

Etude phylogéographique des lignées de truites communes *Salmo trutta* en Corse

Renouard Manon, encadrée par : Christelle Tougard, chargée de recherche CNRS et Patrick Berrebi, directeur de recherche CNRS

2013-2014

Master 1 - Génétique et Biodiversité – Université Montpellier II

GMBE235

Introduction :

La truite commune *Salmo trutta* est une espèce très répandue de la famille des Salmonidae. Elle est actuellement répartie en Europe, en Afrique du Nord et en Asie occidentale (MacCrimmon et Gots, 1970). Cette large aire de répartition s'explique par une grande capacité d'adaptation à de nombreux habitats (large gamme de températures, salinité...) (Baglinière, 1991). La truite commune est une espèce très sédentaire dans la plus grande partie de sa répartition (Launey et al., 2003). Cette forte plasticité phénotypique et cette tendance à la sédentarité font qu'il existe chez la truite de nombreuses adaptations locales entraînant de nombreuses différenciations morphologiques (Pakkasmaa et Pironen, 2001). Du point de vue génétique, les séquences mitochondriales ont permis de distinguer sept lignées évolutives principales (Bernatchez et al., 1992) : la lignée dite atlantique (AT) (la plus fréquente dans les souches d'élevage), les lignées méditerranéenne (ME), adriatique (AD) dont dérive la truite « corse », danubienne (DA) et marbrée (MA) (Bernatchez et al., 1992) ainsi que les lignées mineures du Duero (DU) et du Tigris (TI).

L'abondance et surtout la variabilité génétique de la truite commune sont actuellement menacées par la surpêche, la dégradation et la fragmentation de ses habitats, notamment dues aux activités humaines. La fragmentation et les importants déversements délibérés de truites d'élevage (repeuplements) dans tous les milieux réduisent la diversité génétique, car les populations sont souvent de petites tailles avec une diversité allélique limitée (Gauthier et Berrebi, 2007 ; Faure, 2012). En Corse, il y a eu un afflux important de truites domestiques provenant du continent (Guyomard et Krieg, 1986) seulement arrêté depuis quelques années après la décision de la Fédération de pêche de stopper tout repeuplement. La compatibilité génétique entre les souches naturelles et introduites étant totale, se produisent fréquemment des hybridations entre les différentes souches. L'introgression massive des gènes d'une lignée dans une autre est donc préoccupante car elle peut entraîner une réduction de l'adaptation locale et du polymorphisme en Corse (Ryman et Stahl, 1980).

De nombreuses études, particulièrement sur les allozymes et les microsatellites, ont été réalisées pour définir les caractéristiques génétiques des populations de *Salmo trutta* et tenter d'évaluer les effets directs et indirects liés aux échanges génétiques entre les formes allochtones et autochtones (Splendiani et al., 2013). En Corse, les microsatellites ont permis de caractériser des dizaines de populations en termes de proportion de gènes de types « corse ancestral », « méditerranéen » ou « domestique atlantique » (Berrebi, 1995). En France, les études basées sur des marqueurs nucléaires et mitochondriaux ont également permis de confirmer la forte différenciation génétique entre populations de truite et de déterminer leur structuration géographique globale, ainsi que les relations évolutives entre les différentes populations (Guyomard et Krieg 1986 ; Aurelle et Berrebi, 1998 ; Reynaud, 2011). Parmi les marqueurs nucléaires, les allozymes sont peu variables et décrivent des structurations anciennes ayant eu lieu à l'échelle continentale, alors que les microsatellites, hypervariables, sont précis mais ne décrivent que des événements récents et locaux.

En parallèle, d'après Gauthier et Berrebi (2007), la position altitudinale de chaque lignée de truites en Corse nous renseigne sur leur ordre d'arrivée sur l'île. En aval, on trouve des truites méditerranéennes et des hybrides adriatiques x méditerranéens alors qu'en amont, on ne trouve que des truites adriatiques

(si on fait abstraction de truites atlantiques d'origine anthropique). Or, il existe de nombreux obstacles, cascades et dénivelés qui séparent les parties amont et aval des rivières en Corse. La truite adriatique semblerait donc être la première à avoir peuplé les sommets corses. Ce serait donc ensuite la formation de ces obstacles infranchissables qui aurait empêché la remontée des cours d'eau par les truites méditerranéennes, arrivée par la suite (Gauthier et Berrebi, 2007). Les études géologiques ont révélé l'existence d'importants glaciers sur les sommets corses où la vie aquatique n'était pas possible. Au cours de la déglaciation (-15 000 ans), une modification du relief de la Corse à l'origine des nombreuses cascades et dénivelés (compensations orogéniques) a pu empêcher la remontée de la truite méditerranéenne arrivée plus tard vers les sommets montagneux (Durand Delga, 1978). L'arrivée des truites méditerranéennes sur les hautes montagnes serait donc postérieure au début de la fonte de ces glaciers, c'est-à-dire il y a moins de 15 000 ans environ. Il est également possible d'envisager des effets de priorité aux pionniers, soit des adaptations locales qui auraient restreint les flux génétiques et donc empêché la colonisation du centre de la Corse par la lignée ME.

Les études sur la région de contrôle de l'ADN mitochondrial (CR) ont permis de définir la structure génétique des truites communes (Bernatchez et al., 1992). En Corse, il a été mis en évidence quatre lignées de truites : AT, AD, MA et ME, dont les dates d'arrivée sur l'île n'ont pas pu toutefois être établies (Tougaard, 2013). Parmi ces quatre lignées, trois sont probablement arrivés en Corse de façon naturelle (lignées ME, AD et MA) et la quatrième (lignée AT) aurait été introduite par l'homme pour repeupler les rivières (Gauthier et Berrebi, 2007).

L'origine des lignées et leur arrivée en Corse ne sont donc pas totalement comprises par l'étude des allozymes, des microsatellites et de la CR. Nous nous proposons, dans la présente étude, de rajouter l'analyse du gène du cytochrome *b* (*cyt b*) à celles de la CR. Cela devrait permettre de renforcer le jeu de données déjà acquis afin de dégager une phylogénie mieux résolue de *Salmo trutta*, l'origine des différentes lignées et la date de leur apparition en Corse.

Matériels et méthodes :

Intérêt du marqueur : L'intérêt du *cyt b* est qu'il est considéré comme l'un des gènes les plus utiles pour l'étude phylogénétique, et est probablement le gène mitochondrial le plus connu en ce qui concerne sa structure et la fonction de la protéine qu'il produit. En outre, comme tout marqueur mitochondrial, le *cyt b* possède une hérédité maternelle. N'étant donc pas soumis au brassage génétique, il est donc relativement constant et évolue donc plutôt lentement. Ainsi, l'étude du *cyt b* est capable de décrire l'histoire ancienne des peuplements. Il pourrait donc en effet, nous renseigner sur l'origine des diverses lignées de truites en Corse.

Choix des populations : Les truites étudiées proviennent de 38 populations localisées en Corse et pêchées dans le cadre des programmes européen LIFE et local OEC ainsi que de 24 populations provenant du sud de la France continentale, d'Italie, des Balkans et d'Espagne qui ont déjà fait l'objet de séquençage de la CR (Reynaud, 2011 ; Tougaard, 2012, 2013). L'échantillonnage est constitué d'environ deux à neuf

individus par population, soit 388 individus au total dont les fragments de nageoire sont en collection dans l'alcool à l'ISEM (*Annexes 1-3*).

Extraction et purification de l'ADN : La première étape avant le séquençage est d'extraire l'ADN total contenu dans les fragments de nageoires par la méthode dite Chelex/protéinase k qui consiste à laisser incubé à 56°C les échantillons dans une solution contenant 2µl de protéinase k (20 mM) et 200µl de Chelex (5%) (Walsh et al., 1991). Cette méthode présente l'avantage d'être rapide mais ne permet pas d'assurer la purification de l'ADN. C'est pourquoi, une purification grâce au kit « GenElute » (Sigma) a été rajoutée au protocole, permettant d'obtenir des échantillons d'ADN plus purs.

Amplification de l'ADN : Le mélange réactionnel pour la PCR (volume final : 25µL) est constitué de 1,5µL d'ADN (200-500 ng/ul), 5µL de chaque amorce (2mM) (*Tableau 1*), 0,125µL d'ADN polymérase « GoTaq » (Promega) (50u/µL), de 2,5µL d'une solution contenant les quatre dNTP (0,25mM pour chaque dNTP), 2µL de MgCl₂ (2mM), ainsi que 5µL de tampon 5X. Le programme thermique utilisé comprend un premier cycle avec 5 minutes à 94°C, 1 minute à 51°C et 2 minutes à 72°C, puis 35 cycles alternant 45 secondes à 94°C, 1 minute à 51°C et 2 minutes à 72°C et une élongation terminale de 10 minutes à 72°C. La présence de pseudogènes dans les amplifications de longueur proche du fragment d'intérêt (1140 paires de bases ; pb), nous a conduit par la suite à rajouter deux amorces internes (*Tableau 1*). Ainsi, deux couples d'amorces ont été utilisés, pour amplifier deux fragments chevauchant d'environ 650 pb.

Séquençage : Les produits de la PCR sont ensuite purifiés grâce à un robot automatisé BioMeck 300 qui utilise le kit « CleanPCR » (Agencourt) basé sur l'utilisation de billes magnétiques. La réaction de séquençage peut ainsi avoir lieu. Le volume réactionnel (volume final : 10µL) est composé de 0,5µL de « Big Dye » (Applied Biosystems), 1µL d'une amorce (3,3mM) ainsi qu'1 µL de tampon 10X. La quantité de produits de PCR et d'eau ajoutés sont estimés à partir d'une quantification de l'ADN après la purification grâce à un spectrophotomètre de type « Nanodrop ». Le programme PCR utilisé par la suite comprend 45 cycles alternant 10 secondes à 96°C, 5 secondes à 50°C et 4 minutes à 60°C. Les séquences sont par la suite purifiées grâce, à nouveau, au robot de purification avec le kit « CleanSEQ » (Agencourt), reposant aussi sur une technologie rapide à base de billes magnétiques. Le séquençage par électrophorèse capillaire (méthode dérivée de Sanger) a eu lieu dans un séquenceur ABI Prism (Applied Biosystems). La plupart de ces manipulations a été faite dans la Plateforme Génomique Environnementale du Labex CeMEB à l'UM2.

Reconstructions phylogénétiques : Une fois les séquences obtenues, le logiciel MEGA v 6,6 (Tamura et al., 2013) a permis de réaliser leur alignement. Les phylogénies ont été réalisées avec les données obtenues pour le *cyt b* mais aussi en les combinant avec celles obtenues pour la CR pour les mêmes individus lors d'études antérieures (Reynaud, 2011 ; Tougard, 2012, 2013). Elles ont été réalisées dans un premier temps pour les données corses uniquement afin de caractériser la structuration des populations (nombre de lignées) au sein de l'île, puis dans un second temps pour l'ensemble des données, dans le but de replacer l'étude dans un contexte plus général et d'émettre des hypothèses sur l'origine des différentes lignées. Les jeux de données ont également été enrichis par des séquences provenant de GenBank, appartenant à des truites continentales de lignées connues de *Salmo trutta* et des séquences du saumon, *Salmo salar*, utilisées pour enracer les arbres (*Annexe 1*).

Les arbres ont été obtenus à l'aide de deux méthodes probabilistes : la méthode du maximum de vraisemblance (ML) grâce au logiciel PhyML v 3.0 (Guindon et al., 2010) et la méthode bayésienne (BA) grâce au logiciel MrBayes v 3.1.2 (Ronquist et Huelsenbeck, 2003). Ces méthodes sont basées sur le concept de vraisemblance mais elles diffèrent par l'estimation de la probabilité (Delsuc, 2002). Pour être statistiquement robuste et performantes, ces méthodes nécessitent l'incorporation de modèles qui décrivent les processus d'évolution des séquences nucléotidiques. Le choix des modèles à appliquer a été déterminé en utilisant le logiciel MrModelTest v 2.3 (Nylander, 2008) qui choisit comme modèle le plus approprié celui avec le critère d'information Akaike (AIC) le plus faible. Dans le cas de l'analyse des données combinées, seule la méthode BA a permis d'appliquer à chaque jeu de donnée (cyt *b* et CR) des modèles d'évolution différents. La robustesse des nœuds en ML a été estimée grâce aux pourcentages de bootstraps (BP) après 100 pseudo-répliques. Pour l'analyse BA, trois runs constitués de trois chaînes de Markov avec trois millions de générations ont été lancées. Le degré de confiance de chaque nœud a été estimé par une probabilité postérieure (PP). Le choix de l'arbre consensus, c'est-à-dire celui dont la probabilité postérieure est maximale, se base sur la règle du consensus majoritaire à 50%. Cette règle est appliquée sur tous les arbres échantillonnés toutes les 100 générations, après une étape d'allumage ou « burn-in » de 25 000 itérations.

Indices de divergence, de diversité et indices démographiques : Afin d'évaluer la variabilité génétique des truites corses et, de manière plus générale, en Europe, des distances génétiques intra- et interlignées ont été calculées sous le modèle de Kimura à deux paramètres à l'aide du logiciel MEGA, ainsi que des indices de diversité (nombre de séquences (N), diversité nucléotidique (π), diversité haplotypique (hd), nombre moyen de différences en comparant deux à deux les séquences (k)) obtenus grâce au logiciel DNAsp v 5.10 (Librado et Rozas, 2009).

Dans le but de décrire l'histoire démographique des lignées, ce même logiciel a également permis de réaliser des tests de neutralité : le test F_s de Fu (1997), la statistique D de Tajima (1989) et le test R_2 de Ramos Onsins et Rozas (2002). Dans le cas d'une population en expansion, les valeurs de D et F_s sont significativement négatives et les valeurs de R_2 sont significativement positives. Des distributions de disparités (mismatch distribution) sont également construites sous un modèle de population en croissance-déclin (Rogers et Harpending, 1992). Une expansion y est représentée par une distribution unimodale alors qu'une démographie stable y est représentée par une courbe bimodale. En outre, les distributions observées et théoriques sont comparées pour vérifier le modèle en utilisant la valeur de la somme des écarts au carré (SSD) et l'indice de raggedness (r) (Harpending 1994) obtenus grâce au logiciel Arlequin v 3.1.2 (Excoffier et Lischer, 2010). Pour ces indices, l'hypothèse d'expansion est conservée pour $P > 0,05$. On admet que la démographie est en expansion lorsqu'au moins trois des indices présentés ci-dessus indiquent une expansion, dans le cas contraire la population est dite stable.

Structure des populations : La structure des populations a ensuite été étudiée en réalisant un réseau d'haplotypes à l'échelle européenne, basé sur la méthode du maximum de parcimonie selon l'algorithme du « Median Joining » grâce au logiciel NETWORK v 4.5.1.6 (www.fluxus-engineering.com;

Bandelt et al., 1999). Cette méthode consiste à minimiser le nombre de pas évolutifs d'un haplotype à un autre.

Datation moléculaire : Afin d'obtenir une estimation de la date d'arrivée de chaque lignée en Corse, la datation des embranchements a été estimée grâce au programme BASEML du logiciel PAML X (Xu et Yang, 2013) en utilisant une horloge moléculaire locale et l'arbre obtenu en ML avec PhyML pour les haplotypes uniquement endémiques de Corse. Ainsi cette approche va nous permettre d'estimer les âges d'arrivée des lignées en Corse. On estime que l'âge d'arrivée d'une lignée est antérieur à la valeur donnée par la datation : il est donc compris dans un intervalle semi-ouvert. Pour permettre l'estimation des âges, la date de 11,6 Ma pour l'origine du genre *Salmo* (Crête Lafrenière et al., 2012) a été utilisé comme point de calibration.

Résultats :

Analyses phylogénétiques : L'alignement du *cyt b* comporte 1140 nucléotides dont 114 sites phylogénétiquement informatifs. Lorsque les données de la CR (1022 nucléotides) sont ajoutées, on obtient un alignement de 2162 nucléotides dont 184 sites phylogénétiquement informatifs.

Les arbres phylogénétiques pour le *cyt b* et les données combinées (*Figures 1-4*) ont été réalisés par la méthode ML et la méthode BA. Les modèles d'évolution utilisés en ML et BA pour les arbres représentant les données du *cyt b* en Corse et en Europe sont les modèles GTR+I+G. En ML, pour les données combinées du *cyt b* et de la CR, le modèle choisi est le modèle GTR alors qu'en BA, les modèles GTR+I+G et HKY+I+G sont respectivement utilisés pour le *cyt b* et la CR.

La monophylie de *Salmo trutta* est hautement supportée pour l'arbre des données du *cyt b* en Corse (BP=100%, PP=1), pour l'arbre du *cyt b* à l'échelle européenne (BP=100%, PP=0,89), pour l'arbre des données combinées en Corse (BP=100%, PP=0,99) et pour l'arbre des données combinées en Europe (BP=100%, PP=0,97). Pour les arbres obtenus avec les données corses combinées et non combinées, on observe quatre clades distincts (*Figures 1 et 3*). En replaçant les données dans un contexte plus général, on identifie, tout comme dans les études antérieures sur la CR, quatre lignées évolutives de truites communes en Europe : AD, ME, MA et AT. La lignée AT est non supportée (BP<50% ; PP<0,7) pour les données du *cyt b* européennes, moyennement supportée pour les données du *cyt b* corses et les données combinées européennes (respectivement, BP=68% ; PP=0,77 et BP=63% ; PP=0,73) et fortement supportée (BP=98% ; PP=0,98) pour les données combinées corses. La lignée MA est moyennement supporté en ML et hautement supporté en BA (BP=73% ; PP=0,97) pour les données du *cyt b* européennes mais hautement supportée pour les données du *cyt b* corses, les données du *cyt b* européennes et les données combinées corses (respectivement, BP=91% ; PP=0,98, BP=95% ; PP=0,99 et BP=91% ; PP=0,96). La lignée ME est faiblement supportée en ML et non supportée en BA (BP=52% ; PP<0,7) pour les données combinées européennes, moyennement supportée (BP=89% ; PP=0,89) pour les données du *cyt b* européennes et hautement supportée en ML et moyennement en BA pour les données du *cyt b* ou combinées corses (respectivement, BP=89% ; PP=0,86 et BP=81% ; PP=0,75). La lignée AD est non supportée (BP<50% ; PP<0,7) pour les données combinées corses, non supportée en ML et moyennement supportée en BA pour

les données du *cyt b* européennes et pour les données combinées européennes (BP<50% ; PP=0,85 et BP<50% ; PP=0,85 respectivement), ainsi que non supportée en ML et hautement supportée en BA (BP<50% ; PP=0,91) pour les données du *cyt b* corses. En outre, quelque soit l'arbre phylogénétique étudié (Figures 1-4), les relations de parentés entre les lignées sont incertaines.

Les arbres phylogénétiques représentant l'ensemble des individus échantillonnés pour les données combinées et non combinées (Figures 2 et 4), montrent qu'il n'y a pas de répartition localisée des individus corses au sein des lignées ME, MA et AT. Contrairement aux études précédentes sur la CR seule (Tougard, 2013), les arbres ne montrent pas une répartition basale des truites de Corse dans la lignée adriatique mais il apparaît une séparation évidente entre les truites adriatiques corses et continentales qui sera détaillée dans le paragraphe "structure des populations". On remarque également que les arbres représentant les données du *cyt b* seul et les arbres représentant les données combinées du *cyt b* et de la CR ont une topologie similaire, ce qui indique probablement que les données du *cyt b* sont davantage structurantes et discriminantes que les données de la CR.

Diversité génétique et démographie : Les divergences génétiques entre les lignées et au sein des lignées en Corse et en Europe ont été calculées (Tableaux 2 et 3) et sont plutôt homogènes et faibles. Chez *Salmo trutta*, les divergences génétiques sont comprises entre $0,0012 \pm 0,0060$ (ME) et $0,0029 \pm 0,0007$ (AT) en Corse ainsi que $0,0012 \pm 0,0040$ (ME) et $0,0029 \pm 0,0008$ (AT) en Europe. Les divergences génétiques intralignées sont identiques pour les lignées AT et ME en Corse et en Europe mais plus forte en Europe pour les lignées AD et MA (Dans les études précédentes sur la CR en Corse, les divergences génétiques étaient plus faibles, comprises entre $0,0003 \pm 0,0000$ (MA) et $0,0017 \pm 0,0007$ (ME)).

Du côté des divergences interlignées en Corse (Tableau 2), la divergence génétique est comprise entre $0,010 \pm 0,002$ (ME-MA) et $0,013 \pm 0,003$ (MA-AT). L'étude de la CR en Corse, avait montré des divergences légèrement plus faibles, la divergence était comprise entre $0,0050 \pm 0,0023$ (ME-MA) et $0,0071 \pm 0,0024$ (MA-AT). En Europe (Tableau 3), la divergence génétique est comprise entre $0,009 \pm 0,002$ (ME-AD) et $0,012 \pm 0,003$ (AT-ME et AT-MA). Pour les données du *cyt b*, les divergences génétiques sont très similaires en Corse et en Europe.

En Corse et en Europe, concernant les indices de diversité (Tableaux 4 et 5), l'indice π est compris entre $0,00124 \pm 0,00026$ (ME) et $0,00277 \pm 0,00082$ (AT) en Corse et $0,00123 \pm 0,00021$ (ME) et $0,00286 \pm 0,00043$ (AT) en Europe. L'indice h_d est compris entre $0,242 \pm 0,135$ (MA) et $0,789 \pm 0,03$ (AD) en Corse et $0,519 \pm 0,105$ (MA) et $0,887 \pm 0,012$ (AD) en Europe. La différence k est comprise entre 1,412 (ME) et 3,159 (AT) en Corse et 1,401 (ME) et 3,258 (AT) en Europe. On obtient des valeurs similaires et plutôt basses quelque soit les lignées, et homogènes entre la Corse et l'Europe. Dans l'étude sur la CR, tous les indices présentent les valeurs les plus faibles pour la lignée AT ($\pi = 0,0003$; $h_d = 0,077 \pm 0,070$; $k = 0,077$) et les valeurs les plus fortes pour la lignée ME ($\pi = 0,0016$; $h_d = 0,739$; $k = 1,532$). D'une manière générale, pour les données du *cyt b* uniquement, les lignées ME et MA semblent être en Europe et en Corse moins diversifiées que les lignées AT et AD. *A contrario*, les études sur la CR en Corse suggéreraient que la lignée ME soit la plus diversifiée (et la lignée AT la moins diversifiée).

Trois tests de neutralité et des distributions de disparités ont été réalisés (*Tableaux 6 et 7, Figures 5 et 6*) pour déterminer l'histoire démographique des lignées. Ici, les courbes de disparités sont difficiles à interpréter : en Europe et en Corse, les lignées AD et ME semblent être en expansion (distribution unimodale), alors que l'interprétation pour les lignées AT et MA se montre plus complexe. Les tests de neutralité ainsi que les statistiques liées aux distributions de disparités, ont donc permis de mettre en évidence les modèles démographiques des lignées de façon plus claire en considérant qu'une population est en expansion si au moins trois résultats significatifs sur cinq vont dans le même sens. Ainsi, les tests montrent qu'en Europe, les populations sont toutes en expansion, alors qu'en Corse, elles semblent toutes stables. Les différences entre les tests obtenues entre les distributions de disparités et les tests réalisés peuvent être expliquées par leurs différences de pouvoir statistique pour détecter la croissance des populations, ou du fait de la faible taille des échantillons.

Structure des populations : Le réseau d'haplotypes des truites en Europe a été construit à partir du jeu de données combinées (*Figure 7*). Il a permis d'identifier 48 haplotypes différents chez les individus échantillonnés en Europe et 25 en Corse dont 21 endémiques de l'île (*Annexe 1*). Il confirme que les populations se structurent en quatre grandes lignées : AD, ME, AT et MA. Les réseaux d'haplotypes avaient permis, lors de l'étude précédente sur la CR (Tougaard, 2013), d'estimer les origines des différentes lignées : la Corse pour la lignée AD, l'ouest de la Slovénie pour la lignée MA, le sud-ouest de la France ou le nord-est de l'Italie pour la lignée ME et le sud de la France pour la lignée AT. Pour le *cyt b*, l'haplotype ancestral (haplotype majoritaire situé au centre du réseau) pour la lignée AT (AT_2) est représenté en France, en Espagne, en Italie, en Corse et dans les Balkans. L'haplotype ancestral de la lignée ME (ME_2) est représenté en Corse, en Espagne et en France. Pour la lignée MA, peu d'haplotypes (4) ont été identifiés mais l'haplotype majoritairement représenté (MA_3) ne comporte que des individus corses (Prunelli, Golu et Fiume Grossu). Contrairement aux études sur la CR, il est donc difficile ici de se prononcer sur l'origine continentale des différentes lignées ME, MA et AT. Au niveau des haplotypes AD, une configuration particulière s'observe à nouveau. En effet, on remarque deux pools d'haplotypes bien distincts. Un premier est constitué d'un haplotype ancestral (AD_5) autour duquel gravitent des haplotypes adriatiques dérivés. Ces haplotypes ne sont présents qu'en Corse. Le second possède un haplotype ancestral (AD_10) uniquement représenté dans les Balkans et autour duquel se ramifient des haplotypes dérivés continentaux. Tout comme dans le cas des arbres phylogénétiques, la lignée AD possède un pool corse et un pool continental bien distincts génétiquement au niveau du *cyt b*.

Datation moléculaire : Les analyses de datation moléculaire réalisées à partir d'une horloge moléculaire locale ont permis d'estimer la date d'arrivée des lignées en Corse. Les taux de substitutions obtenues pour *Salmo trutta* et *Salmo salar* sont respectivement de 0,113734 substitutions/Ma et de 0,293154 substitutions/Ma. Les âges d'apparition des lignées en Corse sont donc antérieurs à 0,937 Ma pour la lignée AD, antérieurs à 0,843 Ma pour la lignée ME, antérieurs à 0,448 Ma pour la lignée MA et antérieurs à 0,324 Ma pour la lignée AT.

Discussion :

Concernant l'histoire des lignées en Corse, Gauthier et Berrebi (2007) proposent que la lignée AD est arrivée la première en Corse, à une époque ancienne indéterminée, et a peuplé l'ensemble de l'île. Par la suite, la dernière déglaciation (-10 000 ans) aurait entraîné des cassures de terrains par compensation orogénique créant dans les cours d'eau des cascades séparant le centre de la Corse de son pourtour. La formation de ces obstacles aurait ensuite empêché la remontée des cours d'eau par la lignée ME arrivée par la suite. Ainsi l'arrivée de la lignée ME serait postérieure au dernier maximum glaciaire. La lignée AT serait arrivée très récemment, via l'introduction de truites de pisciculture en Corse. L'étude de Tougard (2013) avait également fourni des hypothèses sur l'origine spatiale des différentes lignées grâce aux données de la CR. L'origine des lignées ME, MA et AT est supposée continentale alors que l'origine de la lignée AD est corse.

Cependant, contrairement à l'étude précédente sur la CR seule (Tougaard, 2013), l'étude de la CR et du *cyt b* combinés indique une origine continentale (Balkans) de la lignée AD mais ne fournit pas d'informations sur les origines spatiales des lignées ME, MA et AT, car pour ces lignées, dans les arbres phylogénétiques et le réseau d'haplotypes, les individus et les haplotypes corses sont entremêlés avec ceux du continent. Conjointement, les deux études semblent indiquer que les quatre lignées trouvées en Corse seraient d'origines continentales.

Les datations moléculaires effectuées à partir des données de la CR et du *cyt b* indiquent que les différentes lignées de truites sont arrivées en Corse durant le Pléistocène, période marquée par de nombreuses périodes glaciaires et interglaciaires (Tzedakis et al., 2006). Or, lors des événements de froids intenses, le stress occasionné par le froid peut modifier les comportements migratoires des organismes (Provan et Bennett, 2008 ; Taberlet et al., 1998 ; Hewitt, 2011), les incitant à descendre vers les zones les plus au sud, aux températures moins extrêmes (Stewart et al., 2010). Cela pourrait être le cas des populations de truites continentales qui bien que sédentaires, ont pu adopter un comportement migratoire durant les glaciations (Bendall et al., 2012; Boel et al., 2014) et migrer ainsi vers la Corse. Si les datations sont exactes, l'île aurait donc pu être un refuge glaciaire pour les truites du continent.

D'après les reconstructions géologiques, au Pléistocène la Corse était déjà bien séparée du continent européen (Garcia Castellanos et al., 2009), le passage du continent à la Corse par voie terrestre semblait donc impossible. Gauthier et Berrebi (2007) avaient proposé un passage par le canal de Corse et les îles toscanes. En effet, l'estimation des topologies anciennes par les géologues s'appuie sur l'étude des topologies récentes. Or depuis la fin du Pléistocène, des accidents tectoniques ont pu modifier le canal de Corse et en changer la bathymétrie, induisant donc en erreur les estimations des géologues. Si les topologies reconstruites décrivent un canal de Corse large durant le Pléistocène, nous pouvons envisager qu'elles soient potentiellement biaisées et donc qu'un passage par le canal de Corse ait pu être possible. Le passage par voie marine peut être également envisageable. Les dates d'arrivée des lignées en Corse étant données sous forme d'intervalles semi-ouverts et le Pléistocène étant marqué par de nombreuses phases glaciaires (>40), elles ne permettent pas de définir clairement si les lignées sont arrivées en Corse en période glaciaire ou interglaciaire (Tzedakis et al., 2006). Cependant, l'hypothèse d'une colonisation de

l'île par traversée de la Méditerranée durant une régression marine lors d'une période glaciaire semble plus probable car elle correspond à l'hypothèse évoquée précédemment d'un changement comportemental de populations actuellement fortement sédentaires.

Les dates d'origine des lignées en Corse estimées dans cette étude suggèrent que la lignée AD aurait colonisé la première l'île (avant -0,937 Ma). Cet isolement ancien de truites adriatiques originaires du continent en Corse pourrait expliquer la forte différenciation observée entre les truites adriatiques continentales et les truites adriatiques corses (Figures 2, 4 et 7). La lignée AD aurait ainsi colonisé l'ensemble de l'île notamment les parties hautes des cours d'eau. L'absence de colonisation de ces dernières par la lignée ME arrivée plus tard (avant -0,843 Ma) peut être expliquée par une barrière physique. Gauthier et Berrebi (2007) proposent que les cassures de terrains provoquées par la dernière déglaciation (-10 000 ans) ont empêché la colonisation des parties hautes des cours d'eau par la lignée ME, arrivée ensuite. Néanmoins, la date d'arrivée de la lignée ME (avant -0,843 Ma) est très antérieure au dernier maximum glaciaire. Ainsi, il est possible que les cassures de terrains ayant empêché la remontée des cours d'eau de la lignée ME soient plus anciennes. La formation de ces obstacles se serait alors produite après la colonisation des sommets corses par la lignée AD et avant la colonisation de l'île par la lignée ME, c'est-à-dire au milieu du Pléistocène, via les nombreuses déglaciations qui se sont produites à cette période. La lignée MA (avant -0,448 Ma) est arrivée en Corse après la lignée ME, suivie par la lignée AT (avant -0,324 Ma). Cette étude conteste la seule origine anthropique récente de la lignée AT en Corse. Il est possible que l'arrivée de la lignée AT soit donc aussi naturelle, comme observé en Sicile (Schöffmann et al., 2007).

Au moment des cassures de terrains, les truites adriatiques peuplant la partie haute des cours d'eau se seraient retrouvées piégées au centre de la Corse. Or au Pléistocène et jusqu'à la fin du dernier maximum glaciaire, les sommets du centre de la Corse étaient recouverts de glace (Conchon, 1998). Ces truites adriatiques peuplant le sommet des cours d'eau auraient donc eu un espace étroit entre ces obstacles infranchissables et ces glaciers, d'autant plus limité durant les glaciations. De plus, pour ces populations piégées en altitude, le Pléistocène aurait été sujet à de nombreuses phases de contraction et d'expansion (Provan et Benett, 2008). En effet, les glaciations peuvent réduire les taux de survie des organismes et agir comme des goulots d'étranglement sur les diversités (Maggs et al., 2008) alors que les phases interglaciaires où le niveau de l'eau remonte peuvent favoriser les expansions des populations et la diversification (Marko et al., 2010).

Après la dernière déglaciation, le réchauffement pourrait avoir permis une expansion et une diversification des populations. Actuellement pour les données du *cyt b*, la démographie de la Corse semblerait être stable alors que la démographie en Europe serait en expansion. La population de truite en Corse est donc stable et structurée. Cette différence de démographie entre le continent et l'île peut hypothétiquement s'expliquer par l'isolation de l'île par rapport au continent.

Les scénarios exposés ci-dessus reposent en partie sur l'utilisation des datations moléculaires de la présente étude. Néanmoins, le modèle d'évolution utilisé pour la datation est le modèle HKY (utilisé pour l'analyse limitée aux haplotypes endémiques) ne différencie pas tous les types de transition ou de transversion. De plus, l'horloge moléculaire appliquée est une horloge moléculaire locale qui assume que

le taux de substitutions est le même pour des groupes étroitement liés mais différent dans d'autres parties de l'arbre (Yoder et Yang, 2000). Dans la mesure où les transitions et les transversions peuvent survenir selon des probabilités différentes et que les lignées ont pu évoluer avec des taux d'évolution différents, ces simplifications peuvent avoir biaisé la datation. De nombreuses études ont également montré l'importance du point de calibration dans la chronologie des événements évolutifs (Sauquet et al., 2012). Le point de calibration utilisé ici (apparition de *Salmo trutta* : 11,6 Ma : Crête Lafrenière et al., 2012) peut être imprécis ou biaisé, il serait probablement utile de rajouter d'autres points de calibration provenant du registre fossile. D'un autre côté, les dates d'arrivée des lignées en Corse sont données sous forme d'intervalles ouverts, qui de ce fait ne permettent pas une estimation réelle des dates d'arrivée.

Outre la méthode, les données utilisées sont également une source d'erreur pour la datation. L'augmentation du jeu de données notamment continental pourrait nous amener à reconsidérer le nombre d'haplotypes définis comme endémiques de Corse. De plus, au niveau des haplotypes de la lignée AT, aucune référence n'a permis de distinguer les haplotypes atlantiques de pisciculture et les haplotypes atlantiques méditerranéens provenant de GenBank. Or les truites de pisciculture ne sont pas arrivées naturellement mais par voie anthropique. Ainsi des haplotypes atlantiques de pisciculture ont pu se retrouver comme haplotypes endémiques de Corse et ainsi biaiser la datation.

Conclusion :

Dans un contexte général, cette étude a permis de confirmer la plupart des informations déjà acquises dans les études précédentes (Gauthier et Berrebi, 2007 ; Tougard, 2013), notamment la présence de quatre lignées de truites en Corse. Elle a également permis de dégager de nouvelles informations en particulier au niveau de l'histoire des lignées. Il semblerait en effet, que les quatre lignées soient d'origine continentale et qu'elles soient arrivées en Corse durant le Pléistocène, marqué par l'alternance entre périodes glaciaires et interglaciaires. Néanmoins, l'étude n'a pas abouti à une phylogénie suffisamment résolue pour qu'elle fournisse des informations indubitables sur les relations entre les lignées, leurs origines et leur date d'arrivée. Ainsi, une augmentation du jeu de données et un perfectionnement des méthodes employées pour les arbres phylogénétiques (augmentation des bootstraps, du nombre de chaînes de Markov et du nombre de générations,...) et pour les datations (modification du modèle d'évolution et de l'horloge moléculaire utilisés, calibration,...) permettrait probablement d'améliorer les résultats obtenus.

Figures et Tableaux :

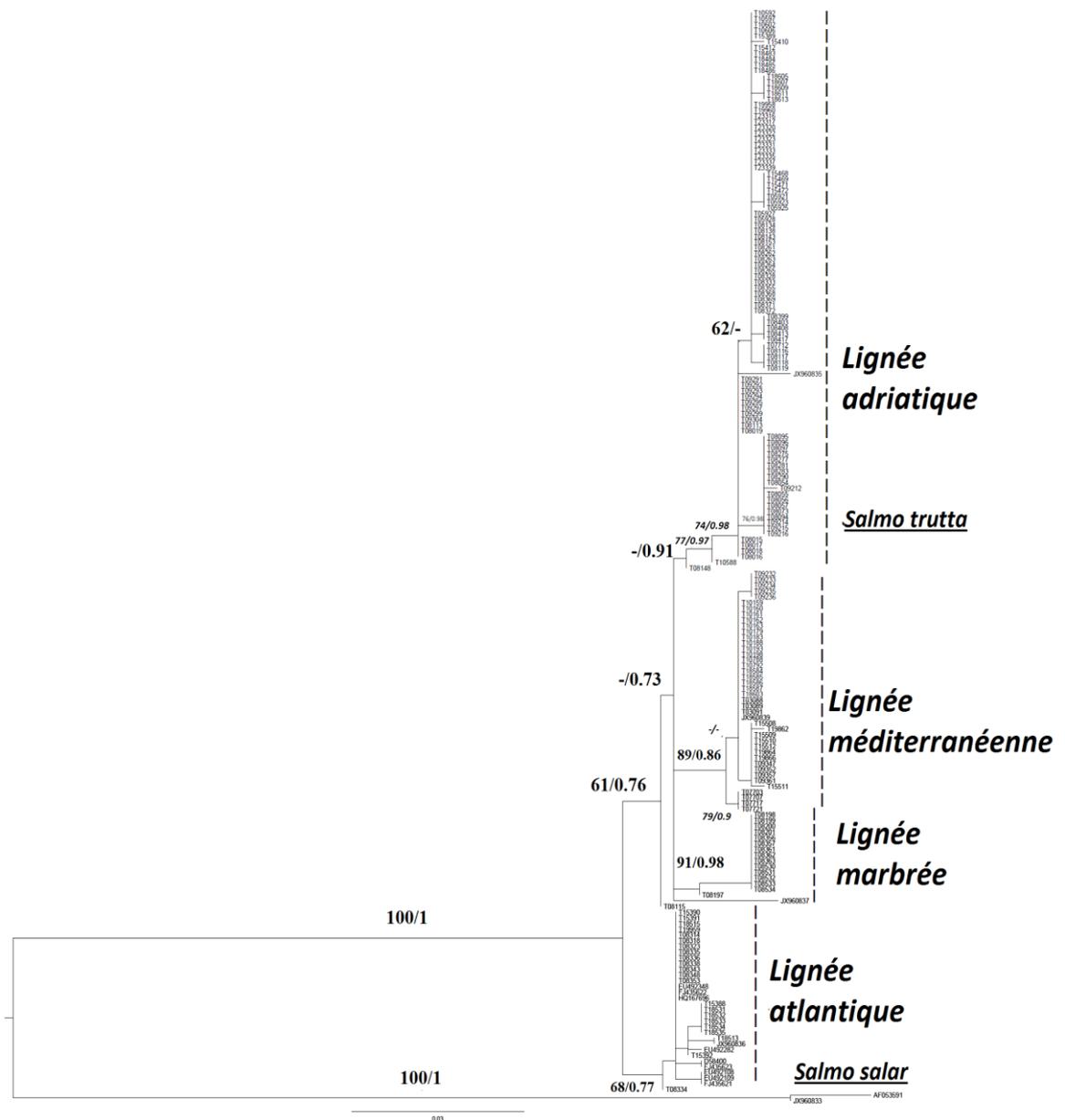


Figure 1 Arbre obtenu en maximum de vraisemblance pour les séquences du *cyt b* des individus corses. Les valeurs aux nœuds correspondent aux pourcentages de bootstraps en maximum de vraisemblance (>50%) et aux probabilités postérieures bayésiennes (>0,7). Les lignées sont indiquées à droite.

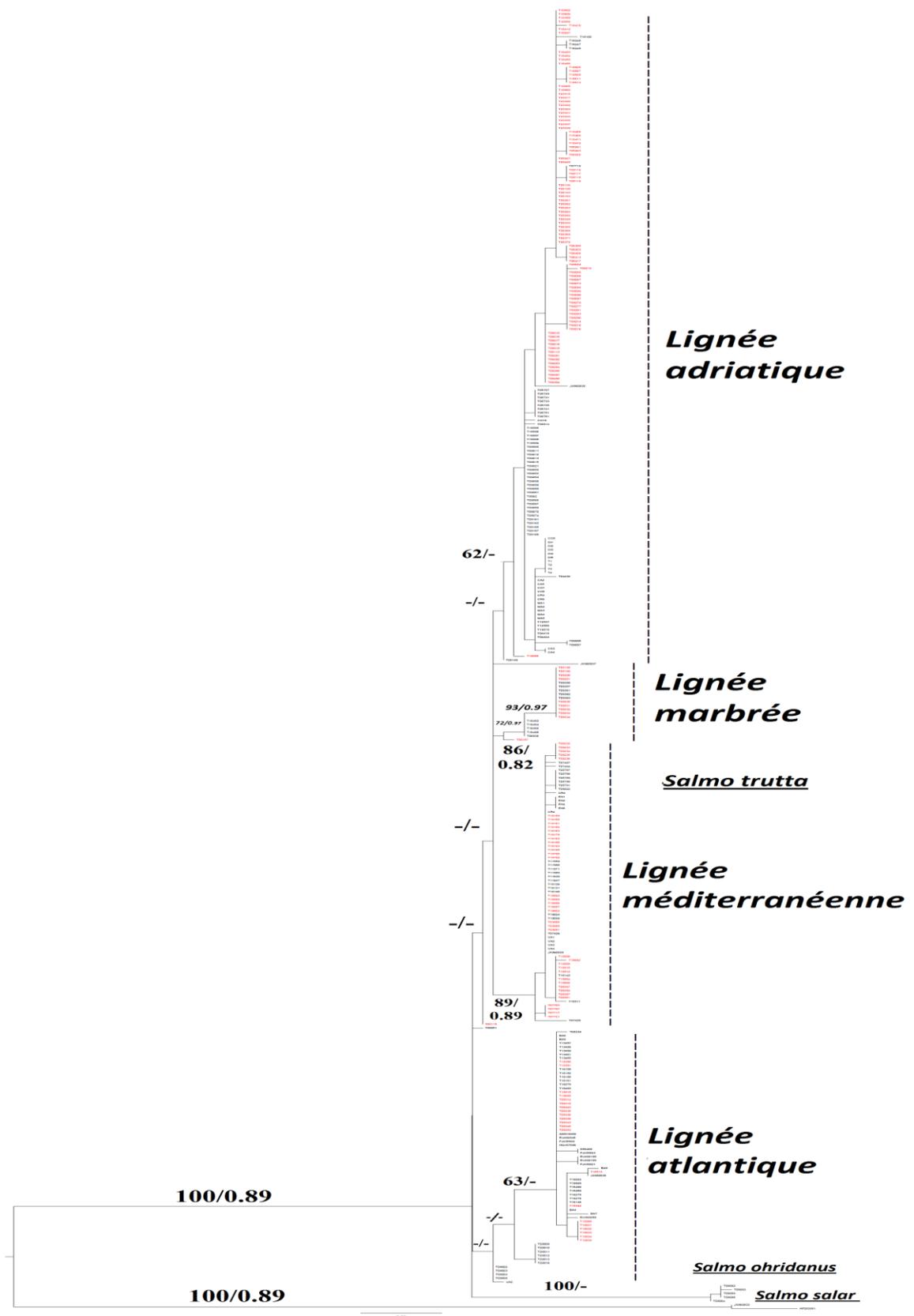


Figure 2 Arbre obtenu en maximum de vraisemblance pour les séquences de cyt b pour tous les individus d'Europe analysés. Les valeurs aux nœuds correspondent aux pourcentages de bootstraps en maximum de vraisemblance (>50%) et aux probabilités postérieures bayésiennes (>0,7). Les lignées sont indiquées à

droite. Malgré sa faible lisibilité, cet arbre met bien en évidence les lignées et la position des séquences corses (en rouge) par rapport aux séquences continentales (en noir).

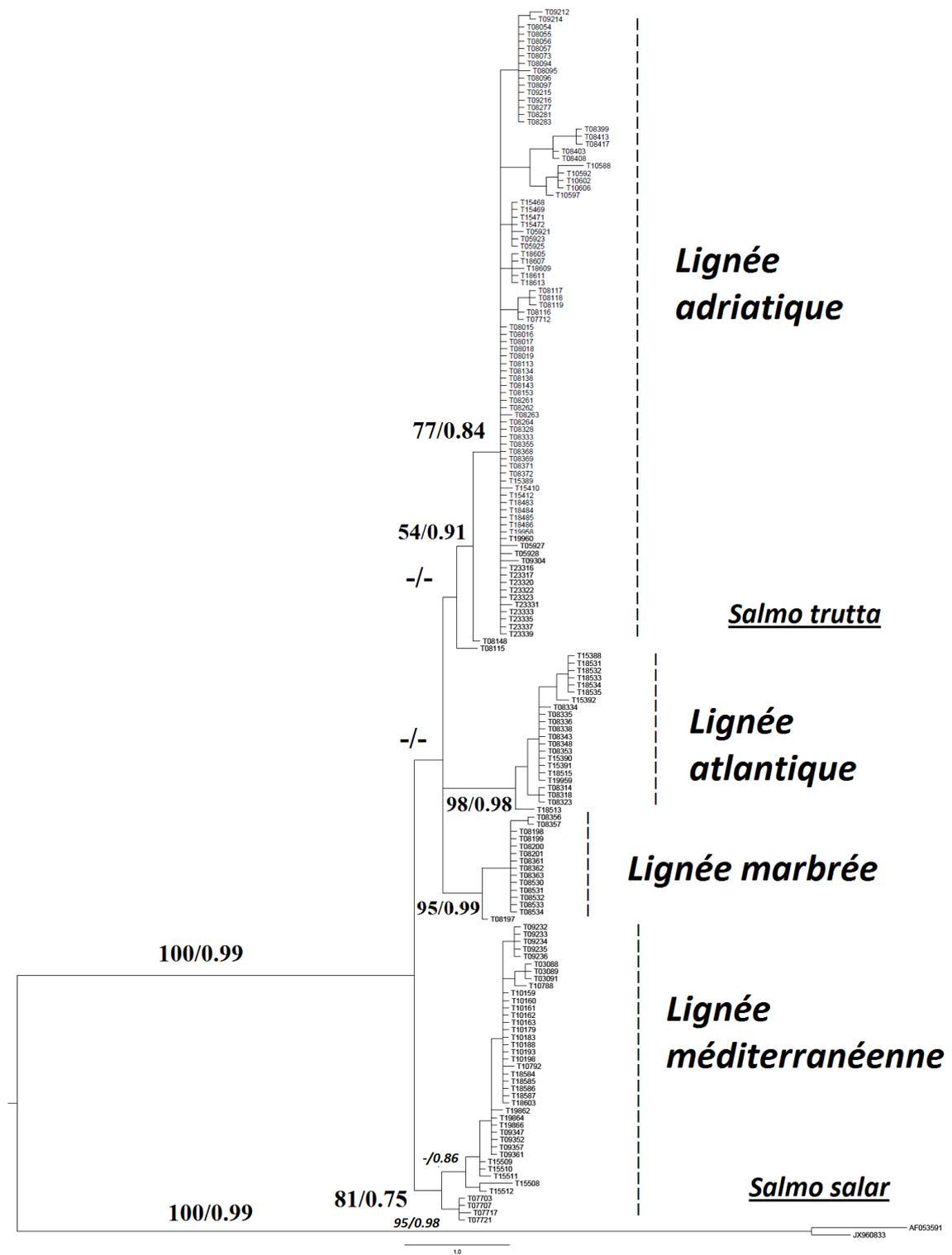


Figure 3 Arbre obtenu en approche bayésienne pour les séquences combinées de la CR et du *cyt b* en Corse. Les valeurs aux nœuds correspondent aux pourcentages de bootstraps en maximum de vraisemblance (>50%) et aux probabilités postérieures bayésiennes (>0,7). Les lignées sont indiquées à droite.

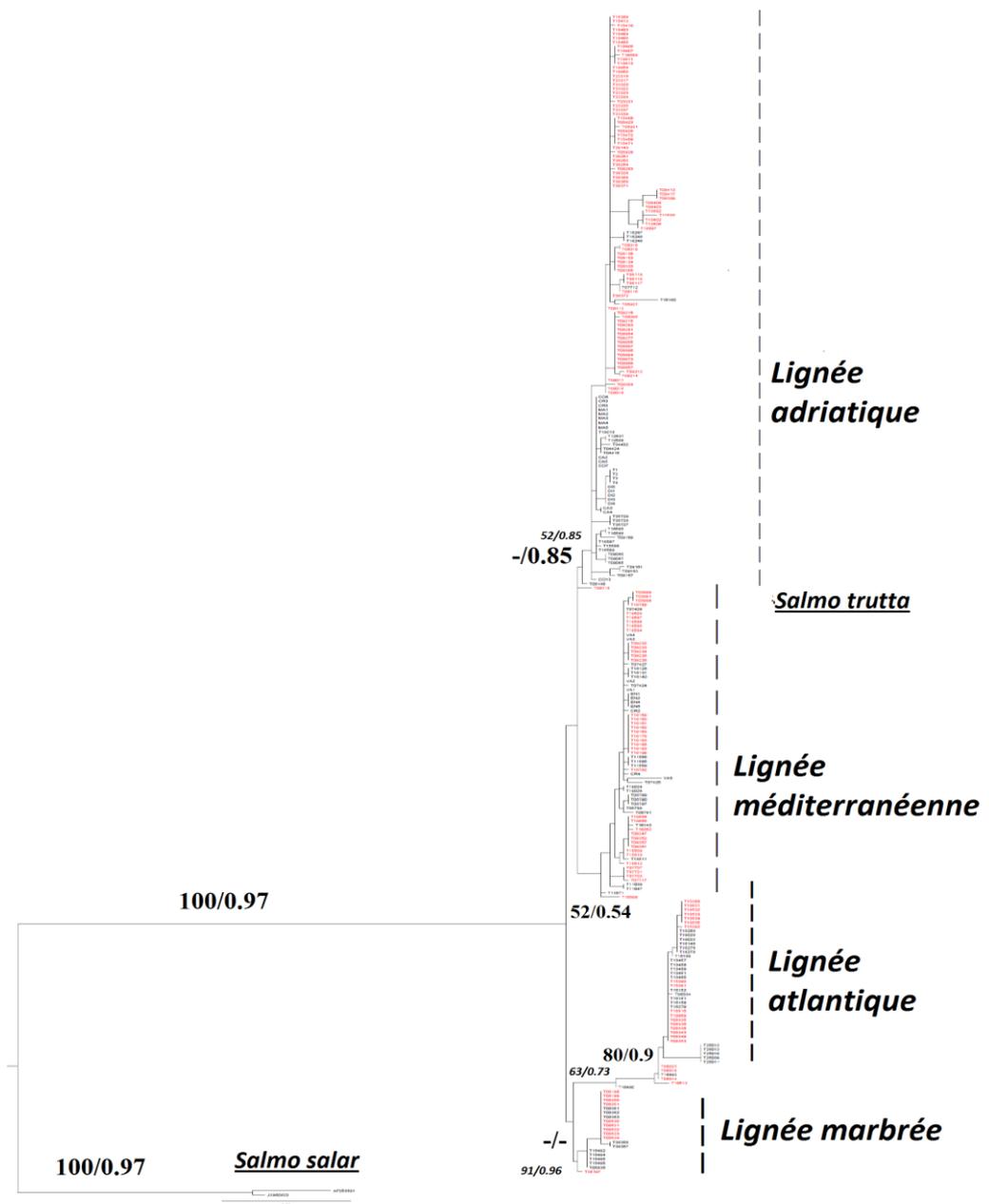


Figure 4 Arbre obtenu en maximum de vraisemblance à partir des données combinées de la CR et du cyt b en Europe. Les valeurs aux nœuds correspondent aux pourcentages de bootstraps en maximum de vraisemblance (>50%) et aux probabilités postérieures bayésiennes (>0,7). Les lignées sont indiquées à droite. Certes peu lisible dans le détail, cet arbre met néanmoins bien en évidence les lignées et la position des séquences corses (en rouge) par rapport aux séquences continentales (en noir).

Tableau 1 Séquences des amorces utilisées

| Amorces | Séquences | Références |
|-------------------|--|--------------------------------------|
| Salmo CBF | 5' – CAT AAT TCC TGC CCG GAC TCT AAC C- 3' | Crête Lafrenière et al., 2012 |
| Salmo CB1F | 5' – CGA CAA CGC CAC CCT AAC ACG A -3' | Cette étude |

| | | |
|-------------------|--|--------------------------------------|
| Salmo CB1R | 5' – GGT TAT TAG AGC CGG TTT CAT G – 3' | Cette étude |
| Salmo CBR | 5' – TTT AAC CTC CGA TCT CCG CAT TAC A -3' | Crête Lafrenière et al., 2012 |

Tableau 2 Divergences génétiques (*d*) au sein et entre les lignées de truites en Corse pour le *cyt b* (\pm erreurs standards)

| Lignées | <i>d</i> | AT | AD | ME | MA | <i>S. salar</i> |
|-----------------|------------------------|-------|----------------|----------------|----------------|-----------------|
| AT | 0,0029 (\pm 0,0007) | - | (\pm 0,003) | (\pm 0,003) | (\pm 0,003) | (\pm 0,009) |
| AD | 0,0016 (\pm 0,0005) | 0,012 | - | (\pm 0,003) | (\pm 0,003) | (\pm 0,009) |
| ME | 0,0012 (\pm 0,0006) | 0,012 | 0,010 | - | (\pm 0,002) | (\pm 0,009) |
| MA | 0,0019 (\pm 0,0005) | 0,013 | 0,011 | 0,010 | - | (\pm 0,010) |
| <i>S. salar</i> | 0,0053 (\pm 0,002) | 0,066 | 0,068 | 0,067 | 0,073 | - |

Tableau 3 Divergences génétiques (*d*) au sein et entre les lignées de truites en Europe pour le *cyt b* (\pm erreurs standards)

| Lignées | <i>d</i> | AT | AD | ME | MA | <i>S. salar</i> |
|-----------------|------------------------|-------|----------------|----------------|----------------|-----------------|
| AT | 0,0029 (\pm 0,0008) | - | (\pm 0,003) | (\pm 0,003) | (\pm 0,003) | (\pm 0,007) |
| AD | 0,0025 (\pm 0,0009) | 0,011 | - | (\pm 0,002) | (\pm 0,003) | (\pm 0,007) |
| ME | 0,0012 (\pm 0,0004) | 0,012 | 0,009 | - | (\pm 0,003) | (\pm 0,007) |
| MA | 0,0023 (\pm 0,0007) | 0,012 | 0,009 | 0,009 | - | (\pm 0,007) |
| <i>S. salar</i> | 0,0053 (\pm 0,002) | 0,066 | 0,068 | 0,067 | 0,072 | - |

Tableau 4 Indices de diversité génétique pour chaque lignée en Corse pour les données du *cyt b* (*N* : nombre de séquences, *h* : nombre d'haplotypes, π : diversité nucléotidique, *hd* : diversité haplotypique, *k* : nombre moyen de différences en comparant deux à deux les séquences) (\pm erreurs standards)

| | <i>N</i> | <i>h</i> | π | <i>hd</i> | <i>k</i> |
|----|----------|----------|---------------------------|----------------------|----------|
| AT | 34 | 9 | 0,00277 (\pm 0,00082) | 0,749 (\pm 0,067) | 3,159 |
| AD | 95 | 12 | 0,00156 (\pm 0,000014) | 0,789 (\pm 0,030) | 1,781 |
| ME | 40 | 7 | 0,00124 (\pm 0,00026) | 0,726 (\pm 0,051) | 1,412 |
| MA | 16 | 3 | 0,00192 (\pm 0,00127) | 0,242 (\pm 0,135) | 2,183 |

Tableau 5 Indices de diversité génétique pour chaque lignée en Europe pour les données du *cyt b* (*N* : nombre de séquences, *h* : nombre d'haplotypes, π : diversité nucléotidique, *hd* : diversité haplotypique, *k* : nombre moyen de différences en comparant deux à deux les séquences) (\pm erreurs standards)

| | <i>N</i> | <i>h</i> | π | <i>hd</i> | <i>k</i> |
|----|----------|----------|--------------------------|----------------------|----------|
| AT | 68 | 14 | 0,00286 (\pm 0,00043) | 0,775 (\pm 0,045) | 3,258 |
| AD | 170 | 22 | 0,00249 (\pm 0,00010) | 0,887 (\pm 0,012) | 2,842 |
| ME | 76 | 14 | 0,00123 (\pm 0,00021) | 0,721 (\pm 0,050) | 1,401 |

| | | | | | |
|----|----|---|---------------------------|-----------------------|-------|
| MA | 21 | 4 | 0,00226 ($\pm 0,00089$) | 0,519 ($\pm 0,105$) | 2,581 |
|----|----|---|---------------------------|-----------------------|-------|

Tableau 6 Indices démographiques pour chaque lignée en Corse pour les données du cytb (Fs : statistique de Fu, D : D de Tajima, R₂ : R₂ de Ramos Onsins et Rozas, SSD : Somme des écarts au carré, r : indice de raggedness) (\pm erreurs standards)

Significativité des tests: *** p.value<0,001 ** p.value<0,01 * p.value<0,05

Valeurs en gras : Valeurs indiquant une expansion

| | Test de Neutralité | | | Distribution de disparités | | |
|----|--------------------|------------------|----------------|----------------------------|----------------|--------|
| | Fs | D | R ₂ | SSD | r | Modèle |
| AT | -0,224 | -1,08379 | 0,0821 | 0,05817 | 0,19240* | Stable |
| AD | -2,858 | -1,09116 | 0,0586 | 0,09566*** | 0,03893 | Stable |
| ME | -1,122 | -1,19177 | 0,0864 | 0,00227 | 0,05844 | Stable |
| MA | 3,109 | -1,87639* | 0,1761 | 0,05222* | 0,62965 | Stable |

Tableau 7 Indices démographiques pour chaque lignée en Europe pour les données du cytb (Fs : statistique de Fu et Li, D : D de Tajima, R₂ : R de Ramos onsins et Rozas, SSD : somme des écarts au carré, r : indice de raggedness) (\pm erreurs standards)

Significativité des tests: *** p.value<0,001 ** p.value<0,01 * p.value<0,05

Valeurs en gras : Valeurs indiquant une expansion

| | Test de Neutralité | | | Distribution de disparités | | |
|----|--------------------|--------------------|----------------|----------------------------|----------------|-----------|
| | Fs | D | R ₂ | SSD | r | Modèle |
| AT | -1,885 | -1,25688** | 0,095 | 0,01580 | 0,03948 | Expansion |
| AD | -6,207* | -1,06183 | 0,0252 | 0,00785 | 0,02519 | Expansion |
| ME | -7,124*** | -1,90264*** | 0,0756* | 0,00455 | 0,07559 | Expansion |
| MA | 2,724 | -1,2177** | 0,4575 | 0,16440 | 0,45753 | Expansion |

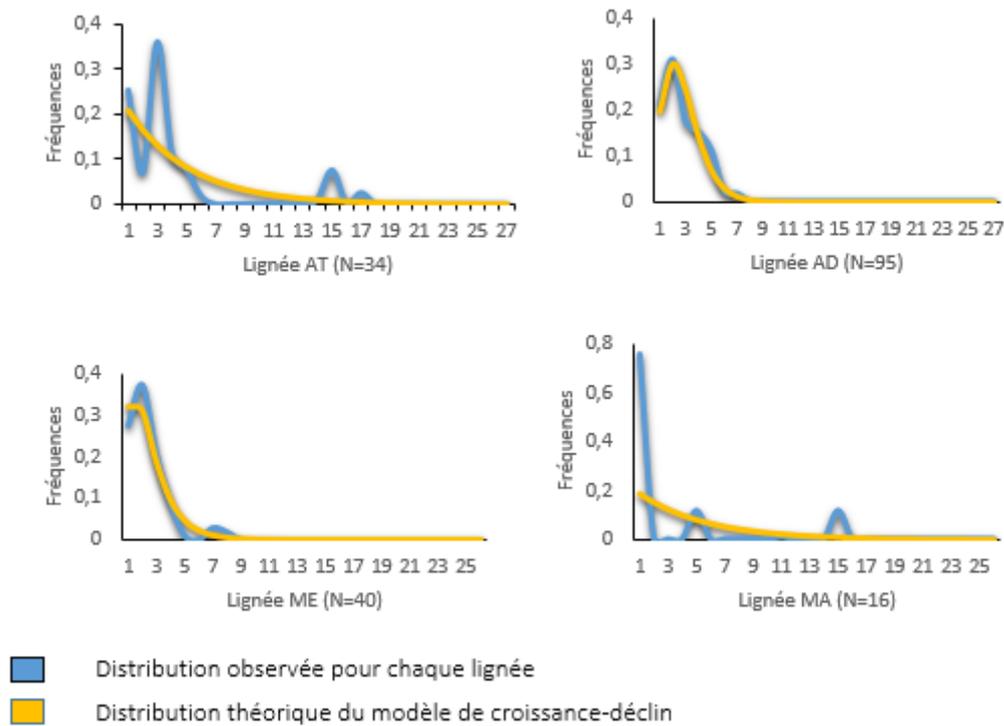


Figure 5 Représentation graphique de l'histoire démographique des différentes lignées de truite en Corse à partir des données du *cyt b* (distribution de disparités des lignées sous un modèle de population en croissance-déclin)

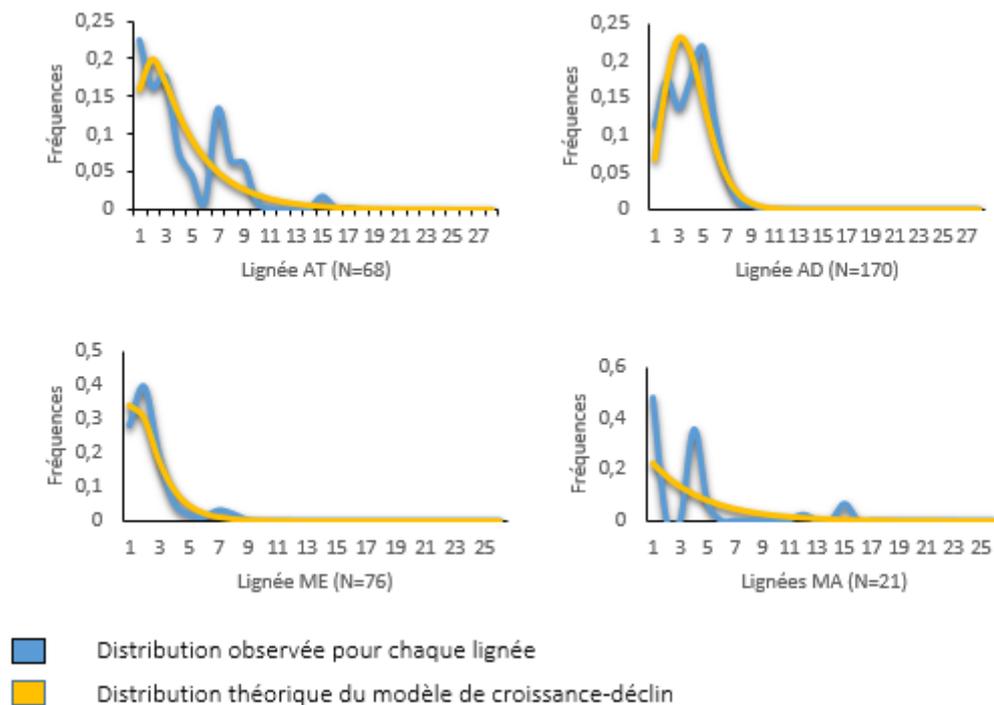


Figure 6 Représentation graphique de l'histoire démographique des différentes lignées de truite en Europe à partir des données du *cyt b* (distribution de disparités des lignées sous un modèle de population en croissance-déclin)

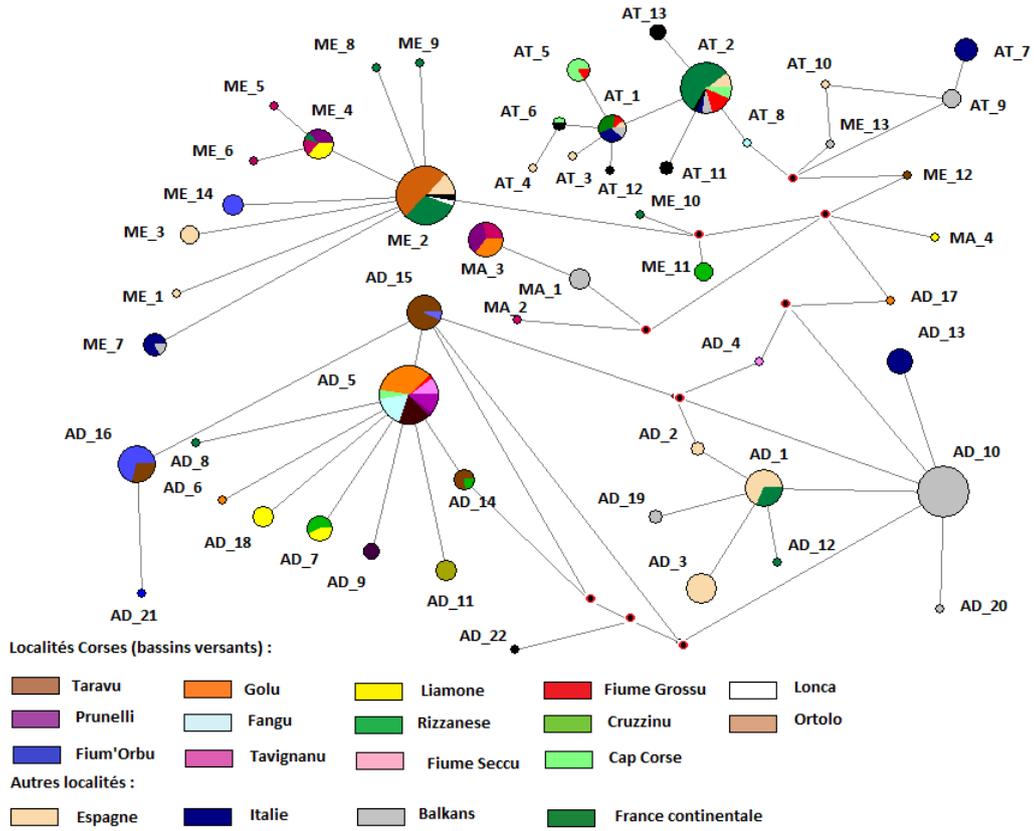


Figure 7 Réseau d'haplotypes de la truite *Salmo trutta* à l'échelle européenne obtenu à partir des données du *cyt b* (voir la correspondance des haplotypes - Annexe 1)

Bibliographie :

Aurelle D et Berrebi P. Microsatellites markers and management of brown trout *Salmo trutta fario* populations in South western France. *Genetics Selection Evolution*. 1998. Volume 30. P 75-90.

Baglinière J.L. La truite commune (*Salmo trutta* L.), son origine, son aire de répartition, ses intérêts économiques et scientifiques. *La truite : biologie et écologie*. INRA, Rennes. 1991. P 11-22.

Bandelt HJ, Forster P, Röhl A. Median-joining network for inferring intraspecific phylogenies. *Molecular Biology and Evolution*. 1999. Volume 16. P 37-48.

Bendall B, Moore A, Maxwell D, Davison P, Edmonds N, Archer D et al. Modelling the migratory behavior of salmonids in relation to environmental and physiological parameters using telemetry data. *Fisheries Management and Ecology*. 2012. Volume 19. P 475-483.

Bernatchez L, Guyomard R et Bonhomme F. DNA sequence variation of the mitochondrial control region among geographically and morphologically remote European brown trout *Salmo trutta* population conservation. *Molecular Ecology*. 1992. Volume 4. P 285-297.

Berrebi P. Etude génétique des truites de Corse. Rapport final. Novembre 1995. Université Montpellier II. 36 p.

Boel M, Aarestrp K, Baktoff H, Larsen T, Sandergaard Madsen S, Malte H et al. The physiological basis of the migration continuum in brown trout (*Salmo trutta*). *Physiological and Biochemical Zoology*. 2014. Volume 87. P 334-345.

Conchon O. Manifestations et chronologie de la déglaciation fini-Würmienne en Corse. *Bulletin de l'Association Française pour l'Etude du Quaternaire*. 1988. Volume 25. P 91-96.

Crête Lafrenière A, Weir L et Bernatchez L. Framing the Salmonidae family phylogenetic portrait: a more complete picture from increased taxon sampling. *PLoS One*. 2012. Volume 7.

Durand Delga M. Corse. Masson. 1978. 297 p.

Delsuc F. Les méthodes probabilistes en phylogénie moléculaire. in : *Phylogénie moléculaire des xénarthres (tatous, fourmiliers et paresseux) : application des méthodes probabilistes à la reconstruction de leur histoire évolutive au sein des mammifères placentaires*. 2002. Thèse non publiée Université Montpellier 2. P 13-56.

Excoffier L et Lischer HEL. Arlequin suite version 3.5: a new series of programs to perform population genetics analyses under Linux and windows. *Molecular Ecology Resources*. 2010. Volume 10. P 564-567.

Faure JP. Etude génétique des populations de truites du département du Rhône. INRA. Rapport final. 2012. 60 p.

Fu YX. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics*. 1997. Volume 147. P 915-925.

Garcia Castellanos D, Estrada F, Jiménez Munt I, Gorini C, Fernández M, Vergés J et De Vicente R. Catastrophic flood of the Mediterranean after the Messinian salinity crisis. *Nature*. 2009. Volume 462. P 778-781.

Gauthier A et Berrebi P. Contribution à la gestion des populations de truites en Corse : la colonisation de l'île par la truite. Guide réalisé dans le cadre du programme LIFE « Conservation de la truite macrostigma en Corse ». 2007. VB Graphics. P 6-11.

Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W et Gascuel O. New algorithm and methods to estimate maximum likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic Biology*. 2010. Volume 59. P 307–321.

Guyomard R et Krieg F. Mise en évidence d'un flux génique entre populations naturelles de truite fario et souche de repeuplement dans deux rivières de Corse. *Bulletin français de la pêche de pisciculture*. INRA. 1986. Volume 303. P 134-140.

Harpending HC. Signature of ancient population growth in a low resolution mitochondrial DNA mismatch distribution. *Human Biology*. 1994. Volume 66. P 591-600.

Hewitt G. The genetic legacy of the Quaternary ice ages. *Nature*. 2000. Volume 405. P 907-913

Launey S, Krieg F, Champigneulle A et Guyomard R. Écotypes sympatriques migrants et sédentaires de truite commune (*Salmo trutta*): différenciation génétique et effet des repeuplements. *Les Actes du BRG*. 2003. Volume 4. P 63-78.

Librado P et Rozas J. DNAsp v5 : a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*. 2009. Volume 25. P 1451-1452.

MacCrimmon H.R. et Gots B.L. World distribution of brown trout, *Salmo trutta*: further observations. *Journal of the Fisheries Research Board of Canada*. 1970. Volume 27. P 811-818.

Maggs C, Catilho R, Foltz D, Henzler C, Taimour Jolly M, Kelly J et al. Evaluating signatures of glacial refugia for north Atlantic benthic marine taxa. *Ecology*. 2008. Volume 89. P 108-122.

Marko P, Hoffman JM, Emme SA, McGovern T, Keever CC et Cox N. The expansion-contraction model of Pleistocene biogeography: rocky shores suffer a sea change? *Molecular Ecology*. 2010. Volume 19. P 146-149.

Nylander JAA. MrModeltest 2.3. Program distributed by the author. Evolutionary Biology Centre. 2008. Uppsala University.

Pakkasmaa S et Pironen J. Morphological differentiation among local trout (*Salmo trutta*) populations. *Biological Journal of the Linnean Society*. 2001. Volume 72. Pages 231–239.

Provan J et Bennett KD. Phylogeographic insights into cryptic glacial refugia. *Trends in Ecology and Evolution*. 2008. Volume 23. P 564-571.

Ramos-Onsins SE et Rozas J. Statistical properties of new neutrality tests against population growth. *Molecular Biology and Evolution*. 2002. Volume 19. P 2092-2100.

Reynaud N. Structuration géographique de la truite commune (*Salmo trutta*) en France basée sur le séquençage de la région de contrôle mitochondriale. *Rapport OSU OREME*. 2011. 46 p.

Rogers AR et Harpending H. Population growth makes waves in the distribution of pairwise genetic differences. *Molecular Biology and Evolution*. 1992. Volume 9. P 552-589.

- Ronquist F et Huelsenbeck JP.** MrBayes3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*. 2003. Volume 19. P 1572–1574.
- Ryman N et Stahl G.** Genetic changes in hatchery stocks of brown trout *Salmo trutta*. *Canadian Journal of Fisheries and Aquatic Sciences*. 1980. Volume 37. P 82-87.
- Sauquet H, Ho SYW, Gandolfo MA, Jordan GJ, Wilf P, Cantrill DJ et al.** Testing the impact of calibration on molecular divergence times using a fossil rich group: the case of *Nothofagus* (Fagales). *Systematics Biology*. 2012. Volume 61. P 289-313.
- Schöffmann J, Susnik S et Snoj A.** Phylogenetic origin of *Salmo trutta* L 1758 from Sicily, based on mitochondrial and nuclear DNA analyses. *Hydrobiologia*. 2007. Volume 575. P 51-55.
- Splendiani A, Ruggeri P, Giovannotti M et Barrucchi VC.** Role of environmental factors in the spread of domestic trout in Mediterranean streams. *Freshwater Biology*. 2013. Volume 58. P 2089-2101.
- Stewart JR, Lister AM, Barnes I et Dalen L.** Refugia revisited: individualistic responses of species in space and time. *Proceedings of the Royal Society*. 2010. Volume 277. P 661-671.
- Taberlet P, Fumagalli L, Wust-Saucy AG et Cossons JF.** Comparative phylogeography and postglacial colonization routes in Europe. *Molecular Ecology*. 1998. Volume 7. P 453-464.
- Tajima F.** Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*. 1989. Volume 123. P 585-595.
- Tamura K, Stecher G, Peterson D, Filipinski A et Kumar S.** MEGA6: Molecular evolutionary genetics analysis version 6.0. 2013. *Molecular Biology and Evolution*. Volume 30. P 2725-2729.
- Tougaard C.** Analyse phylogénétique de 30 populations de truites de Corse à partir d'un marqueur mitochondrial (région de contrôle). 2012. Rapport pour la Fédération de la Corse pour la Pêche et la Protection du Milieu Aquatique.
- Tougaard C.** Détermination de l'origine des lignées de truites de Corse par le biais de l'analyse phylogénétique d'un marqueur mitochondrial (région de contrôle). 2013. Rapport pour la Fédération de la Corse pour la Pêche et la Protection du Milieu Aquatique.
- Tzedakis PC, Hooghiemstra H et Palike H.** The last 1.35 million years at Tenagi Philippon: revised chronostratigraphy and long term vegetation trends. *Quaternary Science Reviews*. 2006. Volume 25. P 3416-3430.
- Walsh PS, Metzger DA et Higuchi R.** Chelex 100 as a medium for simple extraction of DNA for PCR based typing from forensic material. *Biotechniques*. 1991. Volume 10. P 506-513.
- Xu B et Yang Z.** PAMLX : A graphical user interface for PAML. *Molecular Biology and Evolution*. 2013. Volume 30. P 2723-2724.
- Yoder AD, Yang Z.** Estimation of Primate speciation dates using local molecular clocks. *Molecular Biology and Evolution*. 2000. Volume 17. P 1081-1090.

Annexes:*Annexe 1 Liste des échantillons étudiés*

| Rivière | Bassin Versant | Echantillons (n° ISEM) | Haplotypes identifiés |
|------------------|----------------|--|--------------------------|
| Corse | | | |
| Uccialinu | Taravu | T08015 - T08019 | 5 AD_15 |
| Val d'Ese | Prunelli | T08054 - T08057, T08073 | 5 AD_16 |
| Rina | Fium'Orbu | T08094 - T08097, T08113 | 4 AD_16, 1 AD_15 |
| Veraculongu | Taravu | T08115 - T08119 | 1 ME_12, 4 AD_15 |
| A Manica | Golu | T08134, T08138, T08143, T08148, T08153 | 4 AD_5 1 AD_17 |
| Paratella | Prunelli | T08197 - T08201 | 1 MA_2, 4 MA_3 |
| Chjuva | Taravu | T08261 - T08265 | 5 AD_5 |
| E Rocce | Fangu | T08314, T08318, T08323, T08328, T08333 | 3 AT_2 2 AD_5 |
| Bocca Bianca | Fangu | T08334 - T08336, T08338, T08343, T08348, T08353 | 1 AT_8 6 AT_2 |
| Carnevalle | Prunelli | T08355 - T08357, T08361 - T08363 | 1 AD_5 5 MA_3 |
| Puzzatelli | Tavignanu | T08368 - T08372 | 4 AD_5 |
| Botaro | Liamone | T08399, T08403, T08408, T8413, T08417 | 5 AD_18 |
| E Ventose | Golu | T08530 - T08534 | 5 MA_3 |
| Chjuvone | Rizzanese | T07703, T07707, T07712, T07717, T07721 | 4 ME_11 1 AD_14 |
| Susinelle-Ruello | Fium'Orbu | T09232 - T09236 | 5 ME_14 |
| E Radule | Golu | T10159 - T10163 | 5 ME_2 |
| U Golu | Golu | T10179, T010183, T10188, T10193, T10198 | 5 ME_2 |
| Lette | Fiume Seccu | T10588, T10592, T10597, T10602, T10606 | 1 AD_4 4 AD_5 |
| Corbica | Golu | T10788 T10792 | 2 ME_2 |
| Belle Bone | Fiume Grossu | T15388 - T15392 | 1 AT_5, 1 AD_5 2 AT_2 |
| Ercu | Golu | T15410, T15412 | 1 AD_6, 1 AD_5 |

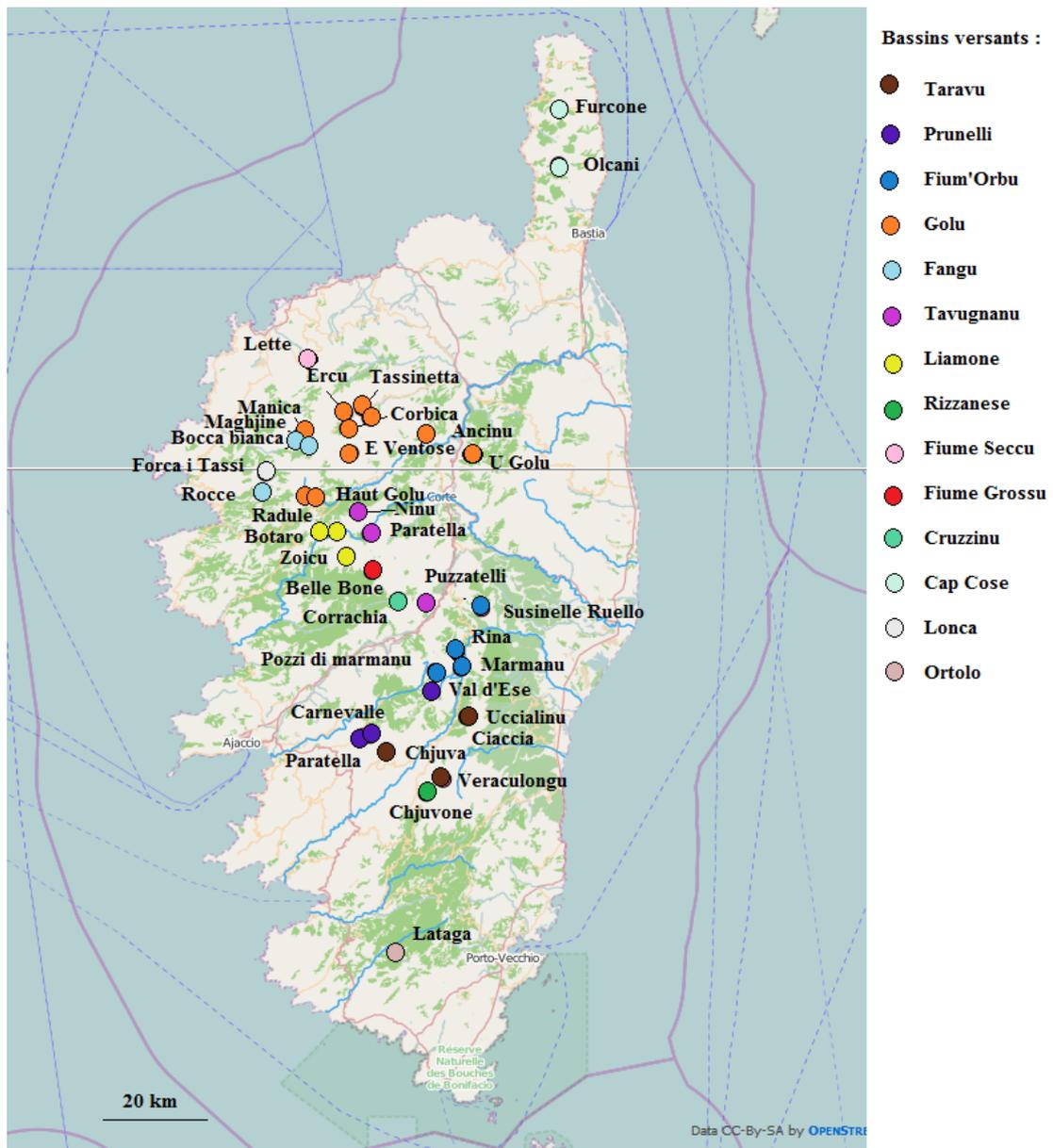
| | | | |
|----------------------------|-------------------|---|---------------------------------------|
| Corrachia | Cruzzinu | T15468, T15469, T15471, T15472, | 4 AD_7 |
| Ancinu | Golu | T18483 - T18486 | 4 AD_5 |
| Olcani | Cap Corse (Ouest) | T18513, T18515 | 1 AT_6, 1 AT_2 |
| Olmèta du Cap | Cap Corse (Ouest) | T18531 - T18535 | 5 AT_5 |
| Forca a i Tassi | Lonca | T18584 - T18587, T18603 | 5 ME_2 |
| Pont de Marmano | Fium'Orbu | T09212, T09214 - T09216 | 1 AD_21, 3 AD_16 |
| Ninu | Tavignanu | T19862, T19864, T19866 | 1 ME_6, 2 ME_4 |
| U Furcone | Cap Corse (Est) | T19958, T19959 | 2 AD_5 |
| Haut Golu | Golu | T03088 - T03091 | 3 ME_2 |
| Pozzi di Marmanu | Fium'Orbu | T08277, T08281, T08283, T08290 | 4 AD_16 |
| Coreccia | Liamone | T05921, T05923, T05925, T05927, T05928 | 3 AD_7 2 AD_5 |
| Ciaccia | Tavaru | T09293, T09294, T09295, T09297, T09299, T09304 | 6 AD_15 |
| Zoïcu | Liamone | T09347, T09352, T09357, T09361 | 4 ME_14 |
| Paratella | Tavignanu | T15508, T15509, T15510, T15511, T15512 | 4 ME_14 1 ME_5 |
| Lataga | Ortolo | T18605, T18607, T18609, T18611, T18613 | 5 AD_11 |
| Tassinetta | Golu | T23316, T23317, T23320, T23322, T23323 | 5 AD_5 |
| Maghjine | Fangu | T23331, T23333, T23335, T23337, T23339 | 5 AD_5 |
| France continentale | | | |
| Loup | Alpes Maritimes | T16126, T16131, T16140, T16143 | 3 ME_2 1 ME_4 |
| Petit Buëch | Durance | T12537, T12556, | 2 AD_1 |
| Biaysse (Amont) | Durance | T11559, T11566, T11571, T11585 | 4 ME_2 |
| Gleizolles | Durance | T04416, T04424, T04432 | 2 AD_1, 1 AD_12 |
| Fontaine Vaucluse | Sorgue | T07424 – T07427 | 1 ME_9 1 ME_10 1 ME_8 1 ME_2 |

| | | | |
|-----------------|--------------------------------|--|------------------------------|
| Fontan (Aval) | Alpes Maritimes | T011639, T011647 | 2 ME_2 |
| Nohèdes (Amont) | Pyrénées Orientales | T13457 - T13461, T13465 | 5 AT_2 |
| Cians | Alpes Maritimes | T16146, T16152, T16158, T16161 | 1 AT_1 3 AT_2 |
| Vidourle | Hérault | T16100, T16109 | 1 AD_8, 1 AT_2 |
| Joncel | Hérault | T19018, T19020, T19022, T19024, T19026 | 1 AD_1 2 AT_1 2 ME_2 |
| Italie | | | |
| Camboni | Sardaigne | T16246 - T16249 | 3 AD_9 |
| Flumensosa | Sardaigne | T16276, T16278-T16280, | 3 AT_1, 1 AT_2 |
| Anapo | Sicile | T25009 - T25013 | 5 AT_7 |
| Fibreno | Latium | T05727, T05729, T05731, T05733, T05735, T05741, T05751, T05761 | 8 AD_133 |
| Volturno | Campanie | T05787 – T05791 | 5 ME_7 |
| Balkans | | | |
| Trebuscica | Soca (Slovénie) | T06836 | 1 MA_1 |
| Volaja | Soca (Slovénie) | T16490, T16492-T16596 | 1 AT_10 4 MA_1 1 AT_2 |
| Valbona | Drin (Albanie) | T09161, T09163, T09165, T09167, T09169 | 5 AD_10 |
| Shala | (Albanie) | T09002 - T09004, T09011 - T09013, T09015, T09021 | 3 AT_9 1 AD_21 5 AD_10 |
| Shkumbini | Pogradec (Albanie) | T09052 - T09054, T09058 – T09062, T09065, T09067, T09069, T09072, T09074 | 13 AD_10 1 ME_13 |
| Lac Ohrid | Dasaretska (Rép. Macédoine) | T16585 - T16589 | 5 AD_10 |
| Espagne | | | |
| Madera | | 1 - 5 | 5 AD_1 |
| Castril | | 2 - 5 | 2 AD_1 2 AD_2 |
| Dilar | | 1 - 5 | 5 AD_3 |
| Endrinales | | 1 - 2, 4, 6 | 4 ME_3 |

| | | |
|---|---------|-------------------------------------|
| Trevelles | 1 - 4 | 4 AD_3 |
| Conangles | 5, 7, 8 | 2 AD_1 1 AD_3 |
| Cardos | 2 - 5 | 2 AD_1 1 ME_1 1 ME_2 |
| Vallauca | 1 - 5 | 4 ME_2 1 AT_10 |
| Baga | 4 - 8 | 1 AT_1 2AT_2 1 AT_3 1 AT_4 |
| Séquences de GenBank <i>Salmo trutta</i> | | |
| AM910409 (Duc et al., non publiée) | | AT_2 |
| EU492348 (Noren et al., non publiée) | | AT_2 |
| FJ435622 (Espineira et al., non publié) | | AT_2 |
| HQ167696 (Keskin, non publié) | | AT_2 |
| JX960839 (Crête Lafrenière et al., non publié) | | ME_2 |
| JX960836 (Crête Lafrenière et al., non publié) | | AT_6 |
| D58400 (Matsuda et al., non publié) | | AT_11 |
| FJ435623 (Espineira et al., non publié) | | AT_11 |
| EU492282 (Noren et al., non publié) | | AT_12 |
| EU492108 (Pelt Heerschap et al., non publié) | | AT_13 |
| EU492109 (Pelt Heerschap et al., non publié., 2008) | | AT_13 |
| FJ435621 (Espineira et al., non publié) | | AT_13 |
| JX960835 (Crête Lafrenière et al., 2012) | | AD_22 |
| JX960837 (Crête Lafrenière et al., 2012) | | MA_4 |
| Séquences de GenBank <i>Salmo salar</i> | | |
| AF053591 (Philips et al., non publiée) | | |
| JX960833 (Crête Lafrenière et al., 2012) | | |

Références pour les séquences de GenBank:

Crete-Lafreniere A, Weir LK et Bernatchez L. Framing the salmonidae family phylogenetic portrait: a more complete picture from increased taxon sampling. Institut de Biologie Integrative et des Systemes, Université Laval, 1030, Avenue de la Medecine, Quebec, Quebec G1V 0A6, Canada



Annexe 3 Positions géographiques des populations en Corse (Via OSU OREME DATA Open street Map : www.openstreetmap.org)