
Portrait de famille du RADseq: cohérence globale et singularités

Régis Debruyne*¹

¹Service de Systématique Moléculaire (UMS 2700 OMSI) – Muséum National d'Histoire Naturelle (MNHN), CNRS : UMS2700 – 43, rue Cuvier, CP26, 3ème étage, 75005 Paris, France

Résumé

L'avènement du RADseq à la fin des années 2000 s'est inscrit dans une démarche globale visant à exploiter les technologies de séquençage de seconde génération (NGS) à des fins d'analyse comparative de banques d'ADN complexes pour des organismes non-modèles par réduction génomique. La simplicité et l'universalité conceptuelles du RADseq ont assuré son succès rapide dans la communauté des sciences de l'évolution et en écologie moléculaire. Elles ont également conduit à un essaimage méthodologique à partir de la technique initiale. Les différents variantes méthodologiques partagent toutes quelques caractéristiques fondamentales: séquençage de la fraction du génome à proximité immédiate des signatures de restriction choisies, indépendance d'une connaissance a priori du génome, génération d'un grand nombre de marqueurs (SNPs) distribués à travers tout le génome. Elles diffèrent donc essentiellement sur les modalités précises de réduction du génome: nombre et type d'enzymes de restriction employées, sous-échantillonnage du catalogue de RADtags générés (par sélection de taille) ou non, etc.

Cette communication présentera les points communs et les singularités expérimentales des principaux représentants de cette famille méthodologique. En présentant ce panel d'approches et en insistant sur les sensibilités et leurs atouts respectifs, elle vise à identifier les contextes favorables à l'application de chacune des alternatives existantes. Elle introduira aussi la multiplicité des objectifs scientifiques pouvant être atteints en fonction des spécificités expérimentales adoptées et des méthodes d'analyse de données mises en œuvre.

*Intervenant

Planification d'expérience RADseq : approche par simulation.

Olivier Lepais^{*1,2}

¹Institut National de la Recherche Agronomique (INRA) – Institut national de la recherche agronomique (INRA) – UMR ECOBIOP Aquapole 64310 Sanit-Pée-sur-Nivelle, France

²Université de Pau et des Pays de l'Adour (UPPA) – Université de Pau et des Pays de l'Adour [UPPA] : UMRECOBIOP – UMR ECOBIOP 1 Allée du Parc Montaury 64600 ANGLET, France

Résumé

De nombreuses approches RADseq ont été développées ces dernières années chacune utilisant différentes stratégies pour réduire la quantité de séquence génomique ciblée. Le choix d'une méthode RADseq et des multiples paramètres techniques devient compliqué du fait de la difficulté d'anticiper le nombre de locus attendu. Or le nombre de locus ciblés conditionne la faisabilité de l'expérience tant au niveau du nombre d'individus qu'il sera possible d'analyser que du nombre de locus nécessaire pour l'application biologique. Il est donc important de planifier préalablement toute expérience de RADseq pour éviter les déconvenues. L'approche par simulation est une méthode flexible permettant d'estimer le nombre de locus produits par les différentes approches RADseq. Après une rapide présentation technique des principales approches RADseq, nous illustrerons l'apport de la simulation à l'aide du package R SimRAD pour la planification d'expérience dans le cas d'espèces procédant ou non un génome de référence. Au-delà de la prédiction du nombre de locus, d'autres apports de la simulation seront également illustrés.

*Intervenant

How and how much does RAD-seq bias genetic diversity estimates?

Sylvain Charlat^{*†1}, Marie Cariou², and Laurent Duret¹

¹Biométrie Biologie Evolutive (LBBE) – CNRS : UMR5558, Université Lyon 1 – Bat. Mendel, 16, rue
Raphael Dubois 69622 Villeurbanne, France

²Unité de recherche en biologie environnementale et évolutive (URBE) – Rue de Bruxelles 61, B-5000
Namur, Belgique

Résumé

RAD-seq is a powerful tool, increasingly used in population genomics. However, earlier studies have raised red flags regarding possible biases associated with this technique. In particular, polymorphism on restriction sites results in preferential sampling of closely related haplotypes, so that RAD data tends to underestimate genetic diversity. Here we (1) clarify the theoretical basis of this bias, highlighting the potential confounding effects of population structure and selection, (2) confront predictions to real data from *in silico* digestion of full genomes and (3) provide a proof of concept toward an ABC-based correction of the RAD-seq bias. Under a neutral and panmictic model, we confirm the previously established relationship between the true polymorphism and its RAD-based estimation, showing a more pronounced bias when polymorphism is high. Using more elaborate models, we show that selection, resulting in heterogeneous levels of polymorphism along the genome, exacerbates the bias and leads to a more pronounced underestimation. On the contrary, spatial genetic structure tends to reduce the bias. We confront the neutral and panmictic model to "ideal" empirical data (*in silico* RAD-sequencing) using full genomes from natural populations of the fruit fly *Drosophila melanogaster* and the fungus *Shizophyllum commune*, harbouring respectively moderate and high genetic diversity. In *D. melanogaster*, predictions fit the model, but the small difference between the true and RAD polymorphism makes this comparison insensitive to deviations from the model. In the highly polymorphic fungus, the model captures a large part of the bias but makes inaccurate predictions. Accordingly, ABC corrections based on this model improve the estimations, albeit with some imprecisions. More elaborate ABC methods might integrate additional parameters, such as population structure and selection, but their opposite effects could hinder accurate

^{*}Intervenant

[†]Auteur correspondant: sylvain.charlat@univ-lyon1.fr

How much can we trust in *in silico* detected RADseq loci and SNPs?

Stéphanie Mariette^{*1}, Myriam Heuertz¹, Henri Caron², Adline Delcamp¹, Olivier De Thier³, Veronica El Mujtar^{1,4}, Alain Franc¹, Pauline Garnier-Géré¹, Frédéric Gévaudant⁵, Erwan Guichoux¹, Quentin Jehanne¹, Yec'han Laizet⁶, Ludivine Lassois³, Philippe Lejeune³, Jean-François Molino⁷, Arnaud Monty³, Annabel Porté¹, Samuel Quevauvillers³, Daniel Sabatier⁷, and Cindy Verdu³

¹UMR BIODIVERSITE GENES COMMUNAUTES (UMR BIOGECO) – Institut national de la recherche agronomique (INRA), Université de Bordeaux (Bordeaux, France) – 69, route d'Arcachon 33610 Cestas, France

²UMR Ecologie des Forêts de Guyane (UMR ECOFOG) – Campus agronomique - BP 316 - F-97379 Kourou, Guyane française

³Gembloux Agro Bio Tech and University of Liège (Gembloux Agro Bio Tech and University of Liège) – Passage des Déportés 2, 5030 Gembloux, Belgique, Belgique

⁴INSTITUTO NACIONAL DE TECNOLOGIA AGROPECUARIA (INTA) – Modesta Victoria N 4450 8400 San Carlos de Bariloche Río Negro, Argentine

⁵UMR BIOLOGIE DU FRUIT ET PATHOLOGIE (UMR BFP) – Institut national de la recherche agronomique (INRA) : UMR1332, Université de Bordeaux (Bordeaux, France) – Bât. IBVM 71, av. Edouard Bourlaux - CS 20032 - 33882 Villenave d'Ornon Cedex - France, France

⁶Institut Bergonié (Institut Bergonié) – Institut Bergonié – 229, cours de l'Argonne CS 61283 33076 Bordeaux Cedex, France

⁷Botanique et Bioinformatique de l'Architecture des Plantes (AMAP) – Institut national de la recherche agronomique (INRA) : UR0931, Institut de recherche pour le développement [IRD], CNRS : UMR5120, Centre de coopération internationale en recherche agronomique pour le développement [CIRAD] : UMR51, Université Montpellier II - Sciences et techniques – Bd de la Lironde TA A-51/ PS 2 34398 Montpellier cedex 5, France

Résumé

The RADseq technology allows researchers to efficiently discover thousands of polymorphic loci across multiple individuals with no prior information on the genome. One important question is whether *in silico* detected single nucleotide polymorphisms (SNPs) from RADseq can confidently be used for direct phylogenetics or population genetics inference, as opposed to a two-step approach with SNP marker development from RADseq and subsequent genotyping in a separate assay. To answer this question, it seems reasonable to check the efficiency of various bioinformatics treatments before implementing large scale *in silico* RADseq analyses. Plant species have two characteristics that deserve specific attention. First, recent genome duplications in many plant clades have created paralogous positions. Paralogy may affect the development of SNP markers and the estimation of genetic diversity. We evaluated

^{*}Intervenant

the impact of putative paralogous loci on genetic diversity estimation during SNP development from a RADseq dataset for the non-model tree species *Robinia pseudoacacia* L. (black locust). For this purpose, we used published bioinformatics tools to quantify the effect of paralogy and efficiently remove paralogous positions during the *in silico* analyses. A series of SNPs was further tested with a genotyping system to confirm the removal of paralogs. Second, hybridization between closely related species leads to the formation of plant species complexes with possible shared polymorphisms among species. In this case, discovering true verifiable polymorphism at the level of the species complex becomes challenging because population genetics assumptions commonly used in SNP filtering do not hold at greater phylogenetic depths. We compared several bioinformatics pipelines on a RADseq dataset acquired on tropical tree species of the Bertholletia clade (Lecythidaceae family), and we verified the existence of RADseq loci and SNPs using a resequencing approach.

Histoire évolutive de la divergence eco-phénotypique parallèle entre paires d'écotypes d'une même espèce

Pierre-Alexandre Gagnaire*¹

¹Institut des Sciences de l'Evolution - Montpellier (ISEM) – CNRS : UMR5554, Institut de recherche pour le développement [IRD] : UMR226, Université Montpellier II - Sciences et techniques – Place E. Bataillon CC 064 34095 Montpellier Cedex 05, France

Résumé

Certaines espèces se subdivisent en écotypes associés à des niches écologiques distinctes, et dont des répliquats naturels offrent des modèles d'étude intéressants pour comprendre le phénomène de spéciation écologique. La divergence éco-phénotypique parallèle entre paires d'écotypes d'une même espèce est souvent vue comme la conséquence de réponses adaptatives similaires et indépendantes face à des contraintes sélectives identiques. Les approches de génotypage par séquençage ont récemment fourni une vision pangénomique des patrons de différenciation génétique associés à la divergence entre écotypes. Certaines régions génomiques fortement différenciées de manière parallèle entre plusieurs répliquats de paires d'écotypes ont ainsi été détectées. Si ces patrons de divergence confirment l'action de la sélection dans ces régions, ils ne disent pas toutefois de quelle manière ni à quel moment la sélection a agi. Les données de polymorphisme RAD-seq peuvent servir à reconstituer l'histoire démographique et sélective associée à la diversification des écotypes. A partir d'exemples d'étude d'écotypes de poissons, j'illustrerai comment des modèles de divergence capturant les variations temporelles et chromosomiques du flux génique permettent d'explorer les multiples facettes du processus de spéciation.

*Intervenant

RAD-sequencing to investigate population genomics, selection footprints and architecture of heritable traits in great and blue tits in heterogeneous habitats

Charles Perrier^{*1} and Charmantier Anne²

¹Centre d'Ecologie Fonctionnelle et Evolutive (CEFE) – CNRS : UMR5175 – 1919 route de Mende
34293 Montpellier Cedex 5, France

²Centre d'Ecologie Fonctionnelle et Evolutive (CEFE) – CNRS : UMR5175 – 1919 route de
Mende;34293;Montpellier Cedex 5, France

Résumé

Striking phenotypic variability at ecologically relevant and heritable traits are exhibited by both blue tits and great tits within and between various habitats, putatively reflecting adaptation to habitats heterogeneity. However, little is known about the genomic bases underlying these traits variability both within and between populations. Using RAD-sequencing over a thousand birds' DNA, we investigate population structure, selection footprints, and genome wide architecture of behavioral traits variability, in blue tits and great tits across heterogeneous habitats. We find an overrepresentation of genes related to neuronal functions among SNPs identified by genome scans as most differentiated between environments or identified by GWAS as associated to behavioral traits. Interestingly, outlier SNPs and related genes differ among species, populations and methodologies used (genome scans *vs* GWAS). These finding opens exciting perspectives for broader investigations of genomic bases related to neuronal functions in relation to adaptation in heterogeneous environment and avian personality, as well as for the coupling of genome scans and GWAS for the study of adaptive variations.

^{*}Intervenant

HyRAD: un nouvel horizon génomique pour l'étude de l'ADN historique et ancien, combinant RAD sequencing et capture par hybridation

Nadir Alvarez^{*1}

¹Département d'Ecologie et Evolution (DEE) – Bâtiment Biophore UNIL-Sorge 1015 Lausanne, Suisse

Résumé

L'analyse génomique d'échantillons dont l'ADN est dégradé, tels que les spécimens de collection, représente un challenge pour les généticiens des populations, en raison de la petite taille des fragments d'ADN qui rend difficile l'amplification par PCR (en combinaison ou non avec la restriction enzymatique). Par conséquent, il n'est pas possible, la plupart du temps, d'appliquer des méthodes de type Genotyping-by-Sequencing ou RAD sequencing.

La combinaison du RAD sequencing et de la capture par hybridation permet de contourner cette limitation. HyRAD et HyRNAD sont des méthodes de capture polyvalentes utilisant le produit de RAD sequencing obtenu à partir d'un ADN ou d'un ARN de qualité, provenant en général de quelques spécimens frais, afin de produire un catalogue. Ce catalogue est ensuite utilisé, tel quel, ou converti en ARN pour améliorer les propriétés de stringence lors de l'hybridation, pour capturer les fragments homologues provenant des extractions d'ADN des spécimens ciblés par l'analyse, préalablement indexés comme dans une procédure de shotgun sequencing classique.

Bien que HyRAD et HyRNAD aient été développés pour l'analyse d'échantillons historiques et anciens, ces méthodes peuvent naturellement être appliqués aux échantillons dont l'ADN n'est pas dégradé. Leur avantage par rapport au RAD sequencing classique, est de produire, à coût légèrement supérieur toutefois, des matrices de SNPs en général mieux remplies, puisque la technique n'est pas sensible aux mutations se produisant sur les sites de restriction dont le maintien au sein des individus analysés, est essentielle dans le RAD classique.

^{*}Intervenant

Phylogénie et délimitation d'espèces du complexe d'escargots terrestres *Pyramidula* par l'analyse de données RADseq

Gontran Sonet^{*1}, Oihana Razkin^{3,2}, Karin Breugelmans⁴, María José Madeira^{2,3}, Benjamín Juan Gómez-Moliner^{3,2}, and Thierry Backeljau^{5,1}

¹Royal Belgian Institute of Natural Sciences (OD Taxonomy and Phylogeny JEMU) (RBINS, JEMU) – Vautierstraat 29, B-1000 Brussels, Belgique

³Biodiversity Research Group (CIEA Lucio Lascaray) – Avda. Miguel de Unamuno 3, 01006 Vitoria-Gasteiz, Alava, Espagne

²Department of Zoology and Animal Cell Biology, Faculty of Pharmacy, University of the Basque Country (UPV/EHU) – Paseo de la Universidad 7, 01006 Vitoria-Gasteiz, Alava, Espagne

⁴Royal Belgian Institute of Natural Sciences (OD Taxonomy and Phylogeny) (RBINS) – Vautierstraat 29, B-1000 Brussels, Belgique

⁵Evolutionary Ecology Group, University of Antwerp (UA) – Groenenborgerlaan 171, B-2020 Antwerp, Belgique

Résumé

L'histoire évolutive de complexes d'espèces proches est souvent difficile à reconstruire à cause d'un manque de données informatives. Les données RADseq (Restriction site-associated DNA sequencing) permettent de collecter une grande quantité de données ADN, y compris pour des organismes non-modèles. Ici nous avons utilisé les données RADseq pour étudier le complexe d'espèces d'escargots terrestres du genre *Pyramidula*. La concaténation des séquences ADN obtenues (environ 1,5 x 10⁶ caractères et > 97 x 10³ sites polymorphiques) a permis de reconstruire les relations phylogénétiques entre les espèces du complexe avec beaucoup plus de résolution que si des marqueurs phylogénétiques standards avaient été utilisés (COI, 16S rARN, 5.8S rARN, ITS2 and 28S rARN). D'autre part, l'analyse de 875 SNPs (single-nucleotide polymorphisms) et de plusieurs scénarios de délimitation d'espèces a soutenu l'hypothèse que neuf espèces de *Pyramidula* étaient à distinguer au sein de notre échantillonnage. Enfin, l'analyse de la répartition des allèles parmi les individus séquencés (D-statistics) n'a pas montré d'évidence d'hybridation interspécifique ancestrale, si ce n'est entre deux espèces, *P. pusilla* and *P. saxatilis*.

*Intervenant

Différentiation allochronique d'une population de processionnaire du pin à cycle décalé: contexte phylogénétique et histoire démographique révélés à l'aide de marqueurs RADseq

Carole Kerdelhué^{*1}, Raphael Leblois¹, Renaud Vitalis¹, Laure Sauné¹, Audrey Rofritsch¹, Julien Foucaud¹, Manuela Branco², Christian Burban³, Astrid Cruaud¹, and Mathieu Gautier¹

¹Centre de biologie et gestion des populations (CBGP) – Institut national de la recherche agronomique (INRA) : UMR1062 – Campus international de Baillarguet - 34398 Montpellier Cedex 5, France

²Instituto Superior de Agronomia (ISA) – Universidade de Lisboa Tapada da Ajuda 1349-017 Lisbon, Portugal, Portugal

³Biodiversité, Gènes Communautés (BioGeCo) – Université de Bordeaux, Institut national de la recherche agronomique (INRA) : UMR1202 – Site de recherche Forêt - Bois de Pierroton - 69, route d'Arcachon F-33612 Cestas Cedex FRANCE, France

Résumé

La processionnaire du pin *Thaumetopoea pityocampa* est un Lépidoptère ravageur des forêts de pin et de cèdres dans l'ouest du bassin méditerranéen. Ses chenilles sont responsables de fortes urtications notamment au moment des processions. Elle forme un complexe d'espèce avec *T. wilkinsoni*, présente en Turquie, au Moyen-Orient, en Crète et à Chypre. Dans toute l'aire de distribution, la reproduction a lieu en été, et les larves se développent dans des nids caractéristiques pendant l'automne et l'hiver. Les chenilles quittent ensuite le nid en procession et vont s'enterrer pour se nymphoser, jusqu'à l'émergence des papillons l'été suivant.

En 1997, une population à cycle décalé a été découverte dans une forêt côtière du Portugal, près de Leiria, en sympatrie avec des individus se développant selon le cycle classique. La population décalée (SP) se reproduit au printemps et ses chenilles se développent en été, la procession ayant lieu en septembre. Des études préliminaires (microsatellites) ont montré une forte différenciation génétique de la SP, qui est par ailleurs très proche des autres populations portugaises d'après les marqueurs mitochondriaux.

Nous avons commencé à développer des ressources génomiques chez *T. pityocampa*, ainsi que des approches de génomique des populations utilisant notamment les marqueurs RAD-seq. Dans cet exposé, nous présenterons les premiers résultats de phylogénomique obtenus sur les espèces de *Thaumetopoea* associés aux conifères, et montrant que la SP appartient au clade ibérique de *T. pityocampa*. Nous montrerons ensuite comment une approche RAD-seq populationnelle a permis de déterminer l'histoire démographique de la SP. Nous avons ainsi pu estimer que cette population a divergé d'un ancêtre commun local il y a environ 65 générations, et a connu deux goulots d'étranglement et deux périodes d'expansion récentes. La migration entre populations au cours de la différenciation a également pu être inférée. La recherche de signatures de sélection est en cours.

*Intervenant

Apport et Comparaison de données génomiques issues de NGS (Shotgun vs ddRADseq) dans l'inférence phylogénétique chez les Rosettes Géantes des paramos

Charles Pouchon^{*1}

¹Laboratoire d'écologie alpine (LECA) – CNRS : UMR5553, Université Joseph Fourier - Grenoble I, Université de Savoie – bat. D - Biologie 2233 Rue de la piscine - BP 53 38041 GRENOBLE CEDEX 9, France

Résumé

Les montagnes des Andes du nord abritent un écosystème de type tropical alpin connu sous le nom de Paramo. Malgré des conditions stressantes, caractéristiques de ces habitats, les paramos contiennent les radiations de plantes alpines les plus rapides avec des exemples emblématiques comme les lupins ou les *Espeletiinae*. En effet, ce dernier complexe de 150 espèces, a su profiter des opportunités écologiques fournies par cet écosystème en en se diversifiant écologiquement et morphologiquement de manière remarquable. Si bien que les *Espeletiinae* constitueraient un exemple majeur de radiation adaptative chez les plantes. Or les reconstructions phylogéniques déjà réalisées chez ce complexe, nécessaires pour vérifier l'hypothèse de spéciation écologique à l'origine de leur radiation, ne traduisent aucune relations entre les espèces. En effet, les marqueurs phylogénétiques classiques apparaissent peu efficace pour étudier les relations phylogénétique chez des groupes ayant radié très récemment comme c'est le cas chez les plantes alpines andines. Le but de cette présentation est de comparer l'utilisation de données génomiques issues de méthodes de génotypage ddRAD-sequencing et d'assemblage de pseudo génome nucléaire par séquençage shotgun pour l'étude des relations phylogénétiques chez ces taxons très récents.

^{*}Intervenant

Should we pursue RAD sequencing for phylogenomics ? a review of the literature

Amelia Viricel^{*1} and Eric Pante¹

¹UMR 7266 LIENSs (LIENSs) – CNRS : UMR7266, Université de La Rochelle – 2 rue Olympe de Gouges, France

Résumé

Sequencing of restriction-site associated DNA (RADseq) has become a popular method in evolutionary biology. Variants of the original RADseq protocol were developed to improve scalability, reduce costs, or fine-tune phylogenetic resolution (e.g ddRAD, 2bRAD, ezRAD). Nevertheless, a major limit of the methods resides in a phenomenon called "locus drop-out:" as divergence increases, conservation of restriction sites erodes and the number of orthologous loci decreases. Two articles were published to test the phylogenetic resolution of RAD-tags by comparing genomes of *Drosophila*, both suggesting that the method would reach its limits when divergence exceeds 60 Mya. Several studies subsequently offered empirical data to test this prediction in oaks, beetles, corals, barnacles, and salmonids, offering contrasting results on the phylogenetic resolution and wide applicability of these methods in phylogenetics. In addition to these empirical tests, significant efforts were invested to develop bioinformatic tools to analyze RAD-tag data at phylogenetic scales (e.g. PyRAD; RADami).

In parallel to the development of RAD-tag sequencing and its testing in phylogenetics, several other genome sampling strategies were proposed to take advantage of next-generation sequencing technologies to boost resolute power in phylogenetic inference. Targeted enrichment (TE) methods such as sequence capture of conserved and ultraconserved elements were proposed. The potential of TE to resolve phylogenetic relationships at deep evolutionary time scales (> 200My divergence) raises the question of whether it remains pertinent to invest significant efforts using RADseq for reconstructing phylogenies. New studies, using empirical and *in silico* data, enlighten this issue by comparing RAD and TE strategies at large evolutionary scales. We were therefore interested in measuring how often RADseq or other reduced representation (RR) methods were successfully used at deep phylogenetic depths, compared to TE methods. In this communication, we will present a review of 64 phylogenetic studies using these approaches above the species level.

^{*}Intervenant

Hybridation et diversification chez des papillons alpins

Thibaut Capblancq*¹

¹Laboratoire d'écologie alpine (LECA) – CNRS : UMR5553, Université Joseph Fourier - Grenoble I, Université de Savoie – bat. D - Biologie 2233 Rue de la piscine - BP 53 38041 GRENOBLE CEDEX 9, France

Résumé

Le rôle de l'hybridation sur la diversification et l'adaptation des espèces est de plus en plus étudié chez les espèces animales. Et le développement des nouvelles techniques de séquençage haut débit a accru notre capacité à identifier et analyser les processus évolutifs en liens avec ce phénomène. En se basant sur des données de ddRADsequencing nous avons cherché à reconstruire l'histoire évolutive d'un complexe d'espèces proches de papillons alpins chez qui l'hybridation était soupçonnée d'avoir joué un rôle important. La comparaison de différents modèles évolutifs par une procédure d'approximate bayesian computation nous a permis de prouver l'origine hybride de deux taxons dont l'apparition serait concomitante avec la fin de la dernière glaciation dans les alpes européennes.

*Intervenant

Développement d'outils génomiques RADseq pour l'étude des flux de gènes intra- et interspécifiques chez les espèces marines.

Claire Daguin Thiebaut^{*†1}, Alan Le Moan¹, Ambre Ribardière¹, Charlotte Roby¹, Gildas Le Corguillé², Thomas Broquet¹, and Frédérique Viard¹

¹Adaptation et diversité en milieu marin [Roscoff] (ADMM) – CNRS : UMR7144, Université Pierre et Marie Curie (UPMC) - Paris VI – Station Biologique de Roscoff Place Georges Teissier -29680 ROSCOFF, France

²Station biologique de Roscoff [Roscoff] (SBR) – Université Pierre et Marie Curie (UPMC) - Paris VI, CNRS : FR2424 – Place Georges Teissier 29680 ROSCOFF, France

Résumé

Les études de l'hybridation et de la spéciation chez les invertébrés marins bénéficient à présent de l'apport des outils de la génomique haut-débit. A l'aide des équipements de la plate-forme génomique de la Station Biologique de Roscoff, nous construisons des banques ddRAD avec un protocole adapté de Brelsford et coll. (2016). Cette technique s'est montrée efficace chez différentes espèces d'algues et d'invertébrés marins, présentant des ADN génomiques de qualité très variable. Pour les espèces avec des génomes de taille modérée, elle permet de multiplexer quelques centaines d'individus, tout en garantissant une représentativité équivalente de chacun dans les données de séquençage. Ces données sont stockées et analysées sur le cluster de calcul de la plate-forme ABIMS. Des résultats obtenus après étude de deux cas contrastés sont présentés: 1- la recherche d'hybridation suite au contact secondaire récent entre les ascidies *Ciona intestinalis* et *C. robusta*. Il s'agissait ici d'analyser conjointement deux espèces très divergentes, en utilisant comme référence le génome publié de *C. robusta*. Malgré cette forte divergence, des paramètres ont pu être trouvés pour un assemblage robuste des reads (validation par comparaison avec les résultats d'un génotypage de SNPs issus de transcriptomes). L'étude de 23000 SNPs issus des séquences RAD a mis en évidence l'absence d'hybridation contemporaine entre ces espèces pourtant syntopiques et non isolées reproductivement ; elle suggère également qu'une espèce acceptée (*C. roulei*) présente en Méditerranée ne soit qu'une population structurée de l'espèce *C. intestinalis* présente dans l'Atlantique. 2- l'étude de la spéciation et de l'isolement reproducteur chez les isopodes du complexe *Jaera albifrons*. Dans ce projet, l'analyse est réalisée *de novo* entre espèces peu divergentes. Dans le complexe *J. albifrons*, jusqu'à 50 000 SNPs ont été génotypés, ce qui permet d'analyser l'hétérogénéité génomique de la divergence entre espèces dans des populations avec et sans hybridation.

^{*}Intervenant

[†]Auteur correspondant: daguin@sb-roscoff.fr

When recurrent hybridization obscures phylogenetic hypotheses, the case of *Chrysocarabus* beetles.

Jean-Yves Rasplus^{*1}, Laure Sauné¹, and Astrid Cruaud¹

¹Centre de Biologie pour la Gestion des Populations (CBGP) – Institut national de la recherche agronomique (INRA) – 755 avenue du campus Agropolis CS 30016, 34988 Montferrier-sur-Lez, France

Résumé

In animal species, interspecific hybridization and subsequent introgression has been documented in an increasing number of taxa. However, few studies have analysed the hidden effect of introgression on phylogenetic hypotheses inferred for well-differentiated and ecologically distinct species. Indeed hybridization can cause significant phylogenetic incongruence between different genomic regions that may insidiously affect phylogenetic reconstruction. When using pangenomic data, topological impact may vary depending on the parameters used in the analysis (ie the number and nature of loci and samples included).

The high incidence of hybridization in *Carabus* (Coleoptera, Carabidae) makes this group an excellent model to test how introgression may affect phylogenetic reconstruction. We used restriction-site-associated DNA (RAD) sequences to i) propose a phylogenetic hypothesis for the subgenus *Chrysocarabus*, ii) examine patterns of hybridization and test whether introgression had occurred between taxa for which we evidenced mitochondrial introgression and iii) analyse the effect of hybridization on topologies.

We will first describe the perl pipeline RADIS we developed to standardize processing of RAD-seq data for phylogenetic inference, allow fast and automated exploration of key options, facilitate comparison among parameters used to form and select sets of loci and assess how they may impact tree topology. We will then give an overview of our main results: 1) depending the parameters we used to analyse our dataset, several topologies were recovered in robust trees with nodes supported by high bootstrap values; 2) nuDNA introgression was substantial among several species of *Chrysocarabus*; 3) High level of nuclear introgression was evidenced in contact zones but also up to 500km away from the contact zone. Finally, we will discuss the confusing effects of clustering parameters, matrix completion and introgression on the phylogenetic hypotheses inferred for *Chrysocarabus* species.

^{*}Intervenant