

THESE

pour obtenir le grade de

DOCTEUR DE L'UNIVERSITE MONTPELLIER II

Discipline : Evolution, Ecologie, Ressources génétiques, Paléontologie

Ecole doctorale : SIBAGHE

Présentée et soutenue publiquement par

Soumaya Oueslati

Soutenue le 19 juillet 2013 à Sète

Caractérisation génétique du complexe d'espèces des anchois du genre *Engraulis* en Méditerranée et dans le proche Atlantique

Jury

Mme L. Bahri-Sfar	Université de Tunis El Manar	Rapporteure
Mme A. Chenuil-Maurel	Université Aix-Marseille	Rapporteure
Mme K Fadhlaoui-Zid	Université de Jendouba	Co-directeur
M. O. Jarboui	I. N. S. T. M. Sfax	Examineur
M. F. Bonhomme	Université de Montpellier 2	Directeur
M. P. Berrebi	Université de Montpellier 2	Président

Avant-propos

Cette thèse s'inscrit dans le cadre d'une cotutelle entre l'Université de Carthage en Tunisie et l'université Montpellier 2.

Les travaux ont été effectués au sein du laboratoire de la station Méditerranéenne de l'environnement littoral à Sète sous la direction de François BONHOMME

.

REMERCIEMENTS

Au terme de la rédaction, je tiens à exprimer ma reconnaissance à toutes les personnes qui, de près ou de loin, ont contribué à l'accomplissement de ce travail.

Je veux, pour commencer, remercier ma famille, sans qui rien de tout ceci n'aurait été possible. Merci maman pour tes encouragements et ton suivie continu c'est à toi que je dois avoir pu faire des études supérieure.

Je tiens à remercier Lilia Bahri et Anne Chenouil d'avoir accepté d'être les rapporteuses de cette thèse. Merci également Anne pour tes participations aux comités de thèse.

Je tiens à exprimer mes profonds remerciements à Othman Jarboui et Karima Fadhlaoui-zid, mes co-directeurs de thèse, pour m'avoir accueillie dans le laboratoire de l'INSTM à Tunis, d'avoir accepté d'encadrer cette thèse et pour leurs aides et leurs conseils.

*Scientifiquement la personne à qui je dois le plus se trouve être mon Directeur de Thèse François Bonhomme. C'est à lui que je dois d'avoir pu faire cette thèse. C'est grâce à lui que l'on a accepté de me faire confiance. Je le remercie sincèrement pour cette confiance. Je le remercie pour toutes ces discussions stimulantes qui ont fait avancer le travail. Je le remercie pour sa patience. **Merci Chef.***

Je tiens également à remercier les membres de mon comité de thèse et particulièrement Nicolas Bierre et Patrick Berrebi pour leurs remarques et leurs conseils avisés. Un grand Merci à Patrick d'avoir accepté de présider le jury de ma thèse.

Les recettes de Marie thérèse, Annie Orth et Frederique Cerqueira m'ont permis de sauver un temps précieux dans les manips, merci Marie pour tes conseils et ton aide pour les commandes de matériels.

Je tiens à exprimer mes sincères remerciements et témoigner de ma reconnaissance aux membres de mon équipe à Sète et aux personnels de la plate forme de séquençage- génotypage de Montpellier

Merci encore aux autres personnes impliquées de près ou de loin dans le collecte d'échantillons.

Je tiens à remercier toutes les personnes de la SMEL que j'ai côtoyées et qui ont rendu mon séjour plus agréable.

Une dernière pensée et un grand Merci à tous mes ami(e)s en Tunisie et en France pour leur soutien.

Sommaire

INTRODUCTION.....	1
Préambule	1
I.1. La notion d'espèce :	1
I.2. Les mécanismes d'isolement reproductif.....	2
I.2.1. Isolement pré-zygotique	3
I.2.2. Isolement post-zygotique	3
II. La notion de population	3
II.1 Définition	3
II.2. Délimitation	4
III. Modèle biologique	6
III.1.Systématique et morphologie	6
III.2. Structure génétique de l'espèce	8
III.3. Objectifs de la thèse.....	9
Chapitre I : Données nucléaires	11
1. Marqueurs utilisés et méthodes analytiques	11
1.1. Microsatellites	11
1.2. Traitement des données nucléaires	12
2. Présentation des résultats.....	13
2.1. Existence d'une différenciation génétique entre échantillons de lagune et de pleine mer	13
2.2. Etude de l'interaction entre les deux entités identifiées	16
2.2.1. Présence d'allèles nuls ?	17
2.2.2. Equilibre de Hardy Weinberg :.....	17
Analyse globale	17
Analyse par échantillon	18
2.2.3 Déséquilibre de liaison	18
2.2.4. Analyse multilocus	19
Chapitre II : Equilibre nucléo-cytoplasmique	22
1. Marqueur mitochondrial utilisé et méthodes analytiques.....	22
1.1 Généralités	22
1.2 Méthodes d'analyse	22
1.3. Calcul des fréquences	22

2. Déséquilibre nucléo-cytoplasmique	25
3. Synthèse sur les résultats de fréquences des gènes nucléaires et des clades mitochondriaux et discussion des résultats.....	26
3.1 Synthèse des résultats	26
3.2 Hybridation ou Introgression ?	27
3.3 Adaptation locale ?	27
Chapitre.III : Coalescence mitochondriale :.....	28
1. Méthode d'analyse:.....	28
2. Principaux résultats et points discutés	28
2.1. Haplotypes du cytochrome b de l'ADNmt	28
2.2. Coalescence et expansion démographique	29
2.2 1. Datation de la divergence des deux clades:	29
2.2 2. Mismatches et expansion démographique.....	31
Discussion générale	33
Scénarios possibles :	33
Conclusion générale	35
REFERENCES BIBLIOGRAPHIQUES	36
ANNEXE A : Article soumis au journal Marine Biology	1
ANNEXE B.....	14
ANNEXE C.....	26

INTRODUCTION

Préambule

La théorie de l'évolution s'attache à rendre compte de la différenciation entre populations au sein d'une même espèce conduisant à la spéciation à long terme. Ceci implique la description de cette différenciation mais aussi la compréhension des mécanismes évolutifs responsables de l'apparition de groupes d'individus partageant un certain nombre de caractéristiques biologiques les distinguant des autres groupes de même nature. Le présent travail aborde ces questions dans le cas particulier d'une espèce marine à forte capacité de dispersion, l'anchois. Avant de rentrer dans les caractéristiques particulières à notre modèle et les questions qui se posent à son sujet, il convient de définir clairement les notions de base que sont la population et l'espèce et les relations qui les lient.

1.1. La notion d'espèce :

A l'origine, deux notions d'espèce ont été citées, une espèce « phénotypique » et une espèce « taxonomique ». La première est une définition empirique qui considère l'espèce comme un ensemble d'organismes vivants se ressemblant plus entre eux qu'à d'autres ensembles équivalents. La seconde notion considère qu'avec un grand nombre de critères, il est possible d'établir une limite claire et objective entre deux espèces différentes. Cette deuxième définition présentant des lacunes, deux nouveaux critères sont venus s'ajouter à la notion d'espèce. Le premier est l'interfécondité selon laquelle deux individus pris au hasard peuvent se reproduire et donner un produit viable et lui-même fécond. C'est la définition la plus répandue, elle correspond à la notion « d'espèce biologique » selon (Mayr, 1942) : les espèces sont des *ensembles de populations naturelles effectivement ou potentiellement interfécondes, qui sont génétiquement isolées d'autres ensembles équivalents*. L'espèce est donc la plus grande unité au sein de laquelle le flux génétique est possible dans des conditions naturelles.

Par la suite, la notion de niche écologique est venue compléter la notion d'espèce (Van Valen, 1976) « *L'espèce est alors une communauté reproductive de populations, reproductivement isolée d'autres communautés et qui occupe une niche particulière dans la nature* ». Cependant, ceci pose le problème de la définition de la niche écologique, qui peut varier selon les endroits occupés par l'espèce.

Plus récemment, l'étude des phylogénies moléculaires a permis d'avancer dans la délimitation des espèces. C'est la notion d'espèce phylogénétique : « *la plus petite lignée d'un ensemble d'organismes vivants pouvant être définie par une combinaison unique de caractères diagnostiques* » (Hewitt, 2001). Néanmoins dans la pratique, toutes les situations ne sont pas forcément claires et bien tranchées. On utilise donc la notion de sous-espèce, qui constitue une forme intermédiaire dans le processus de spéciation. Il s'agit d'un ensemble d'individus qui se trouvent isolés des autres individus de l'espèce pour des raisons écologiques (barrière environnementale ou géographique), anatomiques (ex. pièces génitales ou forme du grain de pollen), ou comportementales (ex. saison de reproduction). Cependant, l'interfécondité est encore possible entre les différentes sous-espèces et les divergences ne sont pas encore suffisamment marquées pour leur donner le statut d'espèce. Ces ensembles peuvent alors évoluer indépendamment les uns des autres et finir par présenter des caractéristiques spécifiques, soit totalement nouvelles (mutations), soit par la fixation de caractères variables (polymorphes) préexistant dans l'espèce source (dérive génétique) mais dans certains cas ils peuvent éventuellement se remélanger si une perturbation écologique ou géographique le permet.

1.2. Les mécanismes d'isolement reproductif

Selon Darwin, la sélection naturelle est une force créatrice d'entités localement adaptées. Depuis (Wright, 1931), il est reconnu que la dérive génétique, due au tirage aléatoire des gamètes à chaque génération, produit elle aussi de la variation génétique entre populations : Deux populations se trouvant isolées l'une de l'autre à la suite de l'apparition d'une barrière physique (apparition d'une chaîne de montagnes, clôture d'un détroit, par exemple) vont voir le flux génique entre elles diminuer et vont dériver à une vitesse inversement proportionnelle à leur taille efficace, et leurs génomes devenir progressivement incompatibles à cause de l'accumulation de mutations. La spéciation résultante est dite allopatrique. Pour constater qu'il y a bien eu spéciation, il faut que les génomes soient remis en contact (sympatrie), on peut alors caractériser et mesurer l'étendue de leur isolement reproductif. L'isolement reproductif peut être post-zygotique ou pré-zygotique selon qu'il intervient avant ou après la fécondation. Les sous-espèces forment souvent des zones d'hybridation (exemples dans (Bierne et al., 2011) ou d'intergradation aux endroits où elles rentrent en contact. Ces zones d'hybridation se maintiennent quand les génotypes hybrides sont maladaptés ou moins bons que les génotypes parentaux. Dans ces régions de contact peuvent se mettre en place des

processus d'isolement reproductif limitant les échanges génétiques et susceptibles d'aboutir à la spéciation.

I.2.1. Isolement pré-zygotique

Il existe de nombreux mécanismes qui interviennent avant la fécondation et empêchent ou ralentissent la formation même des hybrides. On peut citer le choix d'habitat, la philopatrie ou la reconnaissance de partenaires. Dans certain cas, il a été proposé que ces mécanismes puissent permettre d'accélérer le processus de différenciation pour aboutir à la spéciation.

I.2.2. Isolement post-zygotique

On commence à parler d'un isolement post-zygotique entre deux entités taxinomiques lorsque les génotypes hybrides sont créés mais ont une valeur sélective inférieure à celle des génotypes parentaux. Cette mauvaise fitness des hybrides peut-être due à des incompatibilités génétiques qui sont d'origine soit endogène (ex. interactions moléculaires défectueuses entre sous-unités d'un complexe protéique) soit exogènes (ex. adaptation différentielle des deux espèces parentales, les hybrides sont intermédiaires). L'isolement post-zygotique n'est véritablement complet que lorsque tous les hybrides F1 ont une valeur sélective nulle. Lorsque des croisements de générations ultérieures existent (F2, backcross), même en très faible abondance, le flux de gènes entre deux entités taxinomiques est possible et l'isolement est incomplet. L'isolement est également incomplet lorsque la létalité ou la stérilité ne s'expriment que chez certains hybrides (ex. : sexe hétérogamétique principalement affecté selon la loi de Haldane).

II. La notion de population

II.1 Définition

On désigne communément une population comme étant une unité de reproduction : « *un groupe d'individus de la même espèce vivant dans une zone géographique suffisamment restreinte pour que tout membre de la population puisse se croiser avec tout autre (à condition, bien sûr, qu'ils soient de sexes opposés)* » (Hartl, 1994). Ceci sous-entend que les individus d'une population se croisent plus entre eux qu'avec des individus de populations voisines. C'est à cette définition que l'on fait en général référence en biologie des populations. Cette définition ne suppose donc pas que les individus appartenant à une même

population soient nécessairement proches dans l'espace, en dehors des périodes de reproduction.

La délimitation de la zone géographique dans laquelle on trouve des individus appartenant à une même population est évidemment dépendante de l'espèce considérée, celles d'une espèce fortement migratrice comme l'anchois seront potentiellement très étendues, alors qu'un escargot aura un rayon d'action beaucoup plus limité.

La distinction de plusieurs populations chez une même espèce vient du fait que les individus ne sont pratiquement jamais répartis de manière homogène et aléatoire dans le milieu, bien qu'il existe des cas de peuplement continu où la reproduction se fait de proche en proche notamment pour les espèces fixées. La structuration spatio-temporelle des individus d'une espèce dérive en partie de l'action de la structuration du milieu (milieux favorables ou non, barrières environnementales), mais aussi de facteurs intrinsèques à l'espèce (comportements sociaux notamment). C'est pourquoi la délimitation des populations varie d'une espèce à l'autre.

Dans le présent manuscrit, le terme population désigne toujours (sauf dans les traitements statistiques où nous distinguerons l'échantillon de la population qu'il représente) une « population locale », c'est-à-dire un groupe d'individus qui ont une plus forte probabilité de se reproduire ensemble qu'avec des individus d'autres groupes analogues plus distants.

II.2. Délimitation

Afin d'établir des mesures de gestion et de conservation pour les espèces exploitées, il est d'abord nécessaire de définir les limites des différentes populations. Deux approches ont été considérées, l'approche morphologique et l'approche génétique. Depuis longtemps, les caractères morphologiques étaient les seuls outils disponibles aux scientifiques pour caractériser les différences entre populations. Toutefois, cette approche ne rend pas toujours fidèlement compte de l'histoire des populations. En effet, il arrive fréquemment que les variations phénotypiques ne soient pas seulement liées à l'histoire évolutive des populations mais également varient en fonction des conditions environnementales rencontrées localement (plasticité phénotypique). D'un autre côté, l'approche, génétique permet de délimiter les populations au sens défini au paragraphe II.1 ci-dessus (participation à la reproduction), d'estimer leur diversité génétique et d'évaluer le degré de connectivité entre ces populations,

par l'intermédiaire de l'estimation des flux de gènes (Féral, 2002), reflétant en partie les capacités de dispersion des espèces.

En effet, les modèles de la génétique des populations prédisent une relation entre les paramètres de migration et les valeurs de différenciation génétique. C'est l'équilibre entre la dérive génétique qui se produit à l'intérieur de chaque population (si une population est isolée, elle va inmanquablement perdre sa variabilité génétique en l'absence de mutations nouvelles et se différencier des autres populations) et la migration qui introduit des génotypes différents. Cette différenciation est mesurée classiquement par le paramètre F_{st} (indice de fixation) de Wright et le nombre de migrants échangés à chaque génération par $N_e m$ (N_e = effectif génétique de la population, m = taux de migration). Dans le modèle le plus simple dit modèle en îles (considérant un nombre infini de populations équidistantes les unes des autres), plus la migration (dispersion géographique) est importante, plus la divergence génétique entre les populations est faible. On a alors à l'équilibre la relation très connue $F_{st} = 1 / 1 + 4N_e m$ (voir figure 1). La valeur de F_{st} est inversement proportionnelle au nombre d'individus migrants par génération, mais le lien entre la génétique et la démographie est faible, car seules de très petites valeurs de migration vont donner un F_{st} mesurable comme on voit sur la courbe. Comme il suffit d'un nombre limité de migrants entre les différentes populations pour neutraliser le signal de différenciation génétique, en étudiant les valeurs de F_{st} , les généticiens des populations peuvent ne pas voir de différence génétique significative entre les populations, alors que les halieutes remarquent des différences phénotypiques ou démographiques et parlent alors de stocks différents.

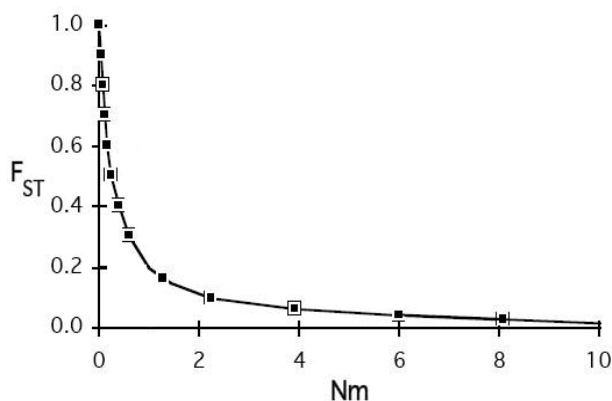


Figure 1 : Distribution des valeurs de F_{st} à l'équilibre en fonction du nombre absolu de migrants par génération (N taille de la population, m taux de migration).

Chez les organismes marins, plusieurs auteurs ont montré la présence de populations différentes au sein de la même espèce, surtout pour les espèces avec des capacités de dispersion limitées (Baus et al., 2005); (Doherty et al., 1995); (Planes, 1998). D'autres études

montrent que dans certaines conditions la structure peut apparaître même pour des espèces pélagiques à fortes capacités de dispersion (Naciri et al., 1999 ; Bahri-Sfar et al., 2000 ; Bargelloni et al., 2005 ; Magoulas et al., 1996 ; Zarraonaindia et al., 2009). Cependant, les raisons en sont la plupart du temps que des clades différents ont été identifiés comme chez l'anchois (Magoulas et al., 1996, 2006) ou le loup (Lemaire et al., 2005) montrant qu'il y a peut-être plusieurs sous-espèces en présence.

Dans le reste de l'introduction, je présenterai la biologie et la distribution de l'espèce, et les hypothèses pouvant expliquer sa structuration.

III. Modèle biologique

III.1. Systématique et morphologie

L'anchois, *Engraulis encrasicolus*, est un poisson téléostéen clupéiforme de la famille des *Engraulidés*. Il est présent dans l'Atlantique Est depuis la Scandinavie jusqu'aux côtes ouest de l'Afrique, en Méditerranée, dans la Mer Noire et la Mer d'Azov. Des individus ont déjà été observés dans le canal et le Golfe de Suez. Des données ont suggéré que sa distribution peut s'étendre aussi loin que l'Afrique du Sud ainsi que dans l'Océan Indien (Grant and Bowen, 1998) . C'est un petit pélagique côtier, grégaire, vivant en bancs très denses. Il pénètre dans les estuaires, les baies et descend au large jusqu'à environ 150 m de profondeur sur des fonds vaseux (Fisher, 1987). Ses déplacements seraient surtout conditionnés par la température de l'eau. Les conditions trophiques, thermiques et de reproduction influent sur les migrations saisonnières des anchois. Vu la fragilité de l'anchois aucune opération de marquage et suivi du déplacement du poisson n'est possible ce qui rend difficile le traçage de sa trajectoire. La situation systématique de cette espèce est confuse. Les recherches antérieures ont proposé des hypothèses de répartition en fonction d'études biologiques et morphologiques. Fage, (1920) a distingué deux races d'anchois *Engraulis encrasicolus*, l'une atlantique et l'autre méditerranéenne qui diffèreraient par leurs caractères morphologiques, mais cette distinction n'a pas été reprise par la suite. Sur la base de critères reproductifs (lieux et période de reproduction), Chashchin, (1996) a parlé de la présence de deux sous-espèces d'anchois en Mer Noire. Avec le même type de données, Spanakis et al. (1989) suggèrent que des populations séparées d'anchois se reproduisent dans la Mer Egée et la Mer Ionienne. Sur la base de critères morphologiques et biologiques, d'autres auteurs ont parlé de l'existence de races ou d'espèces d'anchois qui se trouveraient dans les lagunes. En Sicile, dans la lagune de

Ganzirri (Dulzetto, 1947) et en Tunisie (Quignard et al., 1973 ; Messaoud et al 2011) distinguent les anchois qui se trouvent dans la lagune de Bizerte des anchois du large. Au Maroc, Kada et al. (2009) ont montré que les anchois de la lagune de Nador présentent des différences morphologiques et biologiques par rapport à *Engraulis encrasicolus* capturés en pleine mer.

III.2. Structure génétique de l'espèce

En raison de son importance économique, l'anchois dit « européen » *Engraulis encrasicolus* a fait l'objet de plusieurs études génétiques :

Les premières études allozymiques de (Bembo et al., 1996a) ont trouvé une différence globale faible F_{st} (G_{st}) de 0,034 entre 13 échantillons allant de la Mer Egée au Golfe de Gascogne, avec une différence plus marquée de ce dernier échantillon par rapport à ceux de Méditerranée. A une échelle géographique plus réduite, (Tudela et al., 1999) ont étudié électrophorétiquement 6 échantillons du Golfe du Lion répartis de la côte ligure jusqu'à la Mer Catalane et ne trouvent aucune différenciation F_{st} (G_{st}) de 0,0024, bien que ces mêmes auteurs voyaient des différenciations au niveau méristique et morphologique (Tudela, 1999). Par contre, sur une trentaine d'échantillons de l'Adriatique, (Bembo et al., 1996b) détectent une différenciation relativement marquée sur deux locus enzymatiques (IDH-2 et G3PDH) en fonction de la profondeur de capture. Plus récemment, en complétant l'échantillonnage de Tudela et al., (1999) par des échantillons du Golfe de Gascogne et de mer d'Alboran, (Sanz et al., 2008) montrent une différenciation globale entre les trois mers de F_{st} (G_{st}) = 0,022, alors que cette différenciation atteint 0,046 entre groupes de pontes à l'intérieur de la même région (ceux collectés dans des aires de pontes au nord et au sud du Golfe de Gascogne notamment seraient génétiquement distinguables). Ils montrent également clairement que les anchois de Mer d'Alboran qu'ils ont échantillonnés sont différents de ceux de Méditerranée nord-occidentale et de ceux du Golfe de Gascogne.

Parallèlement, grâce à l'ADN mitochondrial, Magoulas et al. (1996) ont identifié deux clades mitochondriaux (clades A et B) qui coexistent dans des proportions variables dans toutes les populations atlantiques et méditerranéennes. Ces auteurs, ainsi que (Grant et al., 2005) ont proposé que l'actuelle coexistence des deux clades soit le résultat d'un contact secondaire plus ou moins récent entre espèces précédemment isolées. Plus récemment, il a été émis l'hypothèse que le clade A pourrait avoir pour origine la côte ouest-africaine, où il était peut-être limité pendant les périodes glaciaires, et se serait par la suite étendu à la côte ouest du Portugal et dans la Méditerranée et la Mer Noire (Magoulas et al., 2006).

Par ailleurs, Borsa et al. (2004) ont confirmé en étudiant le polymorphisme des introns de la créatine-kinase les hypothèses émises par (Borsa, 2002) portant sur la variabilité morphométrique et les polymorphismes allozymiques selon lesquels une partie de la différenciation génétique observée dans les études électrophorétiques antérieures

correspondrait en fait à l'existence d'une espèce cryptique vivant dans les lagunes, les estuaires et les eaux très côtières, qu'ils ont appelé *E. albidus* pour celle qu'ils ont analysée dans le Golfe du Lion. (Sanz et al., 2008) proposent (sans le démontrer) que l'échantillon de la Mer d'Alboran qu'ils ont analysé correspondrait en partie à cette espèce cryptique. Selon cette hypothèse, les formes côtières et du large (>50m de profondeur) étudiées par (Bembo et al., 1996b) dans l'Adriatique correspondraient également à cette différenciation cryptique.

Enfin, par l'utilisation de locus microsatellites, de SNPs, et d'ADN mitochondrial, (Zarraonaindia et al., 2012) ont proposé l'existence de deux groupes d'anchois, l'un qui serait associé avec les zones de puisage d'eau profonde (upwellings) des plateaux continentaux étroits et comprendrait l'Atlantique ibérique et un deuxième groupe qui serait associé aux plateaux continentaux plus larges, qui comprendrait l'Atlantique NE (y compris la Mer du Nord et le Golfe de Gascogne) et la Méditerranée. Par ailleurs, (Borrell et al., 2012), par l'analyse de l'ADN mitochondrial et de 14 locus microsatellites ont trouvé un F_{st} global de 0,043 leur permettant d'identifier autour de la Péninsule Ibérique trois groupes d'anchois: un groupe méditerranéen, un groupe du Golfe de Gascogne et un troisième à l'ouest de la Mer Cantabrique.

Structuration chez d'autres espèces d'anchois

La famille des Engraulidés compte 17 genres (Bloom and Lovejoy, 2012) décrites dans le genre *Engraulis* (Fishbase, 2011). Parmi celles-ci, quelques unes ont fait également l'objet d'études génétiques. Par exemple Hedgecock et al. (1989) montrent que *E. mordax* de la Californie-Mexique n'est pas subdivisée. Par contre chez *E. japonicus*, (Yu et al., 2002) montrent sur des données microsatellites qu'il existe des différences faibles mais significatives entre sites de pontes de part et d'autre de Taiwan ($F_{st} = 0,007$) et (Chen et al., 2009) montrent sur des données mitochondriales que cette espèce présente également deux clades mitochondriaux distincts et qu'il y aurait au même endroit des différences significatives ($F_{st} = 0,07$) entre stocks pondant au printemps et stocks pondant à l'automne.

III.3. Objectifs de la thèse

Dans le but de fournir des éléments utiles pour la conservation de l'espèce, et plus particulièrement pour l'élaboration de plans de gestion en Tunisie et en France, la finalité de cette thèse est de mettre en relation la structure génétique de l'anchois avec sa répartition géographique. Comme des anchois côtiers ont été décrits en Tunisie sur la base de caractères

morphologiques (Quignard et al., 1973, Messaoud et al., 2011) et au Maroc (Kada et al 2009) que leur statut génétique est encore relativement peu étudié, un effort particulier sera mis sur la détection des différences génétique entre anchois des côtes et anchois du large dans notre échantillonnage et sur les relations génétiques qui existent ou non entre ces entités localisées en plusieurs ponts de la Méditerranée.

Chapitre I : Données nucléaires

1. Marqueurs utilisés et méthodes analytiques

1.1. Microsatellites

Les microsatellites sont des marqueurs d'ADN nucléaires, codominants, hypervariables (taux de mutation avoisinant les 10^{-3} / locus / génération), réputés neutres (situés dans des régions non codantes en général, libres d'évoluer rapidement) et donnant des résultats fiables et reproductibles (Zhang, 2004). Ce sont des séquences d'ADN comprenant un motif répété oligonucléotidique. Ces séquences mutent, offrant un polymorphisme de longueur caractérisé par le nombre de répétitions du motif microsatellite. Les microsatellites sont dispersés dans le génome et situés en général dans des zones inconnues. Il s'agit de déterminer, par électrophorèse sur gel ou à capillaire (génotypeur), la longueur des deux allèles d'un même locus présent chez un individu. Les amorces que nous avons utilisées ont été mises au point à partir d'une banque de microsatellites issue d'un séquençage 454 GS FLX Titanium et détectés selon la méthode développée par Malausa et al. (2011). Nous avons sélectionné 14 couples d'amorces dont les motifs de répétition sont pentanucléotides ou tétranucléotides. Parmi ceux-ci, six se sont avérés produire des génotypes clairs et répétables (Annexe B Table Supp 1). Il nous a été possible de multiplexer nos microsatellites pour l'étape de génotypage. Les allèles des différents microsatellites sont ensuite distingués par la couleur de l'amorce marquée ou, si deux microsatellites ont des amorces marquées de la même couleur, par leurs différentes gammes de tailles. Ainsi, ayant 6 locus mais seulement 2 couleurs possibles (6-FAM, bleu; et HEX, vert), nous avons dû sélectionner trois fois deux couples de microsatellites dont les amorces sont marquées de couleur différentes. La solution d'amplification mise au point est présentée dans l'Annexe C-Tableau1). Les fragments générés par la PCR ont été mis à migrer sur séquenceur automatique *ABI Prism 3130XL* (*Applied Biosystems*) de la plateforme de génotypage du labex CEMEB à Montpellier. Le signal collecté est traduit graphiquement par le logiciel *GeneMapper® v4.5* (*Applied Biosystems*) en pics dont la hauteur et l'aire représentent l'intensité du signal. Comme les locus analysés ont des motifs répétés de 4 ou 5 nucléotides, les pics mineurs ont été assimilés aux pics majeurs (qui étaient en phase avec la périodicité de 4 ou 5) les plus proches (exemple du locus EeA6 en Annexe B.Figure Supp 2). Grâce au logiciel, nous pouvons déterminer la

taille approximative (qui sera arrondie par la suite) des 2 allèles de chaque locus microsatellite pour chaque échantillon analysé.

1.2. Traitement des données nucléaires

Les différentes mesures de la diversité génétique des populations telles que le nombre moyen d'allèle par locus, l'hétérozygotie observée, l'hétérozygotie attendue, le déficit en hétérozygotie ont été calculées à l'aide du logiciel *Genetix* 4.05 (Belkhir et al., 2004). Par ailleurs le diagnostic de *Micro-Checker* 2.2.3 (Van Oosterhout et al., 2004) nous permet de connaître les risques d'avoir des allèles nuls dans nos données. A partir de la matrice de données allèles x individus observées à chaque locus, *Micro-Checker* étudie l'écart entre hétérozygotie observée et attendue et propose plusieurs méthodes d'estimation de la fréquence des allèles nuls. Nous avons choisi l'estimateur de Van Oosterhout, qui tient compte de l'information complète (état homozygote, hétérozygote de chaque allèle) alors que les autres estimateurs disponibles ne font que deux catégories de génotypes : homozygotes ou hétérozygotes quelque soit les allèles considérés. Cet estimateur suppose également qu'il n'y a pas d'homozygotes des allèles nuls ou bien il les considère comme des données manquantes (échantillons non-amplifiés de l'ADN dégradé, l'erreur humaine, etc...).

Le logiciel *Arlequin* 3.0 (Excoffier et al., 2007) est utilisé pour les analyses de variance moléculaire (AMOVA) qui permettent de partitionner la variance totale entre covariances dues à différents niveaux de structure prédéfinis. Les statistiques F suivantes sont calculés : F_{ST} , F_{SC} et F_{CT} qui correspondent à la différenciation entre populations, entre populations à l'intérieur des groupes et entre groupes de populations. La significativité de ces paramètres est testée par 1000 permutations et une distribution est obtenue sous l'hypothèse nulle d'absence de structuration au niveau testé.

Afin de chercher à différencier des populations dans notre jeu de données, les valeurs du coefficient de différenciation par paire d'échantillons (F_{st}) ont été calculées grâce au logiciel *Genetix*. L'estimateur employé est celui de (Weir and Cockerham, 1984). Les déviations de l'équilibre de Hardy-Weinberg sont quantifiées par le coefficient de consanguinité non biaisé de Wright, F_{is} , estimé également selon (Weir and Cockerham, 1984). La significativité des valeurs observées pour ces deux estimateurs est testée par 10 000 permutations.

Le logiciel fournit également une matrice des distances de Reynolds $D_R = -\text{Log}(1-F_{st})$ à partir de laquelle on peut construire un arbre de longueur minimale grâce à la fonction *Neighbour-Joining* du logiciel *Mega* 5.0 (Tamura et al., 2011).

Un autre type de représentation des données consiste à réaliser des traitements de type AFC (analyse factorielle des correspondances), soit standard où tous les individus sont actifs, soit des AFC dites « sur populations » où chaque individu est comparé à l'individu moyen (barycentre) de chaque population. Ce deuxième type d'AFC permet de mieux différencier les populations quand le nombre d'individus est important et que les nuages se chevauchent, ce qui est le cas ici. Le logiciel calcule des axes qui représentent des combinaisons linéaires des variables initiales et qui maximisent l'étirement du nuage de points (donc la différenciation entre populations). Les résultats sont représentés sous forme de projections (en 2 ou 3 dimensions) visualisant ainsi les proximités génétiques entre individus. Cette analyse n'est pas indépendante des mesures de différenciation génétique. En effet les proportions de l'information totale expliquée par chaque axe (inertie), sont équivalentes à des combinaisons linéaires des valeurs de F_{st} monoalléliques (Guinand et al., 1996).

Les analyses de déséquilibre de liaison entre paires de locus sont réalisées grâce à l'estimateur du coefficient de corrélation R_{ij} de Weir calculé selon la méthode de (Black and Krafur, 1985) proposée par *Genetix* et testé par 10 000 permutations. Pour combiner les valeurs de probabilité de déséquilibre obtenues pour plusieurs paires de locus, nous avons calculé une probabilité globale combinée selon la méthode de Fisher où la variable $X^2 = -2 \sum \log_e(p_i)$ suit une loi de χ^2 à $2k$ degrés de liberté, où k est le nombre de tests combinés sous l'hypothèse nulle d'indépendance des variables.

Par ailleurs, nous avons utilisé l'approche d'assignation multilocus réalisée par le logiciel *Structure* (Falush et al., 2007). Par une approche bayésienne, la méthode maximise l'équilibre de H-W et l'équilibre de liaison entre toutes les paires de locus à l'intérieur d'un nombre prédéfini k de groupes. Le meilleur k possible a été cherché selon la méthode d' (Evanno et al., 2005).

2. Présentation des résultats

2.1. Existence d'une différenciation génétique entre échantillons de lagune et de pleine mer

Il apparaît que les valeurs de F_{st} varient fortement d'un locus à l'autre (Tableau 1) : ainsi EeA6 et EeA10 montrent des valeurs de F_{st} fortes et significatives, les valeurs de F_{st} pour les locus EeA2, EeA8, et EeA9 sont faibles mais significatives, alors que le locus EeA7 présente une valeur faible et non significative.

La question qui se pose donc est de savoir quels sont les échantillons qui créent ces fortes valeurs de F_{st} sur les locus EeA6 et EeA10 ? – est-ce que cette différenciation est répartie de manière homogène sur la collection ou pas ?

locus	EeA2	EeA6	EeA7	EeA8	EeA9	EeA10	F_{st} multilocus
F_{st}	<i>0,016</i>	<i>0,065</i>	0,0006	<i>0,0126</i>	<i>0,014</i>	<i>0,191</i>	<i>0,036</i>
P	<0,0001	<0,0001	<0,35	<0,0001	<0,0001	<0,0001	<0,0001

Tableau 1 : F_{st} global obtenu pour chaque locus (en italiques les valeurs significatives après la correction de Bonferroni).

L'arbre global des distances de Reynolds (Figure 2) apporte une réponse à cette question, car les 4 populations lagunaires se présentent sur la même branche et tous les échantillons prélevés en pleine mer sur une deuxième branche. Ceci se traduit aussi sur l'AFC (Figure 3) où l'on voit que se différencie sur l'axe 1 le groupe d'échantillons constitués de TIL-I (Lac Ischkeul), MLN-C04 (Lagune de Nador), TLB-ALB (Lagune de Bizerte), FEO-Z01 (Etang de Mauguio).

Quand on enlève EeA6 et EeA10, on obtient une valeur de $F_{st} = 0,0039$, faible mais toujours significative ($p < 0,0001$). Il apparaît donc que sur nos 6 locus, deux sont fortement associés avec la différenciation mer/lagune. Le F_{st} multilocus pour les échantillons regroupés en deux populations (lagunaires et marines) s'élève à 0,07 ($p < 0,0001$). Quand on regarde dans le détail (Figure 4) des fréquences alléliques des deux locus EeA6 et EeA10) on voit que pour le locus EeA6, l'allèle 166 est majoritaire dans les échantillons lagunaires ; sa fréquence y dépasse 66 %. Pour le locus EeA10, l'allèle 132 est associé aux individus lagunaires à plus que 50 % et l'allèle 128 aux échantillons marins à plus que 50 %. On remarquera que pour EeA6, il n'y a pas d'allèle majoritaire chez les océaniques. Le cas des échantillons TGT-AH et TGG-AS et dans une moindre mesure TGH-AG est intéressant car ils apparaissent comme océaniques pour EeA10 et presque comme les lagunaires pour EeA6.

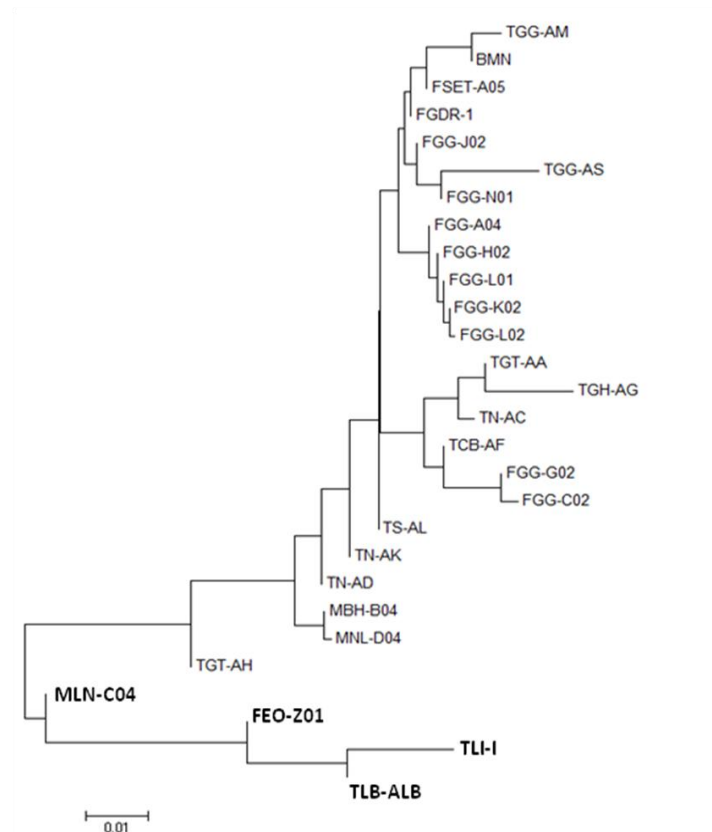


Figure 2 : Arbre des distances de Reynolds entre les populations (populations lagunaires en Gras)

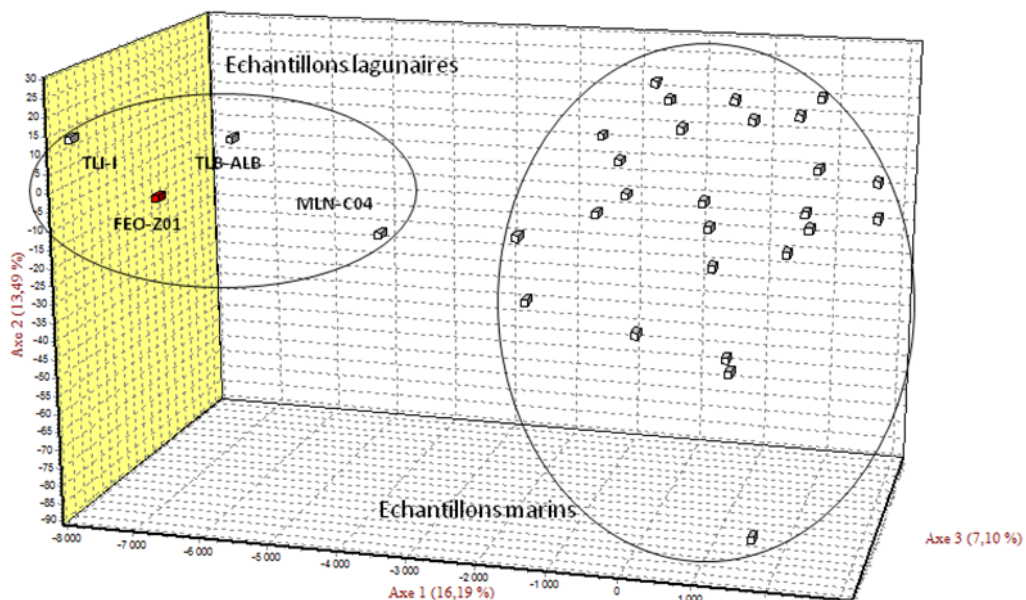


Figure 3 : Analyse factorielle des correspondances aux six locus (populations lagunaires surnommées)

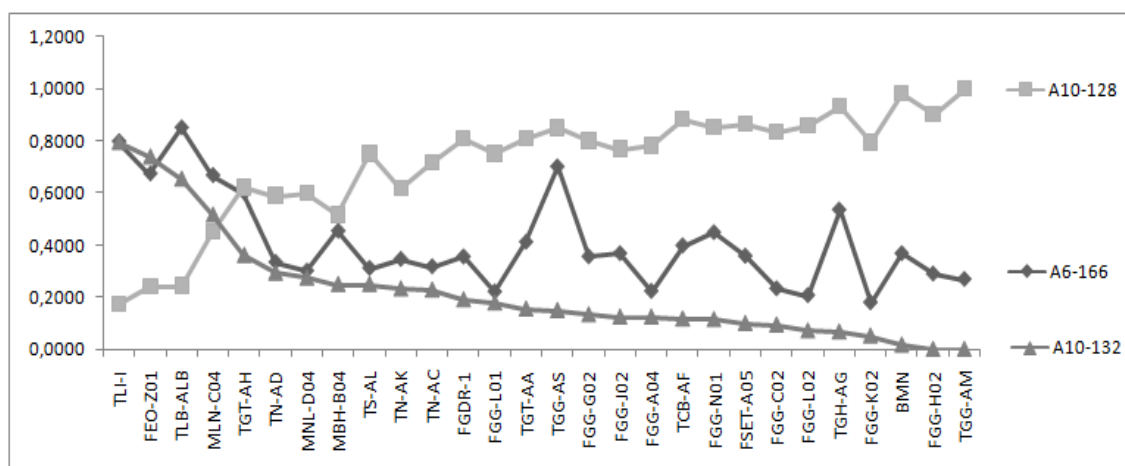


Figure 4 : Distribution des fréquences alléliques aux locus EeA6 (allèles 166 et 128) et EeA10 (allèle 132) ordonnées selon les fréquences décroissante de l'allèle EeA10-132

En conclusion, aux moins deux allèles sont nettement associés avec les échantillons strictement lagunaires; l'allèle 166 pour le locus EeA6 et l'allèle 132 pour le locus EeA10. On peut donc s'attendre à ce qu'il y ait des déséquilibres de liaison entre ces deux locus quand tous les échantillons sont regroupés

2.2. Etude de l'interaction entre les deux entités identifiées

Maintenant que nous savons qu'il existe des populations lagunaires génétiquement différenciées des populations de pleine mer, la question se pose de la nature de leur interaction. Est-ce que les deux entités génétiques sont complètement séparées spatialement ou se recouvrent-elles partiellement ? D'après l'arbre de la figure 2 et les variations observables dans la distribution de certains allèles (figure 4), il apparaît qu'aussi bien les échantillons lagunaires que ceux de pleine mer ne sont pas complètement homogènes génétiquement (on observe un F_{st} multilocus de 0,0183 ($p < 0,0001$) et de 0,0172 ($p < 0,0001$) respectivement dans chaque groupe). Est-ce que cela peut-être du à la présence en proportion variable d'individus des deux types qui se sont retrouvés par hasard capturés au même endroit d'échantillonnage (simple superposition) ? Ou bien les deux entités se côtoient et s'hybrident partiellement, certains échantillons contiendraient alors des hybrides en proportion variable ? Ou bien encore chaque échantillon serait panmictique mais plus ou moins introgressé selon le lieu par des allèles en provenance de l'autre entité suite à un contact secondaire ?

L'étude d'éventuels écarts à la panmixie à travers l'analyse de l'équilibre de Hardy-Weinberg (EHW) et du déséquilibre de liaison (DL) devrait permettre en effet de séparer les deux hypothèses: populations introgressées avec des fréquences d'introgession variables selon les échantillons (on s'attend alors à un équilibre d'Hardy-Weinberg et équilibre de liaison dans chaque échantillon), ou bien coexistence de deux espèces en sympatrie avec ou non présence d'hybrides possible dans chaque échantillon. Dans ce deuxième cas, on attend une absence d'EHW (effet Wahlund) et aussi des DL entre les marqueurs.

2.2.1. Présence d'allèles nuls ?

Mais avant de pouvoir parler valablement d'EHW, il est nécessaire de vérifier d'abord la présence ou non d'allèles nuls dans notre jeu de données. Un déficit en hétérozygotes peut être dû à des facteurs techniques et/ou biologiques. D'une part la présence d'allèles nuls entraîne une sous-estimation des hétérozygotes dans un échantillon car des individus détectés comme étant homozygotes peuvent être en réalité hétérozygotes pour un allèle amplifié et un allèle nul. Il est donc important de pouvoir estimer la fréquence d'allèles nuls éventuels dans les populations avant d'envisager des facteurs biologiques comme causes possibles du déficit en hétérozygotes.

Dans l'Annexe B -Table Supp 9, sont indiquées les populations présentant des allèles nuls en fréquence détectable pour chaque locus ainsi que la comparaison des valeurs de F_{is} corrigées selon *MicroChecker* et non-corrigées selon *Genetix*. La très bonne corrélation des deux séries de valeurs (cf. Tableau 2 ci-dessous) montre que les allèles nuls, s'ils existent, ne sont pas un problème majeur dans nos données, et comme nous avons besoin de travailler sur les valeurs de p , qui ne sont fournies que par *Genetix*, nous avons gardé ces estimations par la suite.

2.2.2. Equilibre de Hardy Weinberg :

Pour l'ensemble des locus, l'hypothèse de l'équilibre panmictique a été testée dans un premier temps pour chaque locus polymorphe sur l'ensemble des individus et dans un deuxième temps pour chaque population.

Analyse globale

Les données de F_{is} estimé sur la totalité des populations (Tableau 2) montrent une déviation globale à l'équilibre panmictique indiquant un déficit en hétérozygote relativement fort et significatif pour chaque locus à l'exception du locus EeA7 si l'on effectue une correction de

Bonferroni (nouveau seuil $p = 0,0083$). On notera cependant que le locus EeA7 ne présentait pas non plus de F_{st} significatifs sur l'ensemble des populations (cf. Tableau 1). On notera également que les deux locus qui montrent le plus fort F_{is} (EeA2 et EeA9) ne sont pourtant pas ceux qui séparent le mieux les lagunaires et les océaniques (comme on aurait pu s'y attendre par effet Wahlund). Les valeurs moyennes estimées après correction pour les allèles nuls de *MicroChecker* ne sont pas très différentes. Par ailleurs quand on décompose en échantillons lagunaires et de pleine mer, on s'aperçoit que les F_{is} multilocus restent significatifs, les valeurs pour les océaniques diminuent un peu sauf pour EeA9, alors que c'est le contraire pour les lagunaires. Nos déficits en hétérozygotes de l'échantillon global ne peuvent donc pas s'expliquer globalement par un effet Wahlund entre les deux types d'anchois.

Locus	EeA2	EeA6	EeA7	EeA8	EeA9	EeA10	Multi locus
F_{is} Global	<i>0,1617</i>	<i>0,0789</i>	<i>0,0295</i>	<i>0,1283</i>	<i>0,2471</i>	<i>0,1058</i>	<i>0,128</i>
P	<0,0001	<0,0001	<0,0070	<0,0001	<0,0001	<0,0002	<0,0001
F_{is} MC Global	0,1181	0,0552	0,0098	0,1204	0,2213	0,0571	0,1049
F_{is} Lagune	<i>0,206</i>	<i>0,1859</i>	<i>0,1051</i>	<i>0,2264</i>	<i>0,093</i>	0,122	<i>0,1571</i>
P	<0,0001	<0,004	<0,003	<0,0001	<0,003	0,052	<0,0001
F_{is} Océan	<i>0,155</i>	<i>0,069</i>	0,019	<i>0,115</i>	<i>0,267</i>	<i>0,103</i>	<i>0,1239</i>
P	<0,0001	<0,0005	<0,006	<0,0001	<0,0001	<0,0003	<0,0001

Tableau 2 : valeurs de F_{is} pour l'échantillon global, pour les anchois lagunaires et ceux de pleine mer (en italiques les valeurs significatives après la correction de Bonferroni).

Analyse par échantillon

Le tableau complet est donné en Annexe B -Table Supp 9. Quand on regarde population par population et locus par locus, on observe une grande variabilité des valeurs, avec certains F_{is} négatifs, mais globalement ils sont positifs. Tous les F_{is} multilocus Annexe B -Table Supp 9 sont positifs et après correction de Bonferroni, 12 échantillons présentent des F_{is} significatifs avec des valeurs allant de 0,131 à 0,254. Il ne semble pas qu'il y ait de logique derrière la liste des populations qui apparaissent hors d'EHW.

2.2.3 Déséquilibre de liaison

Le déséquilibre de liaison (DL) est l'association entre allèles de deux locus différents. La source de cette corrélation peut être due à une liaison physique sur le chromosome (très improbable dans notre cas) ou simplement statistique et qui peut se produire quand deux

entités génétiques différenciées se trouvent mélangées, ce qui est précisément ce qu'on cherche à mettre en évidence sur nos données.

Nous détectons un déséquilibre de liaison sur les différentes combinaisons de locus des populations regroupées et celles océaniques (EeA2-EeA7) et (EeA6-EeA7). Sur les populations lagunaires, nous n'avons détecté aucun déséquilibre. Par contre, on n'a pas de DL significatif entre les deux locus diagnostiques EeA6 et EeA10 pour les deux groupes océaniques et lagunaires comme on aurait pu s'y attendre (Annexe B-Table Supp 10).

Afin d'avoir une image globale de la probabilité de présence de DL dans chaque pop, tous locus confondus, nous avons combiné les valeurs de probabilité de déséquilibre obtenues pour 15 paires de locus selon la méthode de Fisher où la variable $X^2 = -2 \sum \log_e(p_i)$ suit une loi de Khi2 à 30 degrés de liberté si les valeurs des 15 paires sont indépendantes.

L'annexe C-Figure 1 montre la relation entre cette probabilité globale de DL et celle obtenue pour le F_{is} global. On note une absence de corrélation nette, il n'y a donc probablement pas de cause biologique commune à ces deux phénomènes (intra-locus et inter-locus) qui risquent donc d'être des artefacts. On ne peut donc pas à partir de ces seuls résultats, conclure sur les questions posées initialement sur les interactions entre les deux entités.

2.2.4. Analyse multilocus

L'étude sur les locus pris séparément montre une corrélation forte entre la localisation lagunaire des échantillons et certains allèles aux locus EeA6 et EeA10. Cette corrélation forte se traduit par un F_{st} significatif entre ces deux entités génétiques, de l'ordre de 0,07 ($p < 0.0001$). Sur tous les locus regroupés, cette différence se voit directement sur l'arbre phylogénétique ainsi que sur l'AFC. L'étude des locus considérés 2 par 2 (déséquilibres de liaison) a montré très peu de cas significatifs. Sur l'échantillon global, on attendait en fait de voir un DL significatif entre les deux locus diagnostiques EeA6-EeA10 mais c'est seulement dans 3 paires impliquant le locus EeA7 et sur les échantillons regroupés que nous voyons un peu de significativité après correction de Bonferroni. A la lumière de ces résultats, on ne peut mettre en évidence par ces indicateurs de manière claire un effet Wahlund par mélange des deux entités dans le même échantillon. Il est difficile de savoir donc à quel phénomène biologique sont dus les F_{is} observés. Il est possible que les analyses locus par locus ou par paires de locus manquent de puissance. L'hypothèse de l'existence d'entités en mélange doit donc être testée en utilisant l'ensemble de l'information disponible, c'est-à-dire en considérant

tous les locus pris simultanément (et non pas seulement un par un ou deux à deux), c'est ce que nous avons fait grâce au logiciel *Structure*.

La première chose à faire est de décider du nombre k de populations dans lesquelles le logiciel va assigner les en maximisant l'équilibre de H-W et l'équilibre de liaison à l'intérieur de ces populations. La méthode d' (Evanno et al., 2005) montre un gain de probabilité postérieure maximale pour $k = 2$, ce qui est attendu pour la différence mer/lagune. Au delà de $k = 2$, on n'observe pas de saut significatif supplémentaire, donc on en reste sur notre choix de valeur de $k = 2$. Pour tester le niveau de convergence de *Structure*, j'ai refait les mêmes *runs* avec les mêmes conditions (nombre d'itérations, k ...) pour voir si on obtient les mêmes résultats, ceci en étudiant la corrélation entre les différents runs des valeurs d'assignation moyennes obtenues pour chaque échantillon. En outre, nous avons vérifié la stabilité de Log P(D) (Annexe C-Figure 2) pour 50 000 itérations, 100 000 « *burn-in* » (préchauffage), ces derniers étaient visiblement stables jusqu'à la fin du *burn-in* ce qui suggère qu'au delà de ces valeurs il n'y aurait pas de modification des résultats.

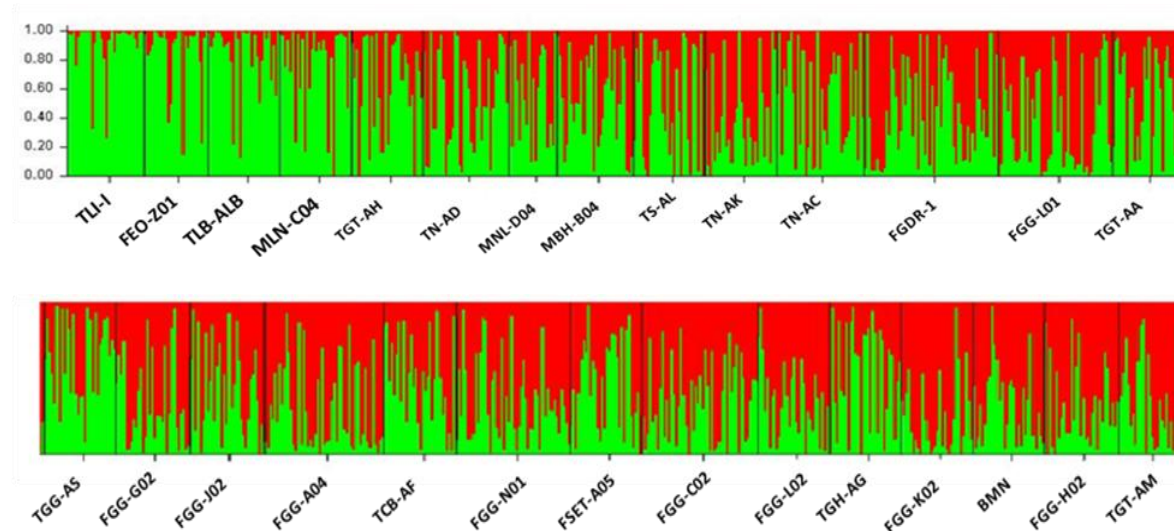


Figure.5: Résultats des tests de regroupement obtenus par méthode d'inférence bayésienne pour $k=2$ (moyenne des 20 répétitions) chaque barre verticale correspond à un individu, une couleur correspond à un regroupement génétique, les sites échantillonnés sont délimités par les barres noires.

Les deux groupes assignés par *Structure* (Figure 5) correspondent relativement bien aux populations lagunaires et aux populations océaniques visibles également sur l'arbre de la Figure 2. On peut différencier les quatre premières populations correspondant aux

échantillons lagunaires dont la couleur verte est majoritaire par rapport au reste des échantillons. Cependant, ces assignations ne sont pas parfaites (aucun échantillon n'apparaît comme « pur »). Néanmoins elles sont concordantes avec l'analyse d'AFC sur les 6 locus microsatellites (Figure 3). En étudiant la régression entre les coordonnées des populations sur l'axe de l'AFC qui porte le plus d'information sur la différenciation mer/lagune (axe1) et l'assignation au cluster lagunaire (Figure 6) dans les différents échantillons, on trouve une corrélation positive quand il s'agit des coordonnées des populations ($R^2 = 0,872$, $p < 0,00001$) ou des individus et $R^2 = 0,944$, $p < 0,00001$) (Figure 6). Ce qui suggère que les résultats convergent très bien selon les deux méthodes.

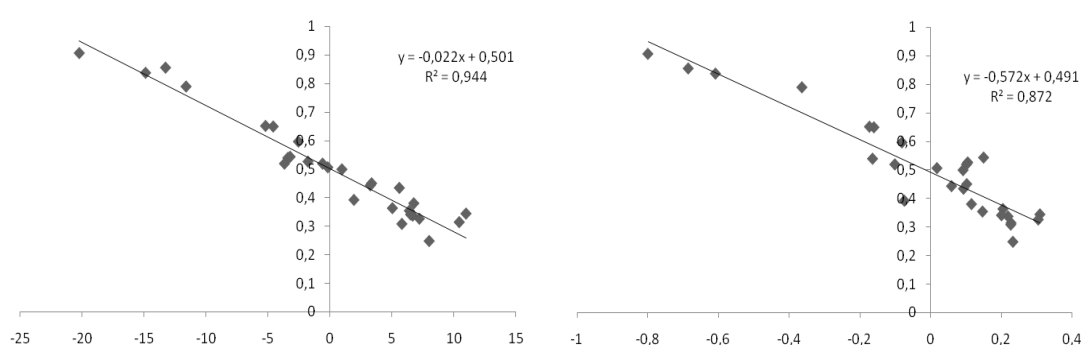


Figure 6 : Corrélation entre le pourcentage d'assignation à un cluster lagunaire (en ordonnées) et les coordonnées de chaque population sur l'axe 1 du AFC (en abscisse : (a) la moyenne de l'échantillon de chaque projection individuelle, (b) échantillons sur les centres de gravité))

On peut néanmoins se demander quelle est la part de ces assignations qui reflètent la réalité ? La présence des traits rouges (individus du cluster océanique), au milieu des traits verts (individus du cluster lagunaire) dans la figure 5 peut être interprétée de deux façons : soit il existe deux populations très bien séparées (une verte en lagune, une rouge en mer) et *Structure* ne les sépare pas très bien parce que la quantité d'information dans les deux principaux locus structurants est trop faible, soit cela reflète la réalité d'un mélange entre les deux entités : notre question est alors de savoir à quel niveau s'effectue ce mélange et avec quelles proportions ?

Chapitre II : Equilibre nucléo-cytoplasmique

1. Marqueur mitochondrial utilisé et méthodes analytiques

1.1 Généralités

Du fait du développement des techniques d'amplification de l'ADN et de séquençage, ce dernier marqueur est couramment utilisé en génétique des populations, notamment pour retracer l'histoire évolutive des taxons. Le génome mitochondrial présente comme avantage majeur d'évoluer rapidement (les systèmes de réparation de l'ADNmt sont très limités ce qui fait que le taux de mutation est en général plus élevé que celui de l'ADN nucléaire codant pour les protéines) donc d'être souvent polymorphe et de présenter, en absence d'effet sélectif, une relation directe entre ce polymorphisme et la taille de la population, d'être non-recombinant (puisque la transmission du génome mitochondrial d'une génération à une autre se fait uniquement par les mitochondries de l'ovule chez de nombreuses espèces) et de présenter un polymorphisme neutre. En outre, et selon (Asmussen and Basten, 1994), en raison des modes d'hérédité distincts entre le génome nucléaire et le génome mitochondrial, la comparaison des résultats obtenus sur les deux types de marqueurs peut permettre de révéler des processus non détectés par l'étude d'un seul type de marqueur, comme par exemple savoir si la migration est équivalente dans les deux sexes. Dans l'étude actuelle le mitochondrial nous permettra *via* les séquences, l'étude de l'expansion démographique.

1.2 Méthodes d'analyse

L'extraction de l'ADN est effectuée selon le protocole de DNAeasy tissu kit (Qiagen) selon les instructions du fabricant. Un couple d'amorce a été défini pour amplifier un fragment de 680pb du gène de cytochrome b de l'ADNmt (Annexe C-Tableau 2). Les séquences sont assemblées et alignées (Clustal) avec les séquences de référence. Puis un arbre de NeighbourJoining (fourni en Annexe B-Figure Supp 3) est produit par le logiciel MEGA 5 (Tamura et al., 2011) afin de déterminer à quel clade (A ou B) les séquences appartiennent.

1.3. Calcul des fréquences

On remarque la présence de chaque clade à l'intérieur des échantillons lagunaires et océanique avec des fréquences très variables comme illustré sur la figure 7. Les échantillons

lagunaires sont majoritairement du clade A à l'exception de celui de la France, qui est du clade B à 92% (Figure 8 et Annexe B-Table Supp 4b)

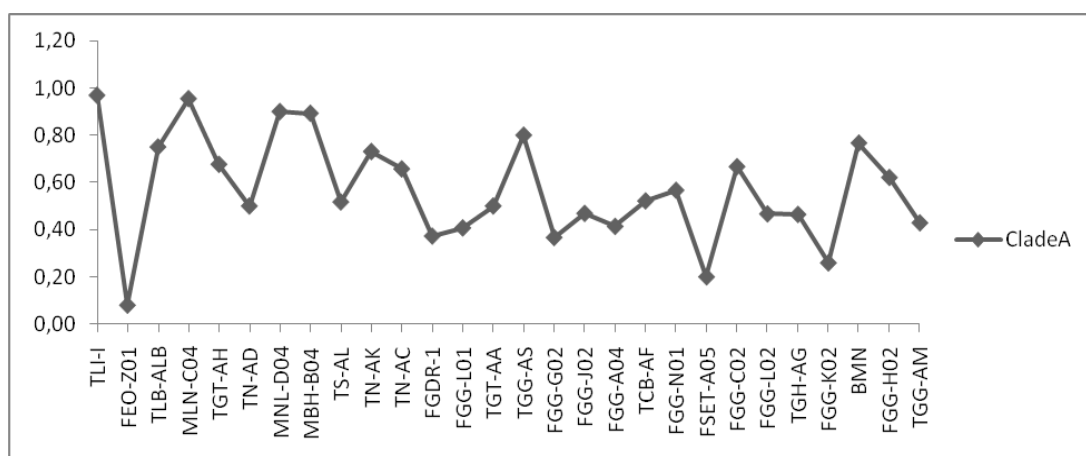


Figure7 : Distribution des fréquences du Clade A pour chaque population

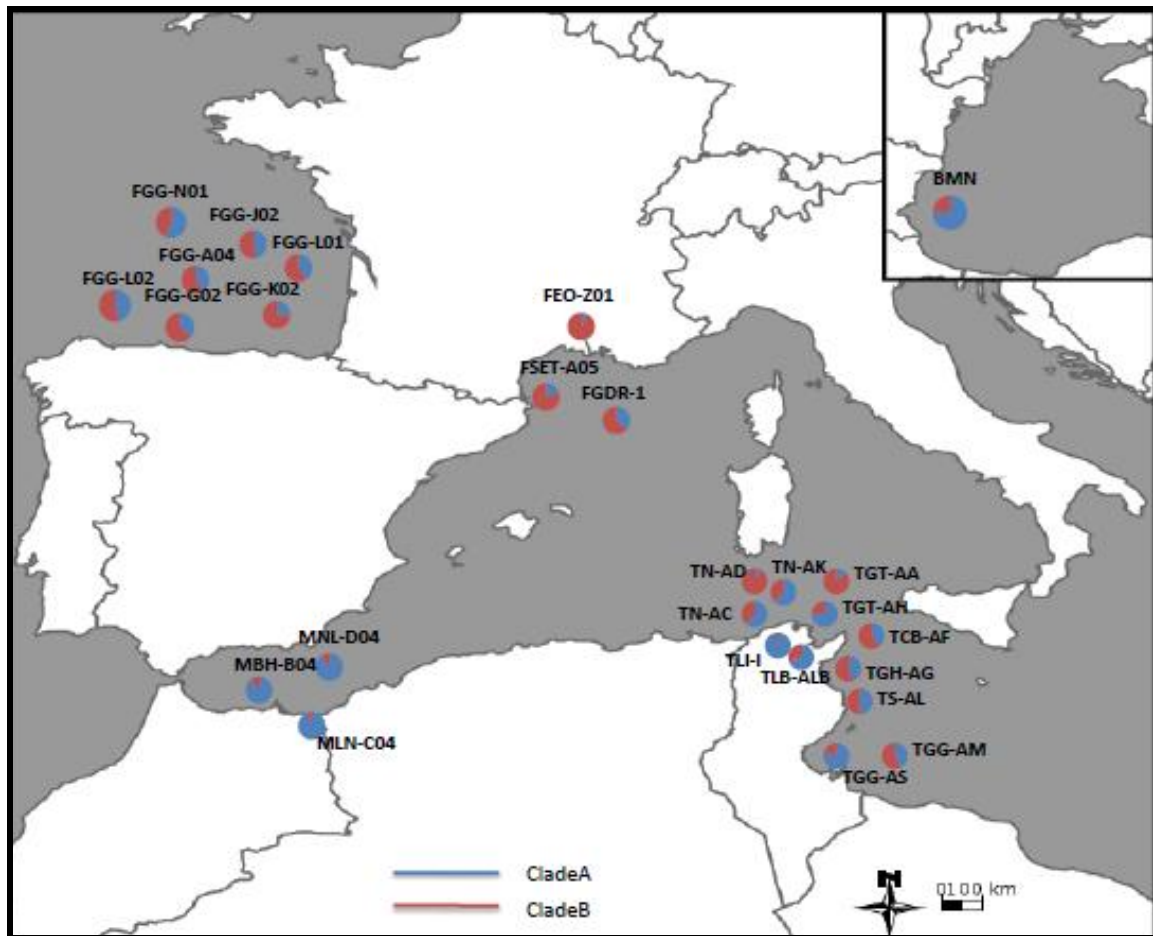


Figure 8: Répartition de la fréquence des clades A et B sur la totalité des données

2. Déséquilibre nucléo-cytoplasmique

Le déséquilibre nucléo-cytoplasmique (ou cyto-nucléaire) est une association non aléatoire des allèles ou génotypes aux locus nucléaires et mitochondriaux. Nous mettons en évidence cette liaison en regardant la corrélation entre le pourcentage d'assignation à un cluster lagunaire ou océanique et la fréquence des clades A ou B de chaque population. Nous aboutissons à une corrélation linéaire positive et significative ($R^2 = 0,138$, $p < 0,05$ sur l'ensemble de la collection, $R^2 = 0,415$, $p < 0,0006$ quand on ôte Z01) (Figure 9).

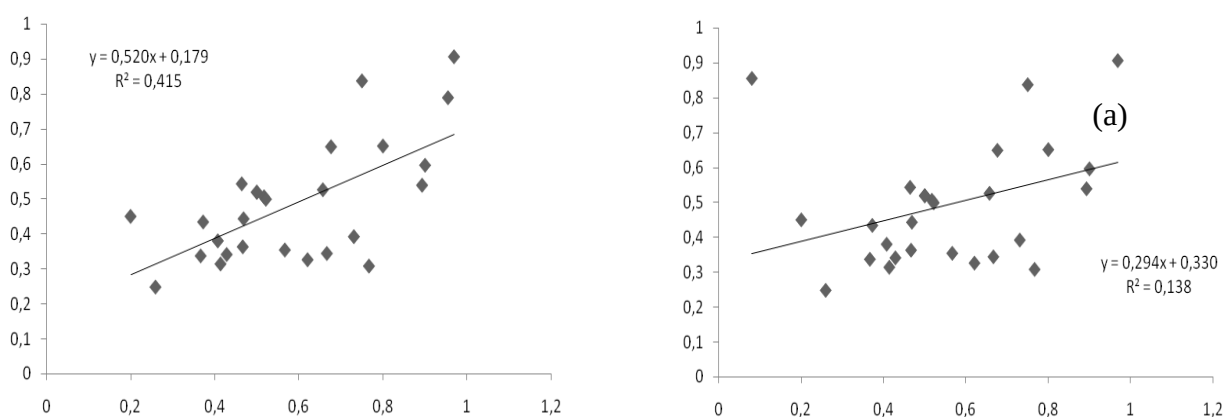


Figure 9 : Corrélation entre les fréquences du clade A par population (abscisse) et l'assignation à un cluster lagunaire (ordonnée) en considérant toutes les populations (a) et toute les populations à l'exception de FEO-Z01 (b)

Nous sommes donc en présence d'un déséquilibre nucléo-cytoplasmique. L'assignation aux clusters étant effectuée par *Structure* de manière complètement indépendante de la fréquence des clades, cette liaison entre mitochondrie et microsatellites ne peut donc pas être un artefact. Qu'il y ait une corrélation entre clades et clusters quand on mélange des lagunaires et des océaniques est normalement attendu, puisque les fréquences sont *a priori* différentes dans l'un et l'autre type de population. Par contre, il est intéressant de noter que ce déséquilibre persiste aussi quand on l'étudie sur les échantillons océaniques de Tunisie pris séparément ($R^2=0,203$, test de Student $p < 0,00001$). Ceci veut donc dire que, même à l'intérieur de ces échantillons de pleine mer, nous n'avons pas à faire à une seule entité homogène, et que donc les génomes des deux entités génétiques s'y confrontent probablement sans atteindre l'équilibre.

3. Synthèse sur les résultats de fréquences des gènes nucléaires et des clades mitochondriaux et discussion des résultats.

3.1 Synthèse des résultats

La littérature récente a souligné l'existence de formes génétiquement différenciées d'anchois au nord de la Méditerranée, ainsi que leur existence probable au sud sur des bases morphologiques ou biologiques. Dans la présente étude, nous confirmons sur nos échantillons de Tunisie ainsi que sur celui de la lagune de Nador au Maroc et de l'étang de l'Or en France l'existence d'anchois lagunaires (ou très côtiers) qui se distinguent génétiquement d'anchois plus océaniques (anchois du large). Sur nos 6 locus microsatellites, seuls 2 sont clairement associés à cette différenciation, l'un associé aux poissons lagunaires avec l'allèle EeA6-166 et l'autre associé aux lagunaires avec un allèle EeA10-132 et aux océaniques avec l'allèle EeA10-128. On note également une association mitochondrie-microsatellite particulièrement nette si on regarde par exemple la régression entre les fréquences du clade A (sans FEO-Z01) avec celles des allèles EeA6-166 et EeA10-128 ($R^2=0,342$, $p<0.0001$ et $R^2=0.404$, $p<0.0001$ respectivement). Ceci correspond à la situation décrite chez les anchois côtiers du nord de la Méditerranée occidentale (Golfe du Lion) par (Borsa et al., 2004), que l'analyse nucléaire montre comme quasiment identiques aux lagunaires tunisiens (Figures 2,3,4). La seule différence entre les deux rives de la Méditerranée est que dans le golfe du Lion, c'est le clade B qui est associé aux échantillons lagunaires, alors qu'en Tunisie c'est le clade A qui est majoritaire chez les lagunaires.

Pour ce qui est des échantillons du large, ils semblent relativement homogènes sur notre aire d'échantillonnage, avec une forte ressemblance entre ceux de Méditerranée Ouest et ceux du Golfe de Gascogne, comme l'ont noté déjà (Bouchenak-Khelladi et al., 2008), et comme cela est visible également dans l'étude de (Zarraonaindia et al., 2009). Les échantillons de Mer d'Alboran et de Mer Noire apparaissent comme très semblables aux précédents, contrairement à ce que proposent (Zarraonaindia et al., 2012) qui voient Alboran et Mer Egée comme légèrement différenciés.

3.2 Hybridation ou Introgression ?

Nos résultats nous permettent également d'aborder la question des mélanges entre les deux entités. Pour une espèce pélagique à forte capacité de dispersion, il semble difficile d'expliquer les différences de fréquences entre populations voisines sans faire intervenir deux sources génétiquement différenciées en proportion variable. Un autre petit pélagique d'écologie similaire, la sardine *Sardina pilchardus* ne présente pas de structure génétique en Méditerranée (Tinti et al., 2002). Dans nos échantillons de mer ouverte de Tunisie, nous observons des F_{st} nucléaires et cytoplasmiques non nuls (Annexe C-Tableau 3), et un déséquilibre nucléo-cytoplasmique persistant. Ceci veut dire que les échantillons capturés au large de la côte ne sont pas tous homogènes et présentent des gènes des deux entités en proportions variables. Il est difficile d'aller plus loin dans l'analyse pour savoir à quel type de mélange, s'il s'agit d'une introgression ou d'une hybridation partielle

3.3 Adaptation locale ?

Ce qui précède est donc la preuve qu'il existe un frein au mélange des deux entités, sinon tout serait beaucoup plus homogène même si au départ il existait deux populations différenciées comme l'indique la présence de deux clades mitochondriaux très différents. Nous sommes donc en présence d'un cas d'interaction entre génomes partiellement incompatibles. Il faut en effet expliquer une homogénéité presque totale aux locus EeA2, EeA7, EeA8, EeA9 à travers tous nos échantillons, et le maintien de différences très significatives pour EeA6 et EeA10 et également au locus mitochondrial.

Quelles sont les barrières génétiques susceptibles d'expliquer ces résultats ? Selon (Bierne et al., 2011), les barrières génétiques endogènes (résultats d'incompatibilités génétiques pré-ou post-zygotiques) peuvent expliquer la majorité des F_{st} anormaux observés (locus *outliers*), même lorsque ceux-ci sont en corrélation avec des variables environnementales. Ces obstacles endogènes sont connus pour pouvoir créer des « zones de tension », qui ont tendance à se retrouver associées avec des barrières exogènes résultant de la sélection écologique. Ce couplage des obstacles endogènes et exogènes peut se produire facilement dans les populations spatialement subdivisées, même si les locus impliqués sont non liés et la sélection écologique est faible.

L'essentiel de ces résultats fait l'objet d'un article soumis à Marine Biology (Annexe A)

Chapitre.III : Coalescence mitochondriale :

1. Méthode d'analyse:

L'inférence de l'histoire démographique des clades identifiés à la suite des analyses phylogénétiques est essentielle pour déterminer les facteurs ayant modelé l'évolution des populations d'anchois. La disparité de distribution des différences nucléotidiques observées entre les paires de séquences nucléotidiques (*mismatch distribution*) a été analysée.

La distribution attendue est unimodale pour les populations en expansion démographique alors qu'elle est multimodale pour les populations à l'équilibre (Slatkin and Hudson, 1991). Nous avons calculé également l'indice D de *Tajima*. Cet indice donne des informations sur la démographie des populations avec une valeur négative dans le cas d'une expansion démographique et une valeur positive dans les cas de populations subdivisées (Tajima, 1989a, 1989b). Tous ces paramètres démographiques ont été estimés avec le logiciel *Arlequin 2.00*.

2. Principaux résultats et points discutés

2.1. Haplotypes du cytochrome b de l'ADNmt

L'analyse de la diversité génétique d'un fragment de 680 paires de bases de long du gène mitochondrial de cytochrome b révèle l'existence de 210 sites polymorphes, permettant ainsi de définir 404 haplotypes sur l'ensemble des 818 individus séquencés. Les niveaux de diversité génétique sont présentés dans l'Annexe C-Tableau 4, Les moyennes de diversité haplotypique (h) et nucléotidique (π) sont respectivement de l'ordre de 0,982 et de 0,01. Les estimations de différenciation génétique (F_{st}) par paire de populations sont présentées dans l'Annexe C-Tableau 3.

L'étude de la diversité des deux clades séparément montre que la diversité haplotypique et nucléotidique est semblable entre les deux clades mitochondriaux (clade A ($P_i = 0.00768$) et clade B ($P_i = 0.007$)).

Des AMOVAs ont été effectuées sur la base du gène mitochondrial pour toutes les populations afin d'estimer la part respective de la variance génétique associée à différents niveaux de subdivisions géographiques : au sein des populations, entre populations et entre les deux groupes définis (mer et lagune) au sein du clade A et clade B. La variance associée à la différenciation entre les groupes s'est révélée faible et non significative (Tableau 3)

Niveau de structure	Indice de fixation	p
Clade A		
Entre Lagunes (sans le seul individu FEO-Z01) et Ocean	FCT = -0,00270	0,578
Entre pops à l'intérieur des lagunes ou océan	FSC = 0,02546	<0,0001
Entre populations	FST = 0,02283	<0,0001
Clade B		
Lagune (avec FEO-Z01 et 8 indiv ALB)-Ocean	FCT = 0,04054	0,081
Entre pop à l'intérieur des lagunes ou océan	FSC = 0,020	0,029
Entre populations	FST = 0,060	0,003

Tableau 3 : Résultats de l'analyse moléculaire de la variance (AMOVA) pour les données mitochondriales

2.2. Coalescence et expansion démographique

2.2 1. Datation de la divergence des deux clades:

Comme tous les gènes codant pour des protéines, le *cyt b* montre des variations dans les taux de substitutions nucléotidique selon la position dans le codon et dans le type de substitutions, les transitions prédominant largement sur les transversions. La redondance du code génétique implique que les mutations sur la troisième position des codons aboutissent rarement à une substitution d'acides aminés. Cette caractéristique permet à cette position d'évoluer plus rapidement que les première et seconde positions, car elle est soumise à de plus faibles contraintes sélectives.

Afin de pouvoir estimer la divergence entre les deux clades mitochondriaux, nous avons eu recours aux données de la bibliographie disponible sur la datation des espèces d'anchois (Grant et al 2010). Nous sommes partis de l'arbre de plus haute probabilité postérieure bayésienne proposé par ces auteurs pour 521 pb de cytochrome b de l'ADN mitochondrial basé sur le modèle de substitution de Tamura–Nei qui calibre la divergence entre *Cetengraulis edentulus* et *C. mysticetus* sur la base de la fermeture de l'Isthme de Panama.

(figure 9). Cette fermeture est connue pour s'être produite aux alentours de -3 millions d'années. La divergence moléculaire du *Cyt b* entre ces deux espèces estimée par MEGA sous le modèle de substitution Tamura–Nei est de 0,01 si l'on prend compte le 1^{er} et le 2^{ème} codon uniquement ; la divergence est de 1,866 dans le cas où l'on considère le 3^{ème} codon seulement (ce qui représente un cas de saturation très fort puisque cette divergence dépasse 100%). Pour les clades A et B de *E. encrasicolus* ces valeurs sont sur nos échantillons entre 0

(1^{er} et 2^{ème} codon) et 0,07 (3^{ème} codon). Puisque nous avons 0 pour la 1^{ère} et 2^{nde} position, nous ne pouvons donc pas établir une simple règle de proportionnalité entre les valeurs observées pour dater la divergence des deux clades de *E. encrasicolus*. Pour la 3^{ème} paire de base cela donnerait $(0,07 / 1,866) * 3 \text{ MA} = 110\,000 \text{ ans}$ mais ceci n'est pas forcément très robuste étant donné que sur cette position des *C. edentulus* et *C. mysticetus*, la divergence exacte est très mal estimée à cause de la saturation. Nous avons donc décidé de dater la divergence entre nos deux clades en nous basant sur une simple interprétation des données de Grant et al. (2010) : la distance entre *E. anchoita* et *E. ringens* qui est de 0,291 sur la 3^{ème} position, correspondant selon leur estimation (peu solide pour les raisons exposées plus haut) à -1,5 MA (intervalle de crédibilité à 95% compris entre -3,1 et -0,25 MA). La distance qui sépare le clade A du clade B étant de 0,07, ceci correspondrait à une divergence moyenne de 360 000 ans (IC à 95% entre - 745 000 et -60 000 ans). Pour les taux de divergence estimés sur la 3^{ème} paire de base, en ramenant ce calcul au nombre de positions susceptibles de muter sur le fragment que nous avons séquencé, ceci correspondrait à $2\mu = 1,94 \cdot 10^{-7} \text{ subst /nucl/an}$ (IC à 95% : $0,94 \cdot 10^{-7} < < 11,6 \cdot 10^{-7} \text{ subst /nucl/an}$). La valeur médiane pour le taux de substitution serait donc de $\mu = 0,97 \cdot 10^{-7} \text{ subst /nucl/an}$, soit ~ 10% par MA, ce qui est 5 fois supérieur au taux global (toutes positions confondues) utilisé par Grant et al (2005). La faible sécurité sur la calibration des données ne permet de savoir réellement si la divergence entre les deux clades date de la précédente glaciation ou bien lui est encore antérieure.

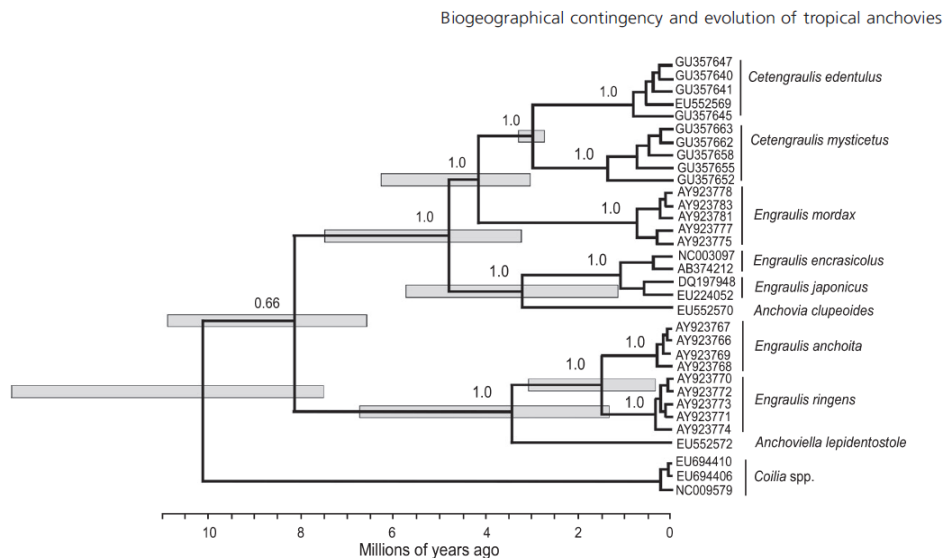


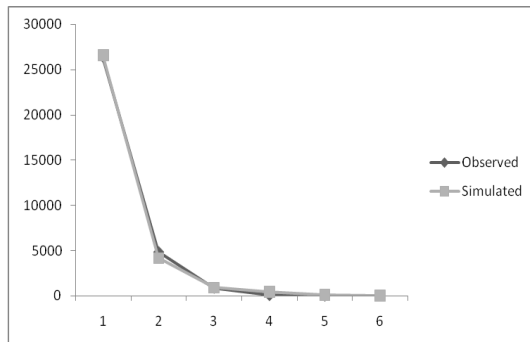
Figure 9.Arbre de plus haute probabilité postérieure bayésienne proposé Grant et al (2010) pour le cytochrome b de l'ADN mitochondrial (521 pb) basé sur le modèle de substitution de Tamura–Nei (calibration de la divergence entre *Cetengraulis edentulus* et *C. mysticetus* sur la base de la fermeture de l'Isthme de Panama)

2.2 2. Mismatches et expansion démographique

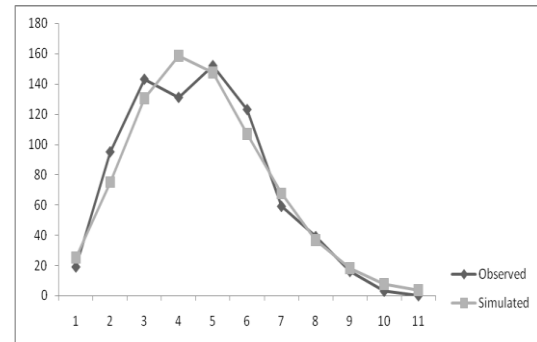
La distribution du nombre observé de différences entre paires d'haplotype sur le troisième codon des séquences, (courbes de Mismatch) (Figure 10) sous le modèle d'expansion soudaine s'est avérée unimodale pour les deux clades. Ceci est le signe d'une expansion démographique récente. Le test sous *Arlequin*, montre qu'aucun des deux modèles d'expansion soudaine ou spatiale, n'est rejeté. D'autre part et en étudiant les Mismatch clade par clade sur le premier et le deuxième codon, on trouve que les valeurs de tau ne sont pas nulles (Tau = 3.865 pour Clade A et Tau= 5.85 Clade B et en l'étudiant sur le 3^{ème} codon Tau est de 3 pour A et pour B, et que la diversité existe au sein de chaque clade, ceci pourra être expliqué par la présence de polymorphisme dans chaque clade et que en comparant les deux ce polymorphisme ne se traduit pas par une divergence.)

Chez l'anchois, (Grant et al., 2005 ; Magoulas et al., 2006) ont déjà appliqué ces méthodes. Ces auteurs mettent en évidence un goulot d'étranglement et une expansion démographique des deux clades, les valeurs du D de Tajima sont de l'ordre de D= -2.180, p=0 pour clade A et D= - 1.544, p=0,034 pour clade B (Grant et al. 2005). Les valeurs de Tajima trouvés dans

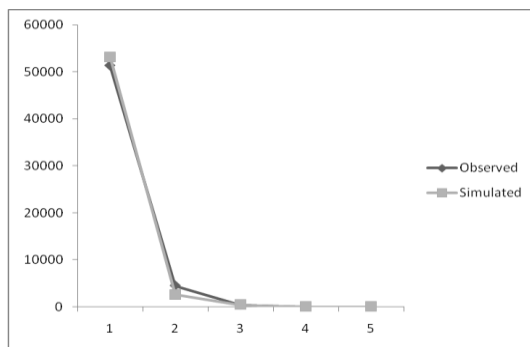
notre étude sont de l'ordre de $D = -2.622$ pour clade A $p < 0.001$ et $D = -2.229$ $p < 0.01$ pour clade B



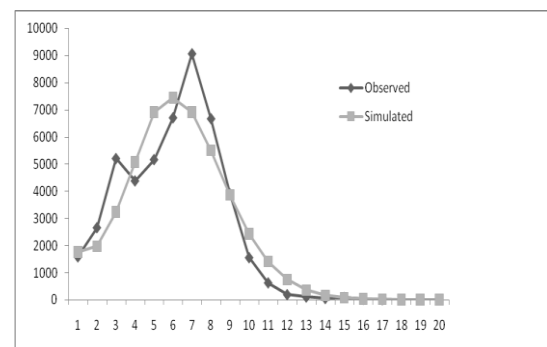
(a)



(b)



(a')



(b')

Figure 10 : distribution du nombre observé de différence nucléotidiques entre paire d'haplotypes pour les clades mitochondriaux A ((a) / sur le 3ème codon sur tous les codons (b) et B (3ème codon (a')/tous les codons (b'))

Discussion générale

Les marqueurs génétiques intègrent de façon générale les histoires passées et présentes des espèces. L'un des objectifs de cette étude spatiale a donc été d'essayer de voir les signatures génétiques présentes des histoires passées. La phylogéographie permet une déduction des phénomènes historiques (génétique et démographique) qui ont conduit à la répartition et à la structuration actuelle des populations (Avice, 2000). En outre, elle permet d'explorer des scénarios évolutifs historiques d'une espèce, et ceci à l'échelle spatiale de son aire de répartition et à l'échelle temporelle. Elle intègre des données issues de domaines différents, comme la biologie moléculaire, la démographie, la biogéographie ou encore la paléontologie et les événements climatiques et environnementales qui ont pu peser sur cette espèce. Dans un passé récent, le Pléistocène (entre -1,87 millions d'années et -10 000 ans AEC*) a été caractérisée par plusieurs oscillations climatiques sévères. Ces cycles glaciaires n'ont pas été sans conséquences sur les êtres vivants et leur distribution géographique (Avice, 2000) (Hewitt, 2000). C'est dans cette perspective phylogéographique que nous avons essayé de reconstituer l'histoire possible de nos deux entités génétiques présentes chez l'anchois.

Scénarios possibles :

Le scénario le plus simple pour expliquer les résultats observés est que les anchois auraient divergé en allopatrie, puis se seraient retrouvés en sympatrie par contact secondaire, tel que proposé notamment par (Grant et al., 2005). Les changements climatiques nombreux qui ont affecté la paléo-océanographie de la Méditerranée au cours des derniers millions d'années permettent souvent pendant les épisodes froids la séparation de populations dans deux zones refuges, une en Méditerranée Est et l'autre par exemple le long de la côte atlantique de l'Afrique.. Au cours de ces phases de séparation un début d'isolement reproductif peut apparaître, et si le temps de séparation est suffisant, les deux populations finiront par devenir partiellement incompatibles. Ce qui fait que quand elles se rencontrent à nouveau, même si des hybrides se produisent si l'isolement prézygotique n'est pas complet, leur fitness est moins bonne que celle des parents et il se crée une zone de tension qui correspond à une zone puits (*hybrid sink*, (Barton and Hewitt, 1985). Ce cas est décrit par exemple pour les soles *Solea aegyptiaca* / *S. senegalensis* (Ouanes et al., 2011), mais aussi pour le loup *D. labrax* (Lemaire et al., 2005). Cette zone va se maintenir si les populations parentales peuvent occuper des espaces différents. A travers cette zone, également appelée barrière semi-

perméable, les gènes compatibles (neutres) vont pouvoir circuler, alors que les locus incompatibles vont créer la barrière et ne pas pouvoir la traverser. La divergence entre ces deux populations est alors irréversible puisque le retour à la compatibilité reproductive devient impossible. Chaque population parentale se retrouve alors « piégée » dans l'environnement où elle peut se maintenir et éventuellement subir des pressions de sélection différentes.

Cela pourrait être le cas pour les anchois. Il est clair que des hybridations ont eu lieu, puisque qu'on retrouve les deux clades dans toutes les populations et il est difficile de dire de quel endroit chacun est originaire. Le clade A est majoritairement présent en Mer Noire mais il est également très fréquent dans l'Océan Atlantique du Portugal au Sénégal (Magoulas et al., 2006) et également en Afrique du Sud (Grant et al 2005), ce qui fait qu'il n'y a pas une origine géographique évidente pour l'un et l'autre clade. Néanmoins, le clade A ayant subi apparemment une expansion plus récente que le clade B (cf. chap. mitoch) on pourrait proposer le scénario hypothétique suivant : la population d'anchois possédant originellement uniquement le clade B était présente en Méditerranée lors du début du réchauffement à la fin de la dernière glaciation, et celle possédant uniquement le clade A était plus au sud dans l'Atlantique. Cette entité, en amorçant sa remontée vers le nord, aurait envahi la Méditerranée et rencontré l'entité possédant le clade B en plusieurs endroits. Des hybrides se seraient produits à chaque fois, avec séparation selon les milieux et création de zones semi-perméables comme décrit plus haut. On peut penser que l'entité présente seule à l'origine en Méditerranée occupait déjà les lagunes au fur et à mesure de leur création lors de la remontée du niveau de la mer et les estuaires, et que c'est elle qui s'est retrouvée naturellement confinée dans ces milieux quand l'autre entité est arrivée, naturellement depuis le large. Les anchois côtiers seraient ainsi les descendants des premiers anchois présents en Méditerranée, où ils se seraient retrouvés fragmentés et isolés les uns des autres en autant de populations qu'il y a de systèmes lagunaires et estuariens capables de maintenir des populations suffisantes de cette espèce. Ils sont maintenant isolés les uns des autres et ont subi des événements d'hybridation initiale qui se sont passés à des endroits différents, ce qui explique des différences génétiques entre eux, par exemple la fréquence du clade B est très différente entre les côtiers du Golfe du Lion et ceux du Golfe de Tunis. Néanmoins ceux du Maroc sont très similaires à ceux de Tunisie

Ce scénario explique assez bien la distribution et la composition actuelle des anchois côtiers de Méditerranée. Pour ce qui est des anchois du large, nos résultats sont basés sur trop peu de

locus et ne nous permettent pas de voir les différences et les ressemblances mises en évidence dans les articles récents. Néanmoins, il semble que les anchois du large de la Mer d'Alboran que nous avons analysés se démarquent un peu des autres anchois du large (cf. leur position dans l'arbre de la fig.2, et il est probable, ainsi que cela semble être le cas dans les travaux de (Zarraonaindia et al., 2012) et (Borrell et al., 2012), que les anchois du large existent aussi sous plusieurs formes, avec plusieurs sous-groupes occupant des espaces géographiques différents, un peu de la même manière que pour les moules du complexe *Mytilus edulis* / *galloprovincialis*, il existe plusieurs formes partiellement isolées, avec par exemple des *galloprovincialis* de Méditerranée légèrement différentes de celles de l'Atlantique (Bierne et al., 2003).

Conclusion générale

Avec peu de marqueurs nos résultats apportent un éclairage supplémentaire sur l'existence de deux formes d'anchois qui coexistent l'une côtière l'autre au large. L'analyse mitochondriale confirme ce qui était déjà connu à savoir l'existence des 2 clades présents chacun dans les 2 écotypes. Les niveaux de mélange variable détecté notamment dans les eaux tunisiennes montrent que l'interaction des 2 écotypes est un phénomène dynamique et que donc il existe un mélange (superposition ou taux de divergence) et que donc les populations locales d'anchois relativement différentes les une des autres du fait du niveau de mélange variable entre entités (qu'on peut qualifier d'écotypes). On ne peut pas donc les assimiler à un stock unique, et ceci bien entendu à des conséquences pour la biologie des pêches.

Pour éclaircir la nature exacte de ces interactions (simple mélange, hybridation, introgression) il est nécessaire d'augmenter considérablement le nombre de locus disponibles.

Abréviation :

AEC : Avant Ere commune

REFERENCES BIBLIOGRAPHIQUES

- Asmussen, M.A., and Basten, C.J. (1994). Sampling theory for cytonuclear disequilibria. *Genetics* 138, 1351–1363.
- Avice, J.C. (2000). *Phylogeography: The History and Formation of Species* (Harvard University Press).
- Bahri-Sfar, L., Lemaire, C., Ben Hassine, O.K., and Bonhomme, F. (2000). Fragmentation of sea bass populations in the western and eastern Mediterranean as revealed by microsatellite polymorphism. *Proc Biol Sci* 267, 929–935.
- Bargelloni, L., Alarcon, J.A., Alvarez, M.C., Penzo, E., Magoulas, A., Palma, J., and Patarnello, T. (2005). The Atlantic–Mediterranean transition: Discordant genetic patterns in two seabream species, *Diplodus puntazzo* (Cetti) and *Diplodus sargus* (L.). *Molecular Phylogenetics and Evolution* 36, 523–535.
- Barton, N.H., and Hewitt, G.M. (1985). Analysis of Hybrid Zones. *Annual Review of Ecology and Systematics* 16, 113–148.
- Baus, E., Darrock, D.J., and Bruford, M.W. (2005). Gene-flow patterns in Atlantic and Mediterranean populations of the Lusitanian sea star *Asterina gibbosa*. *Molecular Ecology* 14, 3373–3382.
- Belkhir, K., Borsa, P., Chikhi, L., Raufaste, N., and Bonhomme, F. (2004). GENETIX 4.05, logiciel sous Windows TM pour la génétique des populations. Laboratoire Génome, Populations, Interactions, CNRS UMR 5000, Université de Montpellier II, Montpellier (France).
- Bembo, D.G., Carvalho, G.R., Cingolani, N., and Pitcher, T.J. (1996a). Electrophoretic analysis of stock structure in Northern Mediterranean anchovies, *Engraulis encrasicolus*. *ICES J. Mar. Sci.* 53, 115–128.
- Bembo, D.G., Carvalho, G.R., Cingolani, N., Arneri, E., Giannetti, G., and Pitcher, T.J. (1996b). Allozymic and morphometric evidence for two stocks of the European anchovy *Engraulis encrasicolus* in Adriatic waters. *Marine Biology* 126, 529–538.
- Bierne, N., Daguin, C., Bonhomme, F., David, P., and Borsa, P. (2003). Direct selection on allozymes is not required to explain heterogeneity among marker loci across a *Mytilus* hybrid zone. *Molecular Ecology* 12, 2505–2510.
- Bierne, N., Welch, J., Loire, E., Bonhomme, F., and David, P. (2011). The coupling hypothesis: why genome scans may fail to map local adaptation genes. *Mol. Ecol.* 20, 2044–2072.
- Black, W.C., and Krafur, E.S. (1985). A FORTRAN program for the calculation and analysis of two-locus linkage disequilibrium coefficients. *Theoretical and Applied Genetics* 70, 491–496.

- Bloom, D.D., and Lovejoy, N.R. (2012). Molecular phylogenetics reveals a pattern of biome conservatism in New World anchovies (family Engraulidae). *Journal of Evolutionary Biology* 25, 701–715.
- Borrell, Y.J., Piñera, J.A., Prado, J.A.S., and Blanco, G. (2012). Mitochondrial DNA and microsatellite genetic differentiation in the European anchovy *Engraulis encrasicolus* L. *ICES J. Mar. Sci.* 69, 1357–1371.
- Borsa, P. (2002). Allozyme, mitochondrial-DNA, and morphometric variability indicate cryptic species of anchovy (*Engraulis encrasicolus*). *Biological Journal of the Linnean Society* 75, 261–269.
- Borsa, P., Collet, A., and Durand, J.-D. (2004). Nuclear-DNA markers confirm the presence of two anchovy species in the Mediterranean. *Comptes Rendus Biologies* 327, 1113–1123.
- Bouchenak-Khelladi, Y., Durand, J.-D., Magoulas, A., and Borsa, P. (2008). Geographic structure of European anchovy: A nuclear-DNA study. *Journal of Sea Research* 59, 269–278.
- Chashchin, A.K. (1996). The black Sea populations of anchovy. *Scientia Marina* 60, 219–225.
- Chen, C.-S., Tzeng, C.-H., and Chiu, T.-S. (2009). Morphological and Molecular Analyses Reveal Separations among Spatiotemporal Populations of Anchovy (*Engraulis japonicus*) in the Southern East China Sea. *Zoological Studies* 49(2), 270–282.
- Doherty, P.J., Planes, S., and Mather, P. (1995). Gene Flow and Larval Duration in Seven Species of Fish from the Great Barrier Reef. *Ecology* 76, 2373–2391.
- Dulzetto, F. (1947). *L'Engraulis dei laghi di Ganzirri e del Faro*. Mem. Soc. It. Delle Scienze (dei XL), Serie III 26.
- Evanno, G., Regnaut, S., and Goudet, J. (2005). Detecting the number of clusters of individuals using the software structure: a simulation study. *Molecular Ecology* 14, 2611–2620.
- Excoffier, L., Laval, G., and Schneider, S. (2007). Arlequin (version 3.0): An integrated software package for population genetics data analysis. *Evol Bioinform Online* 1, 47–50.
- Fage (1920). (Report on the Danish oceanographical Expedition 1908-1910 to the Mediterranean and adjacent Seas) (Biology).
- Falush, D., Stephens, M., and Pritchard, J.K. (2007). Inference of population structure using multilocus genotype data: dominant markers and null alleles. *Molecular Ecology Notes* 7, 574–578.
- Féral, J.-P. (2002). How useful are the genetic markers in attempts to understand and manage marine biodiversity? *Journal of Experimental Marine Biology and Ecology* 268, 121–145.
- Fisher, D.H. (1987). Knowledge acquisition via incremental conceptual clustering. *Machine Learning* 2, 139–172.
- Froese, R. and D. Pauly. Editors. 2013. FishBase.

- Grant, W. a. S., and Bowen, B.W. (1998). Shallow population histories in deep evolutionary lineages of marine fishes: insights from sardines and anchovies and lessons for conservation. *J Hered* 89, 415–426.
- Grant, W.S., Leslie, R.W., and Bowen, B.W. (2005). Molecular genetic assessment of bipolarity in the anchovy genus *Engraulis*. *Journal of Fish Biology* 67, 1242–1265.
- Guinand, B., Bouvet, Y., and Brohon, B. (1996). Spatial aspects of genetic differentiation of the European chub in the Rhone River basin. *Journal of Fish Biology* 49, 714–726.
- Hartel, D.. (1994). *Génétique des populations*.
- Hewitt, G. (2000). The genetic legacy of the Quaternary ice ages. *Nature* 405, 907–913.
- Hewitt, G.M. (2001). Speciation, hybrid zones and phylogeography — or seeing genes in space and time. *Molecular Ecology* 10, 537–549.
- Lemaire, C., Versini, J.-J., and Bonhomme, F. (2005). Maintenance of genetic differentiation across a transition zone in the sea: discordance between nuclear and cytoplasmic markers. *Journal of Evolutionary Biology* 18, 70–80.
- Magoulas, A., Tsimenides, N., and Zouros, E. (1996). Mitochondrial DNA phylogeny and the reconstruction of the population history of a species: the case of the European anchovy (*Engraulis encrasicolus*). *Mol Biol Evol* 13, 178–190.
- Magoulas, A., Castilho, R., Caetano, S., Marcato, S., and Patarnello, T. (2006). Mitochondrial DNA reveals a mosaic pattern of phylogeographical structure in Atlantic and Mediterranean populations of anchovy (*Engraulis encrasicolus*). *Molecular Phylogenetics and Evolution* 39, 734–746.
- Mayr, E. (1942). *Systematics and Origin of P* (Harvard University Press).
- Naciri, M., Lemaire, C., Borsa, P., and Bonhomme, F. (1999). Genetic study of the Atlantic/Mediterranean transition in sea bass (*Dicentrarchus labrax*). *J Hered* 90, 591–596.
- Van Oosterhout, C., Hutchinson, W.F., Wills, D.P.M., and Shipley, P. (2004). micro-checker: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes* 4, 535–538.
- Ouanes, K., Bahri-Sfar, L., Ben Hassine, O.K., and Bonhomme, F. (2011). Expanding hybrid zone between *Solea aegyptiaca* and *Solea senegalensis*: genetic evidence over two decades. *Mol. Ecol.* 20, 1717–1728.
- Planes, S. (1998). Genetic diversity and dispersal capabilities in marine fish. *Evol.Biol* 30, 253–298.
- Quignard, J.P., Hamdouni, T., and Zaouali, J. (1973). Données préliminaires sur les caractères biométriques des anchois *Engraulis encrasicolus* (Linne, 1758) des côtes de Tunisie et du Lac Ichkeul. *Revue des Travaux de l’Institut des Pêches Maritimes* 37, 191–196.

- Sanz, N., García-Marín, J.-L., Viñas, J., Roldán, M., and Pla, C. (2008). Spawning groups of European anchovy: population structure and management implications. *ICES J. Mar. Sci.* 65, 1635–1644.
- Slatkin, M., and Hudson, R.R. (1991). Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. *Genetics* 129, 555–562.
- Tajima, F. (1989a). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123, 585–595.
- Tajima, F. (1989b). The effect of change in population size on DNA polymorphism. *Genetics* 123, 597–601.
- Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. (2011). MEGA5: Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Mol Biol Evol* 28, 2731–2739.
- Tinti, F., Di Nunno, C., Guarniero, I., Talenti, M., Tommasini, S., Fabbri, E., and Piccinetti, C. (2002). Mitochondrial DNA Sequence Variation Suggests the Lack of Genetic Heterogeneity in the Adriatic and Ionian Stocks of *Sardina pilchardus*. *Marine Biotechnology* 4, 163–172.
- Tudela, S. (1999). Morphological variability in a Mediterranean, genetically homogeneous population of the European anchovy, *Engraulis encrasicolus*. *Fisheries Research* 42, 229–243.
- Tudela, S., García-Marín, J., and Pla, C. (1999). Genetic structure of the European anchovy, *Engraulis encrasicolus* L., in the north-west Mediterranean. *Journal of Experimental Marine Biology and Ecology* 234, 95–109.
- Valen, L.V. (1976). Ecological Species, Multispecies, and Oaks. *Taxon* 25, 233–239.
- Weir, B.S., and Cockerham, C.C. (1984). Estimating F-Statistics for the Analysis of Population Structure. *Evolution* 38, 1358–1370.
- Yu, H.-T., Lee, Y.-J., Huang, S.-W., and Chiu, T.-S. (2002). Genetic analysis of the populations of Japanese anchovy (*Engraulidae*: *Engraulis japonicus*) using microsatellite DNA. *Mar. Biotechnol.* 4, 471–479.
- Zarraonaindia, I., Pardo, M.A., Iriondo, M., Manzano, C., and Estonba, A. (2009). Microsatellite variability in European anchovy (*Engraulis encrasicolus*) calls for further investigation of its genetic structure and biogeography. *ICES J. Mar. Sci.* 66, 2176–2182.
- Zarraonaindia, I., Iriondo, M., Albaina, A., Pardo, M.A., Manzano, C., Grant, W.S., Irigoien, X., and Estonba, A. (2012). Multiple SNP Markers Reveal Fine-Scale Population and Deep Phylogeographic Structure in European Anchovy (*Engraulis encrasicolus* L.). *PLoS ONE* 7, e42201.
- Zhang, D.-X. (2004). Lepidopteran microsatellite DNA: redundant but promising. *Trends in Ecology & Evolution* 19, 507–509.

ANNEXE A : Article soumis au journal *Marine Biology*

Existence of two widespread semi-isolated genetic entities within Mediterranean anchovies

S Oueslati^{1,2,3}, K Fadhlou-Zid², O Kada⁴, MT Augé¹, JP Quignard⁵, F Bonhomme¹

¹: Institut des Sciences de l'Evolution, CNRS UMR554, Université de Montpellier II, place Eugène Bataillon, 34095 Montpellier cedex 5 France

²: Institut National des Sciences et Technologie de la Mer, Port de pêche 2060, La Goulette, Tunisie

³: Institut National Agronomique de Tunisie, 43 Avenue Charles Nicolle 1082, Tunis-Mahrajène, Tunisie

⁴: Institut National de Recherches Halieutiques, 13 boulevard Zerkouni 62 000, Nador, Maroc

⁵: Laboratoire d'Ichtyologie, 1 allée de l'Hermitage, 34170 Castelnau-Le-Lez, France

Abstract

We investigated genetic transitions in European anchovies (*Engraulidae*) by analysing one mitochondrial and six nuclear microsatellite loci in samples from 28 geographical locations, primarily in the Western Mediterranean with external samples from the Black Sea in the East, and the Gulf of Biscay in the West. Four samples were collected from lagoons or shallow waters in France, Morocco and Tunisia. Two strongly differentiated loci and nucleo-cytoplasmic disequilibrium allowed us unambiguously to identify a coastal genetic entity that is distinct from the more offshore marine populations. By contrast, four microsatellites were undifferentiated within the entire set of samples, testifying to a variable permeability of the two interacting genomes. Our findings strengthen and extend previous reports of the existence of a coastal anchovy entity that is widespread in Mediterranean lagoons.

Introduction

The so-called European anchovy, *Engraulis encrasicolus* L., is a clupeoid fish distributed in the Mediterranean, Black, and Azov Seas, and along the eastern Atlantic coastline from Norway to Angola (Whitehead et al. 1988), with one further southern hemisphere population off the coast of South-Africa (Grant 2005). Like other pelagic clupeoids, it can achieve considerable biomass and undergoes crash-flush cycles (e.g see (Uriarte et al. 1996)). In recent decades, several studies have attempted to delineate the genetic structure of this marine fish species, characterised by high gene flow. These surveys have reported conflicting results depending on the sampling scheme and the genetic markers used, but have nonetheless found surprisingly high overall heterogeneity of Mediterranean populations, detected with allozymes and mitochondrial DNA (reviewed in Borsa 2002). Atlanto-Mediterranean anchovy are peculiar in harbouring two highly divergent and geographically widespread mitochondrial clades whose frequency varies within samples (Magoulas et al. 1996, 2006). Borsa (2002) revealed linkage disequilibrium in nuclear data, suggesting the presence of two entities, one coastal corresponding to what Gulf of Lions' fishermen call the "white anchovies", and a more offshore one, corresponding to what locally is called "blue anchovy". The existence of coastal forms of anchovies in the Mediterranean had already been reported in the literature on the basis of morpho-anatomical and behavioural differences. Dulzetto (1947) described anchovy from the Ganzirri lagoon in Sicily as a separate species (*E. rissoi*). Quignard et al. (1973) and Messaoud et al. (2011) found a coastal form in Lake Ischkeul and Bizerte lagoon in Tunisia, on the basis of morphological criteria, and considered it identical to *E. rissoi* from Sicily. Kada et al. (2009) showed that an anchovy of Nador lagoon in Morocco displays some

clear morphological and biological differences from marine *Engraulis encrasicolus*. Bembo et al. (1996) identified offshore and inshore stocks in the Adriatic based on electrophoretic and morphological differences that Borsa (2002) possibly equated to two biological species present elsewhere in the Mediterranean. Borsa et al. (2004) proposed the name *E. albidus* for lagoonal anchovies in the Gulf of Lions. The existence of a coastal form is further supported by the results of Borsa et al. (2004) and Bouchenak-Khelladi et al. (2008) based primarily on the nuclear EPIC marker CK6-2. On this basis, these authors showed that coastal anchovies from the Gulf of Lions are genetically differentiated from their offshore counterparts. So far very few studies, other than those mentioned above, have included both kind of samples in the same survey to investigate the population genetics of *E. encrasicolus* and specifically identified inshore and offshore samples.

Nevertheless, it can be assumed that such samples were incidentally included in some previous studies. Based on allozymes, Sanz et al. (2008) proposed without demonstration that the sample of the Alboran Sea they analyzed corresponded in part to the coastal species. However, it does not appear that the recent studies of Borrell et al. (2012) and Zarraonaindia et al. (2012) based on microsatellites, SNPs and mitochondrial DNA, sampled coastal forms, although the latter authors did propose the existence of two groups of anchovies. One group would be associated with areas of deep water upwelling on narrow continental shelves, such as that of the Iberian Atlantic; the second group associated with wide continental shelves, which would include the NE Atlantic (comprising the North Sea and the Bay of Biscay) and the Mediterranean. Nevertheless, this putative distinction does not overlap with the analysis of lagoon vs. marine forms that we specifically address in this study. We extended the genetic characterisation of coastal anchovies to the southern shore of the Mediterranean by analysing inshore and offshore samples from Morocco (Nador lagoon and offshore samples from Alboran Sea) and Tunisia (Bizerte lagoon, Lake Ischkeul and offshore waters). Using mitochondrial DNA cytochrome b and six microsatellite nuclear loci, we tested the hypothesis of the widespread occurrence of two genetically distinct entities with partially incompatible gene pools interacting with each other, probably through secondary contact

Material and Methods

Sample collection

Individuals were collected from various sources (local fishermen, sampling cruises) from 28 samples representing 28 localities from Tunisia, France, Bulgaria and Morocco. Details of sample sizes and locations are given in Table 1. and Figure1

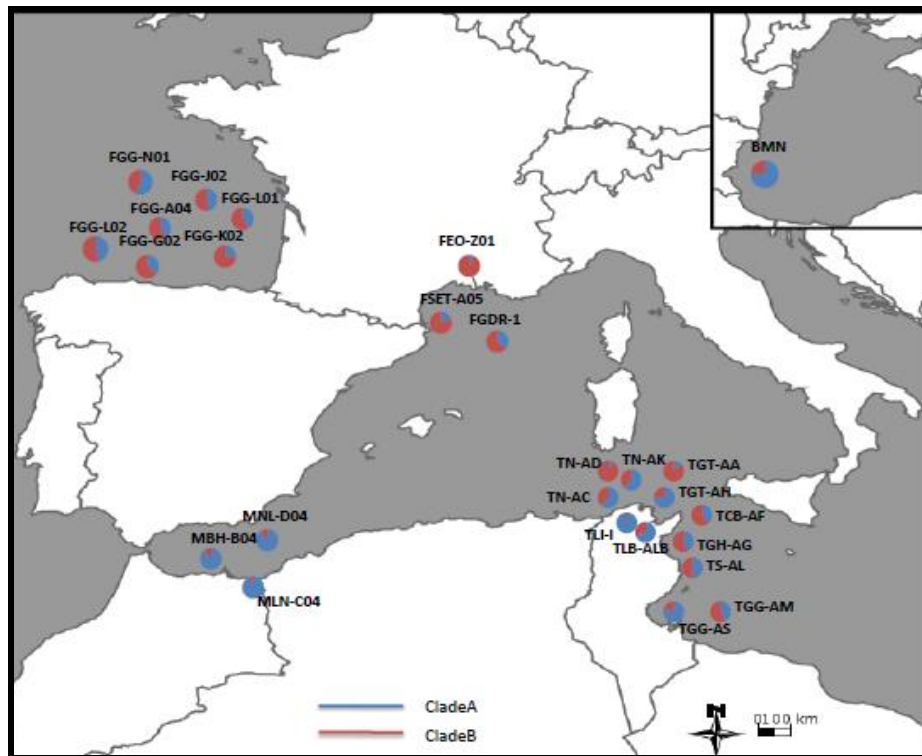


Figure 1: Sampling localities and mitochondrial frequencies (clade A in blue, clade B in red)

Country	Locality	Approximate coordinates	Code	n	Year	Collector	Habitat
France	Mauguio lagoon	43°34'N 4°2'E	FEO-Z01	27	2004	QUIGNARD Jean Pierre	lagoon
France	Bay of Biscay	45°10'N 2°5'W	FGG-L01	48	2001	PELGAS	sea
France	Bay of Biscay	45°30'N 2°10'W	FGG-J02	32	2002	PELGAS	sea
France	Bay of Biscay	43°25'N 1°42'W	FGG-A04	50	2004	MILLY David	sea
France	Bay of Biscay	44°10'N 2°20'W	FGG-K02	30	2002	PELGAS cruise	sea
France	Bay of Biscay	44°5'N 2°5'W	FGG-G02	31	2002	PELGAS cruise	sea
France	Bay of Biscay	44°10'N 2°10'W	FGG-C02	48	2002	PELGAS cruise	sea
France	Bay of Biscay	45°5'N 2°10'W	FGG-N01	48	2001	PELGAS cruise	sea
France	Bay of Biscay	45°10'N 2°5'W	FGG-L02	30	2002	PELGAS cruise	sea
France	Bay of Biscay	45°15'N 3°10'W	FGG-H02	31	2002	PELGAS cruise	sea
France	Sète	43°22'N 3°40' E	FSET-05	30	2005	CANTOU Michel	sea
France	Grau-du-Roi	43°32'N 4°08'E	FGDR-1	56	2006	PARY Béatrice	sea
Moroccoc	Al Hoceima bay	35°14'N 3°50'W	MBH-B04	32	2004	KADA Omar	sea
Morocco	Nador lagoon	35 09'N 2°50'W	MLN-C04	30	2004	KADA Omar	lagoon
Morocco	Nador offshore	35°22'N 2°36'W	MNL-D04	20	2004	KADA Omar	sea
Romania	Bulgarian border	43°40'N 28°48'E	BMN	30	2010	BANARU Daniela	sea
Tunisia	Northern coast	37°3'N 9°3'E	TN-AC	37	2009	INSTM	sea
Tunisia	Northern coast	37°4'N 9°E	TN-AD	36	2009	INSTM	sea
Tunisia	Northern coast	37°3'N 9°2'E	TN-AK	30	2007	INSTM	sea
Tunisia	Gulf of Tunis	37°2'N 10°4'E	TGT-AH	30	2009	INSTM	sea
Tunisia	Gulf of Tunis	37°4'N 10°6'E	TGT-AA	29	2007	INSTM	sea
Tunisia	Lake Ichkeul	37°10'N 9°40'E	TLI-I	32	2012	OUESLATI Soumaya	lagoon
Tunisia	Bizerte lagoon	37°11'N 9°51'E	TLB-ALB	30	2011	INSTM	lagoon
Tunisia	Cap Bon	37°2'N 11°4'E	TCB-AF	30	2007	INSTM	sea
Tunisia	Gulf of Hammamet	36°N 11°4'E	TGH-AG	30	2009	INSTM	sea
Tunisia	Sousse	36°4'N 11°3'E	TS-AL	30	2007	INSTM	sea
Tunisia	Gulf of Gabes	34°N 10°27'E	TGG-AS	30	2009	INSTM	sea
Tunisia	Gulf of Gabes	34°17'N 12°28'E	TGG-AM	28	2009	INSTM	sea

Table1 : Samples description

Laboratory analyses

A piece of muscle or dorsal fin was stored in 90% EtOH. Genomic DNA was extracted from 25 mg of tissue using the QIAGEN DNAeasy Blood & Tissue kit following the manufacturer's recommendations and subsequently quantified in a NanoDrop 8000 Spectrophotometer (Thermo Scientific). The primers for six microsatellite loci were isolated following the 454 GS-FLX Titanium pyrosequencing approach developed by Malausa et al. (2011). The sequence motifs and primers are provided in Table supp 1. PCR reactions with an initial denaturation at 95°C for 1 min, followed by 35 cycles (95°C 15 s; 52°C 20s, 72°C 20s) with a final extension at 72°C for 5 min were performed in a volume of 18 µl / 30 ng DNA, 1.5 mM MgCl₂, 0.3–0.5 mM of each primer, 10 x Taq Buffer 0.2–0.25 mM dNTP, and 1.5 U Taq DNA polymerase. PCR products were run on ABI Prism 3130XL 16-capillary at the Génomique Environnementale platform of Labex CeMEB (Montpellier) and reads processed with the GeneMapper® v4.5 software (Applied Biosystems). Since the repeat motifs of the 6 loci were either tetra- or penta-nucleotide, allele calling was organised in such a way that all minor alleles differing from the main peak by only one or two bp were binned into the main peak. An example of this is given in Figure supp 2.

For cytochrome b, a fragment of 680 bp was amplified with an initial denaturation at 95°C for 10 min, followed by 35 cycles (95°C 60 s; 52°C 1 min, 72°C 1 min) with a final extension at 72°C for 10 min with primers 5'-AACGACGCAGTAGTAGACC' and 5'-GAGGAAGTATCACTCAGGC-defined in positions 42 and 825 of the Cyt-b locus of *E. encrasicolus*, respectively. The amplification products were sequenced by the Sanger method at the CeMEB Platform and 680 bp manually aligned for individuals between positions 108 and 787 of the gene.

Data treatment

For microsatellite dataset, the presence and frequency of null alleles was tested using Micro-Checker (Van Oosterhout et al. 2004). The Genetix 4.05 software (Belkhir et al. 2004) was used to test for departures from Hardy–Weinberg expectation (HWE), to estimate population differentiation (F_{st}) and to perform Correspondence Analysis (CA) either on individual genotypes or on samples' centroids. A Neighbour-Joining tree of Reynold's distance was produced with Mega 5.0 (Tamura et al. 2011). Additionally, an individual assignment to reproductively independent putatively panmictic units was performed using the Bayesian clustering algorithm implemented in STRUCTURE 2.3.2 (Falush et al. 2007), with 100,000 "burn-in" iterations and a check of the stability of Log P (D) after 50,000 iterations. Cytochrome b sequences were aligned to the reference sequences available in Genbank and unambiguously classified either as clade A or clade B according to their position in the haplotypic tree (Fig. supp 3).

Results

Population differentiation

Nuclear loci

F_{st} values over the entire sample collection varied greatly among loci (Table 2). Loci EeA6 and EeA10 showed large and significant values, while those for EeA2, EeA8 and EeA9 were low but significant and non-significant for EeA7. When we look in detail at allelic

frequencies (Fig.2 and Table supp 4a) we see that allele EeA6-166 prevails in the lagoonal samples TLI-I (Lake Ischkeul), MLN-C04 (Nador lagoon), TLB-ALB (Bizerte lagoon), FEO-Z01 (or lagoon), its frequency exceeding 66%. In the same manner, allele EeA10-132 is associated with lagoons at a frequency higher than 50% (reciprocally, EeA10-128 predominates in marine samples). It therefore appears that among our 6 loci, two are strongly associated with a sea / lagoon differentiation while four are not. Interestingly, samples TGT-AH, TGG-AS and, to a lesser extent, TGH-AG appear to carry the lagoonal allele for EeA6 and the marine one for EeA10 (Fig.2). This association is also clearly visible in the unrooted distance tree, where the four lagoon populations cluster in the same part of the tree (Fig.3). This is also reflected in the CA, which shows that axis 1 differentiates lagoon from offshore samples (Fig.4) and also to a large extent in the individual assignments performed by STRUCTURE (Fig.5). The method of Evanno et al. (2005) revealed that the number of partitions k maximising the posterior probability was $k = 2$ (not shown). The green cluster predominates in the first four populations corresponding to the lagoon samples. These assignments are clearly imperfect (no sample appears as "pure"). Nevertheless, they are consistent with the CA analysis. By regressing the coordinates on the first canonical axis and the frequency of assignation to the clusters in the different samples when taking into account individual genotypes, we found a strong correlation ($R^2 = 0.942$, $p < 10^{-5}$) between the two analyses. This correlation amounted to $R^2 = 0.870$, ($p < 10^{-5}$) when populations' centroids only were considered (Fig. supp 5)

locus	EeA2	EeA6	EeA7	EeA8	EeA9	EeA10	multilocus
F_{st}	0,016	0,065	0,0006	0,0126	0,014	0,191	0,036
P	<0,0001	<0,0001	<0,35	<0,0001	<0,0001	<0,0001	<0,0001

Table 2 : F_{st} values in 6 locus

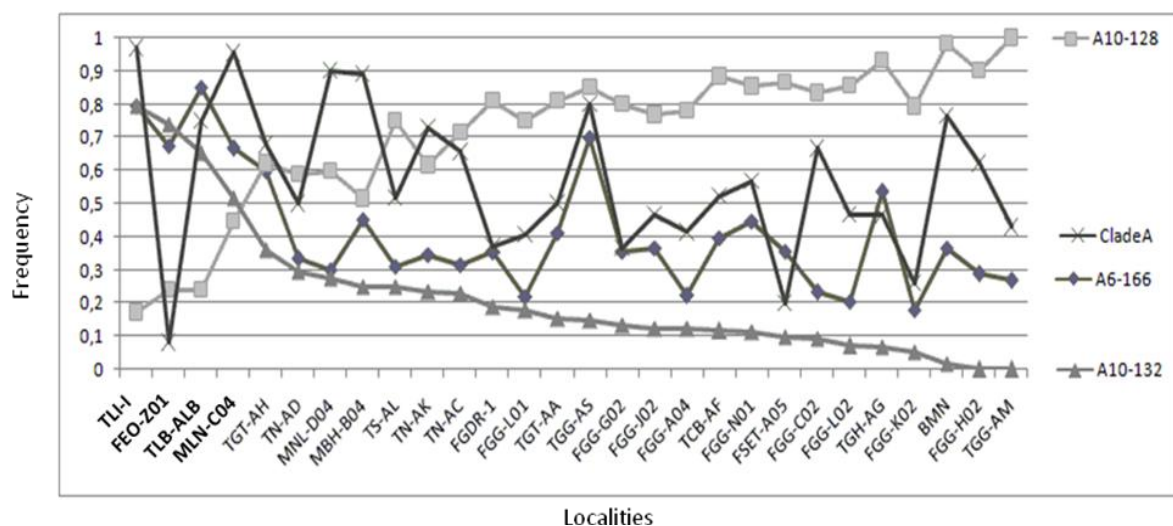


Figure 2: Distribution of allelic frequencies for locus EeA6 (alleles 166 and 128) and EeA10 (allele 132) and for mitochondrial cladeA. Lagoon samples in bold.

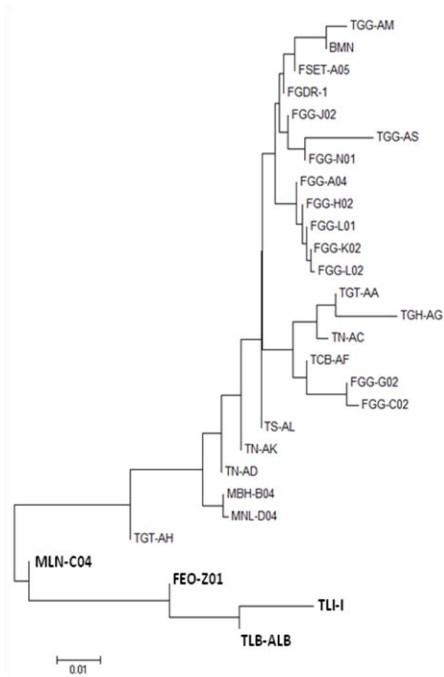


Figure 3: Reynold's distances between populations. Lagoon samples in bold

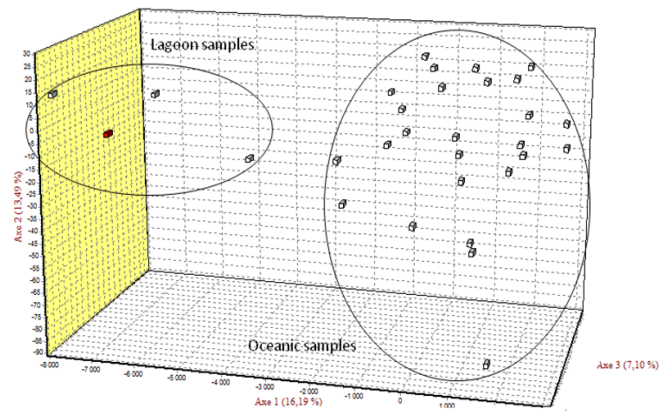


Figure 4: Correspondence analysis for 28 samples' centroid at six microsatellite loci. Lagoon samples indicated

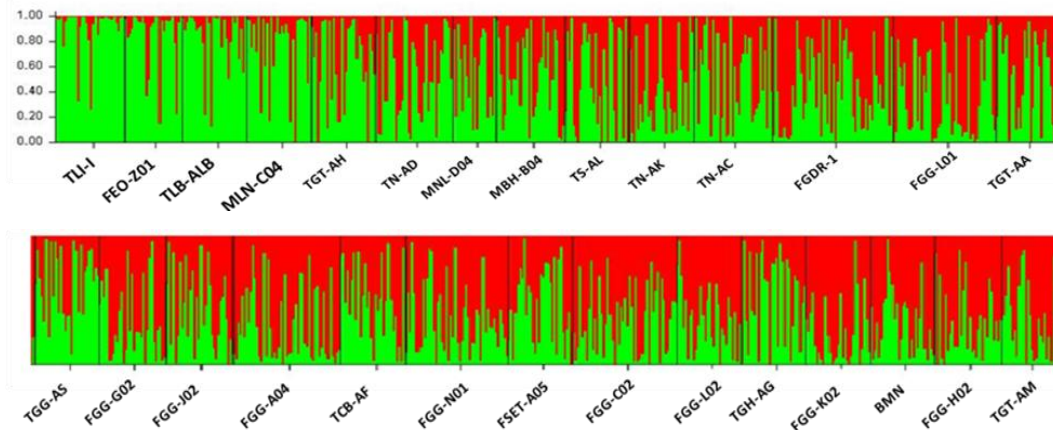


Figure 5: Individual assignment (STRUCTURE method) for $k = 2$. Each vertical bar represents an individual, samples are delimited by black bars. Lagoon samples in bold

Mitochondrial DNA

The two mitochondrial clades reported in the literature were present in all samples, albeit at highly variable frequencies. Southern lagoons were predominantly harbouring clade A (75%-97%) while the northern lagoon FEO-Z01 harboured only 8% of this clade. In the offshore samples the frequency of clade A oscillated between 20% and 90% (Fig. 2 and Table supp.4b).

Nucleo-cytoplasmic disequilibrium

There is a clear trend for parallel variation in the frequency of clade A and that of allele EeA6-166 and to a lesser extent with that of EeA10-128, with the noticeable exception of sample FEO-Z01 (Fig. 2). This translates into significantly positive correlations $R^2 = 0.152$, $p < 0.0006$ and $R^2 = 0.098$, $p < 0.0001$ for the two alleles respectively on the entire collection and $R^2 = 0.342$, $p < 0.0001$ and $R^2 = 0.36$, $p < 0.0001$ after removing FEO-Z01 (Fig. supp.6). We also addressed this question by looking at the link between the percentage of assignment to the lagoon cluster, as estimated by STRUCTURE, and the frequency of clade A. Here the correlation was also positive and significant ($R^2 = 0.138$, $p < 0.05$, $R^2 = 0.415$, $p < 0.0006$, FEO-Z01 removed) this last figure being higher than for each microsatellite locus considered alone (Fig. supp.7). As assignment was performed independently of the frequency of clades, this link between mitochondrial and microsatellite cannot be an artefact and so is the signature of a global nucleo-cytoplasmic disequilibrium. Such correlation between clades and clusters when lagoon and offshore samples are mixed is expected a priori, since the frequencies of nuclear and mitochondrial markers are different in both types of populations. It is interesting to note, however, that the correlation persists when studied on certain samples taken separately. For the offshore samples from Tunisia, we find $R^2 = 0.562$, $p < 10^{-5}$ (Fig supp.8). This means that even within these samples captured at sea, we did not sample a single and genetically homogeneous entity.

Discussion

Two widespread entities

In the recent literature, few papers have addressed the question of the status of lagoonal and estuarine anchovies as a whole, and only a single case was analysed from a genetic stand point. We add two more lagoons to the list of studied cases, both located on the southern bank of the Mediterranean, Nador in Morocco and Ischkeul/Bizerte in Tunisia, where we confirm the existence of a local form genetically close to the one originally described by Borsa et al. (2004) from Manguio lagoon in the Gulf of Lions, and differentiated from the more offshore and marine entity.

This duality translates into the segregation of alleles at two microsatellite loci as well as a variable level of nucleo-cytoplasmic disequilibrium, albeit expressed in opposite directions on the two opposing shores of the Mediterranean, with clade B associated primarily with the inshore entity in France while it was clade A in Tunisia. Hence, it is probable that every Mediterranean lagoonal system, even very small ones like the Ganzirri lagoon in Sicily where *E. rissoi* was described (Dulzetto 1947), may harbour lagoonal anchovies. As to offshore (or oceanic) samples, they are relatively homogeneous over the study area, with a strong resemblance between those of the Gulf of Lion and Bay of Biscay, as already noted by

Bouchenak-Khelladi et al. (2008), and as can also be seen in Zarraonaindia et al. (2009). Samples from Black Sea appear, with our nuclear markers, to be very similar to the previous ones, while the two offshore Alboran Sea samples appear somewhat differentiated (Fig. 3 and Table supp.2) in keeping with Zarraonaindia et al. (2012) who found Alboran and, to a lesser extent Aegean, seas samples differentiated from other marine samples with their more complete data set. Nevertheless, whatever the level of structuring among offshore samples, our data highlight the repeated occurrence of a strongly differentiated coastal entity around the Mediterranean, which can be identified with as few as 2 / 6 randomly chosen microsatellite loci. One remaining question is that of the interaction between the two entities.

Introgression, admixture, mixed stocks

In our open sea anchovies from Tunisia, several lines of evidence indicate that samples are not homogeneous, despite not clustering with the coastal entity in the tree of Figure 3. Firstly, there is a detectable level of nuclear and cytoplasmic F_{st} ($F_{st}=0.021$, $p<10^{-5}$ and $F_{st}=0.048$, $p<0.093$ respectively) and a global nucleo-cytoplasmic disequilibrium (Table 2 and Figure 2). This is unexpected for pure populations of a widespread pelagic species, which should be much more homogeneous at this scale (see for instance what prevails in sardines (Tinti et al. 2002) where at the same geographical scale $F_{st} = 0.0047$). By comparison, the nine offshore samples from the Bay of Biscay were much more homogeneous than the Tunisian ones, ($F_{st}=0.0023$, $p<0.0125$ and $F_{st}=0.046$, $p<0.069$ respectively). This correlates well with results of the morphological study of Messaoud et al. (2011) who, through a canonical analysis of the shape of otoliths, identified a clear differentiation between two offshore samples from northern Tunisia, albeit smaller than what prevailed with Bizerte lagoon anchovies. The simplest explanation is that the Tunisian offshore samples have genes from both entities in different proportions. Theoretically this could result either from simple mixing of animals from the two entities (mixed stock hypothesis) or the existence of recent hybridization, or any combination of these phenomena in various proportions between samples. Past hybridization may be followed by introgression of some genes from one entity into the other, but this in itself does not create inter-sample heterogeneity when gene flow is high enough.

The quality and quantity of our data are not, however, sufficient to distinguish among the various possibilities. For instance, F_{is} estimates on the entire collection showed an overall departure from panmictic equilibrium (heterozygote deficit) that was quite large and significant for each locus, with the exception of EeA7 (see Table supp.9). Estimated mean values after correction for null alleles via MicroChecker were not very different (Table supp.9). However, when our entire data set was grouped into two samples only (lagoon and sea), a significant multilocus F_{is} remained, indicating that the overall heterozygote deficiency cannot be explained by a Wahlund effect alone. The same lack of power applies to linkage disequilibrium (LD) (Table supp.10). There was no significant LD between the two diagnostic loci EeA6 and EeA10 for both ocean and lagoon groups as might have been expected. This last statistic, however, notoriously lacks power, especially with multiallelic and partially diagnostic loci. We also calculated the genotypic nucleo-cytoplasmic disequilibria inside each sample, after pooling alleles for the two diagnostic loci (allele EeA6-166 against all the others for locus EeA6, and a grouping with allele Ee132 or Ee128 according to size for locus EeA10). The resulting 2x3 contingency tables however did not reveal significant departure from random according to an extended Fisher's exact test (not shown), even when several samples from the same geographic locations were pooled.

Barrier to gene flow and secondary contact

At any rate, these three differentiated loci (2 nuclear and a mitochondrial one) over a total of 7 are sufficient proof that there is a barrier to the mixing of two entities, although initially there were two differentiated populations as indicated by the presence of two very different mitochondrial clades (see Grant 2005) for an appraisal of their phylogeographic history). We are thus facing a case of interaction between partially incompatible genomes, whereby some loci are freely exchanged among entities (almost complete homogeneity of loci EeA2, EeA7, EeA8, EeA9 through all our samples) and others retain highly significant differences (EeA6, EeA10 and Cyt-b). According to Bierne et al. (2011), genetic barriers (resulting from pre-or post-zygotic genetic incompatibilities) are widespread and can explain the majority of abnormally elevated F_{st} values, whether or not these loci are directly correlated with environmental variables. These incompatibilities are known to be able to create "tension zones", which tend to be associated with exogenous barriers resulting from ecological selection. This is probably what we are confronted with in the case of the two anchovies, the inshore and offshore habitats clearly being able to capture differential adaptations, as reflected by a suite of morphological and biological differences reported in Quignard et al. (1973), Bembo et al. (1996), Kada et al. (2009), Messaoud et al. (2011).

However, despite their ecological differentiation, the two forms meet and are in continuous interaction. The consequences of this for fishery management may be that mixed stocks are more common than previously thought, although the exact nature of the levels of admixture in each sample remain to be addressed with more numerous markers. Variation of the contribution of each genetic background on a relatively small scale, as exemplified on our Tunisian offshore samples, indicates a locally complex dynamic for this interaction.

Legends to table and figure

Tables

Table 1: *Engraulis encrasicolus* sample description

Table 2: Global F_{st} for each loci (italics: values significant after *Bonferroni* correction)

Tables 3: Clade A percent

Figures legends:

Figure 1: Sampling localities and mitochondrial frequencies (clade A in blue, clade B in red)

Figure 2: Distribution of allelic frequencies for locus EeA6 (alleles 166 and 128) and EeA10 (allele 132) and for mitochondrial cladeA. Lagoon samples in bold.

Figure 3: Reynold's distances between populations. Lagoon samples in bold

Figure 4: Correspondence analysis for 28 samples' centroid at six microsatellite loci. Lagoon samples indicated

Figure 5: Individual assignation (STRUCTURE method) for $k = 2$. Each vertical bar represents an individual, samples are delimited by black bars. Lagoon samples in bold.

Acknowledgements

The authors wish to thank Emeline Lambert for having initiated the mtDNA sequencing, Jacques Dietrich for providing the Gulf of Lion's samples and for unpublished data, Daniela Banaru for Black Sea sample, Pierre-Alexandre Gagnaire and Nicolas Bierne for helpful discussion, Othman Jarboui for his support of the project and David McKenzie for linguistic revision. The analytical part of this work was financed by basic institutional funding to François Bonhomme. During her stay in Sète, Soumaya Oueslati was supported by French Embassy (Tunisia) and Program Hubert-Curien 27961XE fellowships.

References

- Belkhir K, Borsa P, Chikhi L, Raufaste N, Bonhomme F (2004) GENETIX 4.05, logiciel sous Windows TM pour la génétique des populations. Laboratoire Génome, Populations, Interactions, CNRS UMR 5000, Université de Montpellier II, Montpellier (France).
- Bembo DG, Carvalho GR, Cingolani N, Arneri E, Gianneti G, Pitcher T.J (1996) Allozymic and morphometric evidence for two stocks of the European anchovy *Engraulis encrasicolus* in Adriatic waters. *Mar Biol* 126:529–538. doi: 10.1007/BF00354635
- Bierne N, Welch J, Loire E, Bonhomme F, David P (2011) The coupling hypothesis: why genome scans may fail to map local adaptation genes. *Mol Ecol* 20:2044–2072. doi: 10.1111/j.1365-294X.2011.05080.x
- Borrell, Y.J., Piñera, J.A., Prado, J.A.S., and Blanco, G. (2012). Mitochondrial DNA and microsatellite genetic differentiation in the European anchovy *Engraulis encrasicolus* L. *ICES J. Mar Sci* 69, 1357–1371.
- Borsa P (2002) Allozyme, mitochondrial-DNA, and morphometric variability indicate cryptic species of anchovy (*Engraulis encrasicolus*). *Biological Journal of the Linnean Society* 75:261–269. doi: 10.1046/j.1095-8312.2002.00018.x
- Borsa P, Collet A, Durand J-D (2004) Nuclear-DNA markers confirm the presence of two anchovy species in the Mediterranean. *C R Biol* 327:1113–1123.
- Bouchenak-Khelladi Y, Durand J-D, Magoulas A, Borsa P (2008) Geographic structure of European anchovy: A nuclear-DNA study. *Journal of Sea Research* 59:269–278. doi: 10.1016/j.seares.2008.03.001
- Dulzetto F (1947) L'Engraulis dei laghi di Ganzirri e del Faro. *Mem. Soc. It. delle Scienze (dei XL)*, serie III 26
- Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software structure: a simulation study. *Mol Ecol* 14:2611–2620. doi: 10.1111/j.1365-294X.2005.02553.x
- Falush D, Stephens M, Pritchard JK (2007) Inference of population structure using multilocus genotype data: dominant markers and null alleles. *Mol Ecol Notes* 7:574–578. doi: 10.1111/j.1471-8286.2007.01758.x

Grant WS (2005) A Second Look at Mitochondrial DNA Variability in European Anchovy (*Engraulis encrasicolus*): Assessing Models of Population Structure and the Black Sea Isolation Hypothesis. *Genetica* 125:293–309. doi: 10.1007/s10709-005-0717-z

Kada O, Abdellaoui S, Ramdani M, Nachit D (2009) Contribution à l'identification et à la caractérisation biologique et dynamique de l'anchois de la lagune de Nador (Maroc). *Bulletin de l'Institut Scientifique (Rabat), section Sciences de la Vie* 31: 91-98

Magoulas A, Castilho R, Caetano S, Marcato S, Paternello T (2006) Mitochondrial DNA reveals a mosaic pattern of phylogeographical structure in Atlantic and Mediterranean populations of anchovy (*Engraulis encrasicolus*). *Mol Phylo and Evol* 39:734–746. doi: 10.1016/j.ympev.2006.01.016

Magoulas A, Tsimenides N, Zouros E (1996) Mitochondrial DNA phylogeny and the reconstruction of the population history of a species: the case of the European anchovy (*Engraulis encrasicolus*). *Mol Biol Evol* 13:178–190.

Malausa T, Gilles A, Meglécz E, Blanquart H, Duthoy S, Costedoat C, Dubut V, Pech N, Castagnone Sereno P, Délye C (2011) High-throughput microsatellite isolation through 454 GS-FLX Titanium pyrosequencing of enriched DNA libraries. *Mol Ecol Resour* 11:638–644. doi: 10.1111/j.1755-0998.2011.02992.x

Messaoud H, Bouriga N, Daly Yahia MN, Boumaiza M, Faure E, Quignard J-P, Trabelsi M (2011) Discrimination de trois populations d'anchois du genre *Engraulis* (Clupeiforme, Engraulidae) des côtes Tunisiennes par analyse de forme des otolithes. *Bull Inst Natn Scien Tech Mer (Salammbô)* 38: 21-27

Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004) micro-checker: software for identifying and correcting genotyping errors in microsatellite data. *Mol Ecol Notes* 4:535–538. doi: 10.1111/j.1471-8286.2004.00684.x

Quignard JP, Hamdouni T, Zaouali J (1973) Données préliminaires sur les caractères biométriques des anchois *Engraulis encrasicolus* (Linne, 1758) des cotes de Tunisie et du Lac Ichkeul. *Revue des Travaux de l'Institut des Pêches Maritimes* 37:191–196.

Sanz N, García-Marín J-L, Viñas J, Roldan M, Pla C. (2008) Spawning groups of European anchovy: population structure and management implications. *ICES J Mar Sci* 65:1635–1644. doi: 10.1093/icesjms/fsn128

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Mol Biol Evol* 28:2731–2739. doi: 10.1093/molbev/msr121

Tinti F, Nunno CD, Guarniero I, Talenti M, Tommasini S, Fabbri E, Piccinetti C (2002) Mitochondrial DNA Sequence Variation Suggests the Lack of Genetic Heterogeneity in the Adriatic and Ionian Stocks of *Sardina pilchardus*. *Mar Biotechnol* 4:163–172. doi: 10.1007/s10126-002-0003-3

Uriarte A, Prouzet P, Villamor B (1996) Bay of Biscay and Ibero Atlantic anchovy populations and their fisheries. *Scientia Marina* 60:237–255.

Whitehead PJ, Nelson GJ, Wongratana T (1988) FAO Species Catalogue. Rome: UNDP,

Zarraonaindia I, Pardo MA, Iriondo M, Manzano C, Estonba A (2009) Microsatellite variability in European anchovy (*Engraulis encrasicolus*) calls for further investigation of its genetic structure and biogeography. ICES J Mar Sci 66:2176–2182. doi: 10.1093/icesjms/fsp187

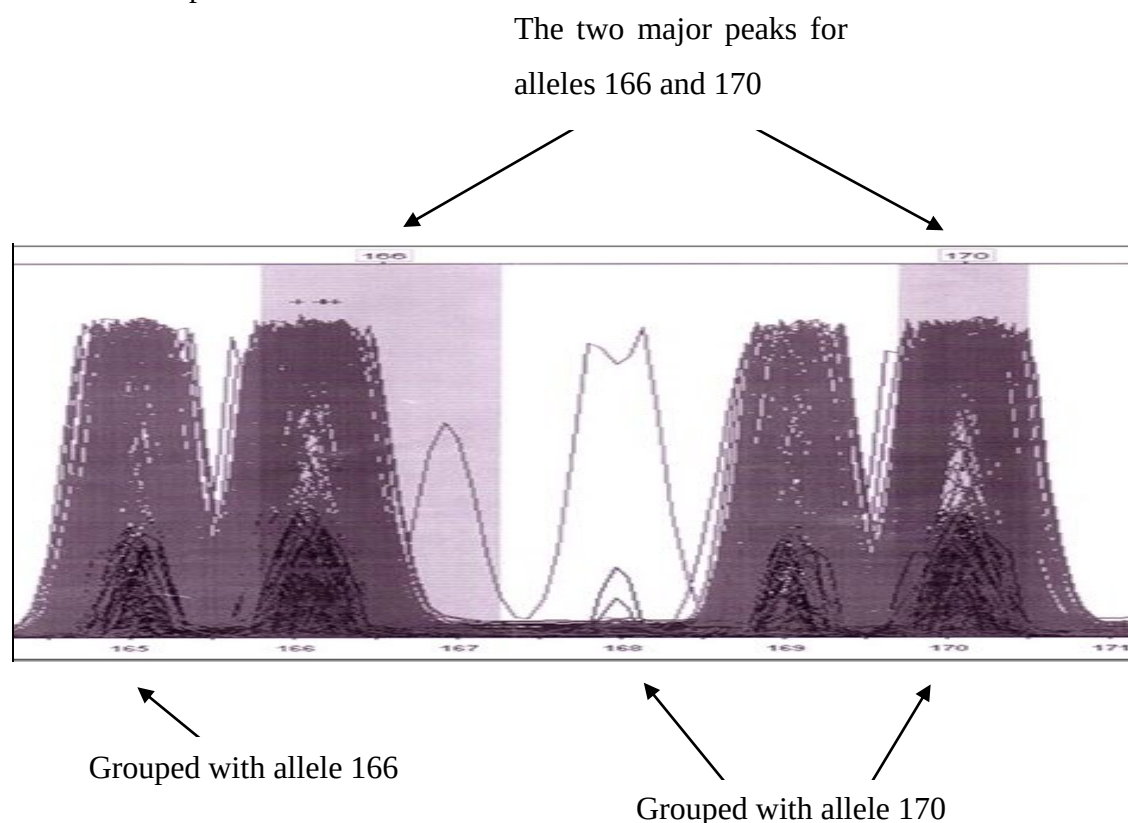
Zarraonaindia I, Iriondo M, Albaina A, Padro Miguel A, Manzano C, Grant WS, Irigoien X, Estonba A (2012) Multiple SNP Markers Reveal Fine-Scale Population and Deep Phylogeographic Structure in European Anchovy (*Engraulis encrasicolus* L.). PLoS ONE 7:e42201. doi: 10.1371/journal.pone.0042201

ANNEXE B

Table Supp. 1: Microsatellite primers & characteristics

CODE	PRIMER LEFT SEQUENCE	PRIMER RIGHT SEQUENCE	MOTIF
PrimerEeA2	CTTCTGCACAGAGCACAGGT	ACATCAACAAAAGGGCCAAG	tacat
Primer EeA6	CCATAACTTTTTGAGTTATCCTGC	TTCAGTTTTTCCACTGAGCGA	acag
PrimerEeA7	GTGGTGGAGGATGGGTGTA	TCATCCTGCCCACTCTTCTC	atgt
PrimerEeA8	AACTCTGCCACAGCCTGAAT	TGTCTCTGTCTGGCTACAGGTG	caga
PrimerEeA9	TCTGTCTGCCTTTCTGTAAC TG	GAGAAAAGACGAGAGGCAGC	ctgt
PrimerEeA10	GCTCTGGTCTGTGAGGAAGG	CCATCATCTATCGCTCCTCC	gatg

Figure Supp. 2: Allele Grouping: Example for tetranucleotidic microsatellite locus EeA6 for alleles 166 and 170. Alleles smaller than the main peak by only one or two bp were binned into the main peak.



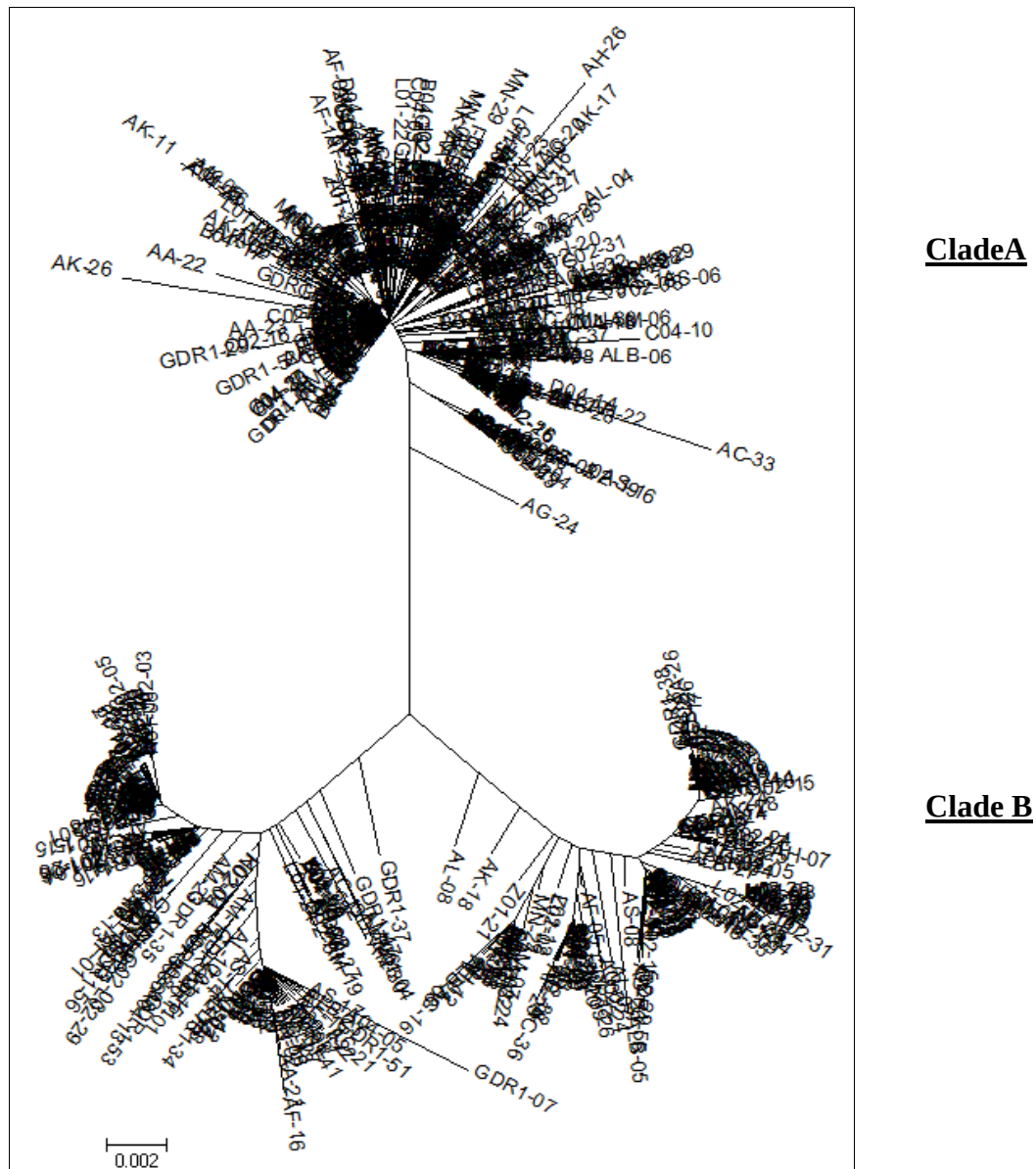


Table Supp. 4a: Alleles frequencies at 6 microsatellite loci

Loci	POP																											
	TLI-I	FEO-Z01	TLB-ALB	MLN-C04	TGT-AH	TN-AD	MNL-D04	MBH-B04	TS-AL	TN-AK	TN-AC	FGDR-1	FGG-L01	TGT-AA	TGG-AS	FGG-G02	FGG-J02	FGG-A04	TCB-AF	FGG-N01	FSET-A05	FGG-C02	FGG-L02	TGH-AG	FGG-K02	BMN	FGG-H02	TGG-AM
EeA2																												
(N)	32	27	30	29	29	30	20	30	23	30	23	54	46	19	30	3	31	40	19	48	30	7	30	24	28	30	22	28
72	0.031	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.033	0.000	0.000	0.022	0.000	0.000	0.000	0.000	0.000	0.010	0.000	0.000	0.000	0.000	0.018	0.000	0.023	0.000	
77	0.031	0.000	0.050	0.052	0.000	0.000	0.050	0.050	0.022	0.067	0.000	0.065	0.033	0.000	0.050	0.000	0.065	0.050	0.000	0.042	0.050	0.000	0.083	0.000	0.071	0.117	0.023	0.018
83	0.453	0.389	0.267	0.379	0.207	0.400	0.375	0.383	0.283	0.283	0.196	0.324	0.315	0.132	0.333	0.000	0.371	0.325	0.158	0.292	0.467	0.143	0.367	0.167	0.321	0.333	0.296	0.500
88	0.031	0.185	0.133	0.017	0.052	0.150	0.075	0.167	0.044	0.100	0.022	0.176	0.174	0.079	0.167	0.000	0.194	0.138	0.000	0.156	0.033	0.000	0.200	0.021	0.143	0.117	0.136	0.054
93	0.297	0.222	0.267	0.207	0.207	0.133	0.225	0.133	0.196	0.183	0.174	0.194	0.239	0.237	0.200	0.167	0.097	0.225	0.158	0.219	0.083	0.000	0.200	0.104	0.214	0.200	0.227	0.232
98	0.047	0.074	0.150	0.086	0.293	0.183	0.225	0.167	0.109	0.100	0.261	0.083	0.087	0.368	0.117	0.333	0.145	0.088	0.290	0.125	0.183	0.143	0.033	0.188	0.071	0.117	0.136	0.161
103	0.078	0.056	0.083	0.190	0.103	0.067	0.000	0.067	0.087	0.117	0.109	0.102	0.065	0.079	0.050	0.500	0.081	0.050	0.211	0.073	0.083	0.429	0.083	0.188	0.036	0.050	0.046	0.018
108	0.016	0.074	0.050	0.052	0.017	0.050	0.025	0.017	0.044	0.083	0.109	0.028	0.065	0.053	0.067	0.000	0.048	0.088	0.026	0.063	0.033	0.000	0.000	0.021	0.107	0.033	0.023	0.018
113	0.016	0.000	0.000	0.017	0.017	0.000	0.025	0.017	0.065	0.000	0.022	0.009	0.000	0.000	0.017	0.000	0.000	0.013	0.079	0.010	0.033	0.000	0.017	0.063	0.000	0.033	0.000	0.000
118	0.000	0.000	0.000	0.000	0.017	0.017	0.000	0.000	0.000	0.017	0.022	0.009	0.000	0.000	0.000	0.000	0.000	0.025	0.053	0.010	0.033	0.000	0.017	0.021	0.000	0.000	0.091	0.000
200	0.000	0.000	0.000	0.000	0.086	0.000	0.000	0.000	0.152	0.017	0.087	0.009	0.000	0.053	0.000	0.000	0.000	0.000	0.026	0.000	0.000	0.286	0.000	0.229	0.018	0.000	0.000	0.000
EeA6																												
(N)	32	26	30	30	30	30	20	32	29	29	30	48	25	28	30	31	15	36	29	19	28	28	17	29	28	30	26	28
154	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.031	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.014	0.000	0.000	0.000	0.018	0.000	0.000	0.000	0.017	0.019	0.000
162	0.000	0.000	0.000	0.000	0.000	0.050	0.025	0.000	0.000	0.052	0.000	0.000	0.020	0.000	0.000	0.016	0.000	0.000	0.035	0.026	0.018	0.018	0.000	0.000	0.018	0.017	0.000	0.000
166	0.797	0.673	0.850	0.667	0.600	0.333	0.300	0.453	0.310	0.345	0.317	0.354	0.220	0.411	0.700	0.355	0.367	0.222	0.397	0.447	0.357	0.232	0.206	0.535	0.179	0.367	0.289	0.268
170	0.141	0.115	0.100	0.233	0.117	0.167	0.500	0.391	0.172	0.207	0.233	0.104	0.380	0.214	0.183	0.226	0.300	0.264	0.293	0.132	0.143	0.268	0.265	0.138	0.375	0.117	0.289	0.179
174	0.016	0.058	0.033	0.000	0.100	0.200	0.125	0.063	0.224	0.035	0.183	0.219	0.080	0.107	0.000	0.081	0.133	0.167	0.138	0.158	0.179	0.054	0.118	0.052	0.071	0.133	0.077	0.196
178	0.047	0.154	0.017	0.083	0.167	0.233	0.050	0.047	0.293	0.293	0.267	0.292	0.280	0.250	0.117	0.323	0.133	0.319	0.121	0.158	0.250	0.321	0.382	0.241	0.268	0.283	0.289	0.268
182	0.000	0.000	0.000	0.017	0.017	0.000	0.000	0.016	0.000	0.069	0.000	0.021	0.020	0.018	0.000	0.000	0.067	0.014	0.017	0.053	0.036	0.071	0.000	0.035	0.054	0.067	0.019	0.089
186	0.000	0.000	0.000	0.000	0.000	0.017	0.000	0.000	0.000	0.000	0.000	0.010	0.000	0.000	0.000	0.000	0.000	0.000	0.026	0.018	0.018	0.018	0.029	0.000	0.036	0.000	0.000	0.000
193	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.019	0.000
EeA7																												
(N)	32	26	30	29	30	30	20	32	30	30	30	56	48	27	29	31	31	49	28	48	30	48	30	29	30	30	28	28
70	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.037	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
74	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.017	0.000	0.000	0.009	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.017	0.000	0.000	0.000	0.000	0.000	0.000	0.000
79	0.000	0.058	0.000	0.000	0.000	0.017	0.000	0.016	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.032	0.000	0.010	0.000	0.010	0.017	0.000	0.000	0.000	0.000	0.000	0.000	0.000
83	0.234	0.192	0.217	0.138	0.200	0.133	0.175	0.172	0.233	0.200	0.150	0.232	0.198	0.148	0.276	0.194	0.226	0.214	0.143	0.188	0.183	0.177	0.133	0.172	0.133	0.283	0.054	0.304
87	0.047	0.000	0.033	0.052	0.083	0.067	0.000	0.000	0.050	0.050	0.033	0.009	0.031	0.037	0.069	0.032	0.065	0.041	0.018	0.073	0.017	0.042	0.017	0.052	0.000	0.017	0.054	0.054
91	0.141	0.308	0.117	0.241	0.217	0.150	0.050	0.203	0.217	0.200	0.150	0.205	0.188	0.148	0.121	0.129	0.210	0.214	0.268	0.104	0.167	0.167	0.150	0.155	0.167	0.117	0.161	0.161
92	0.000	0.058	0.000	0.000	0.000	0.000	0.000	0.000	0.017	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
96	0.125	0.154	0.250	0.172	0.133	0.133	0.175	0.172	0.067	0.183	0.183	0.161	0.104	0.352	0.121	0.242	0.145	0.112	0.161	0.156	0.150	0.156	0.183	0.172	0.200	0.200	0.196	0.125
100	0.188	0.039	0.200	0.172	0.150	0.117	0.175	0.109	0.100	0.083	0.133	0.116	0.115	0.093	0.121	0.129	0.097	0.102	0.107	0.156	0.183	0.146	0.167	0.086	0.183	0.117	0.143	0.089
104	0.156	0.135	0.067	0.103	0.083	0.167	0.200	0.188	0.117	0.167	0.150	0.188	0.167	0.074	0.155	0.129	0.194	0.102	0.161	0.167	0.200	0.198	0.183	0.138	0.150	0.167	0.179	0.196
108	0.063	0.019	0.050	0.017	0.017	0.050	0.100	0.047	0.033	0.033	0.100	0.018	0.063	0.037	0.035	0.048	0.016	0.071	0.018	0.052	0.017	0.010	0.067	0.017	0.050	0.033	0.071	0.000
112	0.031	0.019	0.017	0.052	0.017	0.000	0.050	0.016	0.033	0.000	0.017	0.027	0.031	0.000	0.017	0.000	0.016	0.000	0.054	0.000	0.017	0.021	0.033	0.052	0.000	0.000	0.054	0.000
116	0.000	0.000	0.000	0.017	0.033	0.100	0.000	0.000	0.000	0.050	0.017	0.000	0.010	0.019	0.017	0.016	0.000	0.020	0.036	0.010	0.000	0.000	0.000	0.017	0.017	0.000	0.000	0.018
120	0.000	0.000	0.017	0.000	0.050	0.033	0.050	0.000	0.033	0.000	0.017	0.018	0.052	0.000	0.000	0.032	0.000	0.092	0.018	0.031	0.000	0.021	0.050	0.052	0.050	0.067	0.054	0.036
124	0.016	0.000	0.017	0.000	0.000	0.017	0.000	0.031	0.050	0.033	0.017	0.018	0.021	0.019	0.000	0.016	0.000	0.000	0.018	0.042	0.000	0.010	0.000	0.017	0.017	0.000	0.036	0.018
128	0.000	0.019	0.017	0.000	0.000	0.000	0.025	0.031	0.017	0.000	0.033	0.000	0.010	0.019	0.052	0.000	0.016	0.000	0.000	0.010	0.017	0.042						

Table Supp. 4b: Clade A frequency

Samples	TLI-I	FEO- Z01	TLB- ALB	MLN- C04	TGT- AH	TN-AD	MNL- D04	MBH- B04	TS-AL	TN-AK	TN-AC	FGDR- 1	FGG-L01	TGT-AA
Clade A %	96.88	8	75	95.45	67.65	50	90	89.29	51.72	73.08	65.71	37.25	40.74	50
Samples	TGG- AS	FGG- G02	FGG- J02	FGG- A04	TCB- AF	FGG- N01	FSET- A05	FGG- C02	FGG- L02	TGH-AG	FGG-K02	BMN	FGG-H02	TGG-AM
Clade A %	80	36.67	46.88	41.38	52.17	56.67	20	66.67	46.67	46.43	25.93	76.67	62.07	42.86

Figure Supp. 5a: Correlation between percent assignment to the lagoon cluster and coordinates of each sample on axis 1 of the CA (samples average of each individual projection)

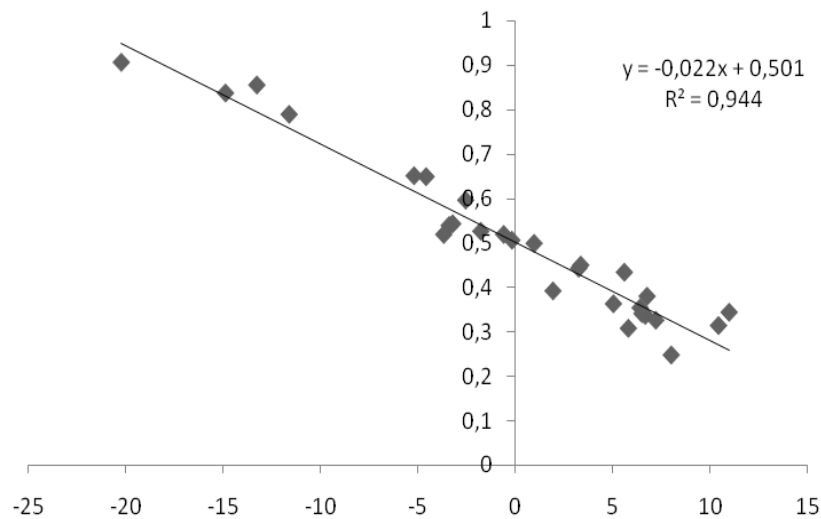


Figure Supp. 5b: Correlation between percent assignment to the lagoon cluster and coordinates of each sample on axis 1 of the CA (samples centroid)

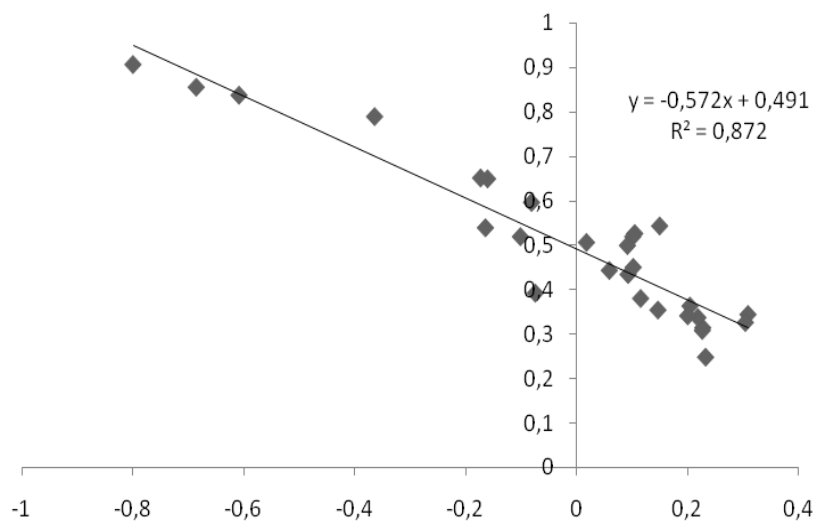


Figure Supp. 6a: Correlation between clade A (abcissa) and allele EeA6-166 (ordinate) frequencies

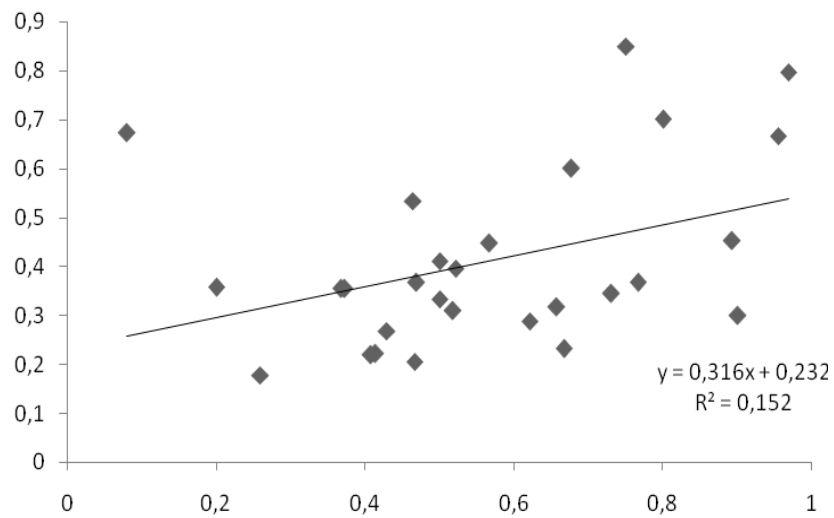


Figure Supp. 6b: Correlation between cladeA (abcissa) and allele EeA10-128 (ordinate) frequencies

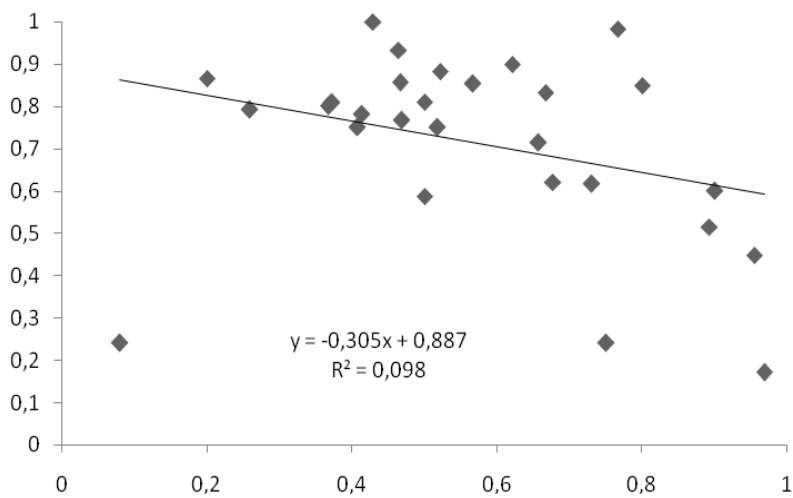


Figure Supp. 6c: Correlation between cladeA (abcissa) and allele EeA6-166 (ordinate) without FEO-Z01

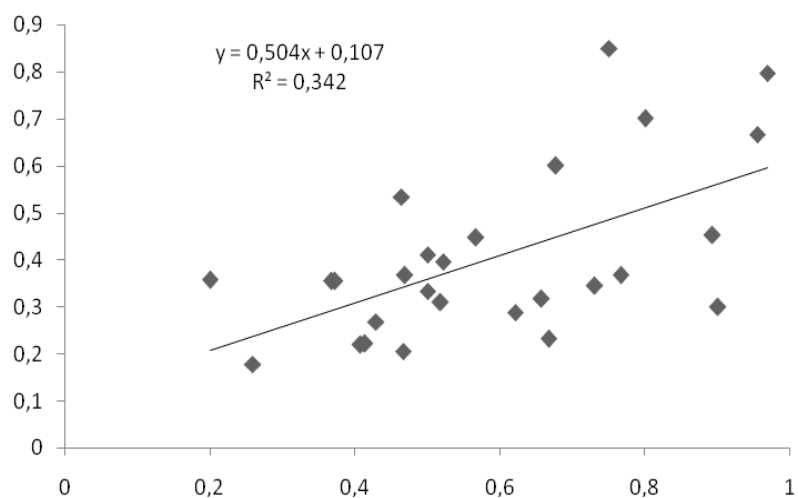


Figure Supp. 6d: Correlation between cladeA (abcissa) and allele EeA10-128 (ordinate) without FEO-Z01

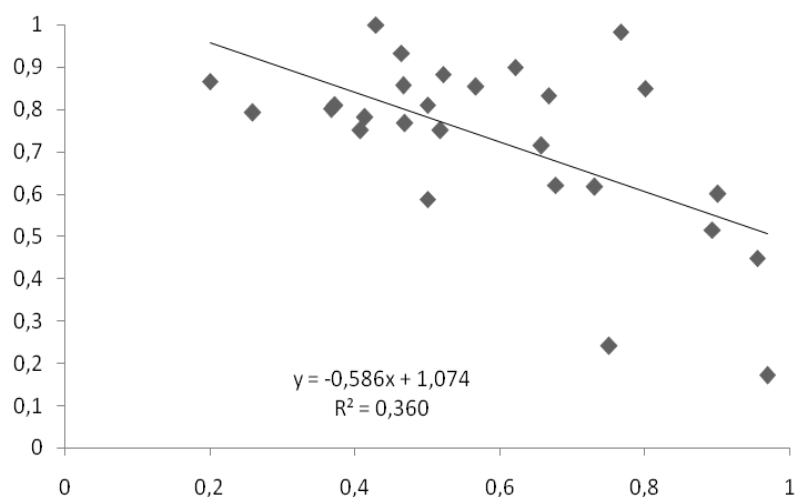


Figure Supp. 7a: Correlation between clade A frequency (abscissa) and percent assignment to the lagoon cluster (ordinate)

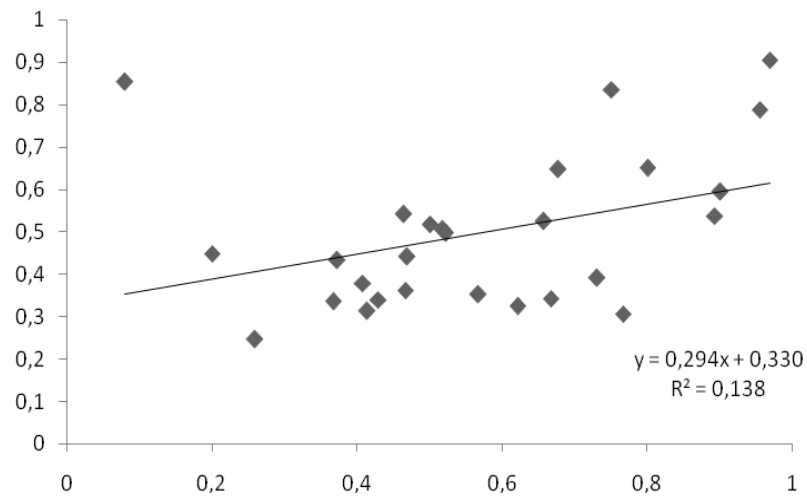


Figure Supp. 7b: Correlation between clade A frequency (abscissa) and percent assignment to the lagoon cluster (ordinate) considering all populations except EFO-Z01

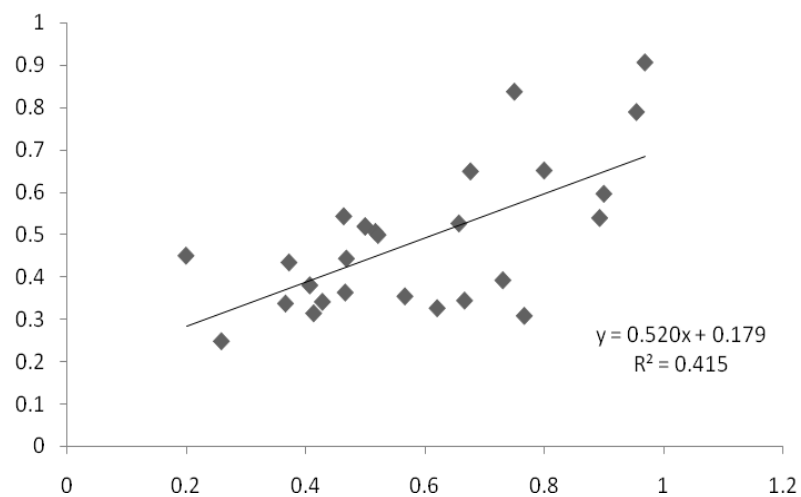


Figure Supp. 8: Correlation between clade A frequency (abscissa) and percent assignment to the lagoon cluster (ordinate) considering offshore Tunisian samples only.

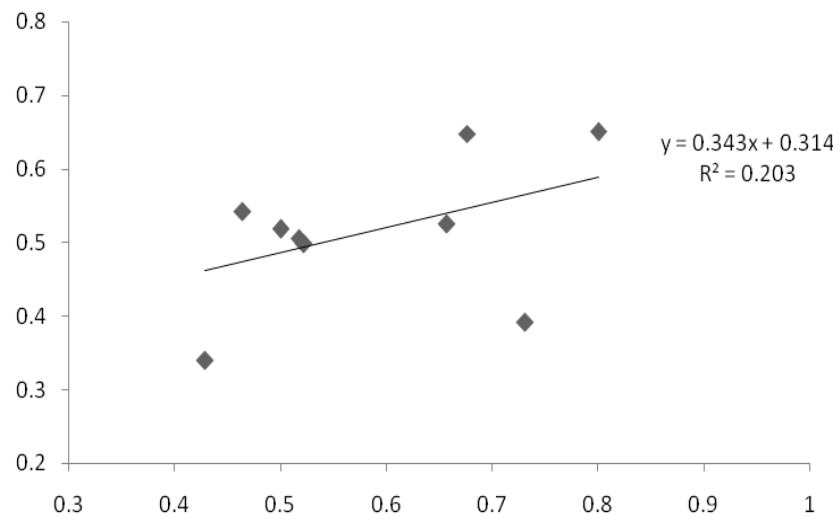


Table Supp. 9: Fis values for 6 microsatellite loci

FisG: according to Weir & Cockerham (*Genetix*), p-value after 1000 permutations; FisMC: after correction for null allele (*MicroChecker*);

PFis=p-value from 1000 permutations

Locus pair	EeA2			EeA6			EeA7			EeA8			EeA9			EeA10			Global Multilocus	
Population	FisMC-A2	FisG-A2	PFis	FisMC-A6	FisG-A6	PFis	FisMC-A7	FisG-A7	PFis	FisMC-A8	FisG-A8	PFis	FisMC-A9	FisG-A9	PFis	FisMC-A10	FisG-A10	PFis	FisG Glob	PFis
FEO-Z01	-0.085	-0.066	0.646	0.224	0.259	0.013	-0.034	-0.016	0.443	0.064	0.083	0.040	0.043	0.074	0.115	0.152	0.170	0.138	0.065	0.047
TLI-I	0.280	0.295	0.001	0.180	0.195	0.046	-0.038	-0.022	0.495	0.255	0.270	0.000	0.041	0.061	0.111	-0.121	-0.106	0.676	0.130	0.000
TLB-ALB	0.131	0.148	0.027	0.374	0.388	0.003	0.235	0.251	0.000	0.116	0.132	0.006	0.023	0.040	0.168	-0.016	0.000	0.398	0.139	0.000
MLN-C04	0.395	0.428	0.000	-0.012	0.005	0.369	0.177	0.201	0.003	0.385	0.399	0.000	0.169	0.232	0.002	0.328	0.365	0.008	0.287	0.000
MBH-B04	0.245	0.282	0.001	0.114	0.129	0.994	0.115	0.131	0.488	0.487	0.238	0.001	0.487	0.499	0.153	0.181	0.197	0.292	0.264	0.000
MNL-D04	0.065	0.091	0.000	-0.403	-0.382	0.105	-0.059	-0.033	0.024	0.213	0.499	0.000	0.213	0.238	0.000	0.016	0.041	0.023	0.020	0.294
TN-AD	0.131	0.182	0.016	0.101	0.144	0.047	0.045	0.072	0.086	0.155	0.169	0.000	0.139	0.171	0.004	0.570	0.639	0.000	0.206	0.000
TN-AK	0.444	0.458	0.000	0.114	0.136	0.064	-0.066	-0.049	0.640	0.203	0.219	0.000	0.224	0.240	0.000	0.516	0.528	0.000	0.241	0.000
TGT-AH	-0.066	-0.051	0.616	0.037	0.054	0.221	-0.092	-0.076	0.788	0.118	0.140	0.005	0.194	0.220	0.003	-0.199	-0.196	0.831	0.029	0.197
TN-AC	-0.057	-0.077	0.710	0.076	0.117	0.100	0.001	0.018	0.263	0.096	0.117	0.025	0.464	0.584	0.000	-0.042	-0.033	0.489	0.116	0.001
TS-AL	0.042	0.080	0.106	0.330	0.361	0.001	-0.161	-0.145	0.984	0.122	0.144	0.004	0.256	0.372	0.000	0.200	0.216	0.047	0.160	0.000
FGG-L01	0.249	0.272	0.000	-0.023	-0.033	0.508	0.036	0.047	0.138	0.001	0.012	0.264	0.207	0.229	0.000	-0.003	0.008	0.362	0.097	0.001
FGG-J02	0.088	0.108	0.064	-0.070	-0.145	0.755	-0.008	0.008	0.346	-0.009	0.007	0.266	0.428	0.457	0.000	-0.154	-0.194	0.831	0.070	0.029
FGG-A04	0.075	0.111	0.045	0.025	0.052	0.222	0.095	0.107	0.020	0.130	0.140	0.001	0.212	0.252	0.000	0.086	0.106	0.103	0.131	0.000
FGG-K02	0.105	0.133	0.034	0.039	0.061	0.189	-0.056	-0.039	0.581	0.137	0.153	0.001	0.297	0.338	0.000	-0.079	-0.069	0.522	0.116	0.000
FGG-G02*	*	0.111	0.197	0.185	0.201	0.022	-0.065	-0.049	0.640	0.010	0.026	0.185	0.259	0.274	0.000	0.192	0.227	0.031	0.117	0.083
FGDR-1	0.295	0.317	0.000	0.024	0.041	0.251	0.072	0.081	0.065	0.154	0.163	0.000	0.353	0.361	0.000	0.333	0.361	0.002	0.205	0.000
FGG-C02	0.106	1.000	0.000	0.057	0.128	0.070	-0.052	-0.042	0.684	0.027	0.039	0.110	0.248	0.279	0.000	0.007	0.018	0.268	0.254	0.000
TGG-AS	-0.086	-0.070	0.706	-0.152	-0.136	0.752	-0.057	-0.041	0.592	0.082	0.099	0.022	0.300	0.356	0.000	-0.176	-0.160	0.507	0.048	0.088
FGG-N01	-0.047	-0.037	0.634	-0.050	-0.129	0.782	-0.010	0.001	0.395	0.031	0.042	0.087	0.157	0.177	0.002	-0.138	-0.128	0.675	0.005	0.422
FGG-L02	-0.082	-0.065	0.674	0.053	0.139	0.096	0.105	0.121	0.030	0.063	0.079	0.029	0.192	0.226	0.001	0.125	0.178	0.044	0.105	0.003
FSET-A05	0.405	0.419	0.000	0.093	0.120	0.082	0.088	0.105	0.053	0.169	0.186	0.001	0.088	0.105	0.038	0.159	0.175	0.050	0.182	0.000
TGT-AA	0.068	0.141	0.057	-0.196	-0.188	0.937	0.039	0.062	0.136	0.069	0.086	0.043	0.438	0.591	0.000	-0.190	-0.173	0.688	0.125	0.001
TCB-AF	0.092	0.183	0.024	0.134	0.157	0.049	0.216	0.251	0.001	0.087	0.104	0.018	0.142	0.165	0.004	0.191	0.208	0.026	0.173	0.000
FGG-H02	0.181	0.294	0.001	-0.069	-0.067	0.634	-0.065	-0.055	0.673	0.018	0.038	0.133	0.259	0.309	0.000	-0.094	-0.094	0.247	0.095	0.004
TGH-AG	0.115	0.171	0.016	-0.033	-0.017	0.454	-0.096	-0.082	0.841	0.400	0.414	0.000	0.172	0.424	0.000	-0.071	-0.055	0.105	0.176	0.000
TGG-AM	0.035	0.054	0.233	0.220	0.237	0.008	-0.095	-0.077	0.742	-0.079	-0.061	0.754	0.131	0.162	0.005	0.000	0.000	0.000	0.060	0.048
BMN	-0.037	-0.020	0.490	0.243	0.259	0.002	-0.058	-0.041	0.584	0.030	0.047	0.137	0.229	0.277	0.000	-0.016	0.277	0.000	0.102	0.002

Table Supp. 10: Linkage Disequilibrium for 15 pairs of 6 microsatellite loci.

R average correlation coefficient according to Black & Krafur (*Genetix*), P = p-value, LD average over 15 pairs, P CF = p-values combined according to Fisher's method

Loci	EeA2-EeA6		EeA2-EeA7		EeA2-EeA8		EeA2-EeA9		EeA2-EeA10		EeA6-EeA7		EeA6-EeA8		EeA6-EeA9		EeA6-EeA10		EeA7-EeA8		EeA7-EeA9		EeA7-EeA10		EeA8-EeA9		EeA8-EeA10		EeA9-EeA10		LD Average	
	R	P	R	P	R	P	R	P	R	P	R	P	R	P	R	P	R	P	R	P	R	P	R	P	R	P	R	P	R	P	DLmoy	P CF
FEO-Z01	0.206	0.148	0.168	0.557	0.175	0.348	0.178	0.646	0.174	0.389	0.165	0.616	0.183	0.231	0.142	0.964	0.217	0.184	0.166	0.476	0.154	0.965	0.158	0.565	0.164	0.938	0.167	0.351	0.171	0.579	0.173	0.805
TLI-I	0.114	0.797	0.162	0.178	0.146	0.270	0.144	0.715	0.148	0.478	0.177	0.113	0.152	0.228	0.131	0.849	0.300	0.011	0.147	0.675	0.156	0.523	0.126	0.817	0.177	0.049	0.173	0.132	0.184	0.167	0.162	0.077
TLB-ALB	0.126	0.918	0.148	0.756	0.157	0.552	0.150	0.786	0.158	0.552	0.120	0.795	0.118	0.881	0.296	0.003	0.163	0.324	0.137	0.822	0.156	0.343	0.138	0.769	0.174	0.169	0.153	0.463	0.151	0.612	0.156	0.600
MLN-C04	0.148	0.553	0.173	0.172	0.156	0.271	0.181	0.357	0.222	0.049	0.145	0.684	0.150	0.496	0.170	0.506	0.080	0.955	0.192	0.029	0.194	0.338	0.214	0.042	0.221	0.040	0.165	0.318	0.190	0.253	0.173	0.030
MBH-B04	0.145	0.600	0.126	0.981	0.140	0.783	0.146	0.644	0.190	0.070	0.117	0.943	0.141	0.467	0.140	0.556	0.143	0.600	0.196	0.022	0.140	0.557	0.168	0.146	0.148	0.293	0.148	0.516	0.130	0.917	0.148	0.559
MNL-D04	0.213	0.266	0.231	0.143	0.190	0.696	0.189	0.701	0.270	0.118	0.336	0.009	0.164	0.918	0.167	0.806	0.180	0.578	0.185	0.937	0.172	0.991	0.184	0.751	0.220	0.429	0.173	0.898	0.219	0.329	0.206	0.559
TN-AD	0.155	0.509	0.157	0.433	0.142	0.895	0.158	0.563	0.175	0.316	0.193	0.099	0.166	0.214	0.154	0.647	0.230	0.053	0.146	0.765	0.165	0.360	0.179	0.226	0.133	0.878	0.163	0.121	0.140	0.653	0.164	0.365
TN-AK	0.157	0.605	0.163	0.222	0.149	0.419	0.142	0.582	0.172	0.194	0.172	0.254	0.183	0.023	0.163	0.409	0.111	0.973	0.144	0.940	0.165	0.152	0.153	0.598	0.138	0.678	0.226	0.004	0.140	0.789	0.159	0.113
TGT-AH	0.178	0.213	0.447	0.000	0.151	0.681	0.175	0.246	0.170	0.339	0.169	0.244	0.174	0.205	0.154	0.559	0.186	0.252	0.149	0.729	0.153	0.549	0.169	0.308	0.153	0.511	0.244	0.128	0.178	0.252	0.190	0.004
TN-AC	0.194	0.434	0.487	0.000	0.198	0.331	0.311	0.137	0.298	0.143	0.171	0.291	0.138	0.937	0.163	0.830	0.219	0.072	0.181	0.243	0.231	0.179	0.225	0.216	0.155	0.317	0.186	0.082	0.326	0.063	0.232	0.000
TS-AL	0.186	0.637	0.365	0.000	0.200	0.266	0.297	0.103	0.209	0.305	0.152	0.849	0.142	0.975	0.184	0.742	0.187	0.330	0.157	0.358	0.254	0.087	0.154	0.587	0.167	0.742	0.124	0.928	0.125	0.962	0.194	0.035
FGG-L01	0.181	0.500	0.108	0.909	0.124	0.479	0.122	0.479	0.137	0.265	0.170	0.535	0.171	0.491	0.160	0.789	0.138	0.886	0.113	0.535	0.131	0.258	0.115	0.584	0.120	0.274	0.133	0.154	0.211	0.026	0.142	0.507
FGG-J02	0.269	0.268	0.143	0.742	0.160	0.364	0.154	0.534	0.111	0.953	0.190	0.982	0.281	0.092	0.255	0.463	0.388	0.077	0.138	0.803	0.208	0.016	0.154	0.744	0.151	0.399	0.169	0.541	0.169	0.565	0.196	0.387
FGG-A04	0.255	0.030	0.128	0.486	0.147	0.223	0.137	0.423	0.137	0.296	0.127	0.761	0.134	0.517	0.130	0.768	0.113	0.836	0.118	0.295	0.131	0.137	0.121	0.249	0.119	0.268	0.112	0.349	0.144	0.101	0.137	0.180
FGG-K02	0.160	0.713	0.163	0.400	0.134	0.999	0.152	0.809	0.184	0.230	0.173	0.219	0.189	0.147	0.223	0.078	0.169	0.426	0.159	0.458	0.159	0.493	0.287	0.047	0.175	0.200	0.150	0.620	0.137	0.903	0.174	0.322
FGG-G02	0.000	0.996	0.000	1.000	0.000	1.000	0.000	0.994	0.000	0.517	0.141	0.734	0.137	0.838	0.148	0.511	0.174	0.321	0.146	0.533	0.151	0.360	0.162	0.389	0.161	0.247	0.143	0.750	0.126	0.925	0.099	0.992
FGDR-1	0.094	0.980	0.113	0.187	0.105	0.326	0.094	0.825	0.139	0.197	0.105	0.794	0.121	0.211	0.086	0.995	0.109	0.620	0.098	0.628	0.098	0.553	0.126	0.306	0.104	0.340	0.098	0.829	0.102	0.698	0.106	0.870
FGG-C02	0.000	0.970	0.844	0.003	0.683	0.532	0.770	0.226	0.370	0.653	0.220	0.151	0.133	0.982	0.190	0.401	0.133	0.764	0.114	0.467	0.107	0.701	0.105	0.621	0.124	0.288	0.165	0.094	0.121	0.307	0.272	0.222
TGG-AS	0.093	0.995	0.142	0.911	0.157	0.452	0.144	0.967	0.198	0.167	0.188	0.185	0.162	0.516	0.224	0.018	0.138	0.582	0.165	0.192	0.200	0.124	0.164	0.504	0.151	0.878	0.138	0.834	0.158	0.680	0.161	0.493
FGG-N01	0.203	0.563	0.119	0.300	0.110	0.595	0.115	0.431	0.116	0.547	0.195	0.715	0.234	0.370	0.294	0.120	0.325	0.065	0.131	0.191	0.118	0.363	0.100	0.841	0.126	0.271	0.130	0.262	0.099	0.917	0.161	0.416
FGG-L02	0.207	0.718	0.150	0.642	0.133	0.971	0.153	0.653	0.173	0.358	0.312	0.123	0.202	0.920	0.260	0.278	0.315	0.085	0.181	0.111	0.153	0.644	0.173	0.341	0.150	0.624	0.151	0.671	0.192	0.145	0.194	0.497
FSET-A05	0.150	0.609	0.175	0.310	0.197	0.016	0.124	0.984	0.192	0.098	0.149	0.694	0.193	0.167	0.198	0.199	0.125	0.819	0.148	0.592	0.143	0.656	0.163	0.263	0.162	0.253	0.184	0.053	0.169	0.214	0.165	0.096
TGT-AA	0.244	0.129	0.536	0.000	0.238	0.057	0.353	0.117	0.191	0.638	0.137	0.955	0.152	0.730	0.221	0.167	0.151	0.624	0.196	0.252	0.225	0.320	0.163	0.408	0.236	0.085	0.143	0.585	0.290	0.151	0.232	0.003
TCB-AF	0.168	0.922	0.647	0.000	0.240	0.212	0.310	0.105	0.181	0.612	0.152	0.674	0.148	0.780	0.142	0.917	0.251	0.056	0.147	0.753	0.168	0.232	0.143	0.658	0.162	0.266	0.171	0.304	0.124	0.925	0.210	0.049
FGG-H02	0.211	0.513	0.190	0.404	0.165	0.911	0.166	0.732	0.236	0.223	0.216	0.058	0.217	0.217	0.154	0.826	0.176	0.430	0.152	0.971	0.160	0.530	0.121	0.953	0.150	0.661	0.173	0.497	0.146	0.782	0.176	0.857
TGH-AG	0.173	0.675	0.420	0.000	0.254	0.072	0.299	0.627	0.172	0.492	0.178	0.160	0.145	0.783	0.389	0.125	0.080	0.964	0.174	0.341	0.308	0.574	0.177	0.303	0.368	0.278	0.166	0.309	0.149	0.831	0.230	0.030
TGG-AM	0.178	0.312	0.125	0.978	0.150	0.580	0.148	0.759	0.000	-----	0.196	0.053	0.185	0.081	0.183	0.292	0.000	-----	0.165	0.411	0.173	0.420	0.000	-----	0.175	0.469	0.000	-----	0.000	-----	0.112	0.320
BMN	0.174	0.154	0.159	0.438	0.172	0.158	0.152	0.862	0.114	0.838	0.148	0.657	0.164	0.283	0.152	0.700	0.134	0.406	0.148	0.755	0.167	0.446	0.231	0.070	0.181	0.291	0.149	0.199	0.083	0.829	0.155	0.508

ANNEXE C

Tableau 1 : Protocole et programme de la PCR mis au point pour les six locus microsatellites

Produit de PCR	1 ech (ul)			
d'eau mQ	12,3			
de solution tampon (10X) fermentas	2.5			
de solution MgCl2 (5 mM)	1.5			
de Mix dNTPsl (25mM)	1			
amorce (10 ulM, une marquée et une non marquée)	0.5			
de Taq polymérase –Taq maison (5U/μL)	0,2			
de solution diluée (30 ng/μL) d'ADN extrait	2			

	T°	Temps	Nb de cycle
Dénaturation initiale	94°	2 min	
Dénaturation initiale	94	30s	34
Hybridation	56	30s	
Elongation	40	72	
Elongation finale	72°C	5 min	

Figure1: Corrélation entre les valeurs de P de Fis du Test Fisher combiné et les valeurs de P de DL de F.combiné

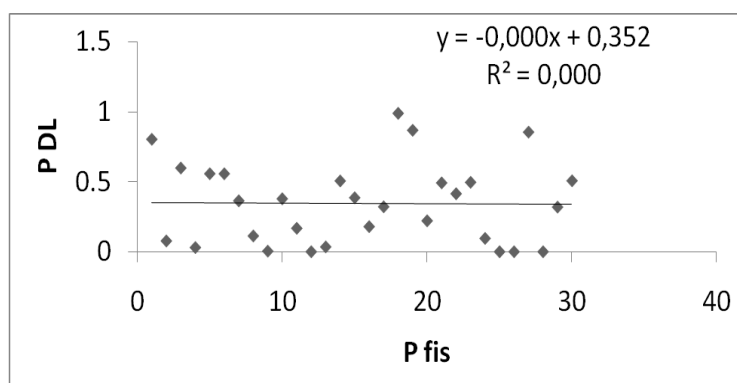


Figure 2 : Analyse de la stabilité de LogP(D) surStructure

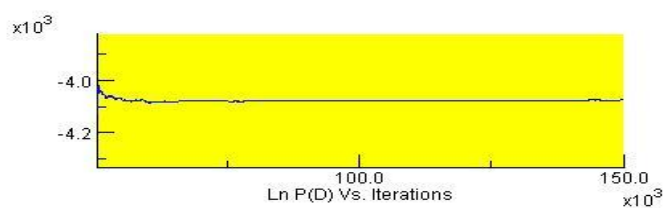


Tableau 2 :

Couple d'amorces utilisé pour l'amplification du gène du cytochrome b de l'ADNmt et la Th choisie.

Séquence	Séquence	Taille	T°
Cytb-F	5'-TAA-CGA-CGC-AGT-AGT-AGA-CC-3'	20	53
Cytb-R	5'- 5'GAG-GAA-GTA-TCA-CTC-AGG-C3'	19	

Produit de PCR	1 ech
Eau mQ	12,5ul
Solution tampon (10X) fermentas	2ul
solution MgCl2 (5 mM)	1,5ul
Mix dNTPsl (10mM)	1ul
Amorce 10 ulM	0,4
Taq polymérase –Taq maison (5U/μL)	0,2
ADN extrait	2ul

Tableau 3 : Matrice des FST par paire d'échantillons au six locus microsat les valeurs en gras et soulignés sont significatives après l'ajustement de *Bonferroni*

	TLI-I	TLB-ALB	FEO-Z01	MLN-C04	TGT-AH	MBH-B04	MNL-D04	FGG-J02	FGG-L02	FGG-K02	FGG-H02	FGG-L01	FGG-A04	TGG-AM	BMN	FSET-A05	FGDR-1	TGG-AS	FGG-N01	TN-AK	TN-AD	TS-AL	TGT-AA	TGH-AG	TN-AC	TCB-AF	FGG-G02	FGG-C02
TLI-I																												
TLB-ALB	0.0061																											
FEO-Z01	0.0069	0.0168																										
MLN-C04	0.0351	0.0171	0.0214																									
TGT-AH	<u>0.0790</u>	<u>0.0486</u>	<u>0.0576</u>	0.0069																								
MBH-B04	<u>0.0871</u>	0.0531	<u>0.0644</u>	0.0146	0.0046																							
MNL-D04	<u>0.0932</u>	<u>0.0591</u>	<u>0.0822</u>	0.0205	0.0048	-0.0034																						
FGG-J02	<u>0.1462</u>	<u>0.1039</u>	<u>0.1224</u>	0.0463	0.0151	0.0118	0.0144																					
FGG-L02	<u>0.1796</u>	<u>0.1291</u>	<u>0.1589</u>	<u>0.0673</u>	0.0270	0.0249	0.0174	-0.0044																				
FGG-K02	<u>0.1758</u>	<u>0.1250</u>	<u>0.1573</u>	<u>0.0662</u>	<u>0.0310</u>	0.0254	0.0222	-0.0006	-0.0097																			
FGG-H02	<u>0.2325</u>	<u>0.1729</u>	<u>0.2109</u>	<u>0.1009</u>	<u>0.0520</u>	<u>0.0461</u>	<u>0.0418</u>	0.0077	-0.0127	-0.0040																		
FGG-L01	<u>0.1290</u>	<u>0.0870</u>	<u>0.1089</u>	0.0396	0.0083	0.0121	0.0138	-0.0026	0.0015	0.0026	0.0137																	
FGG-A04	<u>0.1603</u>	<u>0.1118</u>	<u>0.1343</u>	<u>0.0567</u>	0.0150	0.0232	0.0246	-0.0033	-0.0031	0.0011	0.0057	-0.0048																
TGG-AM	<u>0.2550</u>	<u>0.2109</u>	<u>0.2321</u>	<u>0.1341</u>	<u>0.0697</u>	<u>0.0722</u>	<u>0.0566</u>	0.0273	0.0220	<u>0.0291</u>	<u>0.0290</u>	<u>0.0449</u>	<u>0.0329</u>															
BMN	<u>0.2463</u>	<u>0.1928</u>	<u>0.2234</u>	<u>0.1272</u>	<u>0.0616</u>	<u>0.0703</u>	<u>0.0537</u>	0.0225	0.0059	0.0185	0.0117	<u>0.0291</u>	0.0185	0.0024														
FSET-A05	<u>0.1844</u>	<u>0.1438</u>	<u>0.1663</u>	<u>0.0736</u>	0.0307	0.0342	0.0306	0.0057	0.0001	-0.0007	0.0070	0.0081	0.0098	0.0205	0.0179													
FGDR-1	<u>0.1524</u>	<u>0.1128</u>	<u>0.1281</u>	<u>0.0511</u>	0.0163	0.0230	0.0188	-0.0040	-0.0018	0.0037	0.0132	0.0000	0.0015	0.0214	0.0146	0.0027												
TGG-AS	<u>0.1723</u>	<u>0.1332</u>	<u>0.1530</u>	<u>0.0707</u>	0.0212	0.0299	0.0152	0.0077	0.0080	0.0115	0.0213	0.0126	0.0100	0.0118	0.0129	-0.0008	0.0029											
FGG-N01	<u>0.1801</u>	<u>0.1324</u>	<u>0.1634</u>	<u>0.0760</u>	0.0282	0.0366	0.0248	0.0024	-0.0070	-0.0012	-0.0019	0.0026	0.0022	0.0167	0.0034	0.0008	0.0007	0.0018										
TN-AK	<u>0.1020</u>	<u>0.0631</u>	<u>0.0804</u>	0.0244	-0.0022	-0.0037	0.0077	0.0015	0.0115	0.0134	0.0297	-0.0039	0.0065	<u>0.0534</u>	<u>0.0439</u>	0.0165	0.0063	0.0166	0.0142									
TN-AD	<u>0.0862</u>	0.0538	<u>0.0759</u>	0.0169	-0.0034	0.0006	-0.0020	0.0134	0.0232	0.0263	<u>0.0473</u>	0.0034	0.0178	<u>0.0772</u>	<u>0.0653</u>	0.0285	0.0195	0.0244	0.0241	-0.0059								
TS-AL	<u>0.1287</u>	<u>0.0964</u>	<u>0.1040</u>	0.0347	-0.0024	0.0176	0.0170	0.0045	0.0108	0.0162	0.0271	-0.0020	0.0012	<u>0.0388</u>	<u>0.0311</u>	0.0080	0.0012	0.0033	0.0100	0.0020	0.0065							
TGT-AA	<u>0.1813</u>	<u>0.1279</u>	<u>0.1524</u>	<u>0.0612</u>	0.0269	0.0291	0.0168	0.0224	0.0135	0.0163	0.0214	0.0275	0.0229	<u>0.0385</u>	0.0271	0.0191	0.0173	0.0120	0.0172	0.0228	0.0305	0.0205						
TGH-AG	<u>0.2513</u>	<u>0.1987</u>	<u>0.2207</u>	<u>0.1127</u>	<u>0.0745</u>	<u>0.0718</u>	<u>0.0618</u>	<u>0.0566</u>	<u>0.0443</u>	<u>0.0574</u>	0.0404	<u>0.0597</u>	<u>0.0494</u>	<u>0.0531</u>	<u>0.0505</u>	0.0429	<u>0.0523</u>	0.0417	<u>0.0468</u>	<u>0.0676</u>	<u>0.0720</u>	<u>0.0512</u>	0.0209					
TN-AC	<u>0.1390</u>	<u>0.0991</u>	<u>0.1160</u>	0.0337	0.0074	0.0070	-0.0118	0.0131	0.0127	0.0193	<u>0.0351</u>	0.0151	0.0181	<u>0.0504</u>	0.0433	0.0229	0.0135	0.0094	0.0178	0.0110	0.0108	0.0070	0.0012	0.0350				
TCB-AF	<u>0.1953</u>	<u>0.1497</u>	<u>0.1657</u>	<u>0.0717</u>	0.0267	0.0350	0.0357	0.0016	-0.0046	0.0064	0.0006	0.0095	0.0041	0.0159	0.0134	0.0016	0.0009	0.0122	0.0039	0.0165	0.0376	0.0062	0.0185	0.0356	0.0249			
FGG-G02	<u>0.1632</u>	<u>0.1080</u>	<u>0.1415</u>	<u>0.0586</u>	0.0202	0.0256	0.0187	-0.0016	-0.0077	-0.0044	-0.0005	-0.0022	-0.0020	<u>0.0313</u>	0.0086	0.0065	-0.0009	0.0099	-0.0065	0.0055	0.0160	0.0089	0.0083	<u>0.0487</u>	0.0115	0.0068		
FGG-C02	<u>0.1819</u>	<u>0.1346</u>	<u>0.1599</u>	<u>0.0720</u>	0.0247	0.0296	0.0251	-0.0008	-0.0070	-0.0016	-0.0010	0.0010	0.0016	0.0196	0.0091	-0.0038	0.0001	0.0029	-0.0034	0.0112	0.0235	0.0054	0.0188	<u>0.0480</u>	0.0178	-0.0012	-0.0023	0.0000

Tableau 5 : Diversité génétique (cytb) de chaque population:

N : nombre d'individu séquencés

h : diversité haplotypique

π : diversité nucléotidique

Population	Cytb		
	N	<i>h</i>	π
FEO-Z01	25	0.930	0.08
TLI-I	32	0.965	0.006
TLB-ALB	28	0.986	0.014
MLN-C04	22	0.926	0.007
MBH-B04	28	0.970	0.008
MNL-D04	20	0.942	0.008
TN-AK	26	0.972	0.015
TGT-AH	33	0.998	0.015
FGG-A-A06	45	0.908	0.014
FGG-M-A06	31	0.924	0.013
TN-AC	34	0.992	0.016
TS-AL	14	0.978	0.008
FGG-L01	27	0.968	0.016
FGG-J02	17	0.904	0.006
FGG-A04	29	0.963	0.015
FGG-K02	19	0.888	0.005
FGG-G02	30	0.970	0.016
FGDR-1	51	0.987	0.016
FGG-C02	30	0.970	0.014
TGG-AS	30	0.965	0.013
FGG-N01	29	0.967	0.015
FGG-L02	30	0.963	0.016
TCB-AF	22	0.982	0.018
FGG-H02	29	0.965	0.014
TGH-AG	28	0.984	0.016
TGG-AM	16	0.983	0.006
BMN	30	0.995	0.013

Diversité génétique pour chaque clade

Clade A

Population	Cytb		
	N	<i>h</i>	π
FEO-Z01	0	0	0
TLI-I	31	0.9634	0.0046
TLB-ALB	21	0.9761	0.0055
MLN-C04	21	0.9190	0.0054
MBH-B04	25	0.9633	0.0044
MNL-D04	18	0.9281	0.0046
TN-AK	19	0.9473	0.0073
TGT-AH	23	0.9960	0.0073
TN-AC	23	0.9841	0.0068
TS-AL	15	1	0.0071
FGG-L01	11	0.9454	0.0066
FGG-J02	15	0.9428	0.00658
FGG-A04	12	0.9697	0.0049
FGG-K02	7	0.9454	0.0043
FGG-G02	11	0.9454	0.0062
FGDR-1	19	0.9415	0.0046
FGG-C02	20	0.9421	0.0047
TGG-AS	24	0.9456	0.0068
FGG-N01	17	0.9264	0.0039
FGG-L02	14	0.8901	0.0039
TCB-AF	11	0.9818	0.0059
FGG-H02	18	0.9542	0.0039
TGH-AG	13	0.9615	0.0060
TGG-AM	12	0.9697	0.0050
BMN	23	0.9920	0.0070

CladeB

Population	Cytb		
	N	<i>h</i>	π
FEO-Z01	23	0.917	0.00502
TLB-ALB	7	1	0.0096
TN-AK	7	1	0.0078
TGT-AH	10	1	0.0078
TN-AC	11	1	0.0092
TS-AL	14	0.9780	0.0080
FGG-L01	16	0.9333	0.0069
FGG-J02	17	0.9044	0.0069
FGG-A04	17	0.9044	0.00601
FGG-K02	19	0.8888	0.0057
FGG-G02	18	0.9346	0.0072
FGDR-1	32	0.9879	0.0078
FGG-C02	10	0.9555	0.0068
TGG-AS	6	1	0.0076
FGG-N01	12	0.9545	0.0077
FGG-L02	16	0.95	0.007
TCB-AF	11	0.9454	0.0079
FGG-H02	11	0.8727	0.0055
TGH-AG	15	0.9714	0.00609
TGG-AM	16	0.9833	0.0069
BMN	7	1	0.001

Résumé

On étudie la structure génétique des anchois Atlanto-méditerranéens analysée grâce à 6 locus microsatellites et un gène du cytochrome b de l'ADN mitochondrial sur un échantillonnage dont quatre en provenance de lagunes réparti entre la Tunisie, la France, la Mer Noire et le Maroc. Deux locus microsatellites parmi les six étudiés révèlent une forte différenciation liée à l'habitat lagunaire. L'étude du cytochrome b de l'ADN mitochondrial montre la présence de deux clades mitochondriaux trouvés en sympatrie en fréquence variable dans tous les échantillons. L'analyse combinée montre un déséquilibre nucléo-cytoplasmique significatif, confirmant l'existence d'incompatibilités génétiques partielles entre les deux entités. Ces deux entités génétiques doivent de ce fait être considérées comme des unités de gestion distinctes.

Mots-clés : anchois, introgression, microsatellites, déséquilibre nucléocytoplasmique

Abstract

We study the genetic structure of Mediterranean anchovy by analyzing six microsatellite loci and the cytochrome b gene of mitochondrial DNA on anchovy populations with four lagoons, which are from Tunisia, Morocco, France and Black Sea. Two strongly differentiated loci and nucleo-cytoplasmic disequilibrium revealed a differentiation related to a lagoon habitat. The study of cytochrome b mitochondrial DNA showed the presence of two mitochondrial clades found on sympatric with variable frequency in all samples. By contrast, four microsatellites were undifferentiated within the entire set of samples, testifying to a variable permeability of the two interacting genomes. Our findings confirm the existence of partial genetic incompatibilities between the two entities that should then be considered as distinct from a conservation point of view.

Keywords: anchovy, introgression, microsatellites, linkage disequilibrium