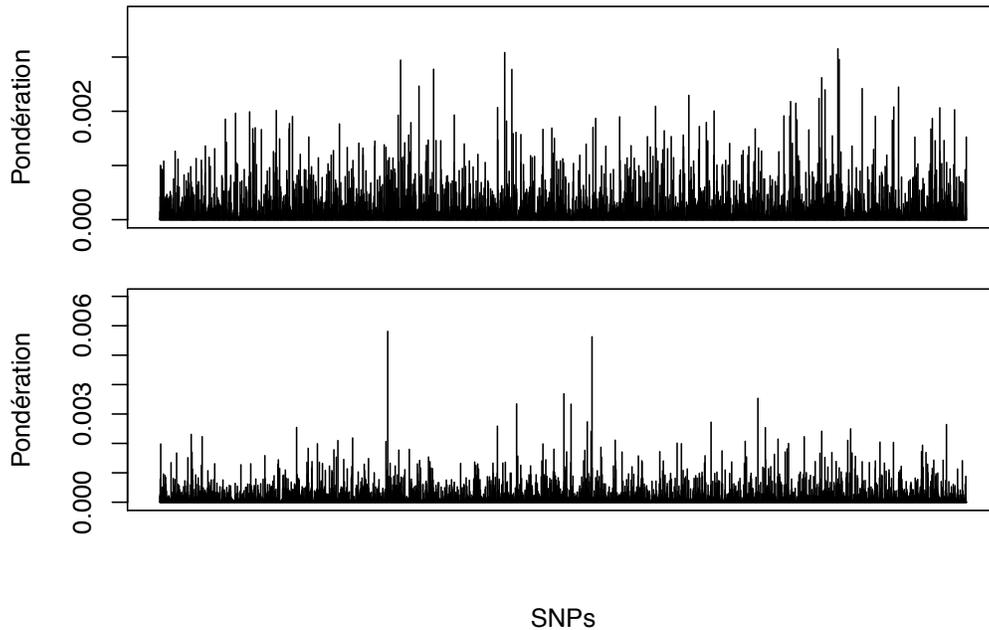


Figure S1: Poids de chaque SNP dans la discrimination des individus selon la première fonction discriminante (haut) et la deuxième fonction discriminante (bas) de la DAPC

La DAPC calcule la probabilité que chaque individu a d'appartenir à chaque zone d'échantillonnage (Figure S2). Ces résultats ont été moyennés et présentés dans le corps de cette étude (Figure 5).

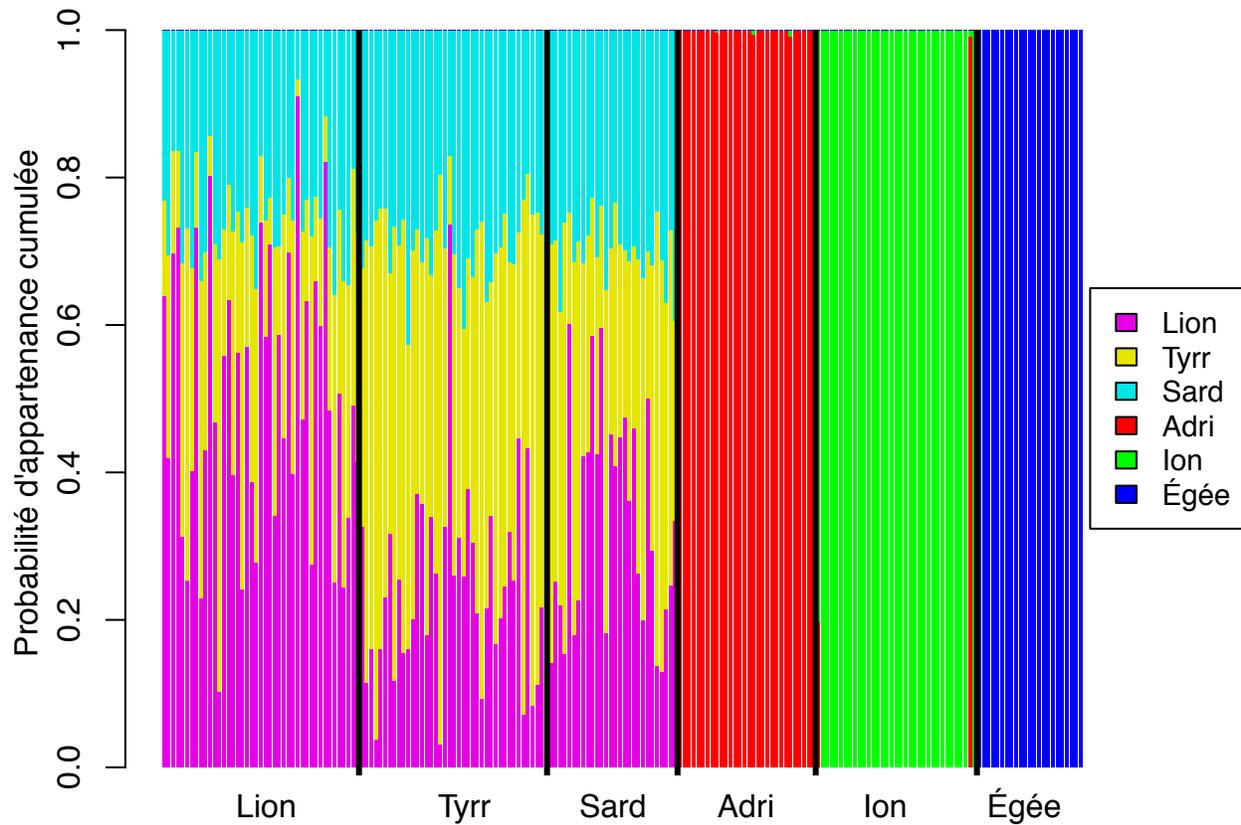


Figure S2: Graphique de mélange présentant la probabilité d'appartenance de chaque individu aux zones d'échantillonnage réalisé par la DAPC - chaque individu est représenté par une barre verticale découpée en segments de longueur proportionnelle à la probabilité d'appartenance à la zone d'échantillonnage correspondante et les individus ont été triés selon leur zone d'échantillonnage d'origine indiquée en abscisse et séparés par les barres noires verticales - Lion : golfe du Lion ; Tyrr : mer Tyrrhénienne nord ; Sard : Sardaigne ; Adri : mer Adriatique ; Ion : mer Ionienne ; Égée : mer Égée

E Compléments sur l'analyse de la différenciation des populations

Table S2: Table des intervalles de confiance autour des estimations moyennes des F_{ST} de Weir et Cockerham entre les quatre populations inférées - Ouest : golfe du Lion, mer Tyrrhénienne nord, Sardaigne ; Adri : mer Adriatique ; Ion : mer Ionienne ; Égée : mer Égée

	Borne inférieure	Moyenne	Borne supérieure
Ouest - Adri	0.015	0.018	0.018
Ouest - Ion	0.021	0.024	0.024
Ouest - Égée	0.049	0.054	0.055
Adri - Ion	0.008	0.009	0.010
Adri - Égée	0.038	0.041	0.042
Ion - Égée	0.025	0.027	0.029

Table S3: Table des intervalles de confiance autour des estimations moyennes des F_{ST} de Weir et Cockerham entre les six zones d'échantillonnage - Lion : golfe du Lion ; Tyrr : mer Tyrrhénienne nord ; Sard : Sardaigne ; Adri : mer Adriatique ; Ion : mer Ionienne ; Égée : mer Égée

	Borne inférieure	Moyenne	Borne supérieure
Lion - Tyrr	0.006	0.006	0.010
Lion - Sard	0.004	0.005	0.006
Lion - Adri	0.018	0.019	0.021
Lion - Ion	0.024	0.027	0.027
Lion - Égée	0.052	0.057	0.058
Tyrr - Sard	0.006	0.006	0.007
Tyrr - Adri	0.017	0.019	0.020
Tyrr - Ion	0.023	0.026	0.026
Tyrr - Égée	0.052	0.055	0.057
Sard - Adri	0.017	0.019	0.020
Sard - Ion	0.023	0.025	0.026
Sard - Égée	0.051	0.055	0.056
Adri - Ion	0.007	0.009	0.010
Adri - Égée	0.038	0.041	0.042
Ion - Égée	0.025	0.027	0.029

F Compléments sur l'analyse de la diversité génétique des populations

Table S4: Table des p-values issues des tests de Student appariés sur 100 ré-échantillonnages bootstrap avec correction de Holm pour les hétérozygosités observée et estimée ainsi que l'indice de consanguinité dans le cas de quatre populations inférées par DAPC de la raie bouclée en Méditerranée - un astérisque (*) indique une différence significative et le sigle ns indique une différence non significative - Ouest : golfe du Lion, mer Tyrrhénienne nord, Sardaigne ; Adri : mer Adriatique ; Ion : mer Ionienne ; Égée : mer Égée

	Hétérozygosité observée	Hétérozygosité attendue	Indice de consanguinité
Ouest - Adri	0.006*	2e-08*	0.49 ns
Ouest - Ion	8.6e-36*	2.4e-52*	0.89 ns
Ouest - Égée	9.2e-10*	1.6e-13*	0.66 ns
Adri - Ion	1e-25*	2e-36*	0.36 ns
Adri - Égée	5.2e-16*	7.4e-30*	0.68 ns
Ion - Égée	3.1e-63*	2.3e-84*	0.45 ns

Table S5: Table des p-values issues des tests de Student appariés sur 100 ré-échantillonnages bootstrap avec correction de Holm pour les hétérozygosités observée et estimée ainsi que l'indice de consanguinité dans le cas de six zones d'échantillonnage de la raie bouclée en Méditerranée - un astérisque (*) indique une différence significative et le sigle ns indique une différence non significative - Lion : golfe du Lion ; Tyrr : mer Tyrrhénienne nord ; Sard : Sardaigne ; Adri : mer Adriatique ; Ion : mer Ionienne ; Égée : mer Égée

	Hétérozygosité observée	Hétérozygosité attendue	Indice de consanguinité
Lion - Tyrr	1.4e-07*	6.3e-10*	1 ns
Lion - Sard	2.1e-12*	4.9e-24*	0.95 ns
Lion - Adri	2e-57*	5.7e-75*	0.75 ns
Lion - Ion	0.00047*	5.6e-12*	0.97 ns
Lion - Égée	0.18 ns	0.27 ns	0.091 ns
Tyrr - Sard	0.014*	5.9e-08*	0.67 ns
Tyrr - Adri	5.2e-35*	3.4e-52*	0.99 ns
Tyrr - Ion	9.9e-17*	2.4e-30*	1 ns
Tyrr - Égée	6.7e-09*	7.3e-07*	0.027*
Sard - Adri	6.4e-26*	2.7e-36*	0.48 ns
Sard - Ion	1.2e-25*	6.2e-50*	0.78 ns
Sard - Égée	8e-16*	1.5e-18*	0.19 ns
Adri - Ion	3.5e-77*	1.3e-105*	0.86 ns
Adri - Égée	6.1e-59*	1.1e-68*	0.0044*
Ion - Égée	0.053 ns	9.9e-14*	0.038*

Table S6: Table des indicateurs de diversité génétique moyens pour les quatre populations inférées (haut) et les six sites d'échantillonnage (bas) de raie bouclée en Méditerranée avec les incertitudes issues de l'estimation des indicateurs sur 100 jeux de données ré-échantillonnés pour 30 individus par population ou zone d'échantillonnage. Aucun ré-échantillonnage n'a été réalisé pour la mer Adriatique et la Sardaigne car 30 individus y ont été échantillonnés - Ouest : golfe du Lion, mer Tyrrhénienne nord, Sardaigne ; Lion : golfe du Lion ; Tyrr : mer Tyrrhénienne nord ; Sard : Sardaigne ; Adri : mer Adriatique ; Ion : mer Ionienne ; Égée : mer Égée

	Hétérozygote observée	Hétérozygote attendue	Indice de consanguinité	Marqueurs monomorphes (%)	Marqueurs privés
Adri	0.295	0.299	0.013	6.96	46.46 (± 2.96)
Ion	0.29 ($\pm 8e-04$)	0.293 ($\pm 3e-04$)	0.008 (± 0.003)	7.71 (± 0.18)	50.39 (± 4.72)
Égée	0.269 ($\pm 5e-04$)	0.273 ($\pm 3e-04$)	0.015 (± 0.002)	12.74 (± 0.26)	258.15 (± 9.39)
Ouest	0.309 (± 0.0017)	0.312 (± 0.001)	0.01 (± 0.005)	3.02 (± 0.31)	13.3 (± 3.26)
Sard	0.306	0.308	0.01	5	21.13 (± 1.71)
Adri	0.295	0.299	0.013	6.96	37.86 (± 2.27)
Ion	0.29 ($\pm 8e-04$)	0.293 ($\pm 3e-04$)	0.007 (± 0.003)	7.7 (± 0.16)	40.94 (± 3.79)
Égée	0.269 ($\pm 6e-04$)	0.273 ($\pm 2e-04$)	0.015 (± 0.002)	12.77 (± 0.29)	227.9 (± 8.67)
Lion	0.313 ($\pm 9e-04$)	0.317 ($\pm 6e-04$)	0.013 (± 0.002)	2.28 (± 0.12)	10.3 (± 2.59)
Tyrr	0.308 (± 0.0012)	0.307 ($\pm 5e-04$)	-0.003 (± 0.004)	4.09 (± 0.21)	10.93 (± 2.54)

Table S7: Table des moyennes, médianes, modes et écarts-types des tailles de population efficaces pour les quatre populations inférées (haut) et les six sites d'échantillonnage (bas) de raie bouclée en Méditerranée - Ouest : golfe du Lion, mer Tyrrhénienne nord, Sardaigne ; Lion : golfe du Lion ; Tyrr : mer Tyrrhénienne nord ; Sard : Sardaigne ; Adri : mer Adriatique ; Ion : mer Ionienne ; Égée : mer Égée

	Moyenne	Médiane	Mode	Écart-type	Quantité de données
Adri	6915.2	6915.2	6900.34	NA	NA
Ion	500.37	361.5	357.24	231.47	100
Égée	12856.23	11175	9073.35	5928.07	74
Ouest	1833.81	1843.05	1914.12	698.9	100
Sard	1073.7	1073.7	1071.39	NA	NA
Adri	6915.2	6915.2	6900.34	NA	NA
Ion	448.66	364.7	354.69	156.19	100
Égée	14178.35	12585.2	10318.28	6091.25	78
Lion	2399.28	2593.15	1350.42	1089.84	100
Tyrr	654.19	171.7	163.96	741.81	100

Bibliographie annexe

1. Benjamini, Y. & Hochberg, Y. (1995). *Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing*. J. R. Stat. Soc. **57** : 289–300.
2. Cattell, R. B. (1966). *The Scree Test For The Number Of Factors*. Multivariate Behav. Res. **1** : 245–276.
3. Devlin, B. & Roeder, K. (1999). *Genomic control for association studies*. Biometrics. **55** : 997–1004.
4. Duforet-Frebourg, N., Bazin, E. & Blum, M. G. B. (2014). *Genome Scans for Detecting Footprints of Local Adaptation Using a Bayesian Factor Model*. Mol. Biol. Evol. **31** : 2483–2495.
5. Dunn, O. J. (1959). *Estimation of the Medians for Dependent Variables*. Ann. Math. Stat. **30** : 192–197.
6. Foll, M. (2012). *BayeScan v2.1 User Manual*. 1–10.
7. Foll, M. & Gaggiotti, O. (2008). *A Genome-Scan Method to Identify Selected Loci Appropriate for Both Dominant and Codominant Markers: A Bayesian Perspective*. Genetics. **180** : 977–993.
8. Jombart, T., Devillard, S. & Balloux, F. (2010). *Discriminant analysis of principal components: a new method for the analysis of genetically structured populations*. BMC Genet. **11** : 94.
9. Lotterhos, K. E. & Whitlock, M. C. (2014). *Evaluation of demographic history and neutral parameterization on the performance of F_{ST} outlier tests*. Mol. Ecol. **23** : 2178–2192.
10. Luu, K., Bazin, E. & Blum, M. G. B. (2017). *pcadapt : an R package to perform genome scans for selection based on principal component analysis*. Mol. Ecol. Resour. **17** : 67–77.
11. R Dev. Core Team. (2020). *R: A language and environment for statistical computing*. Suivre : <https://www.r-project.org/>.
12. Wright, S. (1965). *The interpretation of population structure by F -statistics with special regard to systems of mating*. Evolution. **19** : 395–420.

 	Diplôme : Ingénieur Spécialité : Agronome Spécialisation / option : Sciences halieutiques et aquacoles / Ressources et écosystèmes aquatiques Enseignant référent : Étienne RIVOT
Auteur(s) : Romain LÉCUYER Date de naissance : 06 mars 1997	Organisme d'accueil : Ifremer Adresse : Rue de l'Île d'Yeu – 44300 Nantes
Nb pages : 26 Annexe(s) : 10	Maître de stage : Pascal LORANCE
Année de soutenance : 2020	
Titre français : Structure et diversité génétique des populations de raie bouclée en Méditerranée	
Titre anglais : Structure and genetic diversity of thornback ray populations in the Mediterranean sea	
<p>Résumé : Les chondrichthyens forment un taxon menacé par diverses perturbations et notamment le développement important de la pêche au cours du siècle passé. La Méditerranée est une région fortement exploitée depuis cette période et qui continue de l'être de manière intense par des nations cherchant avant tout à assurer une rentabilité économique au secteur de la pêche ou une sécurité alimentaire à leur population. Ce contexte ne facilite pas les acquisitions et partages de données pour la recherche. Cette étude a pour objectif principal de mieux appréhender la structure génétique et l'état de la raie bouclée afin d'éviter son déclin dû à la pression imposée par la pêche au chalut démersal, difficilement réversible même en cas de soudaine bonne gestion. Dans ce but, un large panel de marqueurs polymorphes issus du génome nucléaire a été développé et analysé afin d'étudier la structuration sous-jacente ainsi que la différenciation et la diversité génétique intrinsèque aux populations de Méditerranée. Une structuration génétique en quatre populations a ainsi été relevée, avec notamment une population semblant panmictique dans le bassin occidental de la Méditerranée. La différenciation génétique précise que les individus de mer Égée sont les plus distants des autres populations, les populations des mers Adriatique et Ionienne et du bassin occidental formant deux sous-groupes au sein desquels la différenciation est plus ténue. Les indicateurs de diversité génétique estimés révèlent une diminution graduelle de la diversité génétique d'ouest en est tandis que la taille de population efficace, indicateur clé en terme d'évolution de la diversité génétique au fil des générations, semble orienter vers une plus grande pérennité de la diversité génétique de la population de mer Égée par rapport aux autres populations plus à l'ouest.</p>	
<p>Abstract: Chondrichthyes are threatened by several disturbances among which is the significant development of fisheries during the 20th century. The Mediterranean sea has long been fished and keeps being intensely so by several nations that seek cost-effectiveness or food security above all. Such a background does not ease data recording and sharing for research purposes. This study mainly aims at assessing the population structure and status of thornback ray so as to avoid its deterioration due to bottom-trawl fisheries pressure, hardly reversible even with a sudden sound management. To do so, an extensive panel of nuclear genetic markers has been developed to analyze underlying structure along with differentiation and genetic diversity of Mediterranean populations. A four-population genetic structure has been identified with potential panmixia within the occidental Mediterranean basin. Individuals sampled in the Egean sea were the most differentiated from the rest in which the Adriatic and Ionian seas and the occidental basin contain two moderately differentiated groups, the latter containing insignificant genetic differences. Estimated indices show a gradual depletion of genetic diversity eastward whereas effective population size estimators point towards a longer sustainability of the genetic diversity in the Egean sea relatively to other western populations.</p>	
Mots-clés : Méditerranée, structure de populations, différenciation, diversité génétique, SNP, <i>Raja clavata</i>	
Key Words: Mediterranean sea, population structure, differentiation, genetic diversity, SNP, <i>Raja clavata</i>	