

Campagne 2020 Contrats Doctoraux Instituts/Initiatives

Proposition de Projet de Recherche Doctoral (PRD)

Appel à projet IBEEs - Initiative Biodiversité Evol Ecolo Société 2020

Intitulé du Projet de Recherche Doctoral : Quantifier le déclin des espèces à partir de données génomiques et d'occurrences.

Directeur de Thèse porteur du projet (titulaire d'une HDR) :

NOM : **ACHAZ** Prénom : **Guillaume**
Titre : Professeur des Universités ou Univ
de Paris
e-mail : guillaume.achaz@mnhn.fr
Adresse professionnelle : Musée de l'Homme
(site, adresse, bât., bureau)

Unité de Recherche :

Intitulé : Eco-Anthropologie
Code (ex. UMR xxxx) : UMR7206

ED227-Sciences vie homme : évolution

Ecole Doctorale de rattachement de l'équipe & d'inscription du doctorant :

Doctorants actuellement encadrés par le directeur de thèse (préciser le nombre de doctorants, leur année de 1ère inscription et la quotité d'encadrement) :

Elise Kerdoncuff (50%, co-encadrement avec A Lambert), 2017-

Karen Salazar (50%, co-encadrement avec R Nattier), 2017-

Co-encadrant :

NOM : **LAMBERT, Amaury** Prénom :
+ FUCHS, Jérôme
Titre : Professeur des Universités ou HDR ☐
à Sorbonne Université
+ Maître de Conférences au
Muséum National d'Histoire
Naturelle
e-mail : amaury.lambert@upmc.fr
+ jerome.fuchs@mnhn.fr

Unité de Recherche :

Intitulé : Laboratoire de Probabilité, Statistiques et Modélisation
+ Institut de Systématique, Evolution, Biodiversité
Code (ex. UMR xxxx) : UMR8001
+ UMR7205

ED227-Sciences vie homme : évolution écologie

Ecole Doctorale de rattachement : Ou si ED non Alliance SU :

Doctorants actuellement encadrés par le co-directeur de thèse (préciser le nombre de

doctorants, leur année de 1^{ère} inscription et la quotité d'encadrement) :

pour Amaury Lambert:

- Elise Kerdoncuff (50%, co-encadrement avec G Achaz), 2017-

- Félix Foutel--Rodier (50%, co-encadrement avec E Schertzer)

Cotutelle internationale : ☒ Non ☐ Oui, précisez Pays et Université :

Description du projet de recherche doctoral (en français ou en anglais)

3 pages maximum – interligne simple – Ce texte sera diffusé en ligne

Détailler le contexte, l'objectif scientifique, la justification de l'approche scientifique ainsi que l'adéquation à l'initiative/l'Institut.

Le cas échéant, préciser le rôle de chaque encadrant ainsi que les compétences scientifiques apportées. Indiquer les publications/productions des encadrants en lien avec le projet.

Préciser le profil d'étudiant(e) recherché.

Le risque d'extinction de 99,9% des espèces vivantes est inconnu, en particulier pour les invertébrés, cœur de la biodiversité multicellulaire. Il est nécessaire d'employer de nouvelles méthodes pour quantifier efficacement la crise en cours de la biodiversité, rapidement et à grande échelle. Nous avons récemment développé de nouvelles méthodes pour inférer la démographie des espèces à très court terme, soit depuis le début du vingtième siècle, date approximative estimée du début de la sixième extinction de masse. Ces méthodes sont basées d'une part sur des données génomiques (génomomes « complets » d'un échantillon d'une dizaine d'individus diploïdes) et d'autre part sur des données d'occurrences (présence/absence/abondance d'une espèce à une localité au cours du temps sur quelques décennies).

En bref, le déclin très récent d'une espèce ne laisse que peu de traces sur les patrons génomiques moyens, tels que la diversité nucléotidique moyenne, ou la distribution de fréquence des mutations, mais change significativement les patrons de recombinaison observés dans un échantillon de génomes (Kerdoncuff et al., 2019, biorxiv). Par ailleurs, des informations sur l'occurrence, ou mieux encore l'abondance, d'espèces dans des localités données peuvent être utilisées pour estimer la probabilité d'extinction, ou le déclin d'espèces à l'échelle locale ou globale (Régner et al., 2015, PNAS).

Le projet s'inscrit dans un programme de recherche porté par un consortium multidisciplinaire porté par Guillaume Achaz qui comprend une dizaine de chercheur.e.s et enseignant.e.s-chercheur.e.s spécialisé.e.s en écologie et taxonomie (J Fuchs, C Fontaine, B Fontaine, L Legall, V Prié, R Rougerie), en bioinformatique (R Vignes-Lebbe, S Brouillet) du Muséum National d'Histoire Naturelle, ainsi qu'en mathématiques et modélisation (A Lambert, E Schertzer) du Collège de France. Il se base aussi sur une collaboration avec deux équipes américaines de Harvard (T Sackton) et Stanford (D Petrov).

Dans le cadre du projet, le consortium génère et assemble des données de génomes complets dont la démographie est connue (p. ex. oiseaux d'Europe, moules d'eaux douces et insectes pollinisateurs) ou inconnues (p. ex. lépidoptères, algues de la côte bretonne). Il rassemble également des données d'occurrences, voire d'abondance quand cela est possible, pour la plupart des espèces dont le

génomique est séquencé.

La projet de cette thèse consiste à faire l'analyse des risques d'extinction à partir de suivis temporels d'occurrences et de séquences des génomes complets de grands singes, d'oiseaux, d'insectes et d'algues pour lesquels nous avons des données soit déjà prêtes soit en cours d'acquisition. Nous appliquerons les méthodes « classiques » de génomique des populations et d'analyse de données d'occurrence mais également les méthodes de vraisemblance maximisée récemment développées par les modélisateurs du consortium. Les méthodes seront comparées et améliorées au vu des connaissances taxonomistes expertes. Les allers-retours entre les experts taxonomistes et les modélisateurs permettront une synergie positive qui assurera la validité des méthodes et des inférences proposées. Bien que cette thèse soit principalement orientée vers l'inférence de scénarios démographiques concrets à partir de méthodes déjà existantes, elle pourra selon le profil du/de la candidat.e également s'intéresser au défi méthodologique qui consiste à intégrer les deux types de données (génomiques et occurrence) dans un unique cadre démographique commun. Il est en effet possible de construire un modèle stochastique ad hoc pour calculer une vraisemblance jointe intégrant les deux sources d'information, en se servant du fait que les deux types de données sont indépendantes conditionnellement à la démographie (inconnue).

Chaque espèce choisie sera dans un premier temps analysée séparément. Lors de la troisième année, le/la candidat.e fera la méta-analyse de tous les jeux de données disponibles. Il/elle testera alors les causes du déclin en contrastant les scénarios démographiques entre (a) zones géographiques différentes (p. ex. continent, pays), (b) entre milieux terrestres et marins et (c) également entre environnements urbains et ruraux.

Cette thèse s'inscrit dans les thématiques de l'Initiatives Biodiversité, Evolution, Ecologie, Société (IBEES) en tant que thématique exploration en amont des activités de l'Institut de la Transition Environnementale (ITE). Le sujet multidisciplinaire (bio-informatique, modélisation, génomique évolutive, écologie) requiert par nature un encadrement de compétences variées possédées collectivement par les encadrants.

Guillaume ACHAZ, encadrant principal possède de solides compétences en génomique des populations et en bioinformatique, Jérôme Fuchs de profondes connaissances naturalistes sur les oiseaux et de solides compétences en séquençage et génomique et Amaury Lambert est expert en modélisation de l'évolution du vivant. Les trois encadrants sont membres actifs du consortium cité ci-dessus et travaillent déjà en collaboration. G Achaz et A Lambert ont co-encadré plusieurs étudiants et post-docs et co-publié une dizaine d'articles en commun.

Cette thèse devrait déboucher sur plusieurs publications. Lors des deux premières années, l'analyse de chacune des espèces ou des clades analysés sera proposée pour publication dans des journaux comme *Molecular Ecology*, *Evolution* ou *Conservation Biology*. La méta-analyse finale de la dernière année de cette thèse fera l'objet d'une publication soumise à un journal à fort impact.

En plus des publications, les résultats de cette thèse seront notamment rendus accessibles sur Redcode of Life, une plateforme en construction qui sera ouverte aux chercheurs et au grand public. La mise en place technique de cette plateforme ne rentre pas dans le cadre de cette thèse.

Le/la candidat/e devra avoir de bonnes connaissances en génomique évolutive, en particulier en génomique des populations. Il/Elle devra également avoir des compétences au moins élémentaires en (bio)-informatique et/ou en modélisation.

En dehors des fonds récurrents des équipes concernées, le consortium bénéficie de fonds propres issus d'un projet financé par la Fondation François Sommer (90 kE sur 3 années) pour, entre autre, générer les données de ce sujet en particulier.

**Merci de nommer votre fichier pdf :
«ACRONYME de l'institut/initiative_2_NOM Porteur Projet_2020 »**

**à envoyer simultanément par e-mail à l'ED de rattachement et au programme :
[cd instituts et initiatives@listes.upmc.fr](mailto:cd_instituts_et_initiatives@listes.upmc.fr) avant le 30 mars.**