Aix*Marseille université

imbe
institut méditerranéen de biodiversité et d'écologie marine et continentale

# Thèse de Doctorat d'Aix-Marseille Université

## Spécialité Océanographie

Présentée par Alexandra Anh-Thu Weber

**Etude écologique et génétique du complexe d'espèces cryptiques *Ophioderma longicauda* (Ophiuroidea : Echinodermata) :**

**Comparaison entre lignées incubantes et lignées produisant des larves planctoniques**

Soutenance prévue le 16 janvier 2015, devant un jury composé de :

**Dr H.A. Lessios**, Senior Scientist, Smithsonian Tropical Research Institute, Panama  Rapporteur

**Dr N. Galtier**, Directeur de recherche CNRS, ISEM – Université de Montpellier 2  Rapporteur

**Dr N. Bierne**, Directeur de recherche CNRS, , ISEM – Université de Montpellier 2  Examinateur

**Dr A. Chenuil**, Chargée de recherche CNRS, IMBE – Aix-Marseille Université  Directrice de thèse

"Seen in the light of evolution, biology is, perhaps, intellectually the most satisfying and inspiring science"

Theodosius Dobzhansky, 1973

# Remerciements

Je tiens tout d'abord à remercier ma directrice de thèse Anne Chenuil, pour m'avoir acceptée en stage de M1 il y a maintenant quatre ans, où j'ai pu faire mes premiers pas dans le monde de la recherche, en apprenant à faire mes premières PCRs (il y en a eu beaucoup depuis) et manipuler pour la première fois de l'ADN d'*Ophioderma*. Merci ensuite de m'avoir acceptée pour ce sujet de thèse qui m'a attirée sur le papier, et qui ne m'a pas déçue. Merci Anne pour avoir fait de moi une chercheuse passionnée et je l'espère, indépendante. Car la recherche, ce n'est pas seulement faire des manips, analyser des données et écrire des articles ; c'est aussi savoir construire des collaborations quand nos compétences sont limitées (je pense à Laurent Abi-Rached, Juan Montoya-Burgos, Nicolas Galtier, Sam Dupont ou encore Sabine Stöhr), savoir profiter des restes de financements (qui m'ont quand même permis de faire une mission de cinq semaines en Crète, dont les échantillons obtenus constituent un pilier fondateur de cette thèse), apprendre à répondre à des appels d'offre (qu'ils aient été fructueux ou non, c'est le jeu), entrevoir l'organisation et l'énergie nécessaires à l'organisation d'un congrès (avec le PPD) ou encore obtenir différents devis de plateformes et entreprises (parce que les fonds ne sont pas illimités).

Merci pour ta patience, car les concepts d'évolution et de génétique des populations ne sont pas forcément évident au premier abord. Merci également pour ton écoute, car j'ai toujours senti que nous fonctionnions comme une véritable collaboration, où les idées et les propositions allaient dans les deux sens, et où le raisonnement scientifique se construisait ensemble. Merci pour ta disponibilité, car je ne me suis jamais fait remballer quand je débarquais à l'improviste dans ton bureau. Merci également pour ta rapidité de correction ; j'avais à peine le temps de t'envoyer un chapitre qu'il était déjà de retour dans ma boîte mail. Finalement, merci pour ta confiance, car c'est en faisant que l'on apprend, et j'ai l'impression d'avoir beaucoup appris.

Je tiens également à remercier chaleureusement Laurent Abi-Rached, pour m'avoir donné beaucoup de son temps précieux au cours de ces trois mois à Saint-Charles, et par mail ou téléphone par la suite. Merci d'être sorti de tes thématiques de recherches pour m'encadrer (la variabilité des gènes du système immunitaire humain étant quand même, à priori, éloignée de l'étude des ophiures). Je te remercie de m'avoir initiée au monde jusqu'alors inconnu de la bioinformatique, et d'avoir partagé, sans retenue, tes nombreuses connaissances et scripts. Merci pour ta patience, et pour ton attitude toujours calme et apaisante. Merci de m'avoir transmis, au moins un peu je l'espère, de ta grande rigueur scientifique, en m'apprenant à me poser les bonnes questions et à choisir une démarche adéquate pour y répondre. Grâce

à toi, les lignes de commande, les scripts et la petite fenêtre noire de Linux ne me font plus (trop) peur, merci pour ça.

Cette collaboration n'aurait pas été possible sans l'aide de Pierre Pontarotti. Merci de m'avoir accueillie au sein de ton équipe, d'avoir partagé ton expérience sur la recherche au cours des nombreuses pauses déjeuner, et surtout de m'avoir donné le conseil de toujours se défier soi-même, d'oser voir grand, et que faire de la bonne science est à la portée de chacun, à condition d'y mettre le temps, l'énergie et la rigueur nécessaires.

Je voudrais aussi remercier tous les membres de mon comité de thèse qui ont suivi mon travail au cours de ces trois années, à savoir Didier Aurelle, François Bonhomme, Juan Montoya-Burgos, Laurent Abi-Rached et Jean-Pierre Féral, pour leurs précieux conseils et idées pour mener à bien ce sujet de thèse. Merci également à tous les membres du jury, Nicolas Galtier, Harilaos Lessios et Nicolas Bierne, pour avoir accepté de donner de leur temps et de leur énergie afin d'évaluer ce travail de thèse.

Ces trois années de thèse n'ont pas seulement inclus de la recherche, mais aussi de l'enseignement. Je voudrais remercier mes collègues de TP et TD, Didier Aurelle, Emmanuelle Renard, Carole Borchiellini et Gwenola Simon, pour m'avoir appris à jongler entre enseignement et recherche (bien plus que 13 km séparent ces deux mondes), à s'adapter à tous type d'étudiants (les sérieux, les motivés, les effacés, les nonchalants, les interrogateurs, les négociateurs et les insolents), à faire preuve de spontanéité lors des dysfonctionnements de l'administration de la fac, et pour m'avoir initiée (même si ce n'était que pour un faible coefficient) à la lourde tâche de noter des copies. J'ai pris goût à la passion de transmettre, et j'espère pouvoir continuer dans cette voie à l'avenir.

Un grand merci à mes collègues de bureau, les passionnés du vélo, Anne Haguenauer et Didier Aurelle, pour l'ambiance de ce bureau, qu'elle soit travailleuse, festive ou blagueuse, mais toujours agrémentée de café et de gâteaux ramenés des quatre coins du monde. Merci Anne pour tes astuces pour faire fonctionner des manips récalcitrantes, pour ton expérience inégalée en biologie moléculaire et aussi pour ta passion commune pour un certain pays où l'on consomme volontiers du cervelas, du cenovis, du rivella et des branches Cailler. Merci Didier pour m'avoir transmis ta passion pour la génétique des populations en master, je n'en serais pas là aujourd'hui si je n'avais pas suivi tes cours. La transition de maître de conférences à collègue de bureau était un peu étrange au début, mais les nombreuses blagues et jeux de mots ont bien contribué à briser la glace. Merci pour tes éclaircissements sur des notions de génétique parfois brumeuses, mais aussi pour les astuces à connaître et les exemples de fichier lorsque l'on utilise des programmes comme Arlequin, Genetix, IMa2 ou encore migrate.

Et bien sûr, dans le bureau d'à côté, je tiens à remercier Joëlle, non seulement une secrétaire motivée, travailleuse et compétente, mais surtout un peu la maman de tous les doctorants et stagiaires de la station. Merci pour ton aide précieuse lors de nombreuses et incompréhensibles (pour moi en tout cas) procédures administratives, mais aussi pour ton sens de l'écoute. Je tiens également à remercier les membres (et anciens membres) du SCBM, Caroline Rocher, Cécile Ribout et Marjorie Selva, pour leur aide dans les laboratoires de biologie moléculaire tout au long de ces trois ans. Un grand merci spécial pour

Marjorie, qui a filtré, chaque semaine pendant trois mois, de nombreux litres d'eau au cours de l'expérience de thermotolérance en aquarium. Je remercie aussi Pascal Mirleau et Marc Bally pour leurs conseils avisés lors de l'expérience de thermotolérance, parce que faire de la bonne science c'est aussi savoir s'arrêter quand nos données sont limitées.

Merci à Frédérique Zuberer, Laurent Vanbostal, Romain David, Helmut Zibrowius (et les plongeurs/plongeuses que j'ai pu oublier) pour les nombreuses plongées pour échantillonner des *Ophioderma*. Merci Fred pour tes magnifiques photos qui m'ont permis de faire la couverture de Comptes Rendus Biologies, et qui illustrent le texte de ce manuscrit. Un grand merci à Christian Marschal pour avoir partagé ses connaissances sur la fixation, l'inclusion et la préparation de coupes histologiques. Tout comme aucun organisme ne résiste à l'araldite et à la scie de Christian, aucune fille ne résiste à ses blagues, souvent douteuses, mais toujours drôles. Merci également à Chantal Mahé-Bézac pour son aide lors de la préparation des échantillons pour le microscope électronique à balayage.

Parce que la thèse, c'est quand même plus fun quand on est entouré, je voudrais remercier tous mes amis, scientifiques ou non, qui ont été présents au cours de ces trois années de thèse. En premier, merci à toi Quentin, parce que même si on est collègues au bureau, colocs à la maison ou adversaires sur le terrain, tu arrives toujours à me supporter au quotidien. Avec toi on ne se prend jamais la tête, c'est toujours facile. Merci pour tes idées folles, tes conseils, ton humour, ta combattivité. Merci de me pousser à donner le meilleur de moi-même, dans toutes les situations. J'espère t'avoir donné au moins autant que tout ce que tu m'as apporté au cours de ces trois ans. En sautant du coq à l'âne, je voudrais te remercier, toi qui te reconnaîtras, pour être gaie comme un pinson, et malgré le fait que tu ne te lèves jamais avec les poules, tu sais toujours prendre le taureau par les cornes. Tu ne prends jamais la mouche, et j'ai toujours apprécié nos nombreux verres pris entre chien et loup. Merci Laura, pour ta bonne humeur, ta générosité, ton assiduité à suivre mes recherches, pour ton écoute et tes conseils, ta sincérité, et pour ton empathie, dans les bons comme les mauvais moments. Et bien sûr, il y a la dernière arrivée, Marine (Pratlong, parce vu le nombre de Marine à la station marine, ça vaut la peine de spécifier). Merci pour ta générosité sans limites, parce que même si j'ai tendance à me moquer (pas trop j'espère), de ton côté « main tendue », tu ne m'as jamais refusé ton aide et ton temps. Merci pour les articles, les scripts perl, les astuces pour débloquer R, et merci aussi pour m'avoir ramené un petit quelque chose de chacun de tes voyages, que ce soit des gâteaux divers et variés, du cidre ou des même ophiures. Merci à Abigail pour avoir corrigé l'anglais du chapitre connectivité, ajouté de précieux commentaires et références, et ce en un temps record ! Merci à tous les potes de la station et d'ailleurs aussi, et j'espère ne pas en oublier, Clémentine, Isabelle, Romain, César, Pauline, Sandrine, Marie (R), Marine (Périot), Marine (Pouget), Anaïs, Mathieu (S) et Chiara. Merci aux anciens doctorants, Marie (LR), Pierre-Alex, Marion, Rosa et Emilie (E) pour m'avoir expliqué les coulisses de la station marine, et partagé leur expérience sur la thèse. Un merci spécial pour Manu et Mathieu, pour avoir supporté sans broncher de nombreuses soirées à parler boulot.

Parce qu'une thèse en génétique des populations, ça requiert un certain nombre d'échantillons à aller chercher ou se faire envoyer, je voudrais remercier chaleureusement Zined Marzouk, pour m'avoir envoyé des *Ophioderma* tellement intéressantes qu'une mission d'échantillonnage en Tunisie s'est avérée

nécessaire. Merci ensuite de m'avoir guidée, accueillie chez toi, accompagnée et d'avoir organisé toutes les plongées au cours de ces deux semaines. En bus ou en louage, nous avons parcouru de bout en bout la Tunisie, en passant par Tabarka, Kelibia, Monastir et Djerba. Merci de m'avoir fait découvrir ton pays, et également de m'avoir ouvert les yeux sur ce que c'est d'être une femme en Tunisie.

I will now switch in English to thank gratefully Christos Arvanitidis, for welcoming me in his lab (HCMR, Crete) for five weeks. This sampling in Crete was essential for my work, as it allowed me to collect precious samples. Thank you for your availability, your constant willingness to help, the dive on the South coast, and for bringing me in excellent restaurants where I could taste typical Cretan food. Thank you to Thanos Dailianis, Elena Sarropoulou and Magdalini Christodoulou for help during the sampling sessions and laboratory work. Many thanks to Panos Grigoriou for giving me the possibility to use the aquarium facilities at the Crete aquarium and for his advice. I also would like to thank Katerina Vasileiadou, Christina Pavloudi and Wanda Plaiti for their help in the lab and availability. I also remember a shorter but not less important sampling trip in Faro and Sagres. I would like to thank Joao Reis for his kindness, his availability and help with the aquarium facilities. Many thanks to Joao Rodrigues for the sampling in Sagres. Although quite difficult to reach, the diving spot was worth it.

Je voudrais remercier tous les (actuels et futurs) co-auteurs des articles qui découlent de cette thèse. Un grand merci à Sam Dupont, pour avoir partagé ses connaissances et conseils pour la mise en place de l'expérience de thermotolérance et régénération, ainsi que pour l'analyse des données. Many thanks to Sabine Stöhr, who patiently cut more than 200 arm pieces of *Ophioderma* from the museum collection, put them in individual tubes, labeled them and sent them to me for genetic analyses. Thank you for teaching me to recognize *Ophioderma* skeletal parts, for your help with the writing of articles, and for finding a very nice name for the new species *Ophioderma vivipara*; I hope we'll find the time to describe it. Un grand merci à Juan Montoya-Burgos et Chloé Hot pour leur temps, et pour m'avoir permis d'utiliser le programme PSG finder, qui a mis en évidence les gènes sous sélection positive. Finalement, je tiens à remercier chaleureusement Nicolas Galtier et Aurélien Bernard, pour avoir traité les données brutes des transcriptomes, pour l'utilisation de leur pipeline et pour ces résultats intéressants sur la diversité génétique. Merci encore pour cette collaboration qui se termine avec un article dans Nature.

Un grand merci à mes parents, à mon frère et à mon grand-père, pour leur soutien continu, tout au long ces huit années d'études supérieures. Proches ou loin, vous étiez toujours là pour moi. Merci à ma mère pour m'avoir transmis sa passion pour la biologie. Je voudrais aussi remercier les copains et copines d'enfance, parce qu'une amitié âgée de plus de 20 ans et qui dure toujours aujourd'hui, ça se chérit. Merci à Paola, Kim, Giusy et Amanda, parce qu'avec vous, qu'on se voie tous les jours ou une fois par année, c'est toujours pareil, c'est facile. Une petite pensée pour Sigi, qui m'a supportée sans râler à la fin de cette thèse, merci pour ça.

Finalement, je voudrais te remercier, Philipp, pour tes constants encouragements et ton soutien inconditionnel. Ca n'a pas été une décision facile de partir pour Marseille, tout d'abord pour le master, puis pour le doctorat. Malgré tout, tu étais toujours là pour effacer les doutes que j'avais, pour structurer mes

idées quand je sentais que je partais dans tous les sens, pour m'encourager quand ma motivation n'était plus là, pour m'apprendre à manger un éléphant, mais aussi pour me dire de profiter quand le succès était au rendez-vous. Tu m'as toujours supportée quand j'étais fâchée, stressée, triste et carrément insupportable, surtout à la fin, quand mon humeur changeait toutes les minutes. Merci pour ta patience inébranlable et pour tes conseils toujours pertinents. Tu crois toujours en moi plus que j'y crois moi-même. Merci d'être toi.

Parce que j'ai peut-être pu oublier certaines personnes, je remercie tous ceux qui, de près ou de loin, ont contribué à l'aboutissement de cette thèse.

# Table des matières

**Remerciements** 5

**Chapitre 1: Introduction générale** 19

1.1 Le concept d'espèce 19

1.1.1 Origines 19

1.1.2 Le concept biologique de l'espèce 19

1.1.3 Le concept unifié de l'espèce 20

1.2 L'outil moléculaire pour délimiter les espèces 22

1.3 Les espèces cryptiques : exemples en milieu marin 25

1.4 Traits d'histoire de vie 27

1.5 Le modèle d'étude 29

1.6 Objectifs de la thèse 32

**Chapitre 2: Les données génétiques, la saison et la stratégie de reproduction soutiennent l'existence d'espèces biologiques chez *Ophioderma longicauda*** 36

2.1 Résumé 36

2.2 Introduction 38

2.3 Methods 39

2.3.1 Sampling 39

2.3.2 DNA extraction, COI sequencing and analyses 40

2.3.3 i51 genotyping and analyses 40

2.3.4 Morphological, reproductive and developmental observations 41

2.4 Results 42

2.4.1 Genetic analyses – COI and i51 42

2.4.2 Morphology, reproduction and development 42

2.5 Discussion 48

2.5.1 Morphological differences 48

2.5.2 Habitat difference to explain L1 distribution 48

2.5.3 Reproduction strategies and reproductive isolation between L1 and L3 49

**Chapitre 3: Délimitation d'espèces et datation de la divergence dans le complexe *Ophioderma longicauda*** 52

3.1 Résumé 52

3.2 Introduction 56

3.3 Material and methods 57

3.3.1 Sampling, DNA extraction and markers development 57

3.3.2 Amplicon sequencing and dataset cleaning 58

3.3.3 Gene-by-gene phylogenetic analyses 58

3.3.4 Concatenated phylogenetic analysis ..... 59
3.3.5 Genetic clustering using Discriminant Analysis of Principal Components (DAPC) .. 59
3.3.6 Divergence model testing and molecular dating ..... 59
3.4 Results ..... 63
3.4.1 Single gene analysis shows incomplete lineage sorting ..... 63
3.4.2 Concatenated dataset provides unresolved tree ..... 69
3.4.3 DAPC analysis identifies six genetic clusters ..... 69
3.4.4 Divergence scenario suggest hybridization event ..... 72
3.5 Discussion ..... 76
3.5.1 Species delimitation in *O. longicauda* ..... 76
3.5.2 Divergence and speciation history in *Ophioderma* genus: a hypothesis ..... 78
3.5.3 COI barcodes: benefits and pitfalls ..... 78
3.5.4 DAPC and DIYABC: an efficient combination of methods for accurate delimitation of recent/incipient species ..... 79
3.6 Extended Methods – COI control ..... 83

**Chapitre 4: Thermotolérance et régénération dans le complexe d'espèces d'ophiures *Ophioderma longicauda* : une étude préliminaire comparant lignées et bassins Méditerranéens ..... 86**
4.1 Résumé ..... 86
4.2 Introduction ..... 89
4.3 Material and methods ..... 91
4.3.1 Specimen collection and genetic analyses ..... 91
4.3.2 Experimental design ..... 92
4.3.3 Animal preparation and measured parameters ..... 92
4.3.4 Statistical analyses ..... 93
4.4 Results ..... 94
4.4.1 Survival and autotomy ..... 94
4.4.2 Regeneration ..... 97
4.5 Discussion ..... 102
4.5.1 Survival and autotomy ..... 102
4.5.2 Regeneration ..... 104

**Chapitre 5: Influence de la phase larvaire sur la structure génétique, la connectivité et la diversité génétique: comparaison entre incubants et dispersants ..... 108**
5.1 Résumé ..... 108
5.2 Introduction ..... 111
5.3 Methods ..... 114
5.3.1 Sampling, DNA extraction, COI sequencing, i51 genotyping and analyses ..... 114
5.3.2 Transcriptome based nuclear markers ..... 118

5.3.3 Comparative connectivity between L1 and L3 in Crete and Rhodes ........................ 118
5.4 Results ........................ 119
5.4.1 Large scale analyses: COI and i51 ........................ 119
5.4.2 Small scale analyses: comparison of genetic structure and migrant estimation in common locations of L1 and L3 using multilocus genotypes ........................ 127
5.4.3 Comparison of genetic diversity between brooders and broadcasters ........................ 127
5.5 Discussion ........................ 131
5.5.1 Within population genetic structure ........................ 131
5.5.2 Influence of life history traits on genetic diversity ........................ 131
5.5.3 Influence of life history traits on genetic structure and connectivity ........................ 132

**Chapitre 6: Transcriptomique comparée entre incubants et dispersants : description des transcriptomes, diversité génétique et gènes sous sélection positive ........................ 138**
6.1 Résumé ........................ 138
6.2 Introduction ........................ 142
6.3 Material and methods ........................ 144
6.3.1 Individual sampling, lineage typing and RNA extraction ........................ 144
6.3.2 Illumina sequencing, assembly and SNP calling ........................ 145
6.3.3 Research and annotation of orthologous genes ........................ 145
6.3.4 Positively selected genes and gamete recognition genes ........................ 146
6.4 Results ........................ 146
6.4.1 Assembly statistics ........................ 146
6.4.2 Annotation of orthologous genes ........................ 147
6.4.3 Comparison of genetic diversity between brooders and broadcasters ........................ 149
6.4.4 Positive selection detected in proteins involved in sperm competition ........................ 149
6.5 Discussion ........................ 154
6.5.1 Genetic diversity is lower in brooders ........................ 154
6.5.2 Mechanisms of prezygotic isolation ........................ 155
6.5.3 Adaptation to brooding and/or reinforcement explains the positive selection observed in Ophioderma longicauda reproduction genes? ........................ 156
6.5.4 Sperm chemotaxis and sperm competition: a common mechanism of species recognition in brittle stars? ........................ 157
6.6 Extended Methods ........................ 159
6.7 Extended methodological discussion: issues in transcriptome generation and polymorphism analysis ........................ 160

**Chapitre 7: Discussion et perspectives ........................ 164**
7.1 Discussion générale ........................ 164
7.1.1 Les espèces d'*Ophioderma longicauda* ........................ 164

7.1.2 Biogéographie, divergence, incubation, hybridation et renforcement : une histoire évolutive complexe chez *Ophioderma longicauda* ..... 166
7.1.3 Avantages et désavantages évolutifs de l'incubation ..... 167
7.1.4 Apports des NGS à l'étude des espèces non-modèles : nouvelles données, nouvelles compétences et nouveaux biais ..... 169
7.2 Perspectives ..... 171
7.2.1 Etude des gènes sous sélection positive dans chaque cluster ..... 171
7.2.2 Génome scans d'adaptation locale ..... 173
7.3 Conclusion ..... 174
**Chapitre 8: Echinoderm phylogeny finally solved: a review of past pitfalls, from biological to methodological ones (Annexe 1) ..... 178**
8.1 Abstract ..... 178
8.2 Introduction ..... 179
8.3 Material & Methods ..... 182
8.3.1 Transcriptome generation and retrieving from databases ..... 182
8.3.2 Identification of orthologous gene sets ..... 182
8.3.3 Gene-by-gene phylogenetic analysis ..... 183
8.3.4 Concatenated phylogenetic analysis ..... 183
8.3.5 Comparison with previous studies ..... 184
8.4 Results ..... 184
8.4.1 Gene-by-gene analysis cannot resolve class-level Echinoderm phylogeny ..... 184
8.4.2 Concatenated dataset analysis fully resolves Echinoderm phylogeny ..... 185
8.4.3 Comparison with previous studies ..... 186
8.5 Discussion ..... 191
8.5.1 Asterozoa strongly supported ..... 191
8.5.2 Difficulties in phylogenetic reconstruction ..... 191
8.5.3 Removal of fast-evolving and constant sites: Not a good idea? ..... 192
8.5.4 Impact of missing data on phylogenetic reconstruction ..... 193
**Annexe 2 ..... 195**
**Annexe 3 ..... 205**
**Annexe 4 ..... 207**
**Références ..... 209**

# Chapitre 1

-

# Introduction générale

# Chapitre 1: **Introduction générale**

## 1.1 Le concept d'espèce

### 1.1.1 *Origines*

De tous temps, l'Homme a cherché à classifier le vivant. Carl von Linné, naturaliste suédois, publia en 1735 *Systema Naturae*, un ouvrage majeur décrivant les bases du système moderne de la nomenclature binomiale. D'après lui, les espèces sont des entités fixes, créées par Dieu lors de la Genèse, qui se différencient sur la base de critères morphologiques. En 1809 Lamarck, naturaliste français, introduit la notion d'évolution des espèces via le transformisme, qui bien que basée sur un mécanisme erroné (la « volonté » d'un organisme est à l'origine de son changement), implique l'adaptation des êtres vivants à leur milieu ; les espèces ne sont donc plus des entités fixes. Ce dernier est d'ailleurs l'un des fondateurs de la science et du terme « biologie », en tant que science de la vie. Par la suite, Darwin, naturaliste et biologiste anglais, publie son ouvrage *De l'origine des espèces* (1859), dans lequel il indique que la diversité du vivant s'explique par l'évolution, et que le moteur de l'évolution adaptative est la sélection naturelle. Il introduit les principes de variation (les individus diffèrent les uns des autres), d'adaptation (les individus les plus adaptés au milieu survivent et se reproduisent davantage) et d'hérédité (pour qu'ils puissent se maintenir à travers les générations, les caractères adaptatifs doivent être héréditaires).

### 1.1.2 *Le concept biologique de l'espèce*

Ernst Mayr, biologiste évolutionniste allemand, énonce le concept biologique de l'espèce (BSC) en 1942, basé sur l'isolement reproducteur. Ce concept est largement repris dans la littérature, lorsque l'on traite d'organismes à reproduction sexuée. Ainsi, les espèces sont des groupes d'individus qui se reproduisent effectivement ou potentiellement entre eux et dont la descendance est féconde. Les limites d'espèces sont définies comme des barrières intrinsèques aux flux de gènes et qui ont une base génétique. Ces barrières peuvent être pré-zygotiques, intervenant donc avant la fécondation entre gamètes d'origine différente, et incluent notamment le choix des partenaires sexuels, des préférences d'habitats (réduisant ainsi les probabilités de rencontre), les systèmes d'attraction et de fécondation des gamètes, ou encore la

période de reproduction. À l'opposé, ces barrières peuvent être post-zygotiques, représentant des incompatibilités génomiques qui se traduisent par une valeur sélective inférieure des hybrides, pouvant aller jusqu'à la mortalité des embryons, ou encore leur stérilité à l'état adulte (Dobzhansky 1934; Gavrilets 2003; Lessios 2007; Bierne *et al.* 2011) lorsque la barrière post-zygotique est totale.

### 1.1.3 *Le concept unifié de l'espèce*

Depuis, de nombreux concepts d'espèces ont été proposés et utilisés; il en existe pas moins de 27 actuellement (Wilkins 2011). Entre autres, il existe le concept écologique, où deux espèces sont considérées différentes si elles ne partagent pas la même niche écologique (Valen 1976; Andersson 1990), le concept phénétique, où les espèces forment des groupes phénétiques différents (avec des différences quantitatives) (Michener 1970; Sokal & Crovello 1970; Sneath *et al.* 1973), ou encore, grâce au développement de l'outil moléculaire, le concept phylogénétique de l'espèce. Les propriétés de ce concept, une seule d'entre elles suffisant à définir un espèce, sont la monophylie (les individus possèdent des caractères ancestraux dérivés) (Rosen 1979; Donoghue 1985; Mishler 1985), la coalescence exclusive des allèles (monophylie réciproque) (Avise & Ball 1990; Baum & Shaw 1995) et la diagnosticabilité (posséder des différences fixées et qualitatives) (Nelson & Platnick 1981; Cracraft 1983; Nixon & Wheeler 1990).

Beaucoup de ces concepts et leur définition associée sont en contradiction, pouvant mener l'analyse d'un même système à différentes conclusions. Ainsi, le concept d'espèce a été de nombreuses fois débattu. La difficulté à définir le concept d'espèces vient en partie du problème de la délimitation des espèces en elle-même ; la question étant de savoir comment déterminer les limites et le nombre d'espèces à partir de données empiriques. De Queiroz (2007) définit alors le concept unifié de l'espèce, incluant les aspects biologique, écologique et phylogénétique. Au lieu de mettre l'accent sur les différences entre concepts, il s'attache à reconnaître ce qu'ont en commun les différents concepts. Une clarification majeure de cette étude a été de séparer le concept d'espèce en soi (la conceptualisation de l'espèce), des critères opérationnels utilisés sur des données empiriques (la délimitation de l'espèce).

En effet, aussi nombreux soient les concepts d'espèces, la majorité d'entre eux s'accordent sur un point (De Queiroz 2007). Les espèces sont des (segments de) lignées de métapopulation qui évoluent séparément. Ceci peut alors être considéré comme une propriété primaire du concept d'espèce. Par contre, les avis diffèrent sur les critères opérationnels à utiliser pour délimiter les espèces, à savoir les propriétés secondaires du concept d'espèce (p. ex isolement reproducteur ; différence phénétique ; différence de niche écologique ; monophylie réciproque). Ces propriétés secondaires supposées nécessaires peuvent mener à des concepts d'espèces incompatibles car elles surviennent à des temps différents dans le processus de spéciation. La figure 1.1 représente schématiquement le processus de spéciation, avec une lignée ancestrale qui diverge en deux lignées filles. Les différents concepts d'espèces (SC) se situent à des endroits différents dans la séparation de ces lignées (zone grise). La raison principale des désaccords concernant le concept d'espèce est l'adoption de différents seuils de différenciation par les différents scientifiques. De plus, ces différents seuils (isolement reproducteur ; niches différentes ; etc) n'apparaissent pas forcément dans le même ordre pour chaque événement de spéciation. Le concept unifié de l'espèce propose que la seule propriété nécessaire pour définir une espèce soit la propriété primaire (lignées de métapopulation qui évoluent séparément). Les propriétés secondaires ne sont pas moins importantes, servant de critères opérationnels pour évaluer la séparation des lignées, et donc la délimitation des espèces. Au lieu d'utiliser un critère opérationnel unique pour délimiter les espèces (ce critère n'étant pas forcément le même dans toutes les disciplines), l'utilisation de n'importe laquelle des propriétés mentionnées ci-dessus permet de délimiter des lignées qui évoluent séparément. Il est forcément plus aisé de déterminer la divergence entre deux lignées si l'on se situe plutôt vers la fin de la « zone grise », où donc plusieurs

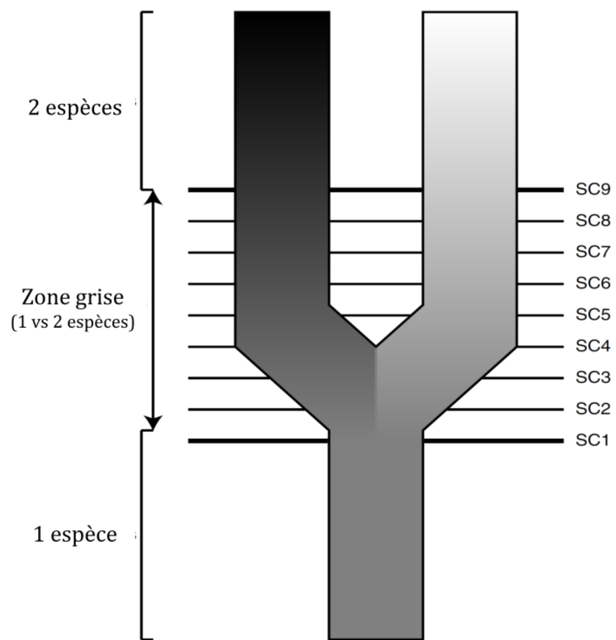



**Figure 1.1:** Représentation schématique du processus de séparation des lignées et divergence. Les différents concepts d'espèce (SC) représentent un stade différent de la séparation. Modifié d'après De Queiroz (2007).

propriétés sont congruentes. Ainsi, les approches multidisciplinaires incluant l'étude de plusieurs propriétés (isolement reproducteur ; différenciation des niches ; monophylie réciproque), communément appelées la taxonomie intégrative, sont actuellement encouragées pour la délimitation d'espèces.

## 1.2 L'outil moléculaire pour délimiter les espèces

L'absence de flux de gènes prolongé entre deux groupes d'individus induit une différenciation génétique de par l'accumulation de mutations fixées et de par la dérive génétique. Des marqueurs génétiques peuvent être alors étudiés, choisis selon la question initiale. Des marqueurs non-codants (p. ex. microsatellites ou introns) sont généralement utilisés pour les études de génétique des populations, au niveau intra-spécifique, où les différences de fréquences des marqueurs sont étudiées. En revanche, pour les études de phylogénie, donc interspécifiques, des marqueurs évoluant plus lentement (p. ex. protéiques) sont généralement choisis (Chenuil 2006).

Le développement de la PCR et du séquençage de l'ADN a ainsi révolutionné la taxonomie moderne. Par exemple, l'arbre phylogénétique des eucaryotes a été complètement remodelé, avec entre autres, l'éclatement des groupes des protozoaires et des algues, regroupant auparavant des taxons non-apparentés (Baldauf 2003). De plus, les animaux et les fungi ne représentent qu'un seul grand groupe d'eucaryotes parmi les cinq actuellement décrits (e.g. Keeling *et al.* (2005)). L'utilisation généralisée de marqueurs universels, dont la variation est censée être représentative des espèces, appelés codes-barres génétiques (p. ex. COI, 16S, 18S), semble alors être une réponse simple (pas besoin d'experts de taxonomie), rapide et efficace pour la classification du vivant (Hebert *et al.* 2003). Elle repose sur la détermination de distance génétique entre taxons. A partir d'un certain seuil de distance, ou divergence, deux taxons sont considérés comme des espèces différentes (2-3% de divergence COI pour les insectes et mammifères, d'après Hebert *et al*. 2003).

Cependant, l'utilisation des codes barres génétiques est sujette à de nombreuses controverses (Will *et al.* 2005; DeSalle *et al.* 2005). L'utilisation d'un seul marqueur, et en particulier tiré du seul génome mitochondrial est particulièrement critiquée (Chenuil 2006; Rubinoff *et al.* 2006; Galtier *et al.* 2009). En effet, cette molécule, bien que facilement amplifiable chez les animaux, et accumulant plus rapidement des mutations

que l'ADN nucléaire, possède des désavantages. Tout d'abord, chez la plupart des animaux, les mitochondries sont héritées de la lignée maternelle, donc tous les processus évolutifs qui n'affectent pas les femelles, ou qui affectent différemment les deux sexes, sont ignorés ou biaisés. Ensuite, l'ADN mitochondrial présente une taille efficace réduite, il est donc plus sensible aux goulots d'étranglement et à l'introgression (Toews & Brelsford 2012; Pons *et al.* 2014). Enfin, comme cette molécule ne recombine pas, un marqueur quelconque a une plus grande probabilité de refléter un signal de balayage sélectif (tout gène étant entrainé par le balayage d'un gène quelconque du génome non recombinant).

Dès lors, tout comme il est douteux d'utiliser un seul caractère morphologique pour la description d'espèces, il peut être risqué d'utiliser un seul marqueur génétique pour définir une espèce. Les code-barres génétiques sont utiles pour la réunion ou le rapprochement de taxons divergent morphologiquement, ou encore comme premier pas vers la découverte de nouvelles espèces, mais soulèvent une confirmation nécessaire en utilisant d'autres marqueurs génétiques, et /ou des données morphologiques / écologiques / comportementales, en s'appuyant sur la notion de taxonomie intégrative (Will *et al.* 2005; DeSalle *et al.* 2005; Rubinoff *et al.* 2006) et en encourageant donc, dans la mesure du possible, les approches multidisciplinaires.

De plus, la notion de seuil de différenciation génétique entre la variabilité intraspécifique et interspécifique est arbitraire, car elle sous-entend qu'il existe un « écart de code-barres » (barcoding gap) entre ces deux variabilités (Meyer & Paulay 2005). Or, il existe un chevauchement entre ces deux variabilités dans la plupart des espèces étudiées (Goldstein *et al.* 2000; Meyer & Paulay 2005). D'un autre côté, lorsque cet écart de code-barres est bien réel, rien n'indique que la valeur seuil de différenciation entre espèces soit la même dans un groupe taxonomique entier, le contraire ayant plutôt été montré (Weber & Pawlowski 2014). Ainsi s'impose l'utilisation de plusieurs marqueurs génétiques pour tenter de définir des espèces au niveau moléculaire.

Lorsque plusieurs marqueurs sont étudiés et que leur topologie est congruente, séparant deux espèces potentielles, on parle de monophylie réciproque. Il s'agit du cas le plus « simple » pour tester si deux espèces potentielles échangent des gènes, si ces dernières sont retrouvées en sympatrie. Chaque marqueur étudié a une histoire qui

reflète l'histoire de la spéciation. En revanche, si les espèces vivent en allopatrie, il devient difficile de déterminer si les espèces montrent vraiment un isolement reproducteur, ou si l'on observe uniquement un effet d'isolement géographique, avec une divergence due à la dérive génétique ou de l'adaptation locale (Bickford *et al.* 2007).

Il existe certains cas où la monophylie réciproque ne s'applique pas, bien que les espèces soient effectivement séparées. Dans un premier cas, si la spéciation est trop récente, il peut y avoir une rétention de polymorphisme ancestral qui se traduit par un tri incomplet des lignées (incomplete lineage sorting) (Nichols 2001). Deuxièmement, les marqueurs étudiés peuvent être sous l'effet de la sélection naturelle, induisant des phénomènes de convergence ou de rétention. Finalement, deux espèces séparées en allopatrie peuvent ne pas avoir développé de mécanismes d'isolement reproducteur. Ainsi, si les espèces se retrouvent par contact secondaire, elles peuvent s'hybrider et induire de l'incongruence dans le signal des gènes étudiés.

Des méthodes robustes utilisant des modèles probabilistes avec des données multilocus simulées ont ensuite été développées afin de délimiter les espèces même en présence d'arbres de gènes non-monophylétiques (Maddison & Knowles 2006; Knowles & Carstens 2007). Ainsi, Maddison et Knowles (2006) ont montré que même en présence de forte rétention de polymorphisme ancestral, le « vrai » arbre d'espèces pouvait être retrouvé dans la majorité des simulations, pour autant qu'un nombre suffisant d'individus ait été échantillonné. Ces simulations sont basées sur la comparaison de différents modèles (différents arbres d'espèces possibles) lors d'une approche basée sur la coalescence des gènes. Les arbres de gènes sont simulés d'après un processus de coalescence neutre, sans flux de gènes. Ensuite, la probabilité de chaque modèle est estimée, indiquant l'arbre d'espèces le plus probable (Knowles & Carstens 2007). Plus récemment, des méthodes en deux temps ont été développées. La première étape consiste à définir sans à priori le nombre de groupes génétiques présent dans l'échantillon observé. Ceci peut se faire, entre autres, avec les méthodes Structurama (Huelsenbeck & Andolfatto 2007; Huelsenbeck *et al.* 2011), Structure (Pritchard *et al.* 2000) ou encore DAPC (Jombart *et al.* 2010) qui utilisent des données de génotypes multilocus. Ensuite, l'arbre d'espèces le plus probable est estimé d'après des simulations d'arbres de gènes, ce qui peut être estimé grâce aux méthodes BPP (Yang & Rannala 2010), spedeSTEM (Ence & Carstens 2011) ou encore ABC (p.ex. Csilléry *et al.* (2010)).

Les méthodes citées précédemment ont été utiles pour la découverte et la délimitation d'espèces cryptiques (Carstens *et al.* 2013), nombreuses en milieu marin.

### 1.3 Les espèces cryptiques : exemples en milieu marin

Les espèces étant originellement décrites sur la base de critères morphologiques, de nombreuses espèces ont été incorrectement décrites sous le même nom de par leur similarité morphologique. Ces dernières sont communément appelées espèces cryptiques, ou espèces jumelles (sibling species). Mayr, à l'origine du concept biologique de l'espèce, fut également le premier à décrire le concept d'espèces cryptiques (Mayr 1947) et à critiquer la classification des espèces uniquement basée sur des critères morphologiques (Mayr 1963). Knowlton (1993) fit une première revue bibliographique des espèces cryptiques marines, et conclut que ces espèces étaient « omniprésentes », pouvant être trouvées chez tous les taxons, dans tous les milieux et à toute profondeur.

La prévalence d'espèces marines cryptiques peut s'expliquer par plusieurs facteurs. Tout d'abord, à l'origine, les espèces marines ont été moins étudiées que les espèces terrestres à cause de leur difficulté d'accès, et de la difficulté d'obtention de nombreux individus nécessaire à une description morphologique adéquate. En effet, l'échantillonnage d'espèces dans certains milieux (p. ex. abysses ou polaires) s'avère en soi complexe. De plus, de nombreux caractères morphologiques potentiellement diagnostiques (tissus mous avec structures particulières ; couleurs) peuvent être perdus à cause des méthodes de conservation des échantillons (éthanol, formaldéhyde ou encore séchés). Finalement, le milieu océanique présentant (de prime abord) moins de barrières physiques à la dispersion que le milieu terrestre, il était communément admis que certaines espèces pouvaient montrer de très larges distributions géographiques, alors qu'une étude plus approfondie aurait révélé l'existence d'espèces cryptiques, supposées allopatriques.

Le développement des études moléculaires, et en particulier l'apparition de la PCR (en 1986), ont induit un développement exponentiel des études sur les espèces cryptiques (Bickford *et al.* 2007). En milieu marin, certains groupes taxonomiques semblent inclure plus d'espèces cryptiques que d'autres. Ainsi, les cas d'espèces cryptiques sont nombreux chez les algues rouges, les éponges ou encore les gastropodes. Dans le cas des algues et des éponges, une explication possible serait la "simplicité"

morphologique et grande plasticité phénotypique de ces organismes (Knowlton 2000). Pourtant, Pfenninger & Schwenk (2007) ont montré, parmi 2'207 cas reportés d'espèces cryptiques, que le nombre d'espèces cryptiques est relativement équitablement distribué entre taxons majeurs de métazoaires et régions biogéographiques (après correction pour la richesse spécifique et le nombre d'études). Deux ans plus tard, Trontelj & Fišer (2009), critiquent cette dernière étude en pointant des erreurs méthodologiques (transformation log des données inadéquate ; usage de groupes non-monophylétiques ; catégories arbitraires p. ex. « autres arthropodes », « autres mollusques »). Après ré-analyse, les auteurs trouvent des différences d'un ordre de magnitude entre taxons majeurs, même en excluant les taxons extrêmes. Ces derniers recommandent ainsi de se concentrer sur des études poussées au niveau du genre (ou inférieur) pour étudier les phénomènes à l'origine de la diversité cryptique au sein de chaque taxon, et non pas généraliser sur les processus reliés à la spéciation qui peuvent s'avérer complexes.

Plusieurs explications peuvent être à l'origine de l'absence de différences morphologiques diagnostiques entre deux espèces cryptiques (Bickford *et al.* 2007). Premièrement, la divergence entre espèces est trop récente pour que des différences diagnostiques morphologiques aient pu se mettre en place. Une seconde explication à ce phénomène est que les signaux liés à la reproduction ne sont pas visuels, mais acoustiques ou chimiques (phéromones ; protéines de reconnaissance gamétique), comme c'est le cas pour la majorité des espèces marines (Knowlton 1993). Troisièmement, il peut exister une pression de sélection stabilisante pour maintenir certaines caractéristiques morphologiques dans des environnements très spécifiques ou extrêmes, on parle alors de stase morphologique (Wake *et al.* 1983). Dans ce cas là, les espèces ne sont pas forcément proches phylogénétiquement. Finalement, des espèces vivant dans des environnements très variables ou hétérogènes peuvent avoir intérêt à préserver une grande variabilité morphologique, qui peut être déterminée par une variabilité génétique ou une forte plasticité phénotypique, montrant alors des morphologies très variables, et donc chevauchantes avec d'autres espèces proches (p. ex. le complexe d'espèces *Echinocardium cordatum*) (Egea (2011); Egea *et al.* soumis). Cela pourrait être le cas de nombreuses espèces marines à forte dispersion larvaire (plutôt passive). Lorsque la fécondité est forte, l'espèce peut supporter le fardeau lié à la

sélection naturelle, sinon la plasticité représente une meilleure solution pour maintenir une variabilité de forme.

L'étude des espèces cryptiques est importante afin d'avoir une meilleure connaissance de la biodiversité en général, mais aussi dans le domaine de la conservation des espèces et de la gestion des écosystèmes (Trontelj & Fišer 2009). Tout d'abord, si des espèces utilisées comme indicateurs de la qualité environnementale s'avèrent être des espèces cryptiques, des conclusions erronées peuvent être basées sur la comparaison d'espèces différentes dans des milieux différents. De même, la connaissance d'espèces cryptiques avec différents besoins de conservation (ayant des statuts IUCN différents) peut permettre de mettre en place des programmes de protection plus efficaces, en définissant des unités de conservation adaptées (Daugherty *et al.* 1990; Bowen *et al.* 1993; Schönrogge *et al.* 2002; Ravaoarimanana *et al.* 2004; Russello *et al.* 2005). En outre, des espèces à valeur marchande (p. ex. les moules *Mytilus* spp. (Rawson & Hilbish 1995); le poisson *Mugil cephalus* (Whitfield *et al.* 2012) qui s'avèrent être des complexes d'espèces cryptiques peuvent poser des difficultés de maintien si elles ne montrent pas les mêmes besoins (p. ex. nourriture, température, salinité). Enfin, les espèces invasives qui s'avèrent cryptiques peuvent masquer le déclin d'une espèce native (Geller 1999) et peuvent également être difficiles à contenir et/ou éradiquer (Laporte *et al.* 2014).

En tant que modèle d'étude, les espèces cryptiques ont un intérêt particulier dans le sens où elles peuvent constituer des « réplicas » naturels pour étudier, par exemple, les phénomènes de spéciation ou d'adaptation locale. Si elles montrent la même écologie, les mêmes traits d'histoire de vie et qu'elles vivent en sympatrie, la variance naturelle de ces espèces peut être étudiée ; elles représentent alors l'importance des phénomènes de contingence (Annexe 2; Romiguier *et al.* (2014), Boissin *et al*. soumis). En revanche, si ces espèces diffèrent pour un seul trait (p.ex. un trait d'histoire de vie), des études comparatives peuvent être menées afin d'étudier l'effet de ce trait sans facteurs confondants (facteurs environnementaux ou facteurs historiques par exemple).

### 1.4 Traits d'histoire de vie

Les traits d'histoire de vie, ou traits biologiques, sont des variables descriptives influençant la croissance, la reproduction ou la survie d'une espèce. Par exemple, on

peut citer l'âge et la taille à maturité sexuelle, le nombre et la taille des descendants, la stratégie de reproduction et l'investissement parental (Stearns 1992). L'étude la variation de ces traits a permis de définir la théorie des traits d'histoire de vie. Cette dernière permet de prédire quel type de stratégie est sélectionné pour produire le nombre de descendants le plus élevé en fonction des différents environnements (Stearns 1977). Dans un environnement stable, les trais favorisés par la sélection naturelle incluent un âge à maturité tardif, un effort reproducteur faible et une forte longévité, favorisant donc la survie (stratèges K) (MacArthur & Wilson 1967). A contrario, dans un environnement fluctuant, les traits favorisés incluent un jeune âge à maturité, un fort effort reproducteur et une longévité faible, favorisant donc la reproduction (stratèges r). Ces différentes stratégies impliquent donc des capacités de résilience (récupération après perturbation) différentes, les stratèges r ayant une forte résilience et les stratèges K une faible résilience.

Ceci peut s'expliquer par le fait que l'adoption de certains traits d'histoire de vie, par exemple la stratégie de reproduction (et la capacité de dispersion), a des conséquences directes sur la connectivité, la taille efficace et la diversité génétique d'une espèce (Romiguier *et al.* 2014). Chez les espèces à forte fécondité et forte capacité de dispersion, une grande aire de répartition est observée avec une forte connectivité entre populations, et donc une faible structuration génétique (Bohonak 1999). De plus, la taille efficace des populations étant élevée, une forte diversité génétique est observée, notamment au niveau local. Au contraire, pour les espèces à faible fécondité et faible capacité de dispersion, on observe une faible connectivité des populations qui implique une forte structuration génétique ce qui a pour conséquences, dans les cas écologiquement réalistes avec variation du succès reproducteur entre dèmes, une taille efficace réduite (Whitlock 2004) même au niveau global. La conséquence d'une faible taille efficace des populations est un effet fort de la dérive génétique. Un faible effectif efficace implique une diversité génétique plus faible chez ces espèces, tant au niveau de l'hétérozygotie attendue He (equ. 1.1 ; Ne = taille efficace ; μ = taux de mutation) ou de la diversité nucléotidique π (equ. 1.2) attendues à l'équilibre.

$$He = \frac{4\ \mathrm{Ne}\ \mu}{1+4\ \mathrm{Ne}\ \mu} \qquad \text{Equation 1.1}$$

$$\pi = 4\ \mathrm{Ne}\ \mu \qquad \text{Equation 1.2}$$

La probabilité de fixation d'une mutation dépend de sa valeur sélective *s* et de l'effectif efficace, Ne. Si ce ratio est négligeable devant 1, la mutation aura le comportement d'une mutation neutre. Dès lors, les mutations génétiques légèrement délétères, pour une valeur de *s* donnée ont une plus grande probabilité de se fixer (Ohta, 1992) si Ne est faible. En conséquence, des ratios $\pi_N/\pi_S$ plus élevés sont attendus chez les espèces à faible fécondité (qui ont souvent effectivement de plus forts Ne).

La diversité génétique est une mesure cruciale en écologie et en évolution (Hughes *et al.* 2008), car elle permet aux organismes de s'adapter aux changements environnementaux (Reusch *et al.* 2005). Si la diversité génétique d'une population est élevée, la probabilité que certains individus possèdent des allèles adaptés à un environnement précis est plus grande. Ces individus vont se reproduire préférentiellement, et leurs descendants porteront ces allèles adaptés à un environnement qui a changé. Au contraire, le risque d'extinction est plus élevé chez les espèces à faible diversité génétique, ne possédant que peu d'allèles « disponibles » pour répondre à des variations environnementales. Ceci contribue à favoriser les stratèges r, à forte diversité génétique, par rapport aux stratèges K, dans les environnements fluctuants.

## 1.5 Le modèle d'étude

Les échinodermes constituent le plus grand phylum exclusivement composé d'organismes marins, ainsi que le deuxième phylum de deutérostomiens, après les chordés, en nombre d'espèces. Il existe cinq classes actuelles d'échinodermes, à savoir les Crinoidea (crinoïdes), les Asteroidea (étoiles de mer), les Ophiuroidea (ophiures), les Holothuroidea (concombres de mer) et les Echinoidea (oursins). Longtemps sujette à discussion, la phylogénie des échinodermes vient d'être clairement

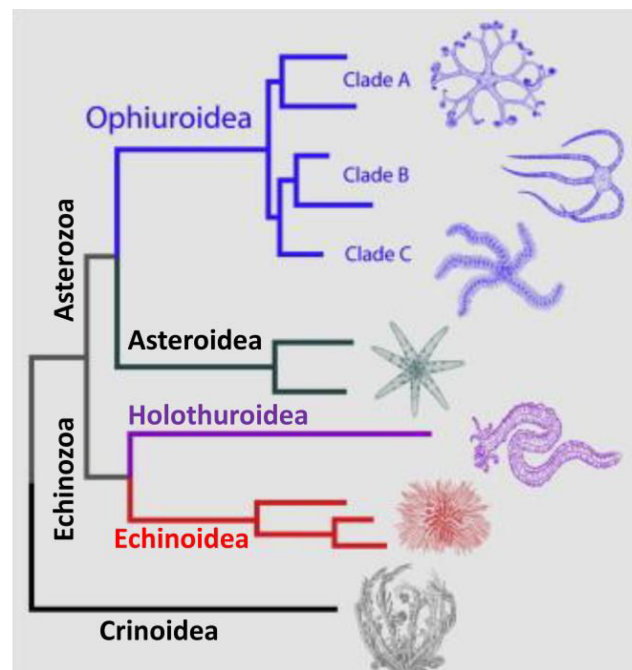


**Figure 1.2**: Phylogénie des échinodermes. Modifié d'après O'Hara *et al* (2014).

établie (Annexe 1; Telford *et al.* 2014; O'Hara *et al.* 2014). Les ophiures et les étoiles de mer forment un groupe monophylétique appelé Asterozoa, tandis que les concombres de mer et les oursins forment le groupe des Echinozoa, les crinoïdes étant le groupe le plus ancestral (Figure 1.2). Parmi les échinodermes, les ophiures constituent la plus grande classe avec plus de 2'000 espèces décrites (Stöhr *et al.* 2012). Elles sont trouvées dans tous les océans, de l'Antarctique aux tropiques, et à toutes les profondeurs, de l'intertidal aux abysses. Le genre *Ophioderma* (Müller & Troschel, 1840) est constitué actuellement de 28 espèces décrites, la majorité étant trouvées en Atlantique Ouest et dans l'Est du Pacifique. Du côté Est de l'Atlantique, seules deux espèces ont été décrites, *Ophioderma wahlbergii* (Müller & Troschel, 1842) en Afrique du Sud et *Ophioderma longicauda* (Bruzelius, 1805) en Atlantique Nord et Méditerranée.

*Ophioderma longicauda* (Figure 1.3) est décrite comme une espèce abondante, de grande taille et largement distribuée, du Sénégal à l'Espagne pour la partie Atlantique, ainsi que dans toute la Méditerranée. Elle est l'une des espèces d'ophiures les mieux décrites de Méditerranée, que ce soit concernant sa biologie (Fenaux 1969, 1972), sa morphologie (Tortonese 1983) ou encore comme modèle de régénération (Biressi *et al.* 2010; Weber *et al.* 2013). Comme de nombreuses autres ophiures, *O. longicauda* craint la lumière (Alonso & Guacimaro 2008), elle reste donc cachée sous des pierres durant la journée et est active durant la nuit. Le régime alimentaire d'*O. longicauda* peut être qualifié d'opportuniste, cette espèce étant décrite comme détritivore (Deschuyteneer & Jangoux 1978) se nourrissant principalement de matière organique déposée sur les sédiments, mais pouvant également être carnivore (Deschuyteneer & Jangoux 1978), nécrophage, voire cannibale selon les conditions environnementales (observations personnelles). Cette espèce est gonochorique (à sexes séparés), et se reproduit une fois par année via le relâchement d'ovocytes et de spermatozoïdes dans la colonne d'eau (Fenaux 1972). Chez deux espèces du même genre, *Ophioderma rubicunda* (Lütken, 1856) et *Ophioderma squamosissima* (Lütken, 1856), que l'on peut trouver dans le Golfe du

**Figure 1.3**: *Ophioderma longicauda*, dans la baie de Marseille. Crédit: Frédérique Zuberer

Mexique et dans la mer des Caraïbes, il a été montré que les femelles relâchent leurs ovocytes environ trente minutes après que les premiers mâles observés aient relâché leurs spermatozoïdes (Hagman & Vize 2003). Les mâles d'*O. rubicunda* formaient des agrégats denses, tandis que les mâles de *O. squamosissima* étaient isolés. Pour les deux espèces, les femelles étaient isolées (Hagman & Vize 2003). Chez *O. longicauda*, aucun comportement de reproduction n'a pour l'instant été observé. Après la fécondation, une larve vitellaria lécithotrophe (donc issue d'un ovocyte contenant des réserves) est formée, qui se développe dans la colonne d'eau durant six jours environ, jusqu'à transformation en juvénile (Fenaux 1969).

Cependant, des individus incubants d'*O. longicauda* ont été récemment récoltés dans l'Est de la Méditerranée (Stöhr *et al.* 2009). L'analyse morphologique a montré des différences de couleurs de corps et de taille maximale entre incubants et non-incubants, toutefois ces caractères étaient largement chevauchants. En revanche, l'analyse du marqueur mitochondrial COI révéla l'existence de quatre groupes génétiques distincts (Stöhr *et al.* 2009). Par la suite, une étude à plus grande échelle (incluant toute la Méditerranée et l'Atlantique) et incluant 218 individus a été réalisée (Boissin *et al.* 2011), où six groupes génétiques, ou lignées, ont été décrites (Figure 1.4). Ces lignées ont montré une forte divergence, allant de 2 à 10% selon les comparaisons effectuées. De la plus basale aux plus récentes, on retrouve les lignées L6, L5, L1, puis un groupe monophylétique incluant les lignées L2-L3-L4. Ces dernières ne sont retrouvées qu'en Méditerranée orientale, et chaque individu incubant appartenait à l'une de ces lignées. De plus, l'analyse du marqueur nucléaire ITS1 sur des individus L1 et L3 échantillonnés en sympatrie a montré de grandes différences de fréquences alléliques, voire la présence d'allèles privés, corrélées à la lignée mitochondriale. Ces résultats confirment ainsi l'existence d'espèces cryptiques chez *Ophioderma longicauda.* Dans ce contexte, *O. longicauda* a été choisie comme modèle pour ma thèse.

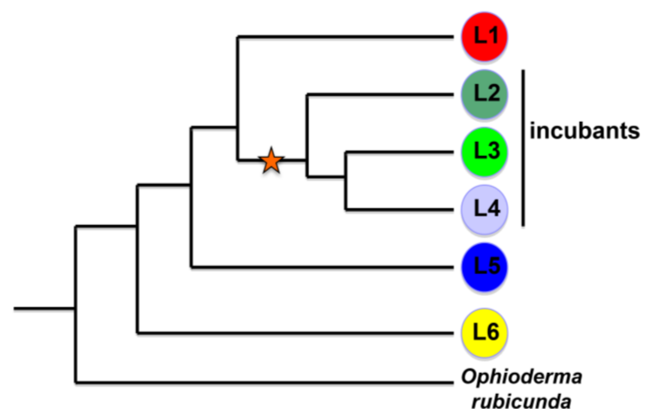



**Figure 1.4**: Arbre phylogénétique de la COI dans le complexe d'espèces *Ophioderma longicauda*. Modifié d'après Boissin *et al* (2011).

## 1.6 Objectifs de la thèse

Cette thèse avait deux objectifs principaux. Avant de procéder à des études comparatives, il a fallu s'assurer que le complexe *O. longicauda* contenait bien des espèces biologiques différentes, et idéalement déterminer leur nombre exact. Ainsi, le premier objectif était de **définir les limites d'espèces** dans le complexe *Ophioderma longicauda.* Le but étant de déterminer, parmi les six lignées mitochondriales connues, combien d'entre elles correspondent réellement à des espèces biologiques différentes (e.g. des entités génétiques séparées). Pour ce faire, je présenterai tout d'abord une comparaison de l'état reproductif des lignées L1 et L3 vivant en sympatrie, durant une période précise de l'année (Mai 2012). Ces résultats seront mis en relation avec les données génétiques obtenues pour deux marqueurs, un mitochondrial (COI) et un nucléaire (i51). La comparaison de la lignée mitochondriale avec les différents allèles nucléaires observés me permettra d'inférer l'existence (ou non) de flux de gènes actuel entre ces deux lignées. (Chapitre 2).

Ensuite, je présenterai l'analyse phylogénétique de 31 marqueurs principalement nucléaires sur des individus appartenant aux six lignées mitochondriales décrites jusqu'alors (Chapitre 3). Leur analyse me permettra de définir plus clairement l'existence de groupes génétiques séparés au sein du complexe *O. longicauda,* et d'inférer le scénario historique le plus probable. L'analyse de groupes externes d'*Ophioderma* me permettra également de dater la divergence du complexe *Ophioderma longicauda*. Je terminerai cette partie sur les limites d'espèces en présentant une étude de thermotolérance comparée entre les lignées L1 et L3, couplée à une étude de régénération (Chapitre 4).

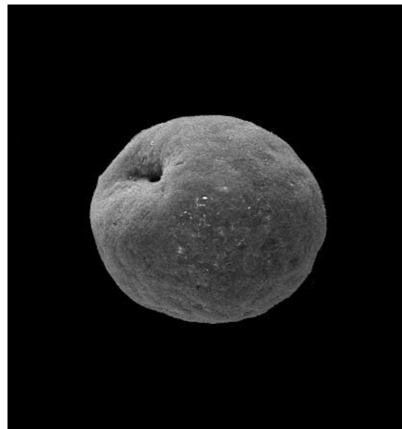
Le deuxième objectif consistait à étudier **l'influence des traits d'histoire de vie sur la connectivité et la diversité génétique**. En effet, grâce à la confirmation de l'existence d'espèces biologiques différentes (L1 et L3), des études comparatives peuvent être menées, et ce sans influence de facteurs confondants. Ainsi, les effets historiques sont minimisés, étant donné que ce sont des espèces proches, et les effets environnementaux sont également minimisés étant donné que ces espèces vivent en sympatrie. Afin de répondre à ces questions, Je comparerai la structure génétique à grande échelle (sur toute l'aire de distribution) entre toutes les lignées (en utilisant le marqueur COI). Je comparerai également la connectivité entre les lignées sympatriques

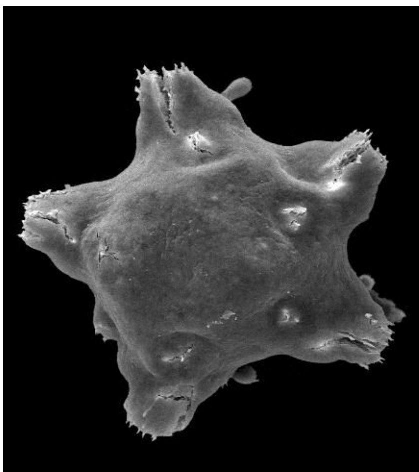
L1 et L3, à l'échelle de la Crète, et la diversité génétique entre lignées, caractérisée dans ce cas pour douze marqueurs de séquences développés sur la base de transcriptomes (Chapitre 5). Par la suite, je présenterai l'analyse comparée des transcriptomes des lignées L1 et L3, qui me permettra de confirmer les résultats obtenus sur la diversité génétique au chapitre précédent. Je présenterai finalement les gènes évoluant sous sélection positive entre ces deux lignées, en me concentrant particulièrement sur les gènes impliqués lors de la reproduction, qui pourraient avoir joué un rôle dans l'isolement reproducteur entre les espèces L1 et L3 (Chapitre 6). Pour finir, une discussion générale sur les recherches menées au cours de cette thèse sera présentée, ainsi que les perspectives de ce sujet d'étude (Chapitre 7).

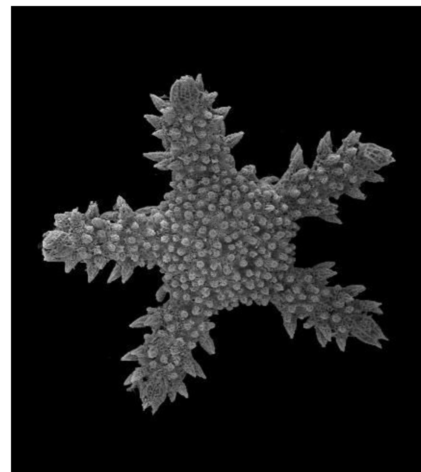
# Chapitre 2

# -

# Les données génétiques, la saison et la stratégie de reproduction soutiennent l'existence d'espèces biologiques chez *Ophioderma longicauda*







# Chapitre 2: Les données génétiques, la saison et la stratégie de reproduction soutiennent l'existence d'espèces biologiques chez Ophioderma longicauda

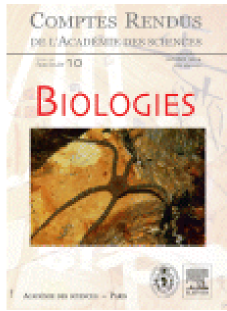
## 2.1 Résumé

L'objectif de ce chapitre était de clarifier les limites d'espèces pour deux lignées sympatriques d'*Ophioderma longicauda*, en utilisant des données de période reproductive, des données écologiques, morphologiques ainsi que des données génétiques indépendantes (nucléaires) aux données génétiques fournies par le génome mitochondrial. Pour ce faire, 169 individus appartenant aux lignées L1 et L3 ont été collectés dans six localités de Crète (Grèce) sur une période de cinq semaines (Mai jusqu'à début juin 2012), cette période étant la période de reproduction présumée des incubants. Des dissections et des coupes histologiques ont été réalisées afin de comparer l'état de reproduction des individus L1 et L3. De plus, des analyses morphologiques ont été effectuées (mesures de la taille du disque des individus et identification de l'état des boucliers radiaux, couvert ou nu). Finalement, tous les individus ont été séquencés pour le gène mitochondrial COI et génotypés pour le marqueur nucléaire codominant i51 (un intron présentant un microsatellite tétranucléotidique).

Bien que chevauchantes, les distributions des tailles des L1 et L3 étaient significativement différentes, les L1 étant en moyenne plus grandes (Tableau 2.1). De plus, l'observation des boucliers radiaux a montré que ce caractère était presque diagnostic pour délimiter les L1 des L3 (100% des L1 montraient des boucliers couverts, tandis que 98% des L3 montraient des boucliers nus). Les femelles incubantes, appartenant toutes à la lignée mitochondriale L3, ont relâché des juvéniles à tout stade de développement, probablement en raison du stress de l'échantillonnage. Le développement des juvéniles a pu être suivi assez finement, et aucun stade larvaire n'a été observé de la gastrula au juvénile. Ces observations soutiennent l'hypothèse que le développement est direct chez *Ophioderma longicauda* L3 (sans phase larvaire, même incubée). Les embryons se sont développés sans apport de nourriture, du stade gastrula au stade juvénile. L'observation des coupes histologiques a montré que chaque femelle L3 était incubante, tandis que tous les individus L1 avaient les gonades pleines. Les mâles L3 ont montré quant à eux, soit des gonades vides soit des gonades pleines. Les L1

n'ont pas montré d'événement de reproduction, même en ayant été stimulées, suggérant que la gamétogénèse n'était pas terminée à cette période de l'année pour L1 (mai-juin). Ceci suggère un isolement pré-zygotique temporel entre les incubants et les individus à larves, les incubants s'étant reproduits fin avril, tandis que les L1 ne s'étaient toujours pas reproduits à mi-juin. Ces observations sont en accord avec la littérature sur la période de reproduction d'*Ophioderma longicauda*, qui avait été reportée en Juillet, sur les côtes françaises, pour la lignée L1.

De plus, des différences statistiques de morphologie et de distribution bathymétrique ont été trouvées entre lignées, les L3 étant retrouvées majoritairement à faible profondeur. Finalement, l'intron i51 était polymorphe chez L1 (60 individus) avec la présence de 12 allèles, tandis qu'il était monomorphe chez L3 (109 individus), confirmant l'absence de flux de gènes entre les incubants et dispersants en Crète. Ce chapitre a montré que les lignées sympatriques à larves L1 et incubante L3 sont des espèces biologiques différentes ; le chapitre suivant permettra de déterminer ce qu'il en est concernant les autres lignées.




Ce chapitre a été modifié sur la base de l'article de Weber A.A-T., Stöhr S. and Chenuil A. (2014) « Genetic data, reproduction season and reproductive strategy data support the existence of biological species in *Ophioderma longicauda* ». Comptes Rendus Biologies. 337 (10) : 553-560. Une illustration d'*Ophioderma longicauda* L1 est en couverture du volume 10 de Comptes Rendus Biologies (Octobre 2014).

## 2.2 Introduction

Cryptic species, or distinct species described as a single one because of morphological similarity, are common (Bickford *et al.* 2007), particularly in the marine environment (Knowlton 1993). The majority of cryptic species was discovered during the last decade thanks to the use of molecular methods for species identification, in particular when using mitochondrial DNA based markers (Knowlton 2000; Ward *et al.* 2008). For example, Hoareau et al (Hoareau *et al.* 2013) investigated the COI and 16S genes in 10 species of brittle star and found seven new lineages , six of which fulfill the criteria to be biological species.

In echinoderms, several examples are known, such as the crinoids *Tropiometra carinata* (Torrence *et al.* 2012), *Promachocrinus kerguelensis* (Wilson *et al.* 2007; Hemery *et al.* 2012), *Cenolia* sp. (Naughton *et al.* 2014), the sea urchin *Echinocardium cordatum* (Chenuil & Féral 2003; Egea 2011), the sea stars *Leptasterias hexactis* (Foltz *et al.* 1996; Hrincevich & Foltz 1996), *Linkia* sp. (Williams 2000), *Patiriella pseudoexigua* (Hart *et al.* 2003), *Parvulastra exigua* (Hart *et al.* 2006), *Cryptasterina pentagona* and *Cryptasterina hystera* (Byrne *et al.* 2003), the brittle stars *Amphipholis squamata* (Sponer & Roy 2002; Le Gac *et al.* 2004; Boissin *et al.* 2008), *Ophiocoma erinaceus* (O'Hara *et al.* 2004), *Ophiothrix fragilis* and *Ophiothrix quinquemaculata* (Baric & Sturmbauer 1999; Muths *et al.* 2009; Pérez-Portela *et al.* 2013) , *Astrotoma agassizii* (Hunter & Halanych 2008; Heimeier *et al.* 2010) and *Acrocnida brachiata* (Muths *et al.* 2006; Stöhr & Muths 2010), amongst others.

The brittle star *Ophioderma longicauda* (Bruzelius, 1805) was known as a common Atlanto-Mediterranean gonochoric species (Tortonese 1983) reproducing via a lecithotrophic vitellaria larva (Fenaux 1969, 1972). However, the discovery of brooding individuals (Stöhr *et al.* 2009) suggested that it is actually a species complex composed of six mitochondrial lineages (Boissin *et al.* 2011). The lineage L1 at the western

Mediterranean French coast reproduces once a year (Fenaux 1972) via lecithotrophic vitellaria larvae (Fenaux 1969), whereas lineages L2-L3-L4, observed in the Eastern Mediterranean, are also gonochoric but brood their offspring (Stöhr *et al.* 2009). Lineages L5 and L6 are expected to reproduce via lecithotrophic larvae as well, but no observation of their reproduction is available yet. The broadcast spawning lineage L1 is found in sympatry with the brooding lineages L2-L3-L4 in the eastern Mediterranean. Brooding females have been reported in the eastern Mediterranean at the end of May/beginning of June (Stöhr *et al.* 2009), whereas the reproduction of lineage L1 was reported at the beginning of July in the western Mediterranean (Fenaux 1969).

In the present study, to infer whether there is potential gene flow between brooding and broadcasting lineages, we compared the "reproductive state" of L1 and L3 specimens occurring in sympatry for one month, in May 2012. In addition, we observed the development of brooded juveniles and made inferences about the fertilization process. Furthermore, we investigated morphological characters (maximal disc size, status of radial shields) and bathymetric distribution of both lineages. Finally, to complement the analysis of the mitochondrial marker COI, we analyzed a codominant nuclear marker (thus independent of the mitochondrial marker), which provides another way of testing reproductive isolation between brooders and broadcasters.

## 2.3 Methods

### 2.3.1 *Sampling*

Specimens of *O. longicauda* from lineages L1 and L3 were collected in six locations of Crete during the month of May 2012 (Fig. 2.1). The sites of collection on the North coast of Crete included: Lygaria Port (LyP) (N 35°23'57.56" / E 25°01'41.24"), Lygaria (Lyg) (N 35°24'08.37" / E 25°02'02.62"), Gouves (Gou) (N 35° 20' 13.61"/ E 25° 18' 18.85") and Agios Nikolaos (AgN) (N 35°11'40.84" / E 25°43'2.37"). The sites on the South coast included: Agios Pavlos (AgP) (N 35°06'06.73" / E 24°33'47.82") and Ierapetra (Ier) (N 35° 0' 13.14"/ E 25° 44' 5.45").

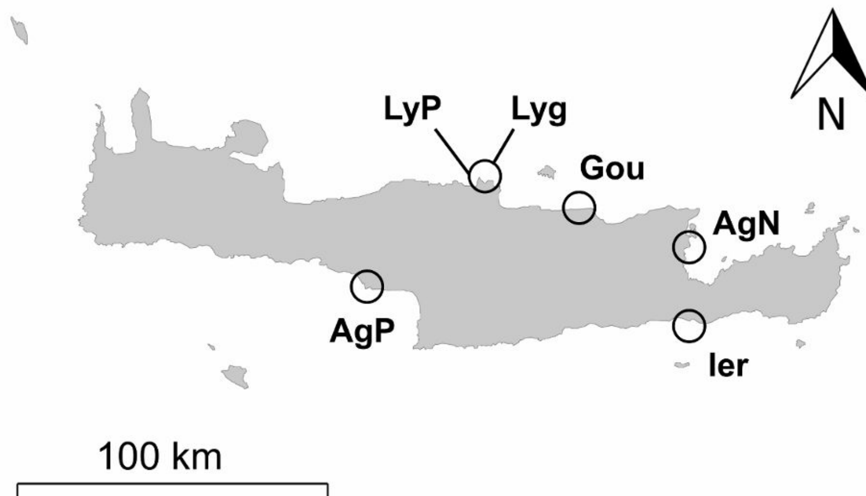



**Figure 2.1**: Map of Crete (Greece) with the six localities where individuals of *Ophioderma longicauda* broadcast spawning lineage L1 and brooding lineage L3 were sampled. LyP = Lygaria Port; Lyg = Lygaria; Gou = Gouves; AgN = Agios Nikolaos ; AgP = Agios Pavlos ; Ier = Ierapetra.

### 2.3.2 *DNA extraction, COI sequencing and analyses*

Each specimen collected (n = 169) was genotyped to determine its mitochondrial lineage. DNA extractions and PCRs targeting a portion of the COI gene were performed as previously described (Stöhr *et al.* 2009; Boissin *et al.* 2011). PCR products were sent for sequencing to a private company (LGC genomics, Berlin, Germany). Sequences were manually edited with the software BioEdit, aligned in the program Seaview (Gouy *et al.* 2010) with MUSCLE v3.8.31 algorithm (Edgar 2004) and manually improved. Number of haplotypes and haplotype diversity were investigated with DNAsp v5 software (Librado & Rozas 2009).

### 2.3.3 *i51 genotyping and analyses*

The intronic codominant marker (EPIC) i51 (Chenuil *et al.* 2010; Gérard *et al.* 2013) was used in this study for the first time. PCRs were performed with the primers OL51F2: 5'CCA TAT ATT TAT GTA TTC CCG TGT 3' and OL51R1: 5'TCA CAG CCA CTT CAT GCT GC AG 3' in a final volume of 20 µl with the following final reagent concentrations: 1X of Go Taq flexi buffer, 1.5 mM of $MgCl_2$, 0.2 mM of dNTPs, 0.2 µM of each primer, 0.1U of Taq Polymerase (Flexi Go Taq, Promega), 1 µl of DNA template diluted ten times and sterile water to complete up to 20 µl. The PCR cycling parameters were as follow: a first step of denaturation at 95°C 3 minutes, followed by 35 cycles of :

30 seconds of denaturation at 95°C, 30 seconds of annealing at 52°C, 30 seconds of elongation at 72°C and final step of elongation at 72°C for 3 minutes. i51 is a codominant marker displaying length heterogeneity with pure microsatellite repetitions of 4 bases (CCAA motif). A new method using high resolution agarose electrophoresis was developed for i51 genotyping. It allowed reading the genotype of each individual on agarose gels. Correct amplification of DNA samples was first verified by migrating the PCR products on a regular agarose 2% TBE 1X gel stained with Ethidium Bromide. Then, a 3.5% high resolution agarose gel (Sigma-Aldrich) was performed; samples were loaded and migrated during 2.5 hours at 120V. The loaded volume of PCR product was adapted for each individual according to the control gel to optimize resolution. This technique permitted to discriminate homozygous from heterozygous individuals and to differentiate alleles differing by 4 bp or higher multiples of 4 bp. For each gel, one sequenced individual was loaded as a reference for sizing the alleles in two lanes (lanes 7 and 19), as well as the DNA ladder (25 bp DNA step ladder, Promega) in three lanes (on lanes 2, 13 and 25, for a total of 26 lanes). Ethidium Bromide staining was done at the end of migration by performing an Ethidium Bromide bath during 10 min. Pictures of the gels were taken with the gel doc (BioRad) and genotypes were determined using the software Quantity One (BioRad).

#### 2.3.4 *Morphological, reproductive and developmental observations*

For each specimen collected, the disc diameter (DD) was measured and the status of the radial shields was investigated (covered with granules or naked). 23 individuals of L1 and 34 individuals of L3 were dissected and the status of the gonads was recorded (full, brooding or spawned) to infer the reproductive status of each lineage during the month of May. In addition, specimens from both lineages were preserved for gonad histology. Entire discs were fixed in EtOH 70% for 24 h, dehydrated in EtOH 95% and embedded in araldite resin by successive baths of acetone 100% , acetone 75% – araldite 25% and araldite 100%. Serial sections were stained with toluidine blue and observed with a Leica DM 2500 microscope.

Due to the stress of collection, brooding females released juveniles at different developmental stages. Juveniles were maintained in filtered sea water in glass Petri dishes at ambient temperature and water was renewed every day. Development of juveniles was monitored and pictures from live specimens were taken at different

developmental stages. In addition, some juveniles were preserved in EtOH 95% and prepared for scanning electron microscopy.

## 2.4 Results

### 2.4.1 *Genetic analyses – COI and i51*

We obtained a total of 169 COI sequences of 491 bp, 60 sequences belonging to L1 and 109 sequences belonging to L3. A higher number of haplotypes was found in L1 compared to L3 (29 vs. 20; Table 2.1), and the average haplotype diversity per population was 65% higher in L1 compared to L3 (Table 2.1). The 169 individuals were genotyped for the i51 marker. Lineage L1 displayed 12 alleles for this marker, each 4 bp intermediate being found between alleles 081 bp and 133 bp, except the alleles 101 bp and 113 bp. In contrast, each L3 individual genotyped was homozygous with the allele 081 bp, indicating that i51 was monomorphic in this brooding lineage (Table 2.1) and constitutes a semi-diagnostic marker of these two lineages.

**Table 2.1**: Comparative summary of morphological characters (disc diameter and status of radial shields) and genetic data (mitochondrial marker COI and nuclear marker i51) in the broadcasting lineage L1 and the brooding lineage L3 of *O. longicauda*.

| | **L1** | **L3** |
|---|---|---|
| N individuals | 60 | 109 |
| N populations | 5 | 5 |
| COI haplotypes (total) | 29 | 20 |
| COI haplotypes (mean per pop) | 8.6 | 5 |
| Hd (mean per pop) | 0.918 | 0.554 |
| i51 alleles (total) | 12 | 1 |
| Radial shields (covered/naked) | 60/0 | 2/107 |
| Disc diameter (mean) | 15.3 mm | 12.3 mm |

### 2.4.2 *Morphology, reproduction and development*

#### 2.4.2.1 Morphological and habitat differences

The radial shields were almost always naked for the L3 specimens (Table 2.1) and always covered with granules for the L1 specimens (Table 2.1). Although size distributions were overlapping, mean of disc diameters (DD) differed significantly between L1 and L3 (Student's t-test; $p<0.001$), with L1 displaying larger specimens than

L3 in Crete (L1: mean = 15.3 mm; max = 25 mm; L3: mean = 12.3 mm; max = 21 mm) (Table 2.1, Fig. 2.2). Bathymetric distribution differed between L1 and L3 (Fisher exact test, $p<0.001$): we found only 15 specimens of L1 and 157 specimens of L3 between 1 m and 3 m depth (including the previously sampled populations of Elounda and Rhodes; see Boissin et al. 2011), whereas we found 48 specimens of L3 and 60 specimens of L1 between 3 m and 12 m.

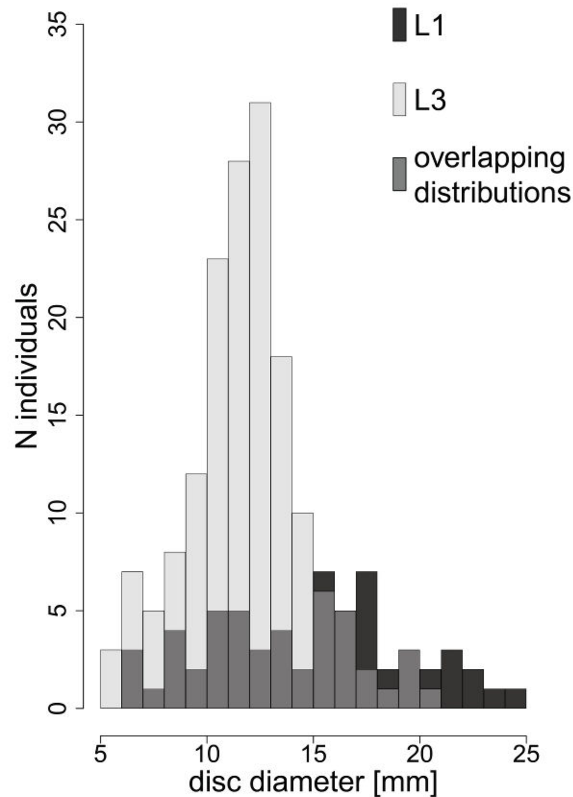



**Figure 2.2:** Size distributions between broadcast spawning (L1, n = 60) and brooding (L3, n = 109) genetic lineages. All individuals were sampled in Crete (Greece).

### 2.4.2.2 Dissections and histology

#### *2.4.2.2.1 Lineage L1*

Among the 23 dissected individuals genotyped as L1, 16 females and seven males were found, each with visible gonads (Table 2.2). All observed sections of females displayed ovaries full of late-vitellogenic oocytes (Fig. 2.3E-F) indicating that the gonads were almost mature. Depending on the section, the number of oocytes per ovary varied between three and 14. No males could be preserved for histology due to their low number. Fertilization trials in aquarium were performed once a week from the

beginning of May to the beginning of June as described in (Selvakumaraswamy & Byrne 2000) but all trials failed, suggesting that L1 individuals were not mature to spawn yet. In addition, we performed dissections of female (Fig. 2.3A, C, D) and male (Fig. 2.3B) L1 individuals from Marseilles (France) collected on the 13th of June 2012. Each specimen displayed full gonads.

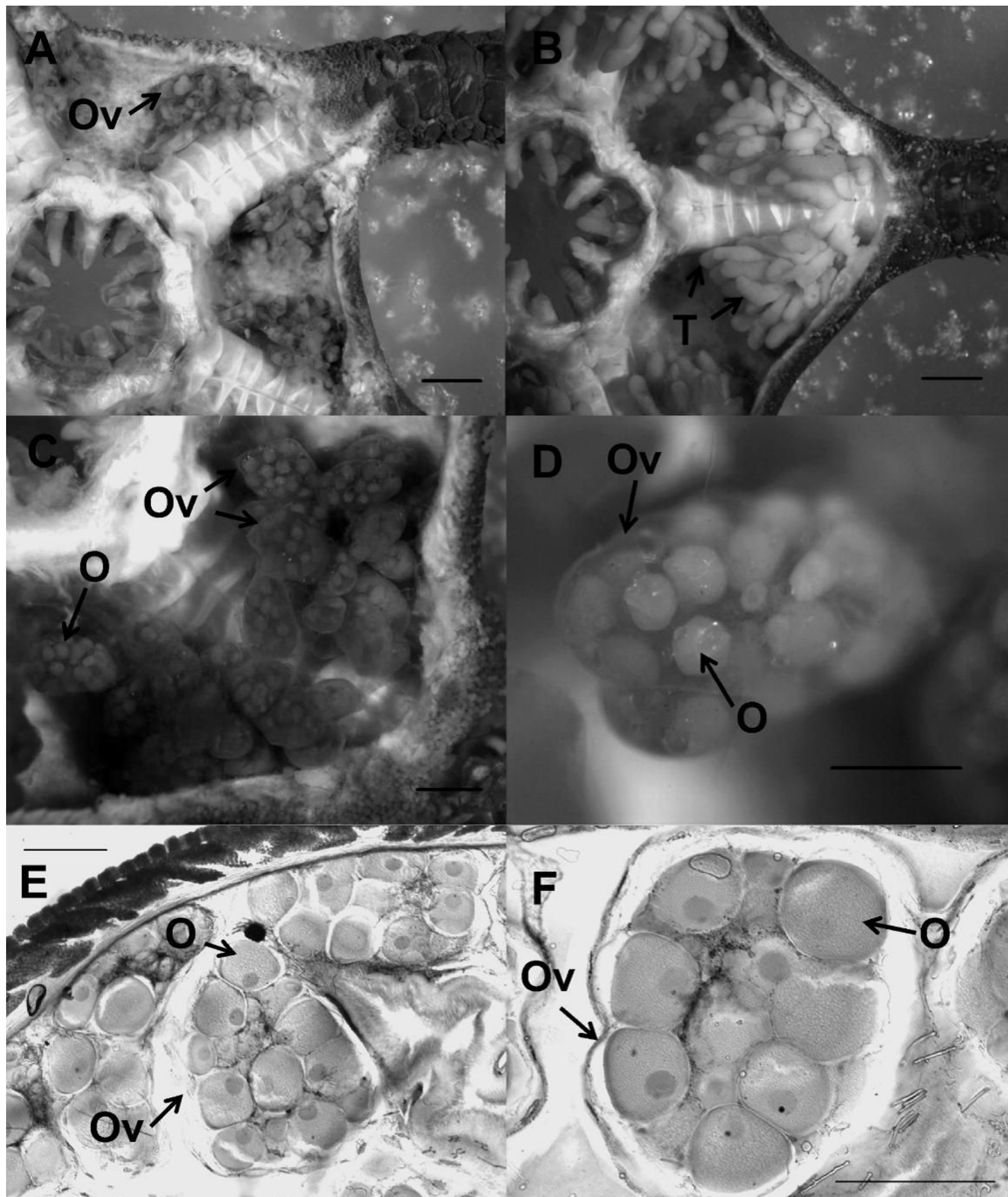



**Figure 2.3**: *Ophioderma longicauda,* broadcast spawning lineage L1. A, C, D: Female with the aboral surface of the disc removed to view the ovaries (Ov) containing oocytes (O). B: Male with the aboral surface of the disc removed to view the testes (T). E, F: Histological sections of ovaries (Ov) containing oocytes (O). Scale bars: A-B: 2 mm; C: 1 mm; D-F: 500 µm.

*2.4.2.2.2 Lineage L3*

Among the 34 dissected individuals genotyped as L3, 24 females and 10 males were found (Table 2.2). Six males displayed spent gonads and four males displayed full gonads, suggesting that the reproduction just occurred for the six individuals and was about to for the four other individuals. Each female was brooding embryos (Fig. 2.4A) at different developmental stages (Fig. 2.4B-H) depending on sampling date (Gastrula for females collected the 3$^{rd}$ of May, juveniles at different stages for the other sampling dates). Immediately after the first sampling on the 3$^{rd}$ of May and due to the stress of collection, brooding females released numerous embryos at the gastrula stage, indicating that fertilization indeed occurred just a few hours ago (Fig. 2.4B). Each brooding female found after this date released juveniles at more advanced developmental stages, indicating that the reproduction temporal window lasted only a few days at the population level (synchrony) and even among populations from northern and southern Crete. Furthermore, each juvenile released by a single female displayed the same developmental stage, suggesting that there was a single fertilization event for each female.

For lineage L3, development was followed in Petri dishes every day and no larval stage was observed between the gastrula and the juvenile stage. The development of embryos first consisted of a rearrangement of cells, as the gastrula diameter and the disc diameter of early juveniles were very similar (380 µm, Fig. 2.4B, 2.4E). After reaching a certain development stage, the size of the juvenile increased as the disc diameter and the arms grew in length. As the arms were growing, new pairs of tube feet were successively added to the arm, allowing the juveniles to move very early in their development, even while still being brooded by the mother. During growth, the juveniles also developed disc and arm spines (Fig. 2.4F). Each embryo/juvenile released developed properly without food supply, supporting the hypothesis that ophiuroids are able to absorb dissolved organic matter through the tube feet (Stephens & Virkar 1966; Clements *et al.* 1988).

**Table 2.2**: Inference of the lineage and the reproductive status of *O. longicauda* individuals collected from the 3rd of May to the 29th of May 2012 in five locations of Crete (Greece).

| Observation | Genetic lineage | Conclusion |
|---|---|---|
| 24 individuals with juveniles | 24 individuals of L3 | Brooding females |
| 16 individuals with eggs | 16 individuals of L1 | Spawning females |
| 11 male individuals with full gonads | 7 individuals of L1<br>4 individuals of L3 | Spawning males<br>Late “brooding” males |
| 6 “empty” individuals | 6 individuals of L3 | “brooding” males post reproduction |

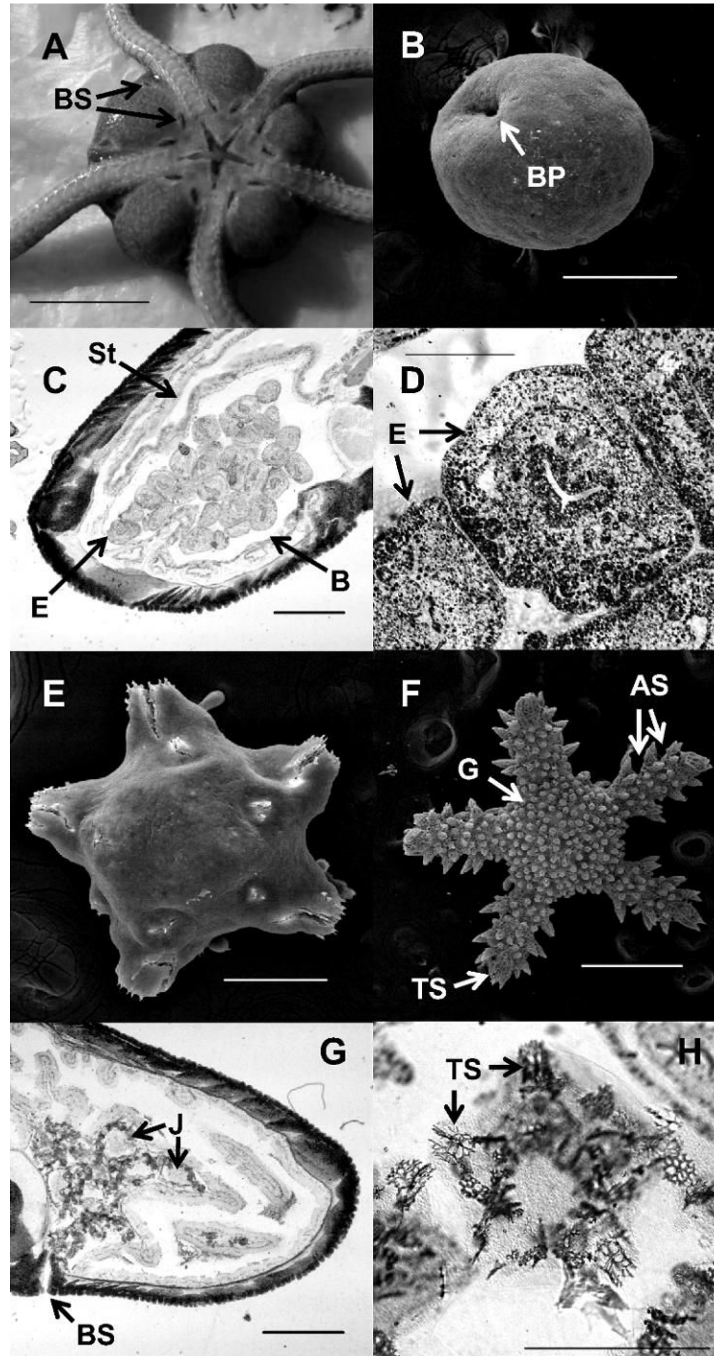



**Figure 2.4**: *Ophioderma longicauda*, brooding lineage L3. A: ventral overview of a brooding female; the gonads are located inside the disc in gonadal chambers (bursae) along both sides of each arm, opening through two bursal slits (BS) each. B: Gastrula with blastopore (BP) (SEM). C: Cross-section of disc of a brooding female with embryos (E); bursal wall (B); stomach (St) D: Embryo (E). E: Juvenile dissected from a bursa, arms with one free segment and the terminal plate, the first disc granules just forming. F: larger juvenile, dissected from a bursa, probably almost ready to hatch, arms with three segments (AS) and the terminal plate (TP), disc and arms covered with granules (G). G: Cross-section of disc of a brooding female with developed juveniles (J), about to be expelled through the bursal slit (BS). H: Mature embryo present in female bursa, probably almost ready to hatch; terminal plate (TP). Scale bars: A: 1 cm; B: 200 µm; C: 1 mm; D-E: 200 µm; F: 500 µm; G: 1 mm; H: 500 µm.

## 2.5 Discussion

### 2.5.1 *Morphological differences*

The status of the radial shields was previously investigated by Stöhr et al (Stöhr *et al.* 2009) who also noticed a tendency for the brooders to display naked radial shields, but this characteristic was not present in every brooding specimen. In contrast, the naked character versus covered radial shields appeared to be diagnostic of the lineage (L1 versus L3) in Crete. Further evidence is needed to determine to what extent this character is reliable on a larger geographic scale and if it is applicable to other brooding lineages L2 and L4. The size distributions of disc diameters overlapped, but very large individuals can be relatively safely diagnosed as belonging to L1. When comparing closely related species in marine invertebrates, the brooding ones are generally smaller than the ones with larvae (Strathmann & Strathmann 1982; Strathmann *et al.* 1984). *Ophioderma longicauda* follows this rule, although size differences are moderate. In this species complex, brooding is most likely the derived character (Stöhr *et al.* 2009), but it is still unclear whether reduced size preceded brooding or not.

### 2.5.2 *Habitat difference to explain L1 distribution*

Although we have to consider habitat differences with caution due to the relatively low number of sampling stations, L1 individuals appeared less abundant than L3 at the shallowest depths (1-3 m). To our current knowledge, L1 individuals are less abundant in the eastern Mediterranean than in the western part, where they largely dominate (Boissin *et al.* 2011) although this may be an artefact because the majority of samples were collected at shallow depths. One possible explanation is that L3 individuals are more adapted to elevated temperatures than L1, as mean temperatures are more elevated in the eastern basin (Durrieu de Madron *et al.* 2011). Indeed, a thermotolerance experiment conducted with L1 and L3 showed that L3 survival was not significantly affected by elevated temperatures contrary to L1 (Weber *et al.*). Another possibility is that L1 might suffer from the high UV penetration in the eastern basin, especially during the larval stage, as seen in several marine ectotherms (Häder *et al.* 2007; Dahms & Lee 2010). As the waters from the eastern Mediterranean basin are known to be very oligotrophic (D'Ortenzio & Ribera D'Alcalã 2009) UV rays penetrate deeper into the water column. During the reproduction period, the larvae of L1 may be

more exposed to UV radiation as they are expected to swim close to the surface (Hendler 1991). They may thus display a lower recruitment compared to brooded individuals that are less exposed to UV due to their strictly benthic life, and their likely habit, once released, of hiding under rocks, as do adults. UV also certainly affects adults (Dahms & Lee 2010), but echinoderms can develop avoidance mechanisms (Verling *et al.* 2002), particularly some brittle stars that live under rocks during the day and are active during the night (Rosenberg & Lundberg 2004). Nevertheless, it is possible that L1 is not less abundant in the eastern basin and that this observation is only a sampling artefact, L1 being as abundant but living deeper. In this case, the fact that L1 lives deeper in the eastern than in the western Mediterranean might also be explained by the UV hypothesis.

### 2.5.3 *Reproduction strategies and reproductive isolation between L1 and L3*

The reproduction of *O. longicauda* was previously described as occurring once a year via lecithotrophic larvae in early July (Fenaux 1969). The specimens of this study were collected in Villefranche-sur-mer (France) and most likely belonged to lineage L1, since all 234 individuals collected from eight locations at the French Mediterranean coast belonged to L1 (unpublished data). Here, we describe the reproduction of lineage L3 that occurred most likely at the beginning of May rather than at the end of May – early June as previously reported (Stöhr *et al.* 2009). The juveniles are then brooded for several weeks, unless a particular stress makes the mother expel the juveniles (Stöhr *et al.* 2009). We did not observe fertilization events, but hypothesize that internal fertilization may occur by intake of sperm into the female bursae, as suggested by Byrne (Byrne 1991) and Byrne et al. (Byrne *et al.* 2008) for *Ophionereis olivacea* (H.L. Clark, 1900) and *Ophiopeza spinosa* (Ljungman, 1867). The development of juveniles was apparently direct, but we cannot rule out the presence of an extremely transient brooded larval stage, which was not observed. Indeed, brooded vitellaria larvae were observed in ophiodermatid brittle stars (Byrne *et al.* 2008).

Our results suggest that, in addition to their divergent reproductive strategies, L1 and L3 may display pre-zygotic isolation as fertilization in L3 individuals occurred before the beginning of May, whereas L1 did not spawn, even when stimulated, until the beginning of June. Although reproduction and gonad development should be followed over a year for both lineages, and in different geographic locations to confirm our

results, the genetic data support reproductive isolation of L1 and L3 in Greece. Indeed, they displayed very divergent mitochondrial lineages, different haplotype diversities and different patterns of polymorphism at nuclear marker i51. The latter was polymorphic for the L1 broadcast spawning lineage, displaying 12 alleles of which 11 were private alleles, whereas i51 was monomorphic for the L3 brooding lineage, indicating a loss of genetic diversity in the brooding lineage. In addition, Boissin et al (Boissin *et al.* 2011) showed that two additional sympatric populations of L1 and L3 in Greece were differentiated for the nuclear marker ITS1.

The morphological as well as genetic and reproductive time differences between the lineages L1 and L3 of *Ophioderma longicauda* establish that those lineages are different biological species. To ensure that there is no possibility of hybridization between L1 and L3 because of a second reproductive period at other times of the year, a year-long study of the gonad condition should be performed. Fenaux (Fenaux 1972), however, observed that there was only one event of reproduction in the year, at least in the lineage L1. In addition, the genetic data allow us to safely state that there is no longer gene flow between L1 and L3 in Crete. Further study is required to test whether the reproductive times can change and overlap in other places of the Mediterranean, due to differing environmental conditions.

In their study, Boissin et al (Boissin *et al.* 2011) described the occurrence of six divergent mitochondrial lineages in the *Ophioderma longicauda* species complex. To fully resolve the speciation events in *O. longicauda*, it remains unknown whether the lineages L2, L4, L5 and L6 belong to a single broadcasting species including L1, or a single brooding species including L3, or if each additional lineage corresponds to another biological species. Only after that, the taxonomy should be revised by properly describing one or several new biological species. However, it is quite clear after these and the previous studies that "*O. longicauda*" includes at least two species. We suggest treating the groups of broadcasting and brooding lineages as a separate morphospecies or a morphologically similar species complex.

# Chapitre 3

-

# Délimitation d'espèces et datation de la divergence dans le complexe *Ophioderma longicauda*

# Chapitre 3: Délimitation d'espèces et datation de la divergence dans le complexe *Ophioderma longicauda*

## 3.1 Résumé

Dans le chapitre précédent, j'ai montré que les lignées sympatriques L1 et L3 étaient des espèces différentes au sens biologique du terme, grâce à un isolement reproducteur pré-zygotique (différences de saison de reproduction), grâce à l'analyse de deux marqueurs génétiques (le marqueur mitochondrial COI et le marqueur nucléaire i51) et finalement grâce à des différences de distribution de taille et de bathymétrie. Dans ce chapitre, j'utilise les méthodes moléculaires afin de différencier les groupes génétiques parmi les six lignées mitochondriales décrites jusqu'ici. Pour ce faire, j'ai développé 21 marqueurs sur la base de transcriptomes L1 et L3 (voir chapitre 5 pour des détails sur la génération des transcriptomes), qui ont été amplifiés chez plusieurs individus appartenant à chaque lignée mitochondriale, ainsi que des marqueurs introniques et l'ITS du rDNA (95 individus analysés). Une grande partie de ces marqueurs ont été choisis parce qu'ils montraient des différences diagnostiques entre L1 et L3. Lorsque cela était possible, des individus sympatriques (et prélevés lors d'une même plongée) appartenant à des lignées différentes ont été analysés. Deux autres espèces d'*Ophioderma* (*O. cinereum* et *O.teres*) ont été utilisées comme groupes externes afin de pouvoir dater la divergence du complexe, ces deux espèces ayant été séparées par la fermeture de l'isthme du Panama il y a 3.5 millions d'années.

Sur les 62 marqueurs testés, 21 ont été amplifiés avec succès. Trois d'entre eux n'ont pas donné de résultat concluant au séquençage. Sur les 18 marqueurs disponibles, 5 ont montré des locus paralogues qui ont pu être séparés avec succès, augmentant le nombre de marqueurs à 33. Un marqueur étant monomorphe dans le complexe *Ophioderma longicauda* et un autre marqueur montrant des paralogues difficilement séparables, ces deux marqueurs ont été retirés des analyses. Au total, 31 marqueurs ont été analysés pour la reconstruction phylogénétique du complexe. Pour la majorité de ces marqueurs, plusieurs lignées, voire toutes partagent de nombreux allèles et ne présentent pas de monophylie réciproque (ILS) (contrairement au marqueur mitochondrial). L'analyse phylogénétique pour chaque marqueur pris séparément n'a donc pas montré de groupes distincts, et n'était pas concluante pour la délimitation des

espèces. La concaténation de plusieurs locus permettant de combiner les informations des marqueurs pouvant sembler plus puissante, j'ai réalisé la concaténation et l'analyse phylogénétique de cette super-matrice en utilisant quatre méthodes de reconstruction phylogénétique classiques (neighbor joining ; parsimonie ; maximum de vraisemblance et analyse bayésienne). Aucune de ces méthodes n'a permis de définir clairement des groupes génétiques distincts, si ce n'est le groupement des L6 ensemble à la base de la divergence des *Ophioderma longicauda*.

Les analyses typiquement utilisées au niveau interspécifique n'étant pas concluantes, principalement en raison de l'ILS, je me suis tournée vers des analyses populationnelles pour définir les groupes génétiques et les dates de divergence du complexe *Ophioderma*. Ces analyses auraient été de toute façon utilisées car elles sont plus adaptées lorsque les entités génétiques présumées sont susceptibles d'être interfécondes. Ainsi, j'ai réalisé une DAPC (Discriminant Analysis of Principal Components) sur les génotypes multi-locus des individus à ces 31 marqueurs. Cette analyse des individus, réalisée sans aucun a priori sur leur appartenance à un groupe, a révélé six groupes génétiques (ou clusters) bien séparés. Il est intéressant de noter que les six clusters ne correspondaient pas exactement aux six lignées mitochondriales. Ainsi, un premier groupe bien défini était constitué de tous les individus de la lignée L6, échantillonnés à Dakar (Sénégal) et Madère (Portugal). Ensuite, les individus de la lignée L5 provenant de Dakar formaient un second groupe. Puis, tous les individus de la lignée L1 la plus abondante et largement distribuée (p. ex. Madère, Tunisie, Grèce, Liban) ainsi que les individus de la lignée L5 échantillonnés hors de Dakar (Madère, Ténériffe, Ceuta) formaient ensemble un troisième groupe. Le quatrième groupe était formé des individus de la lignée L3b de Tunisie. Finalement, les cinquième et sixième groupes étaient formés des individus incubants L2-L3 de Grèce, et des individus incubants L2-L4 de Chypre et Liban, respectivement. On observait donc (i) des séparations nettes entre lignées mitochondriales distinctes en syntopie, et (ii) une congruence imparfaite entre les lignées mitochondriales L1 et L5, d'une part, et L2 et L4, suggérant que de la sélection ou de l'introgression (passée) auraient affecté le marqueur mitochondrial.

Afin de déterminer les temps de divergence, tailles efficaces et éventuels événements d'hybridation entre les différents clusters, différents scénarios ont été explorés grâce à la méthode ABC (Approximate Bayesian Computation). L'arbre de

divergence des clusters le plus vraisemblable a été réalisé en calculant les distances génétiques moyennes entre clusters (K2P moyens entre « lignées-populations »), puis en reconstruisant un arbre phylogénétique d'après la méthode Neighbor-Joining. Cet arbre, constituant le scénario 1, décrit les divergences successives des clusters C6 (L6 Dakar), C3 (L5 Dakar), C1 (L1 et autres L5), C2 (L3b de Tunisie), C4 (L2 et L3 de Grèce) et C5 (L2 et L4 de Chypre et Liban), sans hybridation entre clusters.

Les individus du cluster C2 provenant de Tunisie étaient particuliers, car leur lignée COI (L3b) était très proche de la lignée L3 observée en Grèce, suggérant que ces individus pouvaient pratiquer l'incubation. Or, les génotypes observés du marqueur nucléaire i51, monomorphe chez tous les incubants, étaient polymorphes pour les populations de Tunisie, avec des allèles partagés avec le cluster C1, suggérant la possibilité d'hybridation passée entre ces clusters. Ainsi, le deuxième scénario testé incluait un événement d'hybridation entre le cluster C1 et l'ancêtre commun des incubants (C4 et C5), donnant le cluster actuel C2, sans changements pour les autres clusters. Pour cette analyse, les génotypes des groupes externes d'*Ophioderma* (*O. cinereum* et *O. teres*) étaient nécessaires pour chaque locus afin de pouvoir estimer des dates de divergence. Ainsi, le nombre de marqueurs utilisés pour cette analyse est passé de 31 à 11 (10 marqueurs nucléaires et la COI). Deux analyses séparées ont été réalisées, d'une part les 10 locus nucléaires et d'autre part le locus mitochondrial COI.

L'analyse des 10 marqueurs nucléaires a montré que le scénario 2 (avec hybridation) avait une plus grande probabilité postérieure que le scénario 1. La divergence du cluster C6 a été estimée à environ 2.2 millions d'années ; la divergence du cluster C3 a été estimée à environ 830'000 ans ; la divergence du cluster C1 avec les individus incubants a été estimée à 550'000 ans ; L'événement d'hybridation entre C1 et les incubants a été estimé à 408'000 ans, et finalement la divergence entre les deux clusters d'incubants C4 et C5 a été estimée à 123'000 ans.

L'analyse génétique d'individus *Ophioderma longicauda* appartenant aux six lignées mitochondriales a montré qu'il existait six clusters distincts, un peu différents des lignées mitochondriales. Afin de confirmer si ces clusters correspondent à des espèces biologiques, plus d'individus devraient être échantillonnés, en particulier des populations proches géographiquement, voire sympatrique si cela est possible. Cependant, mes résultats suggèrent qu'en plus de la séparation incubants-dispersants, il

existerait jusqu'à quatre espèces d'*Ophioderma* dispersantes en Méditerranée, à savoir les clusters C6, C3, C2 et C1. Les clusters C1 et C2 ne semblent pas montrer de flux de gène actuel bien que montrant une distribution géographique chevauchante.

Dans ce chapitre, j'ai montré qu'il était possible de délimiter des espèces malgré une séparation incomplète de leurs lignées génétiques à (presque) tous les marqueurs, en utilisant des méthodes généralement utilisées en génétique des populations. Ces dernières sont en effet plus puissantes pour répondre à des questions à l'interface inter et intraspécifique. En perspective de ce travail, il est prévu de décrire, en utilisant ces résultats génétiques et des données morphologiques, une nouvelle espèce d'*Ophioderma*, incubante, *Ophioderma vivipara*. Ce travail sera fait en collaboration avec Sabine Stöhr, du musée d'Histoire naturelle de Stockholm, en Suède.

Cet article est intitulé "Species delimitation and divergence dating in the brittle star species complex *Ophioderma longicauda".* Ses auteures sont Alexandra Anh-Thu Weber et Anne Chenuil. Il est prévu de le soumettre à Cladistics après relecture et modifications éventuelles par toutes les auteures.

## 3.2 Introduction

Accurate species delimitation and description is essential to properly assess biodiversity, but also for management and conservation purposes. Currently, most species are delimited using molecular tools, due to the relative easiness of their utilization, compared to the taxonomic expertise required when morphology-based descriptions are performed. In addition, molecular tools allow delimitating recently diverged species, for which ecological or morphological differences may not have evolved yet. In particular, numerous cases of cryptic species (morphologically undistinguishable species wrongly described as a single one) were discovered since the use of molecular tools to define and delimitate species (Bickford *et al.* 2007). Interestingly, they seem to be found in higher occurrence in the marine environment compared to the terrestrial environment (Knowlton 1993).

The cryptic species complex *Ophioderma longicauda* (Stöhr *et al.* 2009; Boissin *et al.* 2011) encompasses six divergent mitochondrial lineages, three of which are broadcast spawners (L1-L5-L6) whereas the three other lineages brood their offspring (L2-L3-L4). Previous research showed that at least two lineages (L1 and L3) represent different biological species, since they display temporal pre-zygotic isolation and congruent genetic signal between nuclear (i51) and mitochondrial (COI) markers (Chapter 2; Weber *et al.* 2014). In addition, differences of thermotolerance were found between samples from those two lineages, suggesting physiological differences between those two species (Chapter 4; Weber *et al.* 2013). Yet, the species relationships of the other lineages remain unclear. In particular, some populations sampled in Tunisia displayed incongruent genetic signal between the mitochondrial (COI) and the nuclear (i51) markers previously described. Indeed, their COI haplotypes were close from the L3 haplotype found in Greece (this lineage will be called L3b hereafter). In contrast, i51 was polymorphic in L3b, displaying several shared alleles with the broadcasting lineage L1, whereas this marker was monomorphic in the brooding lineages L2-L3-L4. The

populations L3b may therefore have resulted from an ancient hybridization event between L1 and L3.

To infer the species relationships in the *O. longicauda* species complex and determine which scenario of divergence is the most likely (with or without hybridization), we amplified and sequenced 31 transcriptome-based markers in the six *O. longicauda* lineages. In addition, *Ophioderma teres*, *Ophioderma cinereum* and *Ophioderma phoenium* were used as outgroup to estimate the divergence time of the *O. longicauda* species complex. Indeed, the species *O. teres – O. phoenium/ O.cinereum* are geminate pairs that occurred after the closing of Isthmus of Panama, so the divergence between this species pair is at least 3.5 Mya (Harilaos Lessios, pers. comm.).

## 3.3 Material and methods

### 3.3.1 *Sampling, DNA extraction and markers development*

For this study, 96 individuals including the six mitochondrial lineages of *O. longicauda* and five outgroup species (*O. teres, O. cinereum, O. phoenium, Ophiothrix fragilis* and *Ophiomyxa pentagona*) were used. Samples of *O. teres, O. cinereum* and *O. phoenium* were kindly provided by Francisco Alonso Solis-Marin and Harilaos Lessios. Details about individual codes, sampling dates and localities are shown in Table S3.1 and Figure 3.5.

DNA was extracted with MN-Tissue Kit (Macherey-Nagel) using an epimotion robot (Eppendorf) following the protocol of Ribout & Carpentieri (2013), and eluted in 200µl of sterile water. This extracted DNA was then diluted 10x in sterile water prior to PCR. We used orthologous genes from newly sequenced transcriptomes of *O. longicauda* L1 and L3 (see chapter 6 for details) to develop 55 primers pairs to test. The criteria for marker development were: (i) the marker should be polymorphic; (ii) in half of the markers, the presence of at least one diagnostic SNP between L1 and L3; (iii) the length of the PCR product should be between 300-400 bp (due to sequencing technology limitations). In addition, seven already existing markers were used, namely a mitochondrial marker (COI), ribosomal markers (ITS1, ITS2) and four EPIC markers, introns i21, i36, i50 and i54b (Chenuil *et al.* 2010; Penant *et al.* 2013; Gérard *et al.* 2013). On the 55 protein markers tested, 16 amplified correctly in each lineage of *O. longicauda.*

Furthermore, six out of seven existing markers amplified correctly. Finally, 22 markers were PCR amplified in the 96 specimens and sent for sequencing.

### 3.3.2 *Amplicon sequencing and dataset cleaning*

PCR products of different genes belonging to the same individual were pooled and 96 Illumina libraries were constructed. Paired-end (2x 250 pb) sequencing was performed on a MiSeq Sequencing System (Illumina) by the genomic platform Genotoul (www.genotoul.fr). Thirty millions of raw reads were obtained after sequencing. Reads were cleaned, assembled and demultiplexed using the program MOTHUR v 1.31.2 (Schloss *et al.* 2009). On average, between 1000 and 10,000 sequences were obtained per marker and per individual. Then, identical sequences were clustered and the number of reads per marker and per individual was counted.

As the number of reads greatly differed between markers (less than 100 reads to more than 1000 reads), applying a fixed threshold to keep final sequences was not possible. In addition, five markers displayed paralogous genes (e.g. more than two sequences with high and similar number of reads displayed). For this reason, selecting the sequence displaying the highest number of reads could lead to incorrectly selecting paralogous genes. Therefore, for each of the 22 markers, the number of reads obtained for 5-10 individuals was manually checked to determine a threshold to apply to each individual per marker. One to two sequences were kept per individual and per marker when paralogous genes were unambiguously absent, and up to ten sequences per individual and per marker were kept for genes displaying paralogs. On the 22 markers, three could not be used due to a too low number of reads obtained after sequence cleaning. Finally, a total of 18 genes were obtained, and since five of them displayed paralogs, 31 markers were available for further analyses. In total, 87 individuals were used for phylogenetic analyses (Extended Methods; Table S3.1)

### 3.3.3 *Gene-by-gene phylogenetic analyses*

For each marker a maximum likelihood (ML) phylogeny was build using PhyML 3.0 (Guindon *et al.* 2010). Haplotype networks were generated using the median-joining algorithm of Network, version 4.6.1.1 (Bandelt *et al.* 1999).

### 3.3.4 *Concatenated phylogenetic analysis*

The 31 markers of the 87 individuals were concatenated to create a supermatrix of 8899 nucleotides. Both parametric (Bayesian) and non-parametric (Neighbor-Joining; Parsimony and Maximum-likelihood) bootstrap analyses were conducted. Neighbor-Joining analyses were performed using MEGA 6.0.5 (Tamura *et al.* 2013), with 500 replicates. Parsimony analyses were performed using PAUP 4.0 b10 (Swofford 2003) and the tree bisection-reconnection branch swapping algorithm with 500 replicates and a heuristic search. Maximum Likelihood analyses were performed using RaxML 8.0.19 (Stamatakis 2014), with 500 replicates (fast bootstrapping) and the GTR+G model of nucleotide evolution. Bayesian analyses were performed using MrBayes 3.2.1 (Huelsenbeck *et al.* 2001), with a GTR+G model of nucleotide substitution. Sampling was performed through three independent runs (each having one cold chain and three heated chains), which were run for 50,000,000 generations or until the average standard deviation of split frequencies for the three runs was <0.01. Trees were sampled every 200 generations and the first fourth of the trees was discarded as burnin before a consensus tree was generated.

### 3.3.5 *Genetic clustering using Discriminant Analysis of Principal Components (DAPC)*

To determine the number of genetic groups without prior knowledge, we performed a DAPC using the *adegenet* 1.4-1 package from the R software (Jombart *et al.* 2010). The DAPC analysis maximizes the between group variance while minimizing the within groups variance. First, the Bayesian Information Criterion (BIC) was used to determine the optimal number of genetic clusters k. Then, DAPC was performed to define the clusters and visualize their relationships. It also provides membership probabilities, i.e. the probabilities for each individual to belong to a particular cluster. We performed analyses with and without COI to infer whether this marker significantly influenced the genetic clustering or not.

### 3.3.6 *Divergence model testing and molecular dating*

Since genetic clusters appeared as separate genetic entities (see Results) we considered that their relationships could be described by tree-like topologies, possibly assuming gene flow events between some clusters. Based on the genetic clusters found with DAPC, we calculated the average distance between clusters using between group

K2P distances (based on the 31 markers) implemented in MEGA 6.0.5. Then, we reconstructed a phylogenetic tree based on the distance between clusters, using the Neighbor-joining method, to define a first divergence scenario of the species complex (Figure 3.2A). In addition, we defined a second scenario including a hybridization event between the cluster C1 (including individuals of lineage L1) and the common ancestor of brooders (clusters C4 and C5, including individuals of lineages L2-L3-L4) giving rise to the cluster C2. The probability of scenarios, as well as effective sizes (N1-N8), divergence times (t1-t6) and admixture rate (r2) were estimated using Approximate Bayesian Computation (ABC) implemented in DIYABC v2.0.4 (Cornuet *et al.* 2014). Three summary statistics were used to construct posterior probabilities of scenarios: the number of haplotypes (within group), the number of segregating sites (between groups) and the $F_{ST}$ statistics (between groups). Default priors were used in preliminary analyses (20,000 simulated datasets), and were then adjusted using posterior distributions and pre-evaluation verifications. When each posterior probability of parameters fell in the prior range in preliminary analyses, 400,000 simulated datasets were used to accurately estimate posterior probabilities of parameters. In addition, posterior probabilities of each scenario were calculated. Model checking was performed using each available summary statistic, to verify that the parameter values of observed data belonged to posterior distributions.

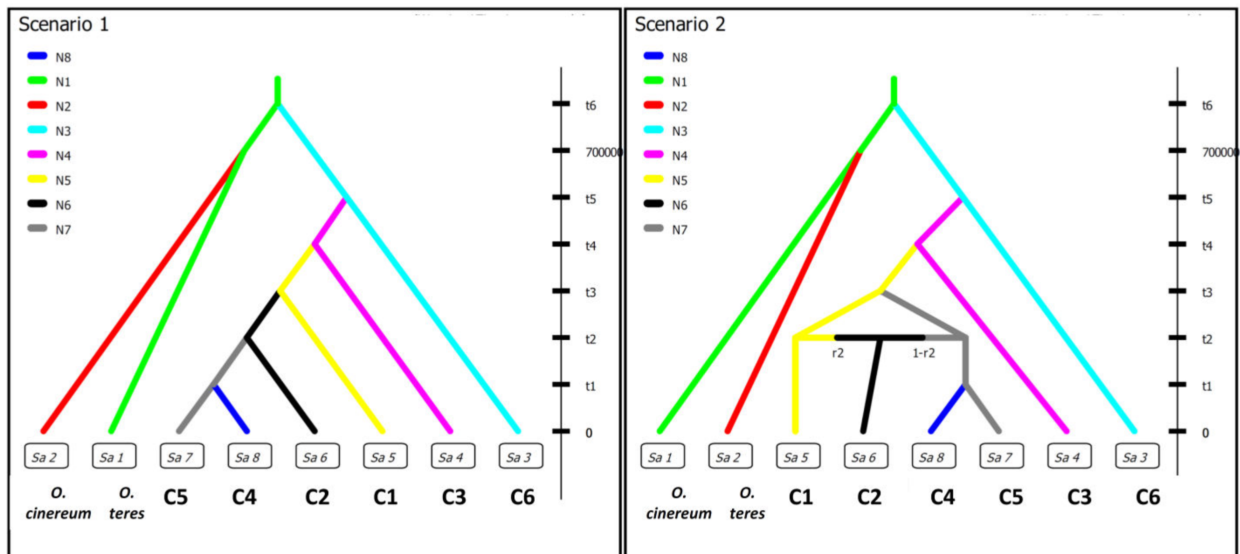



**Figure 3.1:** Divergence scenarios tested to infer evolutionary history of *O. longicauda* species complex divergence. A: Scenario 1: successive divergence without hybridization event. B: Scenario 2: successive divergence with a hybridization event between cluster C1 and the common ancestor of clusters C4 and C5.

**Table 3.1**: Mitochondrial lineages and sampling locations of individuals used in this study. L1-L6: *Ophioderma longicauda* mitochondrial lineages defined in Boissin et al, 2011. C1-C6: genetic clusters or clusters found in DAPC analysis. The population numbers refer to the number indicated in Figure 3.6.

| Locality | population number | L1 | L2 | L3 | L3b | L4 | L5 | L6 |
|---|---|---|---|---|---|---|---|---|
| Cap Vert Peninsula, Dakar, Senegal | 1a | - | - | - | - | - | - | 4 (C6) |
| Cap Manuel, Dakar, Senegal | 1b | - | - | - | - | - | 10 (C3) | - |
| Teneriffe, Canary Islands | 2 | - | - | - | - | - | 1(C1) | - |
| Madeira, Portugal | 3 | - | - | - | - | - | 1 (C1) | 1 (C6) |
| Algarve, Portugal | 4 | 7 (C1) | - | - | - | - | - | - |
| Ceuta, Spain | 5 | - | - | - | - | - | 2(C1) | - |
| Tabarka, Tunisia | 6 | 5 (C1) | - | - | - | - | - | - |
| Kelibia, Tunisia | 7 | - | - | - | 6 (C2) | - | - | - |
| Monastir, Tunisia | 8 | - | - | - | 5 (C2) | - | - | - |
| Agios Pavlos, Crete | 9 | 5 (C1) | - | 4 (C4) | - | - | - | - |
| Symi island, Greece | 10 | 4 (C1) | 5 (C4) | 2 (C4) | - | - | - | - |
| Baths of Aphrodite, Cyprus | 11 | 5 (C1) | - | - | - | 5 (C5) | - | - |
| Ramkine, Lebanon | 12a | 1 (C1) | 4 (C5) | - | - | - | - | - |
| Beirut+Raoucheh, Lebanon | 12b | - | - | - | - | 2 (C5) | - | - |

## 3.4 Results

### 3.4.1 *Single gene analysis shows incomplete lineage sorting*

Using transcriptome based markers, we successfully amplified, sequence and sort 31 informative genetic markers (Table 3.2). On average, between 1000 and 10,000 sequences were obtained per marker and per individual. Most markers provided unsolved topology, or resolved but incongruent between them. Network analyses showed that most markers displayed incomplete lineage sorting, except the markers 68241_I.I, i50_II, 98699 and COI (Figure 3.2).

**Table 3.2:** Markers used in this study. E-value threshold for BLAST search: $10^{-5}$. Sp = *Strongylocentrotus purpuratus*

| marker | length (pb) | used in DIYABC | best hit BLAST (e-value) |
|---|---|---|---|
| 1972 | 342 | yes | No hits found |
| 11915_I | 250 | no | Ubiquitine ($1e^{-35}$) |
| 11915_II | 250 | no | Ubiquitine ($1e^{-35}$) |
| 11915_III | 250 | yes | Ubiquitine ($1e^{-35}$) |
| 11915_V | 250 | no | Ubiquitine ($1e^{-35}$) |
| 41305 | 264 | no | No hits found |
| 50183_I | 319 | no | predicted: β-D-xylosidase 2 ($3e^{-31}$) |
| 50183_II | 319 | yes | predicted: β-D-xylosidase 2 ($3e^{-31}$) |
| 50183_III | 319 | no | predicted: β-D-xylosidase 2 ($3e^{-31}$) |
| 50183_IV | 319 | yes | predicted: β-D-xylosidase 2 ($3e^{-31}$) |
| 55384 | 352 | no | predicted: echinoidin-like isoform 2 ($1e^{-21}$) |
| 68241_I.I | 277 | no | No hits found |
| 68241_I.II | 291 | no | No hits found |
| 68241_II | 288 | no | No hits found |
| 79905 | 284 | yes | No hits found |
| 80488_II.I | 243 | yes | No hits found |
| 80488_II.II | 243 | yes | No hits found |
| 97479 | 364 | no | predicted: uncharacterized protein LOC100893945 Sp ($6e^{-35}$) |
| 98699 | 238 | yes | elongation factor 1-α ($4e^{-40}$) |
| 109458_I | 306 | yes | tubuline α-1 ($2e^{-66}$) |
| 109458_II | 306 | yes | tubuline α-1 ($2e^{-66}$) |
| 109458_III | 306 | no | tubuline α-1 ($2e^{-66}$) |
| 124557 | 249 | no | No hits found |
| 147510 | 324 | no | No hits found |
| 177720 | 241 | no | No hits found |
| COI | 306 | yes | Cytochrome Oxidase I (mitochondrial) |
| i36 | 307 | no | intron 36 |
| i50_I | 250 | no | intron 50 |
| i50_II | 435 | no | intron 50 |
| i54b | 249 | no | intron 54b |
| ITS1 | 360 | no | Internal transcribed spacer I |

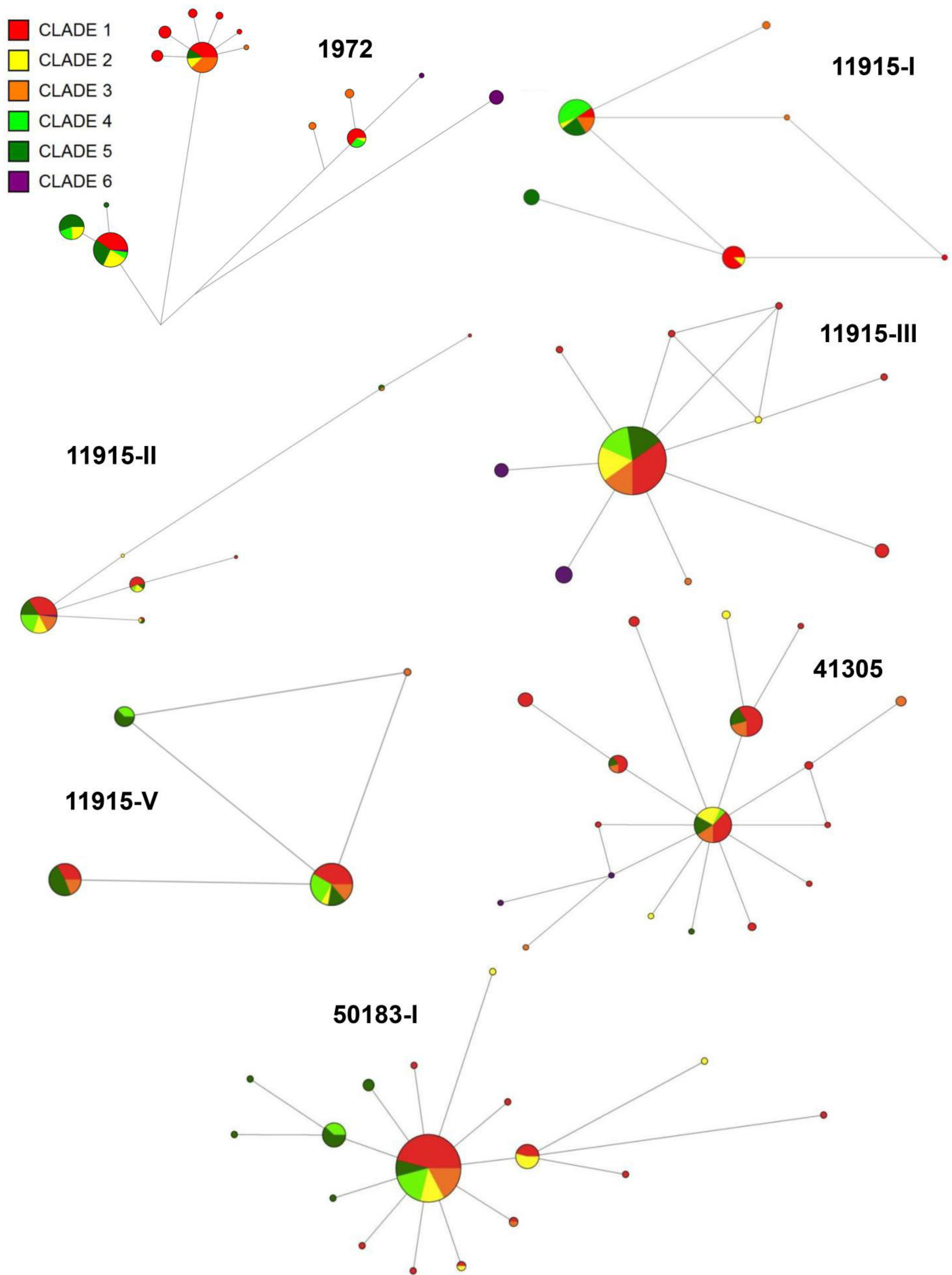



**Figure 3.2**: Haplotype networks of the 31 genetic markers used in this study. The six clusters found in DAPC analysis correspond to the different colors.

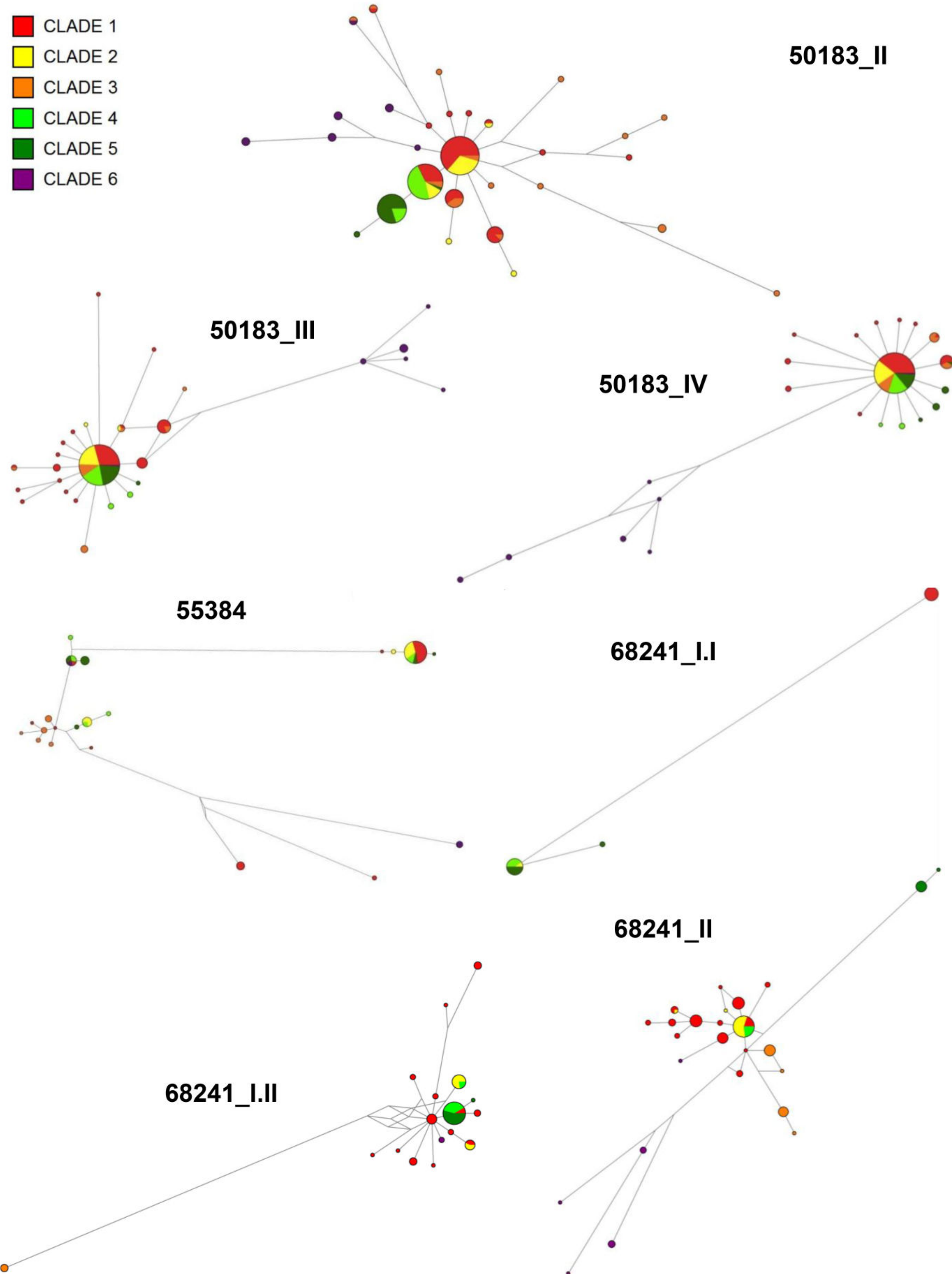



**Figure 3.2**: continued

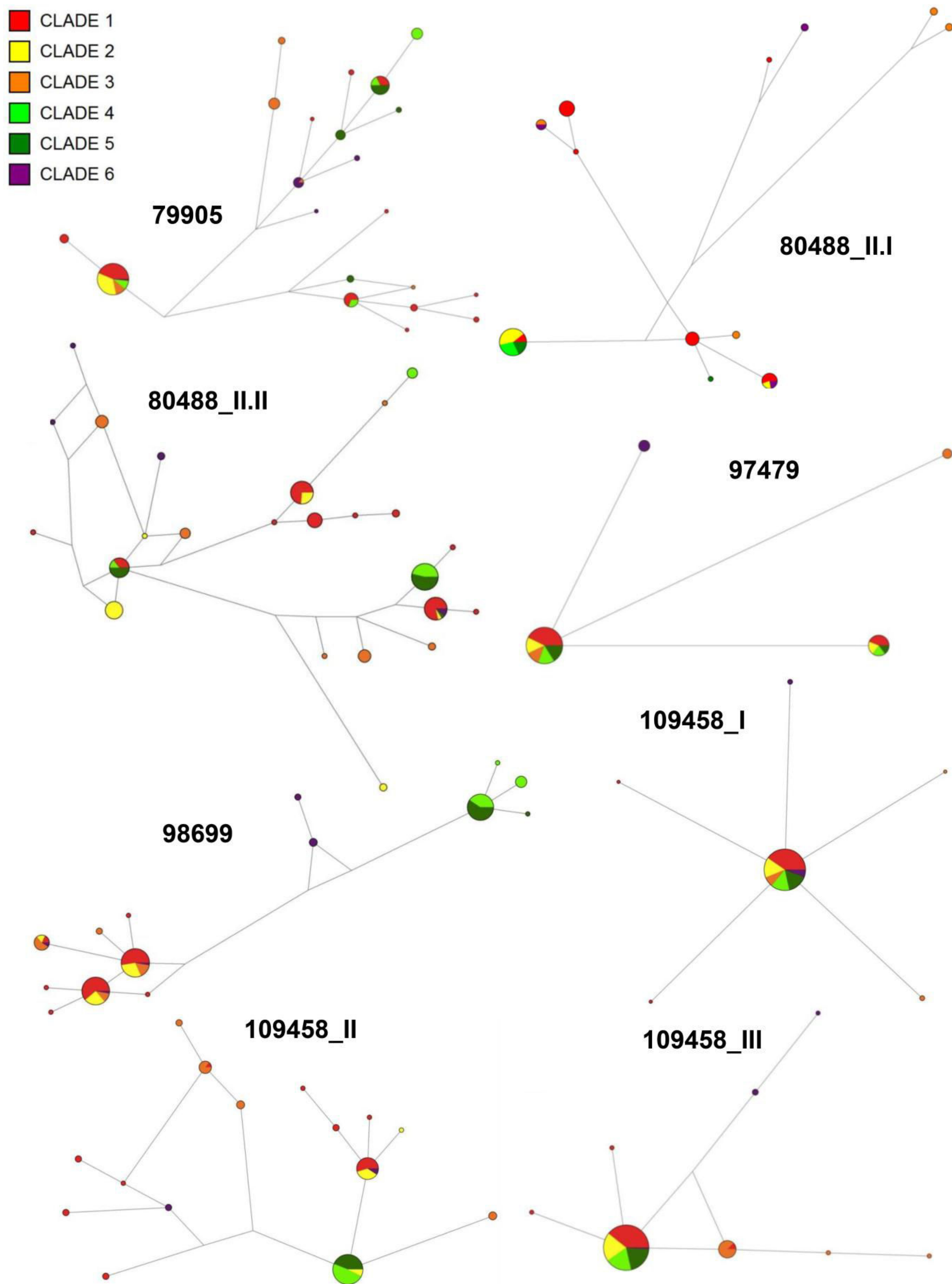



**Figure 3.2**: continued

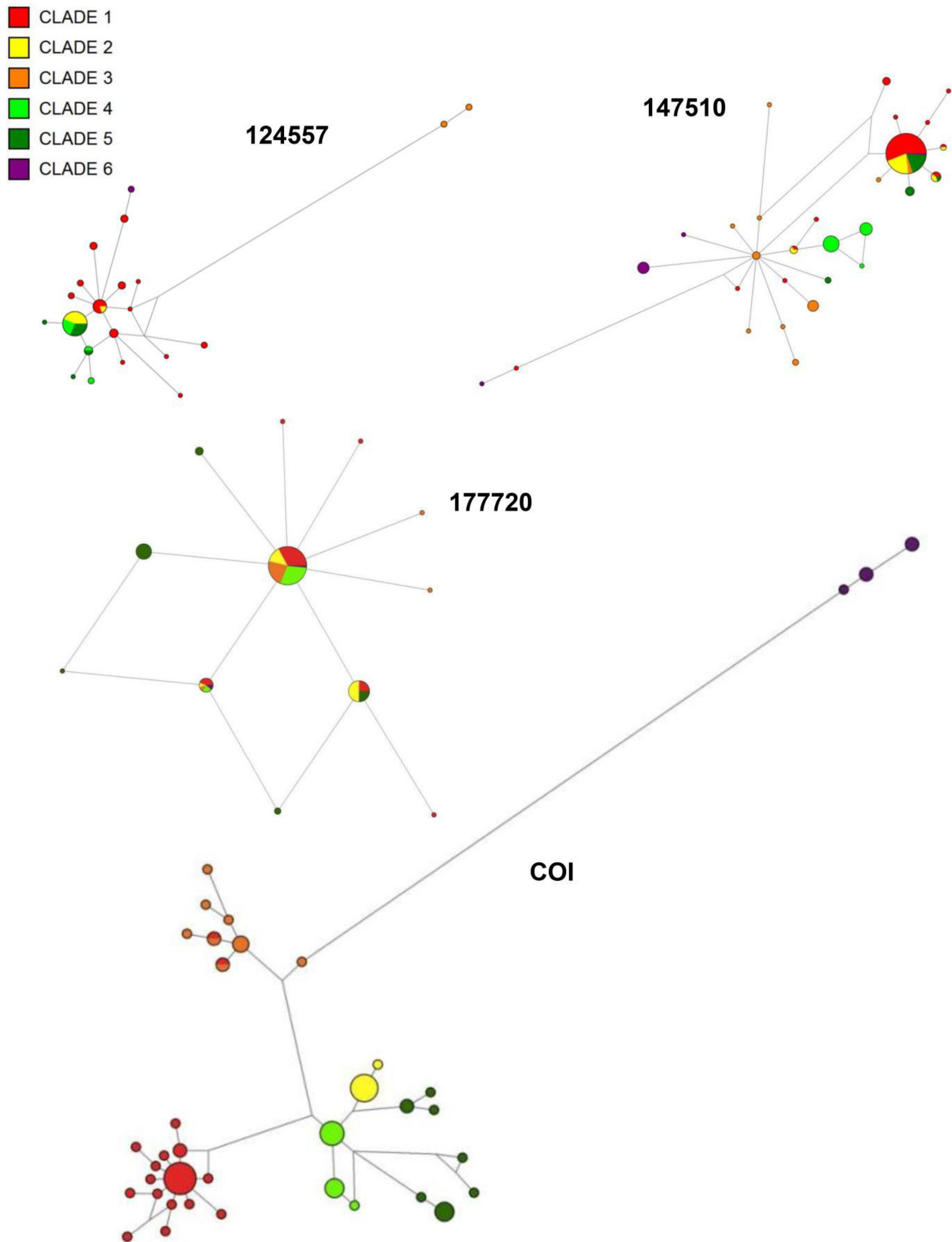



**Figure 3.2**: continued

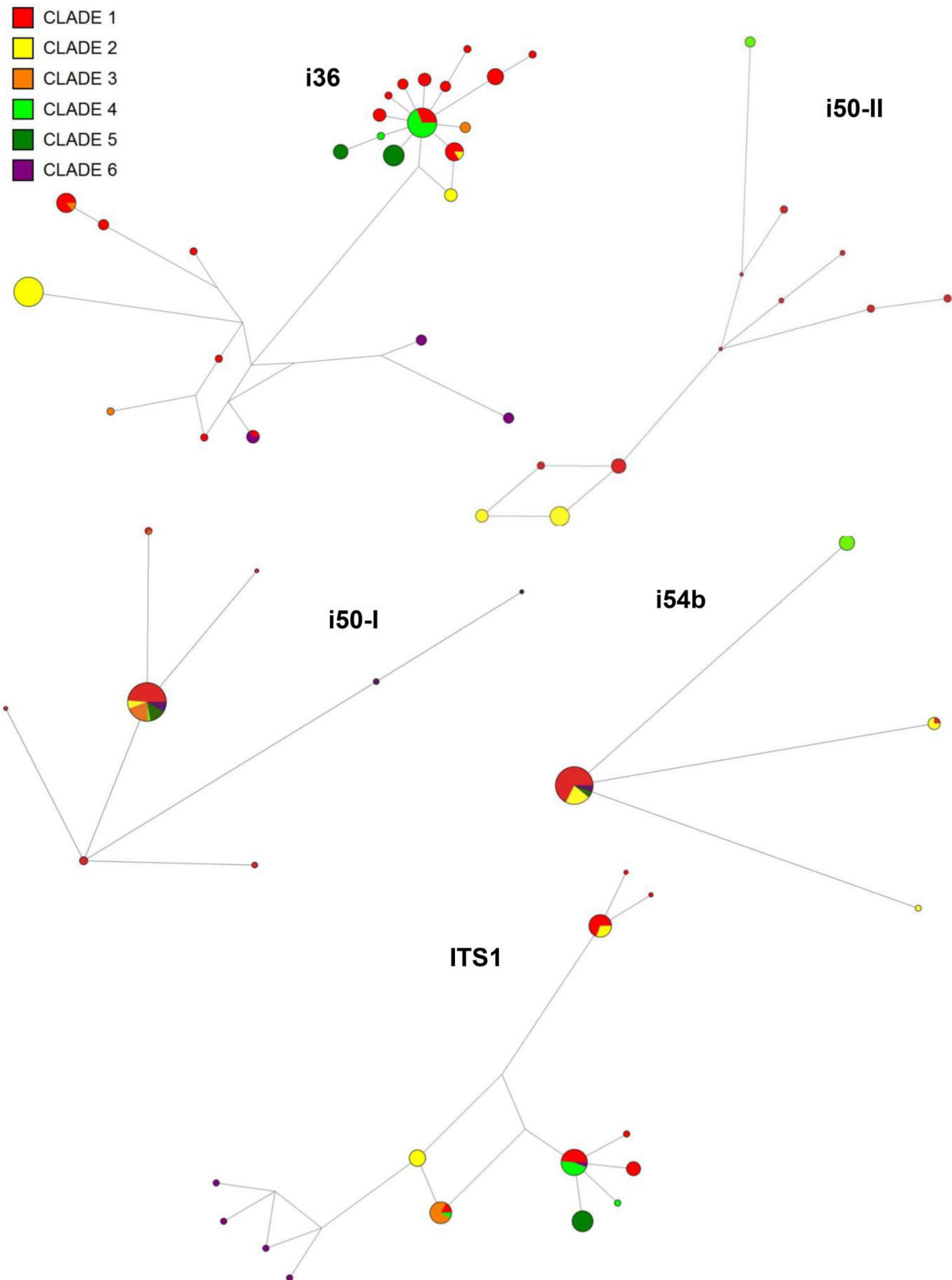



**Figure 3.2**: continued

### 3.4.2 *Concatenated dataset provides unresolved tree*

The Neighbor-Joining analysis provided unresolved topology without particular grouping of lineages or depending on geography, except a group including L5 and L6 individuals from Dakar (bootstrap estimate 84%) and the outgroups (100%). The parsimony analysis identified two monophyletic groups moderately supported, one including each L6 individual (84%) and another one including outgroups (74%). When excluding the only L6 individual from Madeira, all L6 individuals from Dakar form a monophyletic group highly supported (98%). No other group was supported within *O. longicauda*.

The Maximum Likelihood analysis of the concatenated dataset using strongly supported the three *Ophioderma* outgroup species (100% bootstrap) and provided a moderate support for the lineage L6 (90.8%) and the individuals of lineage L5 from Dakar (84.6%). Yet, the other L5 individuals (from Ceuta, Madeira and Teneriffe) did not cluster with individuals from Dakar. A group including brooders (L2-L3-L4) was displayed, yet this group was not supported (32.6%). Another group including individuals from lineages L1-L3b-L5 was also displayed but poorly supported (23%). Other L3b and L1 individuals did not cluster with the described groups. The Bayesian analysis did not converge after 50 million generations, the average standard deviation of split frequencies being at 0.18 (and not <0.1).

### 3.4.3 *DAPC analysis identifies six genetic clusters*

A DAPC analysis was performed to define the different genetic groups in the *O. longicauda* species complex without considering a priori knowledge from mitochondrial lineages. This analysis, contrary to phylogenetic reconstructions, used genotypic information for each individual and each locus. Our results showed that the optimal number of clusters (which corresponds to the minimal BIC value) was six (Figure 3.3A). The six clusters were very distinct with nearly no overlapping in the 2D representation (Figure 3.3B) and 100% probability of memberships for each individual (Figure 3.3C). The cluster C6, including all individuals of mitochondrial lineage L6 (from Dakar and Madeira) forms a well defined group distant from the five other groups. The individuals of lineage L5 are distributed in two different genetic clusters. All individuals from Dakar form the cluster C3, whereas the other L5 individuals (from Madeira, Teneriffe and

Ceuta) join all L1 individuals in cluster C1. Each individual of lineage L3b belongs to cluster C2, closely related to cluster C1. Finally, the brooding individuals are distributed in two different genetic clusters, incongruent with mitochondrial lineages but congruent with geography. Cluster C4 includes all individuals of lineages L2 and L3 sampled in Greece (Crete, Symi Islands and Rhodes), whereas cluster C5 includes all individuals of lineages L2 and L4 sampled in Cyprus and Lebanon. For each analyzed individual, the membership probability to a group was 100%, indicating that no "hybrid" (or admixed) individual between groups could be detected (Figure 3.3C). The DAPC analysis run without the COI marker yielded the same clustering, the only difference being that the clusters C1 and C3 were more distant on the graph (Figure S3.1). Some of the clusters were collected in the same locality, for instance C1 and C4 in Greece, C1 and C5 in Cyprus and Lebanon, C1 and C6 in Madeira (Table 3.1). This proves that there are reproductive barriers among those clusters, and they should therefore be considered as different biological species.

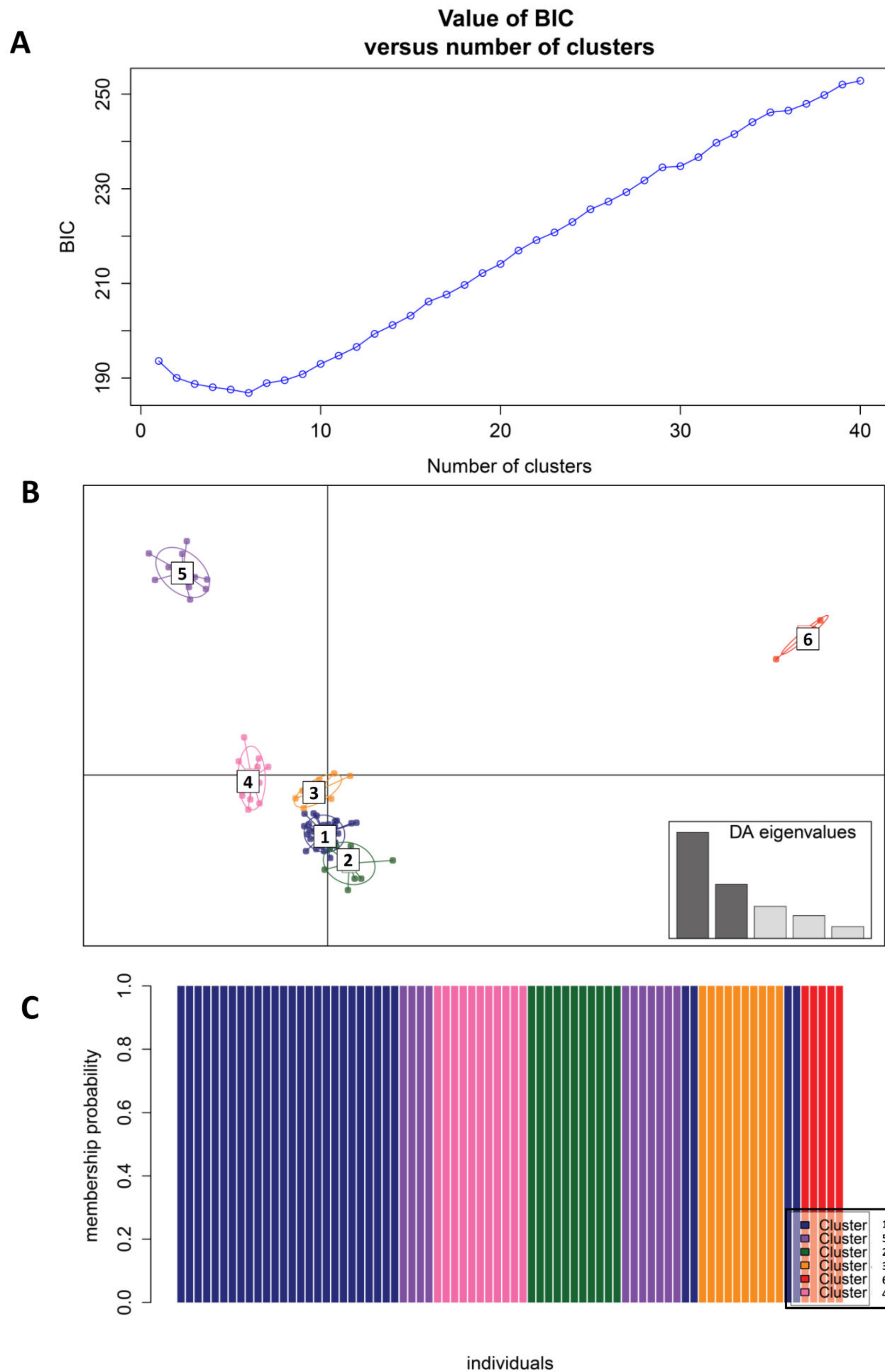



**Figure 3.3:** Results of the Discriminant Analysis of Principal Components. A: The minimal value of Bayesian Information Criterion (BIC) indicates the optimal number of genetic clusters. B: DAPC analysis showing the six different genetic clusters. C: Membership probability of each individual to belong to a particular genetic cluster

### 3.4.4 *Divergence scenario suggest hybridization event*

The DAPC analysis identified six different genetic clusters, some of them being unambiguous biological species. Yet, the historical relationships among them are not known. Based on the average genetic distance (K2P) among them, a Neighbor-Joining tree was reconstructed (scenario 1, Figure 3.1). An alternative scenario to test was proposed, including a hybridization event between the cluster 1 and the common ancestor of clusters 4 and 5, which gave rise to cluster 2 (scenario 2, Figure 3.1). Posterior probabilities of scenarios indicate that the most probable scenario is the scenario 2, including hybridization, both in the direct estimation (Figure 3.4H) and in the estimation based on logistic regression (Figure 3.4I). Model checking was also performed to assess the validity of estimated parameters (Figure 3.4G), and the observed dataset fitted well in the range of simulated datasets.

Species or clusters with larvae displayed 3-7 times larger effective size estimates (outgroup and clusters 1, 3, 6) than brooding clusters (4 and 5) (Table 3.3). Cluster 2, presumably brooder, also displayed a reduced effective size compared to the closely related cluster 1. The divergence time estimations indicated that cluster 6 split from other *O. longicauda* clusters about 2.2 MYA (Figure 3.5). Then, cluster 5 split about 830,000 years ago. The split between the broadcasting cluster 1 and the common ancestor of brooding clusters (clusters 4 and 5) occurred about 550,000 years ago. Then, a hybridization event between cluster 1 and the common ancestor of brooding clusters occurred about 400,000 years ago, giving rise to cluster 2. Finally, cluster 4 and cluster 5 separated about 120,000 years ago. Overall, those divergence events follow a pattern of West to East differentiation (Figure 3.5).

**Table 3.3:** Posterior probability values of estimated parameters. N = effective size; t = divergence times in number of generations, see details in Figure 3.2. Divergence time estimation in years are given assuming generation time of five years). r2 = admixture rate from C1 to C4-C5. Q025 and q975 indicate the range of 95% confidence interval.

| Parameter | Species / cluster | mode | q025 | q975 |
|---|---|---|---|---|
| N1 | *O.teres* | 209,000 | 126,000 | 500,000 |
| N2 | *O. cinereum* | 275,000 | 147,000 | 518,000 |
| N3 | cluster 6 | 242,000 | 132,000 | 366,000 |
| N4 | cluster 3 | 207,000 | 134,000 | 272,000 |
| N5 | cluster 1 | 282,000 | 166,000 | 483,000 |
| N6 | cluster 2 | 84,700 | 32,200 | 171,000 |
| N7 | cluster 4 | 41,700 | 13,800 | 64,900 |
| N8 | cluster 5 | 72,200 | 18,600 | 353,000 |
| t1 (years) | | 24,500 (122,500) | 6,690 | 52,700 |
| t2 (years) | | 81,700 (408,500) | 25,600 | 144,000 |
| r2 | | 0,673 | 0,0816 | 0,943 |
| t3 (years) | | 111,000 (555,000) | 46,400 | 240,000 |
| t4 (years) | | 167,000 (835,000) | 93,400 | 320,000 |
| t5 (years) | | 448,000 (2,240,000) | 253,000 | 631,000 |
| t6 (years) | | 998,000 (4,990,000) | 862,000 | 999,000 |

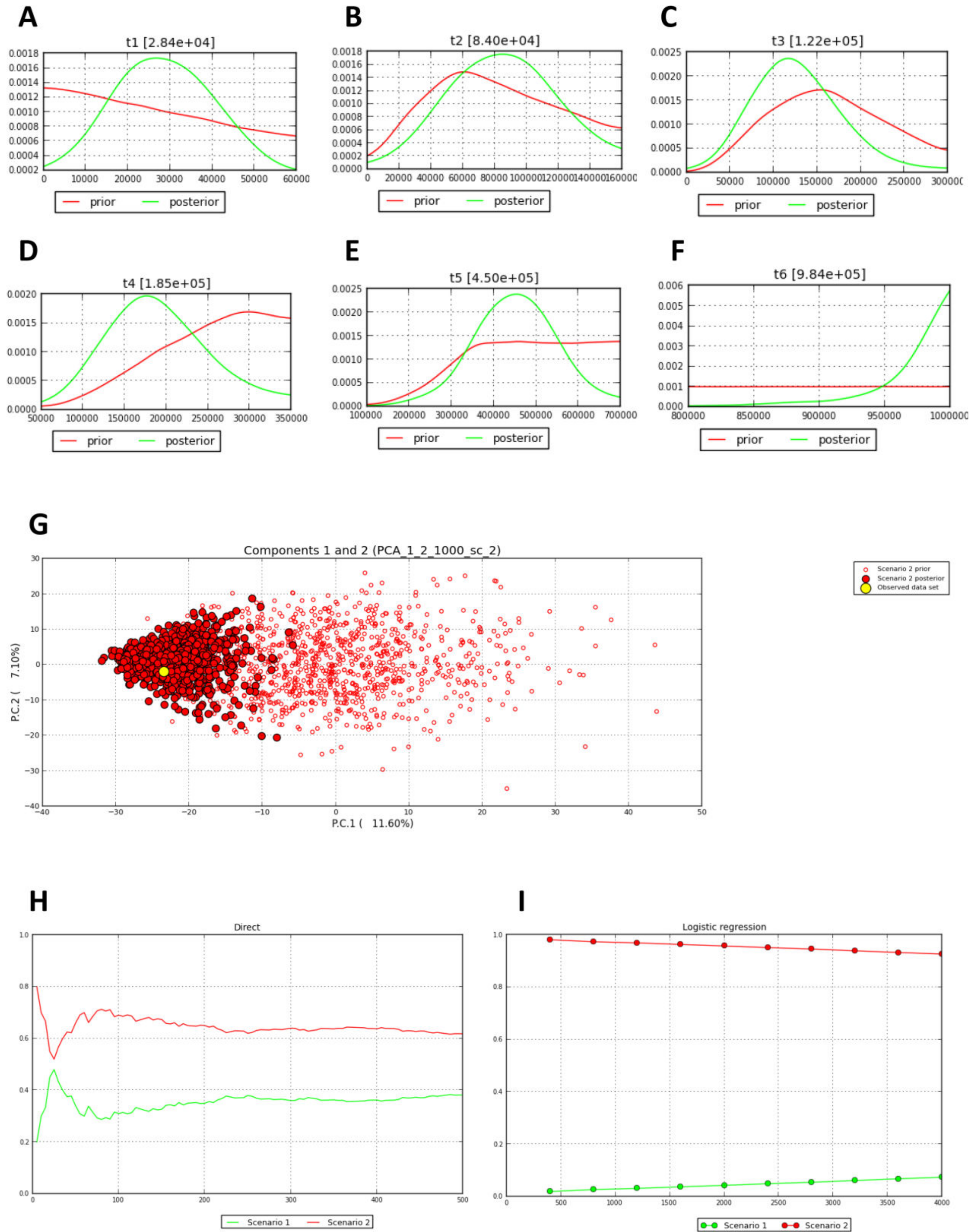



**Figure 3.4:** A-F: posterior probabilities of divergence time estimations performed with DIYABC. G: Model checking for scenario 2, to assess if the observed data fell within the range of estimated datasets. H-I: Probability of scenarios 1 and 2 using the direct estimation (H) and the logistic regression estimation (I).

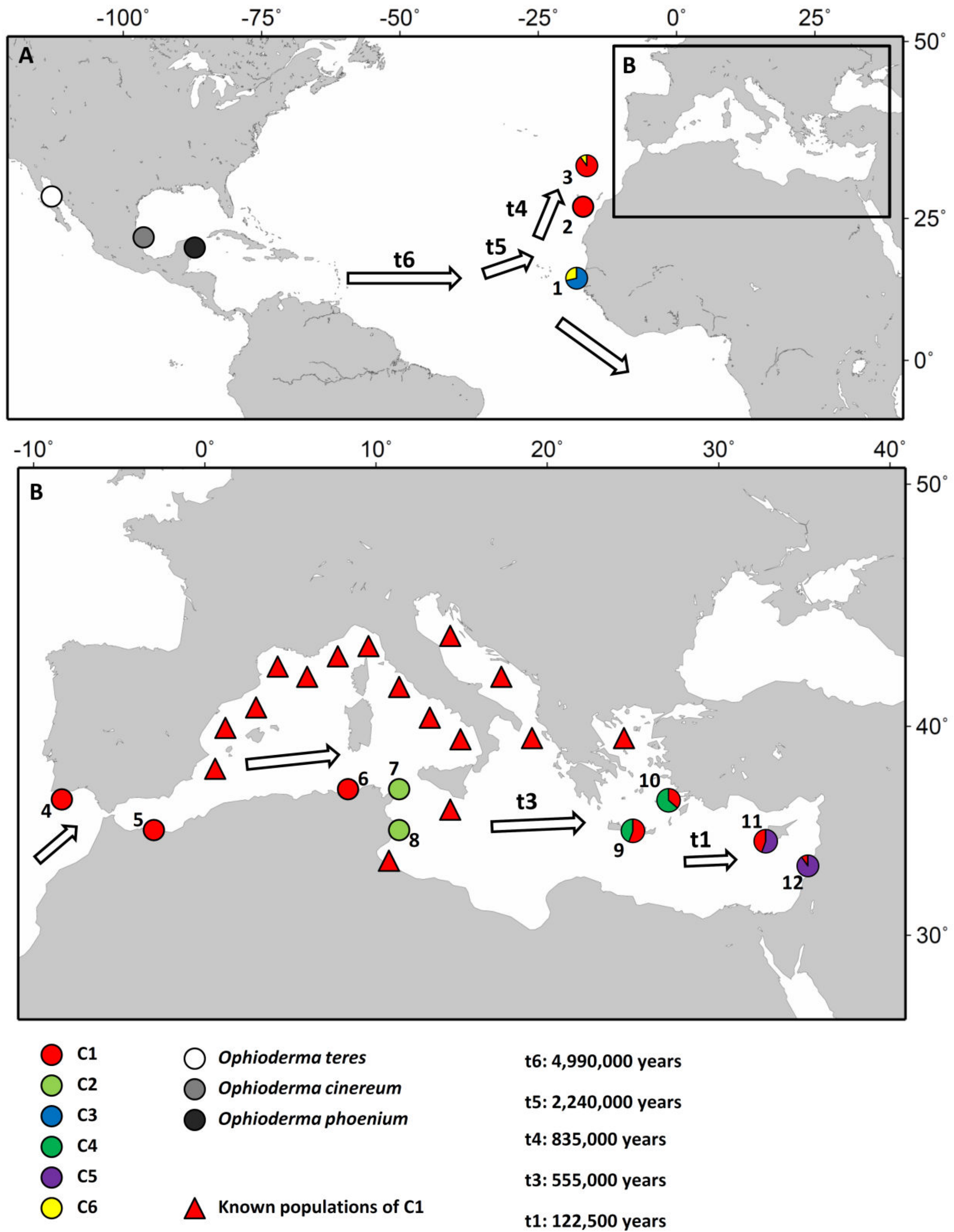



**Figure 3.5:** Biogeographic pattern of divergence in *O. longicauda* species complex. Divergence dates were estimated using DIYABC.

## 3.5 Discussion

### 3.5.1 *Species delimitation in O. longicauda*

In this study, we used more than 30 genetic markers to delimitate species in the *O. longicauda* species complex. The DAPC analysis clearly defined six genetic clusters, and the historical scenario leading to those six clusters and the divergence times could be inferred using Approximate Bayesian Computations. Those results suggest that up to three broadcasting *Ophioderma longicauda* species (the clusters C1, C3 and C6) might occur in the Atlantic and Mediterranean. The cluster C6, including all L6 individuals, represents very likely a new biological species, since this group was well defined in the concatenated phylogenetic analysis and in most of the gene-by-gene analyses. In addition, one individual of cluster C6 was sampled in Madeira in the same locality (during the same dive) than individuals from cluster C1. The genetic analyses showed that there was no gene flow between this C6 individual and the other C1 individuals, suggesting that those two clusters are reproductively isolated and should therefore be considered as different biological species. Furthermore, individuals from cluster C6 were sampled close to the localities where individuals of cluster C3 were sampled (about 11-17 km apart). Although they were not sampled in the very same locations, C6 and C3 individuals do not seem to exchange genes anymore as those two groups were genetically very distant and should therefore be considered as different biological species.

Clusters C1 and C3 may represent two different biological species or a single species that displays genetic differentiation due to geographical distance. Yet, since cluster C1 displays a very large geographic distribution (with individuals from Canaries Islands and Lebanon clustering undoubtedly together), dispersal does not seem limiting in this cluster. If it was present in Senegal (which is less distant and less oceanographically isolated from Canary Islands than Lebanon) its populations should not be in different clusters. Therefore, we believe that clusters C1 and C3 are different biological species, but samples from less distant areas than Dakar are needed to confirm this hypothesis.

Our results confirmed that the brooding clusters C4 and C5 represent different biological species from the cluster C1, since those two clusters were genetically highly

differentiated from C1. In addition, C4 and C5 are found in sympatry with C1, and C1 and C4 reproductive periods are distinct (Weber *et al.* 2014). Furthermore, strong differences in effective size estimations are congruent with dispersal abilities of these clusters, with high effective size for species with dispersing larvae (clusters C1, C3, C6 and outgroups) and low effective size for brooding species with very low dispersal (clusters C4 and C5). Whether clusters C4 and C5 are different brooding species is difficult to say. Those two clusters are genetically differentiated as shown by DAPC analysis, yet this signal might be due to the very low dispersal abilities of these clusters, since they were not sampled in the same locality. Clusters C4 and C5 could therefore be two differentiated populations within a single very structured brooding species. A denser sampling along the eastern Mediterranean coast may allow inferring whether there is a genetic break between C4 and C5, reinforcing the hypothesis that C4 and C5 are different species, or whether genetic differentiation follows a pattern of isolation by distance, for instance.

According to ABC scenario testing, the most likely scenario of divergence is the one including a hybridization event between cluster C1 and the common ancestor of brooders, which gave rise to cluster C2 (found in Tunisia). This scenario is congruent with general observations of this cluster. Indeed, it displays typical characteristics of brooders, for instance a mitochondrial lineage close to the cluster C4, a reduced effective size compared to broadcasters and an ecological preference to low depth, since this species was sampled between 0.3 and 3m in Tunisia (2 populations), like the cluster C4, cluster C1 being sampled between 10m and 20m in Tunisia (2 populations) (chapter 2; Weber *et al.* 2014). In contrast, the DAPC analysis showed that cluster C2 was genetically very close to broadcasting cluster C1. In addition, the nuclear marker i51 was polymorphic in cluster C2, as observed in clusters C1, C3, C6, whereas it was monomorphic in brooding clusters C4 and C5. Therefore, the cluster C2 seems to represent a new biological species ecologically differentiated from cluster C1. Whether this species broods its progeny or not is unknown, since individuals were sampled in July and the gonad observation was inconclusive to answer this question. In addition, brooding individuals were sampled in Tunisia in 1849 and 1924 (Stöhr *et al.* 2009), but no genetic analyses could be performed due to poor DNA quality, despite several extraction and amplification tests. These brooding individuals, sampled at about 100 km from C2 individuals, probably belong to the cluster C2.

### 3.5.2 *Divergence and speciation history in Ophioderma genus: a hypothesis*

From a biogeographic point of view, the genus *Ophioderma* displays a general pattern of speciation from West to East (Figure 3.6). So far, only two *Ophioderma* species have been described in Eastern Atlantic; *Ophioderma longicauda* from Dakar to Spain in the Atlantic and in the Mediterranean, and *Ophioderma wahlbergii* (Müller & Troschel, 1842) in South Africa. The emergence of this genus occurred most likely around the actual Caribbean Sea, before the closing of Panama Isthmus. Indeed, most species (26/28) of this genus described today live in this region (Stöhr *et al.* 2009). Then, the common ancestor of *O. longicauda* species complex and *O. wahlbergii* somehow crossed the Atlantic Ocean. The closest continental coast is the coast of Senegal. From this point, we can hypothesize that the ancestor of *O. wahlbergii* and (the ancestor of cluster C6 of) *O. longicauda* diverged. Interestingly, a "form" of *Ophioderma longicauda* was described in Sao Tomé Island (Equatorial Guinea) by Greeff (1881), under the name of *Ophioderma longicauda f. guineense.* Later, it was included in the species *Ophioderma longicauda*, as the morphological characters fell within the range of *O. longicauda* (Madsen 1970). It is possible that this *O. longicauda* "form" is actually the different biological species that we defined as C6. Yet, samples from this locality are required to confirm this hypothesis.

Then, around Senegal, Clusters C6 and C3 separated about 2.2 MYA. Later, cluster C1 diverged from cluster C3 about 830,000 years ago and colonized the Mediterranean. A plausible explanation for the presence of mitochondrial haplotypes L5 in C1 is that C3 and C1 hybridized. Haplotypes related to L5 are not found in brooders but this should rule out the possibility that the event took place after their divergence, since L5 haplotypes in C1 are not found in the eastern basin. However, hybridization between species is more likely to occur while their divergence is recent. About 550,000 years ago, the common ancestor of all brooders diverged from cluster C1, probably around the central Mediterranean basin. Then, about 400,000 years ago, C1 and the common brooder ancestor hybridized leading to C2. Finally, migrating further eastward, the brooding cluster C5 diverged from the cluster C4 about 120,000 years ago.

### 3.5.3 *COI barcodes: benefits and pitfalls*

Barcodes such as COI in species delimitation and species complexes discovery have been widely used (e.g. Hebert *et al.* 2003). Indeed, it seemed to be an ideal marker

because of its ubiquity, easiness to amplify even in poor quality DNA samples (numerous mitochondria for a single nucleus), high mutation rate and its reduced effective size with respect to nuclear DNA which implies lower coalescence times. In fact, it allowed in the first place discovering that *O. longicauda* was a species complex (Stöhr *et al.* 2009; Boissin *et al.* 2011). Nevertheless, we showed that mitochondrial lineages did not exactly correspond to the genetic groups defined using nuclear data (which most likely correspond to species). This could either be explained by the stochasticity of the coalescent process, or by introgression, which would explain why individuals of mitochondrial lineage L5 cluster in the cluster C1 with L1 individuals. Mitochondrial DNA is particularly prone to both selective and introgression sweeps (Currat *et al.* 2008; Galtier *et al.* 2009; Toews & Brelsford 2012; Pons *et al.* 2014) with respect to nuclear DNA. Finally, selection events may be responsible of the retention of particular mitochondrial haplotypes. This study emphasizes the necessity of using several markers (including nuclear ones) to accurately delimit species.

### 3.5.4 *DAPC and DIYABC: an efficient combination of methods for accurate delimitation of recent/incipient species*

In this study, we used a two-stage approach to species delimitation. First, we inferred the number of genetic clusters using DAPC, which is a multivariate method to infer genetic structure of populations from multilocus genotypes (Jombart *et al.* 2010). Then, we inferred the most probable historical scenario comparing three scenarios, using Approximate Bayesian Computation (e.g. Csilléry *et al.* 2010) implemented in DIYABC v2.0 (Cornuet *et al.* 2014). Such a two-stage approach has already been performed in other studies (e.g. (Leaché & Fujita 2010; Barrett & Freudenstein 2011) using a combination of Structurama (Huelsenbeck & Andolfatto 2007; Huelsenbeck *et al.* 2011) or Structure (Pritchard *et al.* 2000) to find genetic clusters based on Bayesian inference and BPP (Yang & Rannala 2010) to delimitate species based on those clusters, using a initial defined species tree. Besides being orders of magnitude faster than Bayesian clustering methods (Structure or Structurama), DAPC performed better under complex population genetic models (Jombart *et al.* 2010). In addition, the ABC approach allows comparing scenarios including hybridization events, whereas BPP assumes that all of the shared polymorphism is the result of unsorted ancestral polymorphism, not taking into account putative hybridization events. In addition, ABC allows estimating

population genetic parameters such as effective size of populations and divergence times, as well as marker mutation rates. Nevertheless, model based methods such as ABC should be used with caution because if a model has a higher posterior probability than another, it does not mean that it is true. Therefore, several plausible models should be tested. In our case, we think that we explored the most plausible models, remaining in the limits of parameter complexity compatible with the amount of information of our data, and that the scenario selected is very close to the reality.

The combination of DAPC and ABC applied on genetic data from 31 markers revealed unambiguous genetic entities in *Ophioderma longicauda* and dated their divergences and hybridization event. From the 28 studies presented in a recent review on species delimitation, only two used more than 10 genetic markers for species delimitation (Carstens *et al.* 2013). This review emphasized the low number of new species actually described after genetic delimitation. Accordingly, taxonomic revision should be performed for the brooding species including the clusters C4 and C5, which are undoubtedly different species from broadcasting clusters C1, C3 and C6. Current work with taxonomist is ongoing to name this new species *Ophioderma vivipara*.

**Table S3.1**: Sample species names, individual codes and localities used for this study. Individuals excluded from concatenated dataset analyses are in grey.

| Species / lineage | sampling location | individual code | Species / lineage | sampling location | individual code | Species / lineage | sampling location | individual code |
|---|---|---|---|---|---|---|---|---|
| L1 | Algarve, Portugal | Alg001 | L2 | Ramkine Island, Lebanon | Ram023 | L5 | Paul do Mar, Madeira | PdM009 |
| L1 | Algarve, Portugal | Alg002 | L2 | Ramkine Island, Lebanon | Ram024 | L5 | Garachico, Teneriffe | Gar004 |
| L1 | Algarve, Portugal | Alg003 | L2 | Ramkine Island, Lebanon | Ram025 | L5 | Dakar, Senegal | Da1001 |
| L1 | Algarve, Portugal | Alg004 | L2 | Ramkine Island, Lebanon | Ram026 | L5 | Dakar, Senegal | Da1002 |
| L1 | Algarve, Portugal | Alg005 | L3 | Agios Pavlos, Crete, Greece | AgP001 | L5 | Dakar, Senegal | Da1003 |
| L1 | Algarve, Portugal | Alg006 | L3 | Agios Pavlos, Crete, Greece | AgP015 | L5 | Dakar, Senegal | Da1004 |
| L1 | Algarve, Portugal | Alg007 | L3 | Agios Pavlos, Crete, Greece | AgP028 | L5 | Dakar, Senegal | Da1005 |
| L1 | Tabarka, Tunisia | Tab001 | L3 | Agios Pavlos, Crete, Greece | AgP029 | L5 | Dakar, Senegal | Da1006 |
| L1 | Tabarka, Tunisia | Tab002 | L3 | Symi Island, Greece | Sym007 | L5 | Dakar, Senegal | Da2001 |
| L1 | Tabarka, Tunisia | Tab005 | L3 | Symi Island, Greece | Sym009 | L5 | Dakar, Senegal | Da2002 |
| L1 | Tabarka, Tunisia | Tab011 | L3b | Kelibia, Tunisia | Kel001 | L5 | Dakar, Senegal | Da4001 |
| L1 | Tabarka, Tunisia | Tab019 | L3b | Kelibia, Tunisia | Kel002 | L5 | Dakar, Senegal | Da4002 |
| L1 | Agios Pavlos, Crete, Greece | AgP002 | L3b | Kelibia, Tunisia | Kel003 | L6 | Naples, Italy | Nap005 |
| L1 | Agios Pavlos, Crete, Greece | AgP004 | L3b | Kelibia, Tunisia | Kel014 | L6 | Paul do Mar, Madeira | PdM003 |
| L1 | Agios Pavlos, Crete, Greece | AgP007 | L3b | Kelibia, Tunisia | Kel018 | L6 | Dakar, Senegal | Da3001 |
| L1 | Agios Pavlos, Crete, Greece | AgP010 | L3b | Kelibia, Tunisia | Kel039 | L6 | Dakar, Senegal | Da3002 |
| L1 | Agios Pavlos, Crete, Greece | AgP017 | L3b | Monastir, Tunisia | Mon003 | L6 | Dakar, Senegal | Da3003 |
| L1 | Symi Island, Greece | Sym003 | L3b | Monastir, Tunisia | Mon005 | L6 | Dakar, Senegal | Da3004 |
| L1 | Symi Island, Greece | Sym005 | L3b | Monastir, Tunisia | Mon015 | *O. phoenium* | Puerto Morelos, Mexico | O.phoen1 |
| L1 | Symi Island, Greece | Sym010 | L3b | Monastir, Tunisia | Mon016 | *O. phoenium* | Mahahual, Mexico | O.phoen2 |
| L1 | Symi Island, Greece | Sym011 | L3b | Monastir, Tunisia | Mon017 | *O. phoenium* | Belize | O.phoen3 |
| L1 | Baths of Aphrodite, Cyprus | BoA013 | L4 | Baths of Aphrodite, Cyprus | BoA066 | *O. phoenium* | Belize | O.phoen4 |
| L1 | Baths of Aphrodite, Cyprus | BoA016 | L4 | Baths of Aphrodite, Cyprus | BoA067 | *O. cinereum* | Islas Lobos, Mexico | O.cinere1 |
| L1 | Baths of Aphrodite, Cyprus | BoA019 | L4 | Baths of Aphrodite, Cyprus | BoA068 | *O. cinereum* | Islas Lobos, Mexico | O.cinere2 |
| L1 | Baths of Aphrodite, Cyprus | BoA061 | L4 | Baths of Aphrodite, Cyprus | BoA069 | *O. cinereum* | Belize | O.cinere3 |
| L1 | Baths of Aphrodite, Cyprus | BoA063 | L4 | Baths of Aphrodite, Cyprus | BoA070 | *O. cinereum* | St-Vincent, Florida, USA | O.cinere4 |
| L1 | Ramkine Island, Lebanon | Ram027 | L4 | Raoucheh, Lebanon | Rao001 | *O. teres* | Guaymas, Mexico | O.teres1 |
| L2 | Symi Island, Greece | Sym001 | L4 | Beirut, Lebanon | Bei002 | *O. teres* | Guaymas, Mexico | O.teres2 |
| L2 | Symi Island, Greece | Sym002 | L5 | Canaries Islands | Can001 | *O. teres* | Panama | O.teres3 |
| L2 | Symi Island, Greece | Sym004 | L5 | Algero-Tunisian Coast | ATu007 | *O. teres* | Panama | O.teres4 |
| L2 | Symi Island, Greece | Sym006 | L5 | Ceuta, Spain | Ceu001 | *Ophiothrix fragilis* | Crete, Greece | O.fragilis |
| L2 | Symi Island, Greece | Sym008 | L5 | Ceuta, Spain | Ceu003 | *Ophiomyxa pentagona* | Crete, Greece | O.penta |

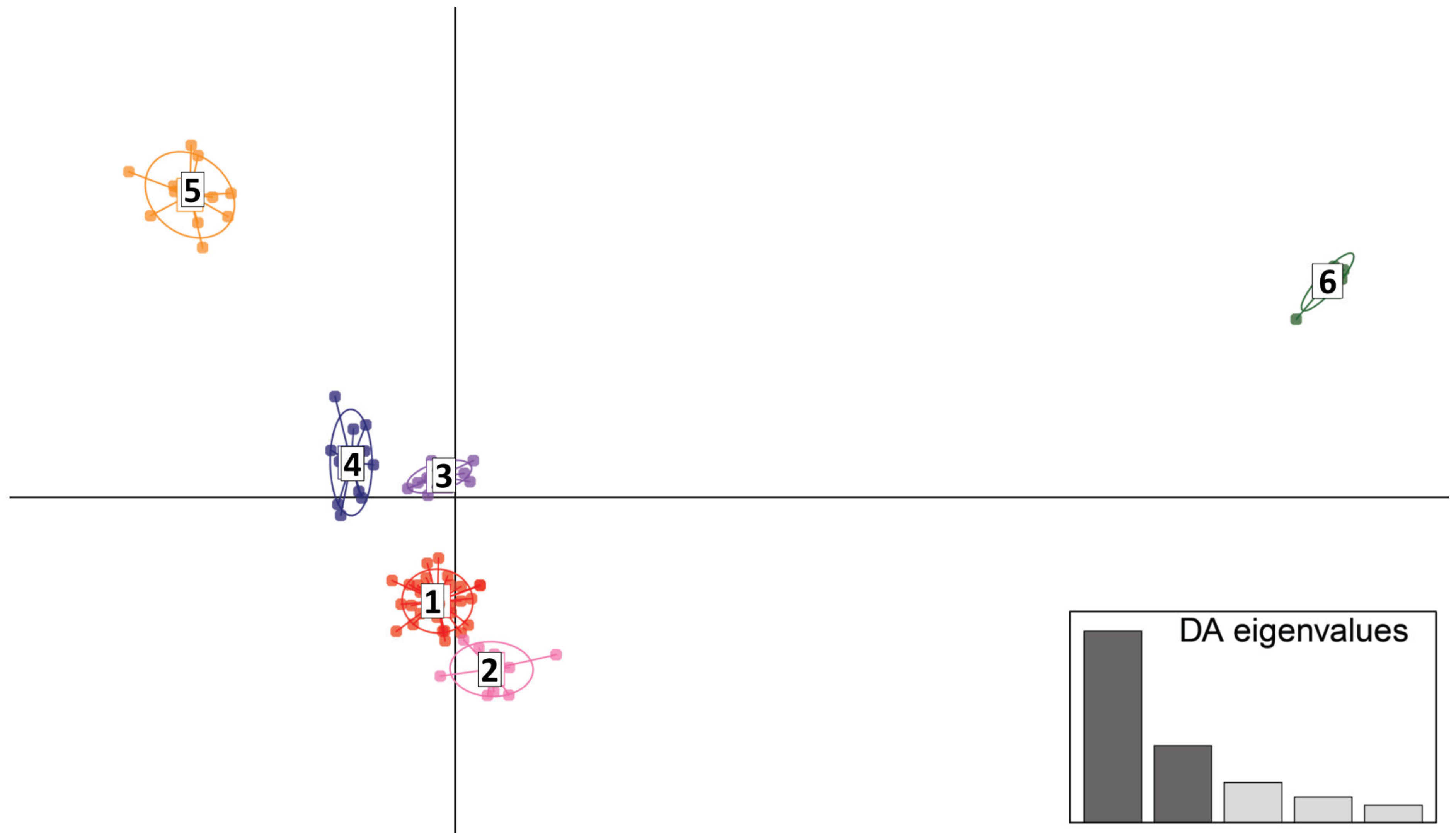



**Figure S3.1**: DAPC results excluding COI marker. The six genetic clusters are also retrieved in the absence of COI marker.

### 3.6 Extended Methods – COI control

Prior to this study, a part of the COI gene (about 700bp) of each individual was sequenced using the Sanger method, to determine the mitochondrial lineage of each *O. longicauda* individual. In addition, all outgroups previously mentioned were also tested, to check DNA quality (by successful amplification) and obtain those sequences in our database. Primers and PCRs conditions are described elsewhere (Stöhr *et al.* 2009; Boissin *et al.* 2011). As a positive control, COI was added in the Miseq sequencing. Surprisingly, 19 individuals clustered wrongly in the COI tree. Two hypotheses may explain this result. First, tubes containing individual tissues may have been wrongly labeled, thus the "true" lineage is the currently observed one; since new DNA extractions were performed for this study. Second, the sequence may be a cross-contamination between samples, during manipulation of PCR products.

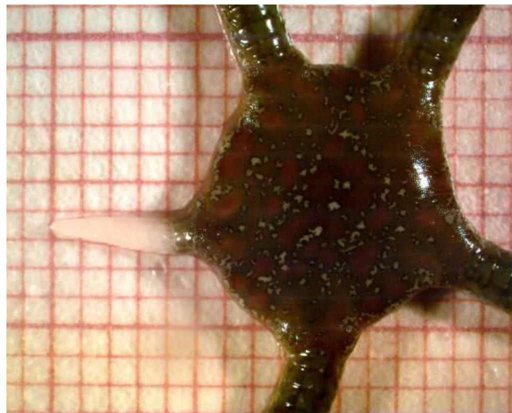
To discriminate between the two hypotheses, the COI fragment (306 bp) used for the Miseq sequencing was re-amplified for the 96 individuals used in this study, and Sanger sequenced. 13 individuals were indeed misclassified, indicating wrong labeling of the tubes containing tissues. Those individuals were then renamed according to their "true" lineage, namely the one found with the Mi-seq sequencing. Five individuals were unsuccessfully sequenced, indicating poor DNA quality, and were thus removed from analyses. Finally, *O. pentagona* provided a wrong sequence (from L1), indicating contamination of the DNA extract, and was thus excluded. Finally, three individuals displayed very low number of amplified markers (<8/31), and were also excluded from analyses. Finally, 87 individuals remained for further analyses (see Table S3.1 for details).

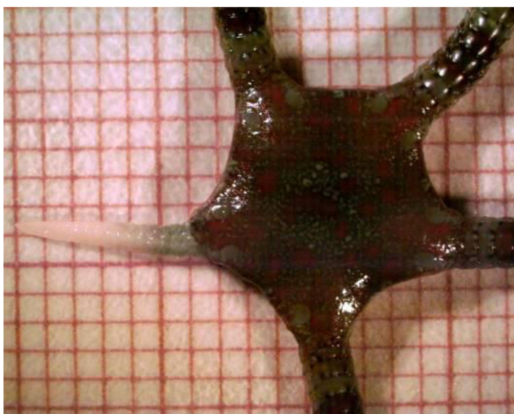
# Chapitre 4

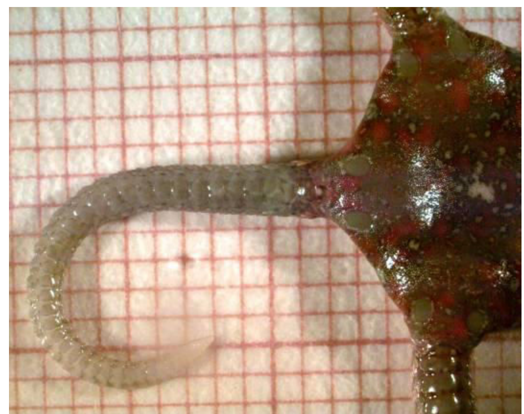
# -

# Thermotolérance et régénération dans le complexe d'espèces d'ophiures *Ophioderma longicauda* :

# Une étude préliminaire comparant lignées et bassins Méditerranéens







# Chapitre 4: Thermotolérance et régénération dans le complexe d'espèces d'ophiures *Ophioderma longicauda* : une étude préliminaire comparant lignées et bassins Méditerranéens

## 4.1 Résumé

Dans les chapitres précédents, j'ai montré que les lignées incubantes et dispersantes du complexe *Ophioderma longicauda* correspondaient effectivement à des espèces biologiques différentes, et ce en comparant les données morphologiques, les données de reproduction (période de l'année) et les données génétiques. Dans ce chapitre, j'aborde la composante écologique dans la comparaison des deux espèces sœurs L1 et L3, à savoir leur thermotolérance respective. En effet, bien que L1 et L3 soient trouvées en sympatrie et syntopie en Méditerranée orientale, il semble y avoir une différence de bathymétrie préférentielle entre ces deux espèces, tout du moins en Méditerranée orientale. Ainsi, l'espèce L3 est plus abondante que L1 à faible profondeur (2m à 5m), tandis que leurs proportions s'équilibrent pour des profondeurs plus élevées (à partir de 10m), cette différence étant significative (test exact de Fisher). Par ailleurs, la lignée L1 a une distribution plus large (des îles Canaries au Liban) que la lignée L3, cette dernière étant limitée à la Méditerranée orientale.

Ces données pourraient s'expliquer, de façon ni exhaustive ni exclusive, par (i) une meilleure tolérance des individus L3 aux températures élevées par rapport aux individus L1 ; (ii) une compétition tournant à l'avantage des individus L3 à faible profondeur dans le choix de l'habitat et l'occupation des roches sous lesquelles elles s'abritent, mais aussi par (iii) un évitement des faibles profondeurs par les L1, qui serait avantageux à la période de ponte, les larves (produites par les L1 et non par les L3) étant potentiellement plus sensibles aux rayons solaires (Häder *et al.* 2007; Dahms & Lee 2010).

Afin de tester l'hypothèse de thermotolérance, des spécimens de la lignée L1 provenant de Marseille et de Crète (L1M et L1C), ainsi que des spécimens de la lignée L3 provenant de Crète (L3C) ont été récoltés. Deux réplicas (contenant 5 à 13 individus par échantillon et par température) ont été réalisés. Après deux semaines d'acclimatation, la survie, l'autotomie (perte, non due à une action mécanique extérieure, d'un segment de bras), utilisée comme proxi de l'état de stress des individus, et la régénération des bras à

17°C, 26°C et 30°C ont été mesurées toutes les trois semaines, sur une période de 14 semaines (le dernier intervalle de temps étant de cinq semaines), 17°C représentant une température contrôle, 26°C un stress léger (valeur haute de température mesurée in situ) et 30°C un stress fort (valeur n'étant jamais observée in situ). Quatre bacs (L3C 17°C, L1M-L1C-L3C 30°C), ayant subi des contaminations bactériennes, ont été exclus des analyses car la mortalité n'était pas directement liée à la température. Pour ces conditions expérimentales, il n'y a donc eu qu'un seul réplica analysé.

Au total, 49 (26 L1M, 10 L1C, 13 L3C), 62 (26 L1M, 10 L1C, 26 L3C) et 31 (13 L1M, 5 L1C, 13 L3C), individus ont été suivis, pour les températures de 17°C, 26°C et 30°C, respectivement. Globalement, les deux espèces L1 et L3 du complexe *O. longicauda* ont montré une bonne résistance aux températures élevées, supérieure à celles reportées pour d'autres ophiures, avec des individus de L1M et L3C encore vivants et intacts après 14 semaines à 30°C (Fig. 4.1). L'échantillon L3 a montré une meilleure thermotolérance que les échantillons L1, suggérant une possible meilleure adaptation de l'espèce L3 aux températures élevées, ce qui est en accord avec les hypothèses de départ. Cependant, des recherches complémentaires sont nécessaires afin de déterminer si ces différences sont effectivement dues à la lignée, à l'origine géographique ou encore aléatoires, en raison du nombre insuffisant de réplicas disponibles.

Cette étude fournit, pour la première fois, des trajectoires de régénération individuelles, et montre que la régénération suit une courbe de Gomperz (phase de latence – phase exponentielle – phase plateau) et est grandement influencée par la température dans les deux lignées (la portion de bras régénéré étant environ 10 fois plus longue à 26°C qu'à 17°C à la fin des 14 semaines) (Fig. 4.2). En revanche, l'origine des échantillons n'entraine pas de différence significative, pour une même température, excepté entre L1M et L3C à 26°C (Table 4.5). Une plus grande diversité de réponse de régénération entre individus été observée chez les L1 comparé aux individus L3, à mettre éventuellement en relation avec la plus grande diversité génétique observée chez L1 (voir chapitre 3). Cependant, des études complémentaires sont nécessaires afin de déterminer si cette diversité de réponse est effectivement due à la diversité génétique des L1, ou simplement à la différence de taille entre individus. Cette étude, bien que préliminaire, met en avant une différence écologique potentielle entre L1 et L3, dont on a montré que ce sont des espèces séparées.

En perspective, afin de répondre plus précisément à la question de la thermotolérance, il faudrait utiliser un plus grand nombre d'individus (15-20 par condition), avec au moins trois réplicas pour chaque température. Cela n'a pu se faire au cours de ma thèse pour des raisons logistiques (nombre d'individus échantillonnés et capacité de la salle aquarium). Il serait également intéressant d'avoir des individus d'une même espèce provenant de plusieurs localités, afin de séparer les effets potentiels de l'espèce et de la localité d'origine. De plus, une période d'acclimatation de minimum un mois, voire plusieurs, serait souhaitable dans le cas où une nouvelle expérience serait mise en place. En effet, il semble qu'à posteriori, la période d'acclimatation de deux semaines ait été trop brève (d'après les commentaires de relecteurs). Finalement, la survie et l'autotomie ne semblent pas être des mesures très précises pour estimer la thermotolérance chez *Ophioderma,* à cause de facteurs confondants comme la contamination bactérienne. Une approche combinée avec la mesure de l'expression de gènes de stress aurait peut être pu mieux renseigner sur les capacités de thermotolérance de ces deux espèces. En ce qui concerne la question plus générale de la différenciation écologique éventuelle entre les lignées sympatriques L1 et L3, d'autres pistes que la thermotolérance paraissent intéressantes à explorer, comme celle consistant à comparer la tolérance aux UV (sur les stades larvaire et juvénile), ou la compétition interspécifique en fonction de la profondeur.

Ce chapitre a été modifié sur la base de l'article de Weber A.A-T., Dupont S. and Chenuil A. (2013) Thermotolerance and regeneration in the brittle star species complex *Ophioderma longicauda*: a preliminary study comparing lineages and Mediterranean basins. Comptes Rendus Biologies. 336 (11): 572-581 publié en novembre 2013.

## 4.2 Introduction

Temperature is the main factor driving marine species distributions (Somero 2002, 2005; Pörtner 2002). Over the last century, the average global temperature has increased by 0.6°C and is projected to increase by a further 1-6°C by 2100 (Solomon 2007). In the Mediterranean Sea, an increase in sea-surface temperature (SST) has already been documented (Vargas-Yáñez *et al.* 2008, 2010; Nykjaer 2009; Coma *et al.* 2009) but no clear warming projection is currently available due to the lack of data in the eastern basin (Durrieu de Madron *et al.* 2011). In the Northern hemisphere, species distributions are expected to shift northward (Barry *et al.* 1995; Pearson & Dawson 2003) or when northward migration is not possible, e.g. for some marine species restricted to the Mediterranean north coast, to acclimate/adapt to new thermal conditions, or, to become extinct (Thomas *et al.* 2004).Thermal tolerance differs between species and even between populations of the same species due to local adaptation (Kuo & Sanford 2009; Kelley *et al.* 2011). Thus, when comparing species and populations it is important to test for phenotypic plasticity and/or local adaptation in order to infer a species' potential response to future climate change (Sorte *et al.* 2011).

The brittle star *Ophioderma longicauda* (Bruzelius, 1805) is a common Atlanto-Mediterranean species (Tortonese 1983) reproducing via the production of a vitellaria larva (Fenaux 1969). However, it was recently shown to be a species complex (Stöhr *et al.* 2009; Boissin *et al.* 2011) comprising at least six divergent mitochondrial lineages with contrasting life-history strategies. Three lineages produce dispersing larvae (lineages L1-L5-L6) while the three others brood their offspring (lineages L2-L3-L4) (Stöhr *et al.* 2009; Boissin *et al.* 2011). A recent study (Weber et al. 2014) confirmed that L1 and L3 are separate species since they occur in sympatry and in syntopy but do not display the same reproductive mode (larvae vs. brooder), do not reproduce at the same period of the year (April vs. July, prezygotic isolation), and are genetically differentiated at a nuclear locus, independent of the mitochondrial genome used to define lineages (Fenaux 1972). L1 is the dominant species in western Mediterranean and Atlantic

basins, but seems less abundant in eastern Mediterranean where the dominant lineages are brooders L2-L3-L4 (Boissin *et al.* 2011). For example, L1 and L3 are the only lineages present in Crete, L3 being more abundant than L1. A differential bathymetric distribution was also found, L1 being predominantly found from five meters depth or more, and L3 being dominant at the shallowest depths (two-three meters). Brooding lineages L2-L3-L4 are absent from the western Mediterranean basin (Boissin *et al.* 2011).

Abiotic conditions in the eastern Mediterranean basin differ from those of the western basin, and include oligotrophic waters (D'Ortenzio & Ribera d'Alcalà 2009), higher salinity (Touratier & Goyet 2011) and higher mean water temperatures (Durrieu de Madron *et al.* 2011). At the study site in Marseilles, monthly average temperatures at five meters depth range between 12°C and 24°C during the year, with maximum temperatures ranging around 26°C in summer, when excluding an exceptional thermal anomaly at 28°C (Bensoussan *et al.* 2010). At the Crete study site, temperatures range between 15°C and 27°C during the year, with a maximum of around 28°C in summer (Durrieu de Madron *et al.* 2011). Short term temperature variations can be very high and differences of 10°C can occur in 24 hours in Marseilles (Bensoussan *et al.* 2010). In contrast, summer seawater temperatures in Crete are much more stable (Thanos Dailianis, unpubl data).

To study the thermotolerance of *Ophioderma longicauda*, we must consider the occurrence of several lineages. Therefore, we used the different lineages L1 and L3, without geographical replicates, for this preliminary study. Based on its geographical and bathymetric distributions, the lineage L3 may be more adapted to elevated temperatures than L1. Thermal tolerances of L1 and L3 were measured using three parameters: survival, spontaneous autotomy as a proxy for thermal stress (Christensen *et al.* 2011) and arm regeneration rate. We assumed that a temperature environment that is favorable to regeneration is also favorable to somatic growth of the not-injured brittle stars. Thus, experimental investigation of regeneration would help us making inference on the thermal tolerances of the lineages. Three different temperatures were tested, covering today (17-26°C) and future (30°C) natural variability. Individuals from the species L1 and L3 were collected from two locations: Marseilles (France; L1) and Crete (Greece; L1 and L3).

## 4.3 Material and methods

### 4.3.1 *Specimen collection and genetic analyses*

*O. longicauda* specimens were collected by scuba diving during May 2012, in Crete (35° 16' N; 25° 49' E), Greece (n=112) and in June 2012 from Marseilles (43° 16' N; 5° 17' E), France (n=80). Specimens from Crete were shipped alive to Marseilles at ambient temperature in 2L non-insulated plastic containers (10 individuals per container) filled with seawater saturated with $O_2$. The transportation lasted 10 hours in total. All individuals survived and did not display any autotomy or other signs of stress. To reduce possible stress due to transportation, all individuals were kept prior to experiment in an open seawater circuit aquarium for two weeks for acclimation (temperature ranging between 17°C and 19°C) and were fed once a week *ad libitum* with commercial fish pellet food.

All individuals used were genotyped to determine their lineage. A small piece of arm was cut from each individual and DNA extracted following a Chelex protocol (Walsh *et al.* 1991). A 689 pb fragment of Cytochrome Oxidase I (COI) was PCR amplified with the primers COIa (5' AGT ATA AGC GTC TGG GTA GTC 3') and COIf (5' CTT GCA GGA GGA GGA GAY CC 3') using the following parameters. An initial denaturation step of 5 min at 95°C, followed by 40 cycles of 30 s of denaturation at 95°C, 30 s of annealing at 52°C and 45 s of extension at 72°C, with a final extension step of 5 min at 72°C. A Restriction Fragment Length Polymorphism (RFLP) protocol was then developed to rapidly recognize the lineage of each individual. BccI restriction enzyme (cutting site: CCATC(N)4 / ) (New England Biolabs) generates different fragments lengths specific of each lineage after digestion. Restriction simulations were done with the online program NEBcutter V2.0 and verified with PCR products of sequenced L1 and L3 individuals.

For each individual, 10 µl of PCR product were added to 1U of Bcc I, 1X of buffer NEB 1, 1X of BSA and digestion was performed for 90 min at 37°C. Restriction fragments were migrated on a 2% agarose gel stained with ethidium bromide. The RFLP analysis indicated that individuals collected from Crete belonged to lineages L3 (L3C; n=80) and L1 (L1C; n=32). All individuals collected in Marseilles belonged to lineage L1 (L1M; n=78). Beside the missing piece of arm used for lineage determination, individuals did not show any sign of a recent wound.

### 4.3.2 *Experimental design*

Three different temperatures were chosen to test the thermal tolerance of *O. longicauda*, based on the natural temperature variation in Crete and Marseilles during spring and summer. 17°C corresponds to the average temperature during the studied period, 26°C is close to the maximal temperature observed at present (slight stress) and 30°C would correspond to extreme conditions projected under a global warming scenario (strong stress). This temperature was chosen to ensure that the stress would be sufficient to cause mortalities. Two replicate tanks in semi-open circuit were used per temperature (17°C, 26°C, 30°C) and per group (L1M, L1C, L3C) each containing five (L1C) or 13 (L1M and L3C) specimens, for a total of 18 tanks. After two weeks of acclimation, temperature was raised ~1°C per day until the desired temperature was reached. Brittle stars were fed once a week *ad libitum* with commercial fish pellet food. At the beginning of the experiment, four tanks (one of L3C at 17°C; L3C, L1C and L1M at 30°C) displayed bacterial contaminations that rapidly killed all individuals; those tanks were excluded from further analyses.

### 4.3.3 *Animal preparation and measured parameters*

Dupont and Thorndyke (Dupont & Thorndyke 2006) showed that length lost (LL) is the key intrinsic parameter regulating regeneration rate (RR). As the specimens of *O. longicauda* displayed heterogeneous sizes, arms were amputated at a fixed LL of 5 cm to standardize the RR. Experimentally induced amputations were performed on the arm already amputated for DNA extraction. The position of amputation was calculated for a LL of 5 cm from the linear relationship between disc diameters (DD) and longest arm lengths (AL), where AL = 0.48 DD + 0.52 ($R^2$ = 0.88). There were no significant differences between regressions constructed with L1 and L3 data (n L1 = 40; n L3 = 42), thus L1 and L3 data were pooled to construct the linear regression (n = 82). For a few small individuals with arm lengths < 5 cm, the whole arm was cut from the base of the disc. Their regeneration trajectories are displayed in Figure 4.2 but they were excluded from further statistical analyses testing the impact of temperature, lineage and populations.

All amputations were performed after anesthesia by immersion in 3.5% w/v $MgCl_2.6H_2O$ in seawater. Experimental arm amputation was achieved by applying a

scalpel blade across a natural inter-vertebral autotomy plane. The disc of each individual was photographed to measure disc diameter and to allow individual identification. Aquaria were monitored daily during 14 weeks and dead individuals were immediately removed. The length of the regenerate was measured after anesthesia by immersion in 3.5% w/v $MgCl_2.6H_2O$ for each individual at weeks 3, 6, 9 and 14. At the end of the 14th week period, remaining individuals were scored as intact or with autotomized arms.

### 4.3.4 *Statistical analyses*

At the end of the experiment, individuals were classified into three categories: intact, showing sign of autotomy or dead. The effects of temperatures (17°C, 26°C, 30°C) and brittle star group of origin (L1C, L1M, L3C) were tested by applying Fisher's exact test on the contingency tables displaying the number of individuals in each category for different temperatures or different groups. Data from replicate tanks were pooled to increase the robustness of the analyses and average the different responses per tank. The kinetics of regeneration for each individual was modeled as a Gomperz growth between regenerate length and time. Specimens for whom regenerate length was not available for the 4 time points (e.g. as a consequence of autotomy) were not considered for this analysis. Each individual trajectory was characterized by three parameters calculated as constants of the Gompertz equation of an asymmetrical growth curve estimated by the least-square method from the original data (Hunt 1982). The calculated parameters were ML, the maximal length of the regenerate (mm); MR, maximum rate of regeneration (the maximum growth rate of the curve at inflection point, ip); *t*ip, time to the inflexion point in weeks (Eqn 1):

$$\mathrm{ML} = \mathrm{ML},\infty \times \mathrm{e}^{-\mathrm{e}^{-\mathrm{MR}\times(t-t\mathrm{ip})}} \quad (1),$$

The effects of brittle star groups and temperature on these parameters (ML, MR, *t*ip) were investigated using the Wilcoxon test as the data were non-normally distributed (unpaired data, unilateral test L1>L3 and L1C>L1M). All statistical analyses were performed with the R software (www.r-project.org).

## 4.4 Results

### 4.4.1 *Survival and autotomy*

Autotomies and deaths occurred throughout the experiment but remained low in numerous cases, therefore we did not build LT50 curves and we displayed the final results only. The proportion of individuals in each category (intact, autotomy, dead) at the end of the 14$^{th}$ week period is presented in Figure 4.1. For L1M, there was no significant difference between 17°C and 26°C. However, a higher mortality was observed at 30°C compared to 17°C and 26°C (statistics in Table 4.1). For L1C, there was no significant difference between 17°C and 26°C, but a 100% mortality was observed at 30°C, all individuals being dead after nine weeks of experiment (Table 4.1). Temperature had no significant effect for L3, each temperature condition displaying around 10 - 20% of mortality, 40% of autotomy and 40 – 50% of intact individuals (Table 4.1).

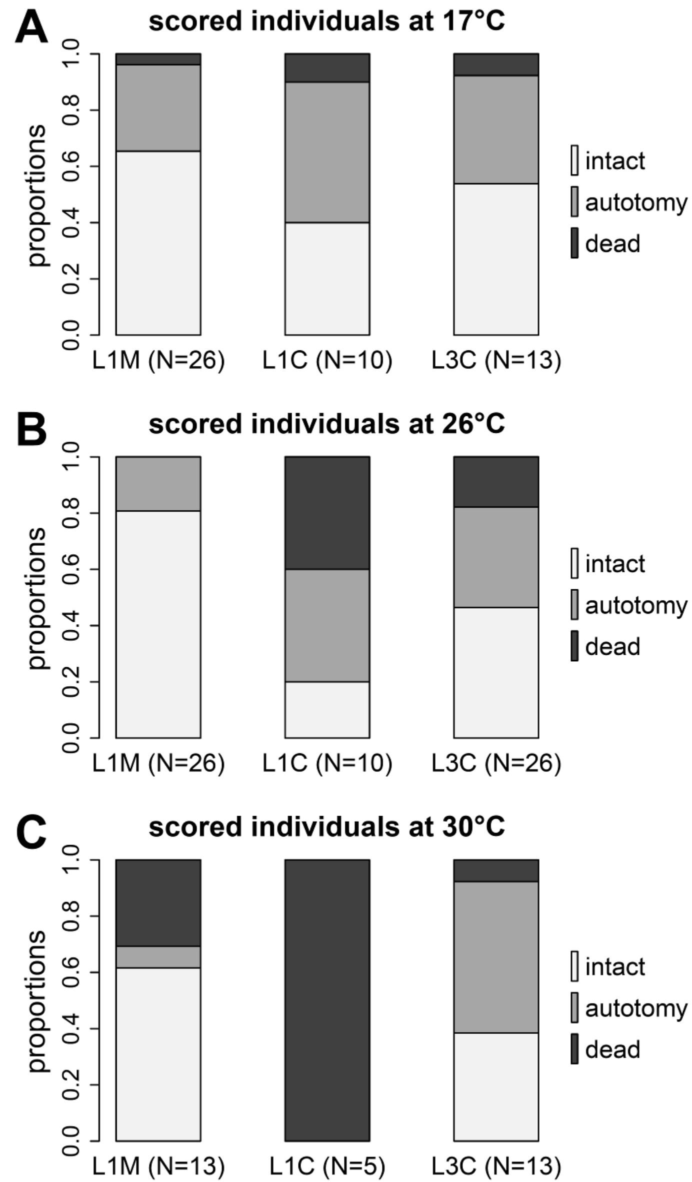



**Figure 4.1:** Proportions of intact, showing evidence of autotomy and dead individuals in each group (L1M, L1C, L3C) and temperature (17°C, 26°C, 30°C) after 14 weeks of experiment. L1M: individuals of lineage L1 collected in Marseilles. L1C: individuals of lineage L1 collected in Crete. L3C: individuals of lineage L3 collected in Crete.

At 17°C (Fig. 4.1A) there were no significant differences between groups (statistics in Table 4.2), each group displaying high survival and few autotomy. At 26°C (Fig. 4.1B) there was no difference between L1C and L3C, but L1M was significantly different from L1C and L3C (Table 4.2). L1M did not display any mortality and only few specimens presented signs of autotomy in comparison to the higher proportion of autotomy observed in L1C and L3C. At 30°C (Fig. 4.1C) all comparisons between groups were significant (Table 4.2). L3C had low mortality and high autotomy, whereas L1M displayed higher mortality but lower autotomy. For both L1M and L3C, at least 40% of the individuals remained alive and intact after 14 weeks at 30°C. Yet, mortality was lower in L3C (10%) compared to L1M (30%) and L1C (100%).

**Table 4.1**: Comparison of survival, autotomy and death between three different temperatures (17°C, 26°C, 30°C) for each group. Proportions of each category were compared by applying a Fisher's exact test. NS: p-value > 0.05; *: 0.01 < p-value < 0.05; **: 0.001 < p-value < 0.01; ***: p-value < 0.001. P-values are in brackets. L1M: individuals of lineage L1 collected in Marseilles. L1C: individuals of lineage L1 collected in Crete. L3C: individuals of lineage L3 collected in Crete.

| | **L1M** | **L1C** | **L3C** |
|---|---|---|---|
| 17°C vs. 26°C | NS (0.35) | NS (0.39) | NS (0.81) |
| 17°C vs. 30°C | * (0.048) | ** (0.006) | NS (0.84) |
| 26°C vs. 30°C | * (0.015) | NS (0.12) | NS (0.54) |

**Table 4.2**: Comparison of survival, autotomy and death between three different groups (L1M, L1C, L3C) for each temperature. Proportions of each category were compared by applying a Fisher's exact test. NS: p-value > 0.05; *: 0.01 < p-value < 0.05; **: 0.001 < p-value < 0.01; ***: p-value < 0.001. P-values are in brackets. L1M: individuals of lineage L1 collected in Marseilles. L1C: individuals of lineage L1 collected in Crete. L3C: individuals of lineage L3 collected in Crete.

| | **17°C** | **26°C** | **30°C** |
|---|---|---|---|
| between sites (L1M vs. L1C) | NS (0.24) | *** ($2.0 \times 10^{-4}$) | * (0.044) |
| between lineages (L1C vs. L3C) | NS (0.83) | NS (0.232) | *** ($8.17 \times 10^{-4}$) |
| between sites & lineages (L1M vs. L3C) | NS (0.64) | * (0.013) | * (0.049) |

### 4.4.2 *Regeneration*

Regeneration rate was estimated from the relationship between regenerate length and time following a Gomperz growth model (Fig. 4.2, Table 4.3). For L1C, all specimens died at 30°C by the end of the experiment (14 weeks) and it was then not possible to measure any regeneration rate. Temperature had a significant impact on regeneration rate parameters (Table 4.4). For the three tested groups, the maximal rate of regeneration (MR) and maximal length of regenerate (ML) significantly increased when temperature increased from 17°C to 26°C and the time to inflexion point (*t*ip) decreased. The only exception was the absence of significant effect between 17°C and 26°C for ML in L3C. No significant difference between any of the regeneration parameters was observed between 26°C and 30°C.

**Table 4.3**: Regeneration rate parameters estimated from a Gomperz model (MR: maximal rate of regeneration; ML: maximal length of regenerate; *t*ip: time to inflexion point). Number of individuals (*n*), mean values and standard deviation (in italic). L1M: individuals of lineage L1 collected in Marseilles. L1C: individuals of lineage L1 collected in Crete. L3C: individuals of lineage L3 collected in Crete.

| | **17°C** | | | **26°C** | | | **30°C** | |
|---|---|---|---|---|---|---|---|---|
| | **L1M** | **L1C** | **L3C** | **L1M** | **L1C** | **L3C** | **L1M** | **L3C** |
| *n* | 14 | 6 | 9 | 22 | 3 | 16 | 6 | 6 |
| MR | 0.21<br>*0.14* | 0.15<br>*0.03* | 0.15<br>*0.07* | 0.27<br>*0.06* | 0.31<br>*0.09* | 0.29<br>*0.06* | 0.30<br>*0.09* | 0.36<br>*0.06* |
| *t*ip | 13.28<br>*4.68* | 12.97<br>*2.63* | 16.89<br>*7.06* | 6.74<br>*0.99* | 6.0<br>*0.8* | 6.94<br>*1.38* | 6.04<br>*1.03* | 5.51<br>*1.01* |
| ML | 11.3<br>*6.55* | 12.06<br>*3.79* | 20.13<br>*14.93* | 23.71<br>*7.83* | 24.54<br>*6.20* | 19.77<br>*4.66* | 23.32<br>*8.54* | 20.25<br>*4.91* |

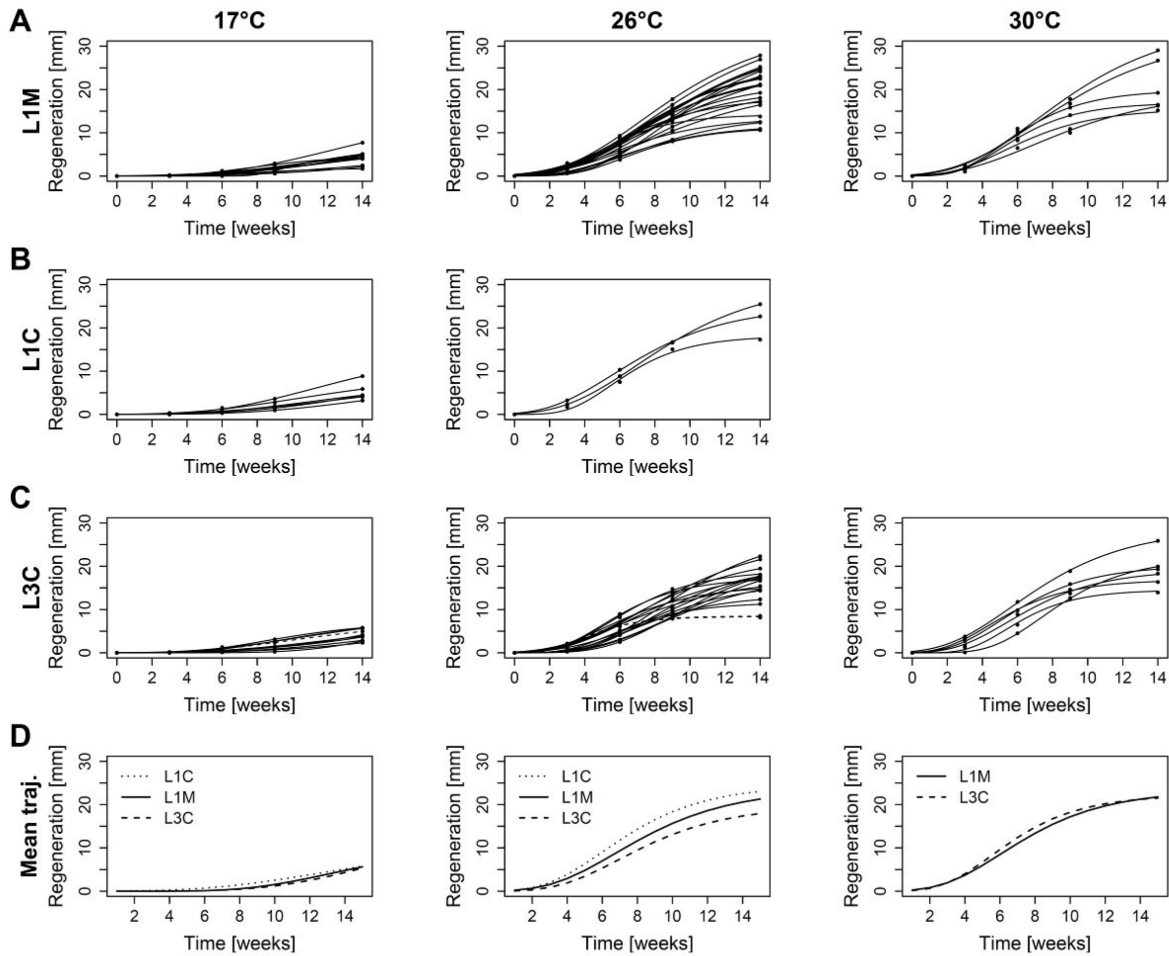



**Figure 4.2A-C** Individual regeneration trajectories in each group (L1M, L1C and L3C) and temperature (17°C, 26°C, 30°C). Dashed lines trajectories correspond to individuals with DD < 10 mm, excluded from analyses as LL < 5 cm. **4.2D**: Mean Gomperz curves, calculated from the mean of estimated parameters ML, MR and *t*ip for each group and temperature. L1M: individuals of lineage L1 collected in Marseilles. L1C: individuals of lineage L1 collected in Crete. L3C: individuals of lineage L3 collected in Crete. DD: disc diameter. LL: length lost. ML: Maximal length of regenerate. MR: Maximal rate of regeneration. *t*ip: Time to the inflexion point.

At the different temperatures tested, there was no significant effect of the group on any of the tested regeneration parameters (Table 4.5) with the exception of ML which was significantly different between L1M and L3C at 26°C. Gompertz curves were only available for individuals that remained intact until week 14. Thus we also compared regenerate lengths at the different sampling points to take advantage of higher sample sizes (Fig. 4.3). These results indicated that L1C and L1M regenerated significantly faster than L3C at 26°C ($p < 0.05$, Wilcoxon test). e. g. at week 14, means of regenerate lengths were 21.8 mm, 19.7 mm and 16.8 mm for L1C, L1M and L3C, respectively. This tendency was also confirmed by the mean regeneration Gompertz curves for each group (Fig. 4.2D). In addition, we observed that standard deviation of ML at 26°C and 30°C was higher for L1M compared with L3C. Inferences about L1C at 30°C could not be done after week 6 because all L1C individuals were dead at week 9. The influence of the arm length lost is illustrated by the few individuals for which LL was lower than 5 cm (Fig. 4.2C, dashed lines). As expected, the plateau was reached sooner and corresponded to a lower length for these individuals.

**Table 4.4**: Temperature effects on the regeneration rate parameters estimated from a Gomperz model (MR: maximal rate of regeneration; ML: maximal length of regenerate; *t*ip: time to inflexion point). Each comparison was performed two by two with a Wilcoxon test (unpaired data, unilateral test, for parameters MR and ML: 17°C<26°C<30°C; for parameter *t*ip: 17°C>26°C>30°C) since the data were non-normally distributed. NS: p-value > 0.05; *: 0.01 < p-value < 0.05; **: 0.001 < p-value < 0.01; ***: p-value < 0.001. P-values are in brackets.

| | **L1M 17-26** | **L1C 17-26** | **L3C 17-26** | **L1M 17-30** | **L3C 17-30** | **L1M 26-30** | **L3C 26-30** |
|---|---|---|---|---|---|---|---|
| MR | ** (0.0028) | * (0.012) | *** ($6.78 \times 10^{-05}$) | * (0.0203) | *** ($1.99 \times 10^{-4}$) | NS (0.31) | NS (0.18) |
| *t*ip | *** ($1.34 \times 10^{-07}$) | * (0.012) | *** ($5.87 \times 10^{-06}$) | *** ($2.58 \times 10^{-05}$) | *** ($1.99 \times 10^{-4}$) | NS (0.16) | NS (0.16) |
| ML | *** ($2.2 \times 10^{-05}$) | * (0.012) | NS (0.26) | ** (0.00232) | NS (0.30) | NS (0.89) | NS (0.67) |

**Table 4.5**: Group effects on the regeneration rate parameters estimated from a Gomperz model ers (MR: maximal rate of regeneration; ML: maximal length of regenerate; *t*ip: time to inflexion point). Each comparison was performed two by two with a Wilcoxon test (unpaired data, unilateral test, for parameters MR and ML: 17°C<26°C<30°C; for parameter *t*ip: 17°C>26°C>30°C) since the data were non-normally distributed. NS: p-value > 0.05; *: 0.01 < p-value < 0.05; **: 0.001 < p-value < 0.01; ***: p-value < 0.001. P-values are in brackets.

| | **L1M-L3C 17°C** | **L1C-L3C 17°C** | **L1M-L1C 17°C** | **L1M-L3C 26°C** | **L1C-L3C 26°C** | **L1M-L1C 26°C** | **L1M-L3C 30°C** |
|---|---|---|---|---|---|---|---|
| MR | NS (0.153) | NS (0.304) | NS (0.516) | NS (0.830) | NS (0.396) | NS (0.303) | NS (0.845) |
| *t*ip | NS (0.138) | NS (0.264) | NS (0.548) | NS (0.358) | NS (0.180) | NS (0.119) | NS (0.650) |
| ML | NS (0.920) | NS (0.612) | NS (0.273) | * (0.0476) | NS (0.086) | NS (0.516) | NS (0.350) |

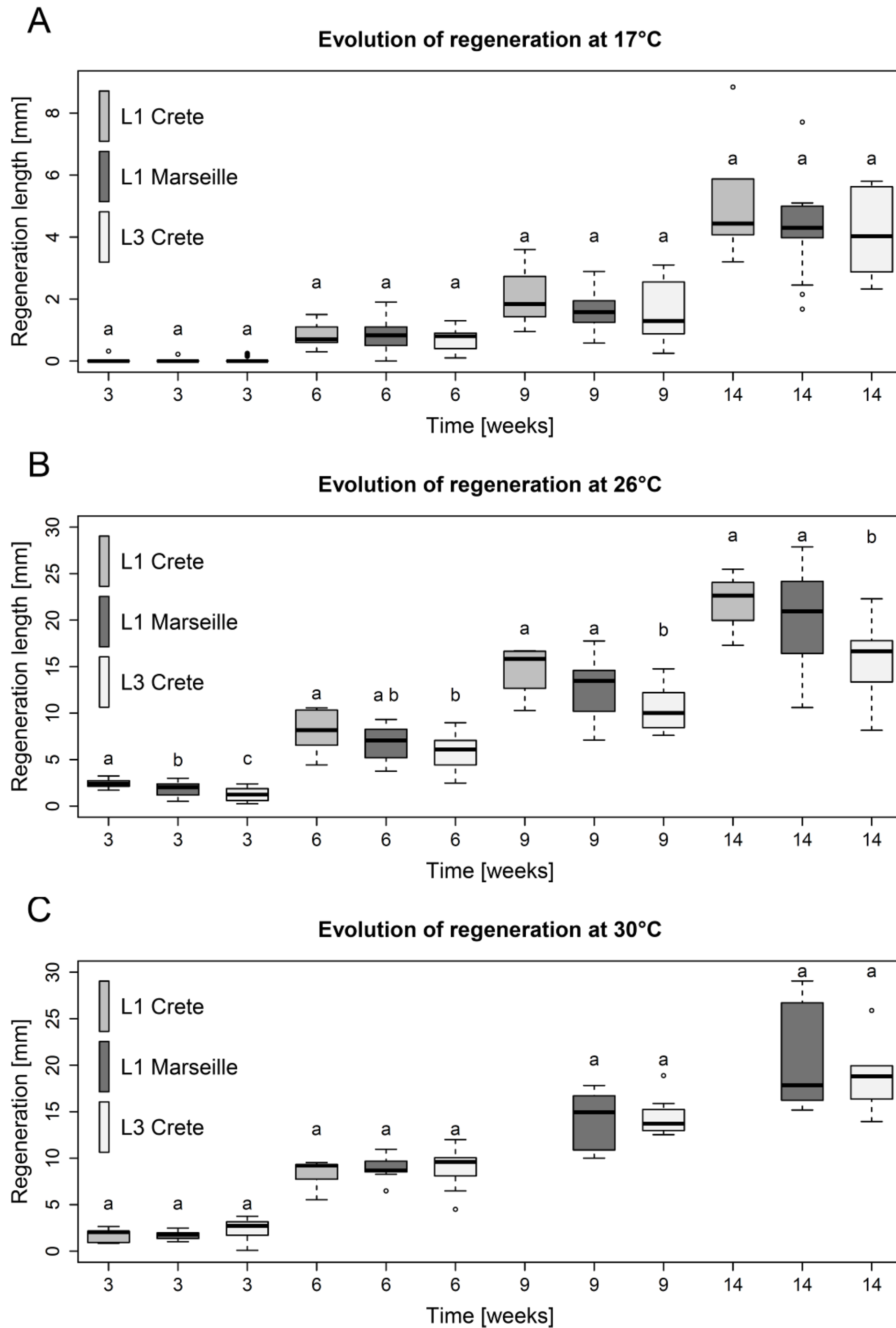



**Figure 4.3** Boxplots of regeneration lengths at each measurement date (3, 6, 9 and 14 weeks) for each group L1M, L1C and L3C at each temperature 17°C, 26°C and 30°C. Each comparison was performed two by two with a Wilcoxon test (unpaired data, unilateral test L1<L3; L1C<L1M) since the data were non-normally distributed. Shared letters mean non-significant comparison at 0.05 level of p-value. L1M: individuals of lineage L1 collected in Marseilles. L1C: individuals of lineage L1 collected in Crete. L3C: individuals of lineage L3collected in Crete.

## 4.5 Discussion

### 4.5.1 *Survival and autotomy*

Our results showed that globally, the species complex *O. longicauda* displayed good resistance to elevated temperatures. Indeed, both *O. longicauda* L1 and L3 lineages could resist for three months a chronic exposure at 26°C with high survival (100%, 60% and 80% for L1M, L1C and L3C, respectively) and more interestingly, there was 70% and 90% survival for L1M and L3C, respectively, after three months at 30°C. Those results contrast with the idea that intertidal marine invertebrates already live at the edge of their thermal limits (Singletary 1971; Somero 2010). Christensen et al. (Christensen *et al.* 2011) studied the brittle star *Ophionereis schayeri* and showed that *O. schayeri* could not resist more than a few days a temperature rise of 3°C compared to the maximum SST experienced by this species. In addition, the Antarctic brittle star *Ophionotus victoriae* could not acclimate to a temperature rise of 2°C compared to 0°C (Peck *et al.* 2009). Thus, we can speculate that adult *O. longicauda*, particularly the L3 lineage, would be little affected by future water temperature rise. However, it is important to consider that larval or juvenile survival may be more impacted by ocean warming (e.g. (Byrne *et al.* 2009, 2013; Nguyen *et al.* 2012)).

When comparing at the intra-complex level, we showed that the L1C sample was more impacted by elevated temperature since a higher mortality was observed between 17°C and 26°C, and at the end of the experiment, all individuals were dead at 30°C. For L1M there was a significant higher mortality at 30°C compared to 26°C, indicating that the specimens from Marseilles were also affected by elevated temperature, but to a lesser extent. In contrast, the L3C sample did not display any difference among 17°C, 26°C and 30°C, indicating that temperature had no effect on mortality and autotomy. Furthermore, the L3C sample displayed lowest mortality at 30°C compared to L1M and L1C. Thus, when estimating thermotolerance by a response difference among temperatures, we showed that our L1 samples were less thermotolerant than our L3C sample. However, to confirm that this is actually an effect of the lineage (L1 versus L3) and not an effect of the local populations considered nor a random effect, further experimentation should be performed including additional populations of L3 (e.g. from North and South coasts of Crete) and additional replicate tanks. Interestingly, we observed that L1 individuals from Marseilles were more resistant to elevated

temperatures (26°C and 30°C) than L1 individuals from Crete despite higher average and maximal seawater temperatures in Crete. One explanation may be local adaptation to the rapid thermal variations typically observed in Marseilles (Bensoussan *et al.* 2010), yet an experiment including short term temperature variations would better test this hypothesis. There might also be a difference in reproductive times between L1C and L1M that could have placed some additional stress on L1C individuals, as the exact reproduction time of L1C is not known yet.

Another possibility is that individuals from Marseilles were in better condition than individuals from Crete, because they were not transported to the experimental laboratory and did not undergo a change in seawater. However, such transportation effects are unlikely since the proportions of intact, autotomized and dead individuals did not differ among the three groups (L1C, L1M and L3C) at the control temperature of 17°C, and in addition, the experiment started after an acclimation period of two weeks and lasted for 14 additional weeks. The difference between L1 individuals from Crete and Marseilles may not reflect a true difference among localities. We cannot exclude random effects affecting whole tanks and biasing the results, since we observed a contamination at the very beginning of the experiment leading to complete mortality in four tanks. When analyzing the L3C response to thermal stress, our data support the hypothesis that L3 may be more resilient to elevated temperatures, as temperature had no significant effect on survival for this population, the samples at 17°C, 26°C and 30°C displaying similar relative proportions of intact, autotomized and dead individuals. This result is in agreement with the hypothesis that temperature may contribute to the geographical distribution of L1 and L3, as the eastern Mediterranean displays higher temperatures.

An interesting aspect to develop in future studies would be to test the importance of local adaptation in thermotolerance, as it is expected to differ between populations of the same species (Kuo & Sanford 2009; Kelley *et al.* 2011). The *Ophioderma longicauda* species complex comprising brooding and broadcasting lineages may then represent an ideal model to conduct research on local adaptation. Indeed, life-history differences among lineages have direct consequences on gene flow and connectivity, as confirmed by population genetic analyses (Boissin *et al.* 2011). Dispersal involves a cost to local adaptation, since maladapted genotypes are likely to be introduced into a non-native

environment, thus species reproducing via larvae might be less prone to local adaptation (Pechenik 1999; Somero 2010). Such species however are able to explore more diverse environments and can produce more offspring, thus they may cope with a rapid change in environmental conditions better than non-dispersing species such as brooders (Sorte *et al.* 2011). Thus, conducting a thermotolerance study with several L3 and L1 populations sampled in eastern Mediterranean, and L1 populations from western Mediterranean (ideally after the reproduction period), would allow us to make inferences about the link between local adaptation and dispersal abilities, by discriminating the effects of species, populations and life-history trait variation.

### 4.5.2 *Regeneration*

To our knowledge, this is the first study comparing regeneration kinetics from different individuals (Fig. 4.2). Regeneration follows a growth curve with first a lag phase, followed by an exponential growth phase and ending with a plateau. Biressi et al. (Biressi *et al.* 2010) have described the cellular processes during regeneration in *Ophioderma longicauda* (most probably belonging to lineage L1 given the sampling location) and *Amphiura filiformis*. They identified four phases: a repair phase, an early regenerative phase, an intermediate regenerative phase and an advanced regenerative phase. During the repair and early regenerative phases, the complete healing of the epithelial layer is accompanied by extensive migration and proliferation of cells leading to the formation of a blastema of undifferentiated cells. This phase is not associated with significant growth and corresponds to our lag phase. Intensive cell proliferation during the intermediate regenerative phase leads to rapid growth and corresponds to our exponential growth phase. The advanced regenerative phase is the time of slow differentiation and leads to a strong reduction in growth rates as indicated by the plateau.

Furthermore, we showed that regeneration increased with temperature, as the time to inflexion point (*t*ip) was divided by two between 17°C and 26°C. This is consistent with previous studies on the effect of temperature on brittle star regeneration (e.g. (Clark *et al.* 2007; Wood *et al.* 2010)). Our data also showed a threshold effect at 30°C, as regeneration did not significantly increase between 26°C and 30°C, this result being congruent with the higher mortality observed at 30°C for the lineage L1 and confirms the cost of survival at elevated temperatures (Clarke 2003).

Overall, our results demonstrate the importance to better define regeneration rate. Often, the regeneration rate is only considered as the linear relationship between regenerate length and time but we clearly show that this rate is not constant during the regeneration process. Our results did not display any difference in regeneration between the three groups (L1M, L1C and L3C), except for the L1M vs. L3C comparison of ML at 26°C, where L1M displayed a significantly higher ML (maximal length of regenerate) than L3C. The fact that only some of the surviving individuals were considered, because the stressed individuals autotomized their regenerate or died, thus reducing the number of analyzed individuals, may have reduced the differences among groups and their statistical significance.

We observed a high inter-individual variation, especially for the plateau parameter ML. These variations were not correlated to the size of the individuals, as DD and ML were compared and no correlation was found (data not shown). In addition, we showed that individual variation was higher for L1 compared to L3, as shown by the standard deviation comparisons. This may be a consequence of the higher genetic diversity of L1 compared to L3 (Boissin *et al.* 2011). Species with high genetic diversity may also display high phenotypic diversity (e. g. [38]). To confirm the correlation between genetic and response diversities, further regeneration studies should be performed on more numerous populations of L3 and L1. Yet, our results highlighted a physiological difference between L1 and L3, supporting the fact that they are truly separated species. The fact that the increase in regeneration rate with temperature is greatly reduced between 26°C and 30°C confirms that the temperature of 30°C induces a significant stress and supports the idea that future experiments with geographic replicates may use the same temperatures tested in this study.

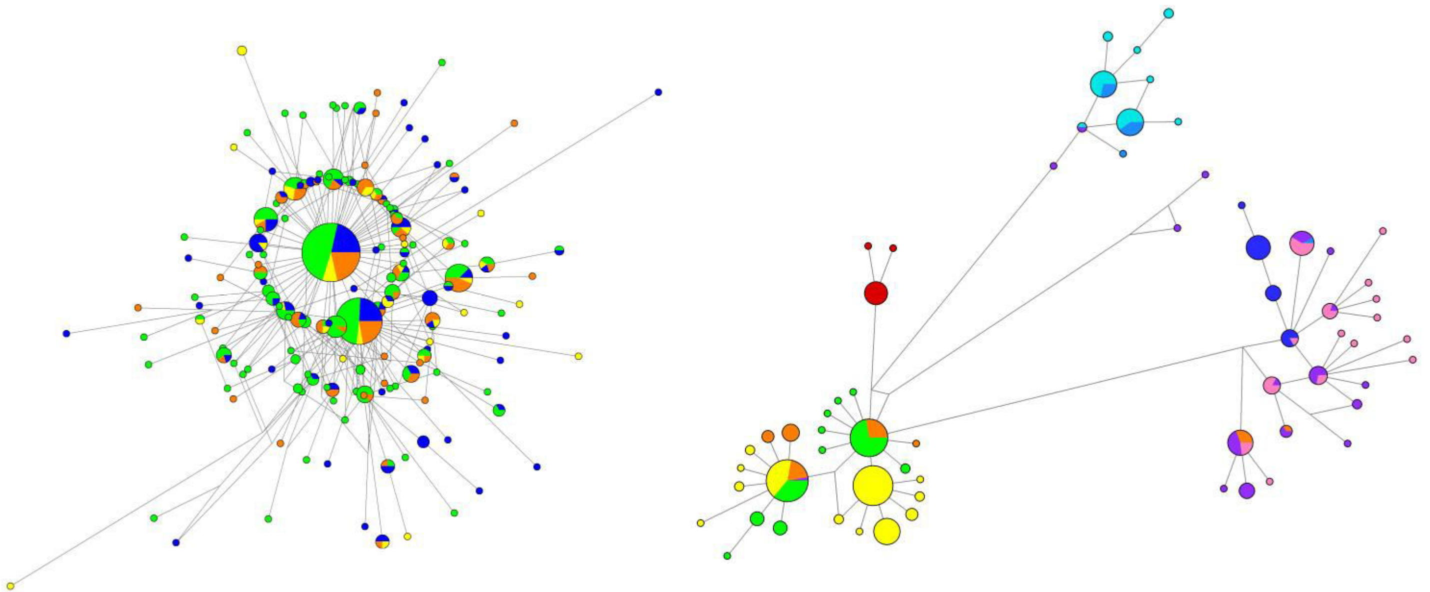
Although preliminary, our results are interesting because we compared the thermotolerance of two still undescribed sibling species and found potential differences. This is in agreement with the hypothesis that L1 may be less thermotolerant than L3 given their distribution across the Mediterranean, L1 being less abundant in eastern Mediterranean where water temperature is more elevated, in particular at shallower depths. If future research confirms our results, the distribution of lineage L3 may not change considering future warming scenarios. In contrast, the distribution of L1 may shift both geographically and in bathymetry. Cryptic species are common in the sea and

often undetected (Knowlton 1993; Bickford *et al.* 2007). Although generally closely related, they may respond differently to environmental changes. Thus, researchers currently studying potential responses to climate change should be aware that different sibling species may occur within a nominal species before predicting global warming impacts on a given species.

# Chapitre 5

# -

# Influence de la phase larvaire sur la structure génétique, la connectivité et la diversité génétique: comparaison entre incubants et dispersants



# Chapitre 5: Influence de la phase larvaire sur la structure génétique, la connectivité et la diversité génétique: comparaison entre incubants et dispersants

## 5.1 Résumé

Dans les chapitres précédents, j'ai montré que les lignées incubantes et dispersantes d'*O. longicauda* correspondaient à des espèces biologiques distinctes, en particulier les clusters C1 (dispersant – lignée mitochondriale L1) et C4 (incubant – lignées mitochondriales L2 et L3 en Grèce). Les recherches présentées dans ce chapitre et dans le suivant se concentreront sur ces deux clusters (=espèces) qui sont trouvés en sympatrie et en abondance suffisante pour mener des études comparatives.

Dans ce chapitre, j'ai réalisé une comparaison de la structure génétique et de la connectivité entre espèces incubantes et dispersantes, sur leurs aires de distribution. En effet, ces deux espèces représentent un modèle idéal pour tester l'influence des traits d'histoire de vie sur la connectivité, et ce sans influence de facteurs confondants. Ainsi, les effets historiques sont minimisés, étant donné que ces espèces ont divergé récemment (environ 500'000 ans) et les effets environnementaux sont également minimisés étant donné que ces espèces vivent en sympatrie. De par sa spécificité (espèces cryptiques sympatriques à modes de reproduction contrastés), le modèle *Ophioderma* est particulièrement intéressant pour l'étude de l'influence des traits d'histoire de vie sur la connectivité, et il n'existe que peu d'études similaires dans la littérature.

Une seconde partie de ce travail a consisté en la comparaison de la diversité génétique (et de sa structuration) entre populations sympatriques des clusters C1 et C4. La structure génétique au sein du complexe *Ophioderma* a été étudiée à deux échelles. Tout d'abord, j'ai étudié la structure génétique des clusters C1, C2, C4 et C5 à grande échelle (des Canaries au Liban), pour un total de 801 individus, en utilisant deux marqueurs génétiques : la COI et l'intron nucléaire i51. La COI a été séquencée grâce à la méthode Sanger, tandis que l'intron i51 a été génotypé sur gel d'agarose haute résolution (ce dernier présentant uniquement un polymorphisme de taille). Par la suite, six populations sympatriques de C1 et C4 échantillonnées tout autour de la Crète (Grèce) ainsi qu'à Rhodes ont été étudiées, afin d'inférer plus finement la structure génétique

ainsi que la connectivité à une échelle géographique plus réduite. Pour ces populations, en plus des deux marqueurs cités précédemment, 10 marqueurs nucléaires nouvellement développés sur la base de transcriptomes ont été ajoutés pour les analyses, ces derniers ayant été séquencés avec la méthode Illumina. Des analyses de diversité génétique (nombre d'haplotypes/allèles ; diversité haplotypique ; diversité nucléotidique) ont également été réalisées sur ces marqueurs.

L'étude de la structure génétique à large échelle a montré une forte structure pour l'espèce incubante C4 ($F_{ST}$: 0.07 à 0.65), qui s'observe également sur les réseaux d'haplotypes, où l'on observe une forte corrélation entre haplogroupes et géographie (Figure 5.2). L'AMOVA chez C4 a montré qu'environ les deux tiers de la variance moléculaire étaient retrouvés au sein des populations (67%), avec l'autre tiers de la composante se trouvant entre populations au sein des groupes (28%) et une faible composante entre groupes (5%). L'étude du marqueur nucléaire i51 n'était pas informative pour la structure étant donné que ce marqueur était monomorphe chez les lignées incubantes. Au contraire, l'espèce dispersante C1 a montré une plus faible structuration génétique, que ce soit pour la COI ($F_{ST}$: 0 à 0.14) ou pour le marqueur nucléaire i51 ($F_{ST}$: 0 à 0.21). Les réseaux d'haplotypes n'ont pas montré de structure phylogéographique particulière, chaque haplotype majoritaire étant trouvé dans les trois bassins méditerranéens ainsi qu'en Atlantique. L'AMOVA a montré que la majorité de la variance moléculaire était retrouvée au sein des populations (99%), avec une très faible composante entre populations au sein des groupes (1%), et aucune composante entre groupes (0%). Cette différence au niveau du pourcentage de variance moléculaire au sein des populations observée entre dispersants et incubants se retrouve également pour d'autres espèces, par exemple pour les holothuries *Cucumaria miniata* (à larves lécithotrophes) et *Cucumaria pseudocurata* (incubant) qui montrent 97.5% et 3.4% de variance moléculaire au sein des populations, respectivement (Arndt et Smith, 1998).

L'analyse multilocus (six marqueurs) à plus petite échelle a également révélé une structure génétique très contrastée entre dispersants et incubants. Ainsi, aucune structure génétique n'était détectée chez les C1 autour de la Crète ($F_{ST}$ moyen : 0.00054) tandis qu'elle était très marquée chez C4 ($F_{ST}$ moyen : 0.49). De plus, l'estimation du nombre de migrants entre populations, réalisée avec le programme migrate-n, a montré qu'en moyenne il y avait environ 50 fois plus de migrants chez les dispersants comparé

aux incubants. Chez les deux espèces étudiées, un déficit en hétérozygotes a été observé. Finalement, les diverses mesures de la diversité génétique employées pour les 12 marqueurs montrent que les incubants (C4) ont une diversité trois à quatre fois plus petite que les dispersants (C1).

Ce chapitre a montré qu'il existait de fortes différences de différentiation et de diversité génétique, à grande et à petite échelle, entre espèces très proches, sympatriques, ne différant, a priori, que pour leur capacité de dispersion. En conclusion, bien qu'elle soit courte (estimée à six jours), la phase larvaire a une influence majeure sur la connectivité et la diversité génétique. Une perspective de ce travail serait de trouver une zone géographique exempte de barrières physiques causant une hétérogénéité de migration, afin de pouvoir observer un pattern d'isolement par la distance (IBD) pour les espèces C1 et C4. Il serait ainsi possible de calculer la distance moyenne parent/enfant pour ces deux espèces, ce qui représenterait une autre et meilleure manière de caractériser l'influence de la larve lécithotrophe sur la connectivité.

Cet article est intitulé "Influence of the larval phase on connectivity: differences in the genetic structure of brooders and broadcasters in the *Ophioderma longicauda* species complex". Ses auteurs sont Alexandra Anh-Thu Weber, Sabine Stöhr, Bastien Mérigot, Laurent Abi-Rached et Anne Chenuil. Il est prévu de le soumettre à Molecular Ecology après relecture par tous les auteurs et modifications subséquentes.

## 5.2 Introduction

Studying connectivity is essential for implementing conservation and management strategies (Palumbi 2003; Shanks *et al.* 2003) but also to answer evolutionary questions such as local adaptation (Sanford & Kelly 2011). Dispersal ability is known to be a key factor in explaining genetic structure (Bohonak 1999). For this reason, closely related species with similar niches but different modes of dispersal are ideal models to study the consequences of dispersal modes on genetic structure. Indeed, when closely related species live in sympatry, comparative connectivity studies can be carried out in the same set of locations, so that environmental conditions and physical connectivity are identical among compared species. Sympatric sister species with overlapping niches are considered to be scarce because of competitive exclusion (Hardin 1960). Yet, this principle does not apply when the resources involved in the overlapping niche component are not limiting. Such situations can occur in the marine environment, where sympatric sister species do not seem to be rare (Knowlton 1993; Boissin *et al.* 2008). Thus, the marine environment may provide suitable models to conduct comparative studies in order to investigate the influence of dispersal on connectivity.

In most marine invertebrates, the adult phase is benthic and has no (or very limited) ability to disperse (Knowlton & Jackson 1993). Planktonic larvae may therefore represent the only significant dispersal phase. Therefore, population genetic theory predicts that species with planktonic larvae display high dispersal, low genetic structure, and high genetic diversity (Emlet 1995). In contrast, species lacking a dispersive phase (laying egg masses or brooding) are expected to display the opposite pattern (Hunt 1993; Duffy 1993). Intuitively, one expects that the longer larvae stay in the plankton, the lower the level of genetic structure, since a correlation between pelagic larval duration (PLD) and dispersal distance has been shown (Siegel *et al.* 2003), in particular for species displaying a short PLD (e.g. less than a day) (Shanks 2009). However, species with a long PLD (e.g. more than a day to a month) did not show this correlation,

highlighting the putative effects of larval behavior (Shanks 2009). Furthermore, the PLD was found to be poorly correlated with genetic structure (Weersing & Toonen 2009). In fact, when species lacking a dispersive stage were removed from analyses, the correlation between $F_{ST}$ and PLD became non-significant (Weersing & Toonen 2009; Riginos *et al.* 2011). Although the correlation between $F_{ST}$ and PLD does exist for some species, the authors caution about generalizing it among marine taxa. In addition, they emphasize the difficulty of properly estimating PLD, based on the fact that PLD can depend on the environment (e.g. delayed metamorphosis; Pechenik 1999), on the sampling design and on the influence of genetic marker choice. Finally, other factors such as past biogeographic events, differences in habitat, larval retention, ecological associations can have stronger effects than PLD on genetic structure (Riginos *et al.* 2011). Nevertheless, when using a larger and more recent set of studies, including a robust sampling design where each study used the same metrics, a significant correlation between genetic distance (measured by the isolation by distance (IBD) slope) and PLD was found ($R^2$=0.34) (Selkoe & Toonen 2011).

The best way to test the influence of the larval phase on genetic structure is to compare population genetic structures among sympatric and closely related species which have contrasting PLD (McMillan *et al.* 1992) or in which one species lacks the larval phase. If the species are not closely related (e.g. different genus) effects of different evolutionary histories and divergent life history strategies may be confounded (Hellberg 1996; Hoskin 1997; Teske *et al.* 2007; Sherman *et al.* 2008; Hoffman *et al.* 2011; Barbosa *et al.* 2013). Similarly, comparing closely related species living in allopatry can confound the effects of different environmental pressures and different physical connectivity among the localities (Puritz *et al.* 2012). Thus, relatively few examples of sympatric and closely related species with contrasting life history strategies are found in the literature (Berger 1973; Hunt 1993; Duffy 1993; Arndt & Smith 1998; Collin 2001; Dawson *et al.* 2002; Lee & Boulding 2009; Steele *et al.* 2009; Tarnowska *et al.* 2012).

The brittle star *Ophioderma longicauda* (Bruzelius, 1805) was known as a common Atlanto-Mediterranean gonochoric species (Tortonese 1983) reproducing via a lecithotrophic vitellaria larva (Fenaux 1969, 1972). However, the discovery of brooding individuals (Stöhr *et al.* 2009), also gonochoric, suggested that it is actually a species

complex composed of six mitochondrial lineages (L1-L6) (Boissin *et al.* 2011). Furthermore, a recent study (Chapter 2; Weber *et al.* 2014) showed that the lineages L1 and L3 are different biological species since they display temporal pre-zygotic isolation, as well as morphological and genetic differences at two loci. In addition, preliminary analyses showed potential differences in thermotolerance between L1 (corresponding to cluster C1) and L3 (cluster C4) (Chapter 4; Weber *et al.* 2013). Finally, an extensive study based on 31 nuclear markers used in individuals belonging to the six mitochondrial lineages showed that the species complex *O. longicauda* included six different genetic clusters, slightly different from the mitochondrial lineages (Chapter 3).

The cluster C1 encompasses each individual from the lineage L1, as well as each individual from the lineage L5, except those sampled in Dakar (Senegal). C1 is therefore present in the Atlantic and in the entire Mediterranean, from the Canary Islands to Lebanon. The cluster C2 includes only individuals sampled in Tunisia, displaying a mitochondrial lineage close to L3 but alleles common with L1 for the nuclear locus i51. This cluster most likely results from an ancient hybridization event between clusters C1 and C4 (Chapter 3). The cluster C4 includes all individuals from lineage L3, and the individuals from lineage L2 sampled in Greece. The cluster C5 includes all individuals from lineage L4, and the individuals of lineage L2 sampled in Cyprus and Lebanon. Finally, the cluster C3 includes all individuals from lineage L5 sampled in Dakar, and the cluster C6 includes all individuals from lineage L6.

Therefore, the *O. longicauda* species complex may represent an excellent model to test hypotheses on life history traits and connectivity as it displays two contrasting reproductive strategies: the cluster C1 reproduces once a year (Fenaux 1972) via lecithotrophic vitellaria larvae (Fenaux 1969), whereas clusters C4, C5 and most likely C2 brood their offspring (Stöhr *et al.* 2009). The broadcast spawning cluster C1 is found in sympatry with the brooding clusters C4 and C5 in the eastern Mediterranean, providing an interesting opportunity to compare genetic structure and connectivity levels among those lineages.

In this study, we investigate how the larval phase contributes to connectivity using two sympatric, ecologically similar species that display contrasting dispersal abilities. We compared the genetic structure within and between the different clusters on a large geographic scale (801 individuals from clusters C1, C2, C4 and C5 from the Canary Islands to Lebanon) and a smaller scale (C1 and C4 from six syntopic populations

in Greece). In addition, we compared the relative distribution of molecular variance within and between populations in those two species. Furthermore, we estimated the difference in number of migrants between those two sympatric species. We expect a strong difference in genetic structure between brooding and broadcast spawning species, since the dispersal of brooders is expected to be close to zero as adults are exclusively benthic and often hidden under rocks. Finally, we compared the levels of genetic diversity between the brooding and broadcasting species. K-strategists or species with high parental investment such as brooders are expected and known to display lower effective size and therefore should experience higher genetic drift, resulting in lower genetic diversity.

## 5.3 Methods

### 5.3.1 *Sampling, DNA extraction, COI sequencing, i51 genotyping and analyses*

Specimens from previous studies (Stöhr *et al.* 2009; Boissin *et al.* 2011) and additional specimens, collected between 2002 and 2013 by scuba diving (Table 5.1, Figure 5.1) were used. All individuals (Table 5.1) were sequenced for COI to determine their lineage. DNA extractions and PCRs targeting a portion of the COI gene were performed as previously described (Stöhr *et al.* 2009; Boissin *et al.* 2011). PCR products were sent to a private company for sequencing (LGC genomics, Berlin, Germany). Sequences were manually edited with the software BioEdit, aligned in the program Seaview (Gouy *et al.* 2010) with MUSCLE v3.8.31 algorithm (Edgar 2004) and manually improved. All sequences produced in this study were submitted to GenBank and are available with the accession numbers KF217260 to KF218086. Average per population haplotype and nucleotide diversities were investigated with DNAsp v5 software (Librado & Rozas 2009) as well as the following neutrality tests (performed on the pooled samples per cluster): Fu and Li's D and F (Fu & Li 1993); Fu's Fs (Fu 1997) and Tajima's D (Tajima 1989). $F_{ST}$ and $\Phi_{ST}$ were performed for clusters C1, C4 and C5. AMOVAs (Excoffier *et al.* 1992) were performed for the clusters C1 and C4, with South Crete, North Crete and Symi/Rhodes as the groups of populations. Haplotype networks were generated using the median-joining algorithm of Network, v.4.6.1.1 (Bandelt *et al.* 1999). Isolation by distance was investigated for clusters C1 and C4 by performing Mantel tests using the software Genetix, v.4.05 (Belkhir *et al.* 2004) using both genetic

distances ($F_{ST}/1-F_{ST}$), and -Log(1-$F_{ST}$)), and using conventional $F_{ST}$ (based on frequency) as well as $\Phi_{ST}$ (for sequence data). The codominant nuclear marker i51, displaying size polymorphism, was genotyped for each individual on high resolution agarose electrophoresis as previously described (chapter 2; Weber *et al.* 2014). $F_{ST}$ and $F_{IS}$ analyses were performed using Genetix.

**Table 5.1**: Sampling dates, locations and number of individuals for each cluster used in this study. See legend of Figure 1 for geographic codes.

| Basin | Code | Locality | Sampling date | C1 | C2 | C4 | C5 |
|---|---|---|---|---|---|---|---|
| Atlantic | Can | Teneriffe / Gran Canaria / La Palma,Canaria | January 1998 / October 2012 | 21 | | | |
| | Mad | Paul do Mar / Reis Magos / Canico, Madeira | February 1998 / October 2002 / February 2012 | 31 | | | |
| | Alg | Algarve, Portugal | March 2001 / March 2008 / July 2012 | 91 | | | |
| West Med | Ceu | Ceuta | March 2007 | 1 | | | |
| | CaB | Cap de Bol, Spain | August 2012 | 7 | | | |
| | Cou | Cap Oulestrell, France | December 2007 | 18 | | | |
| | Ves | La Vesse, France | December 2008 | 29 | | | |
| | Mar | Impériaux du Large / Maïre island, France | September 2004 | 21 | | | |
| | Cca | Cap Caveau, France | December 2009 | 30 | | | |
| | Por | Porquerolles, France | March 2009 | 31 | | | |
| | Sra | Saint-Raphaël, France | March 2010 | 31 | | | |
| | VsM | Villefranche sur mer, France | January 2009 | 32 | | | |
| | Sca | Scandola, Corsica, France | March 2010 | 29 | | | |
| | Ton | Tonnara, Corsica, France | January 2010 | 9 | | | |
| Central Med | Mon | Monastir, Tunisia | September 2012 / July 2013 | 1 | 12 | | |
| | Mal | Mghatab Bay, Malta | September 2005 | 17 | | | |
| | Cre | Cres, Croatia | November 2001 | 8 | | | |
| | Mlj | Mljet island, Grabova, Croatia | June 2002 | 17 | | | |
| | Alb | Saranda, Albania | April 2001 | 4 | | | |
| East Med | Sar | Gulf of Saronikos, Greece | May 2001 | 4 | | | |
| | Sym | Symi island, Greece | August 2005 | 4 | | 7 | |
| | Rho | Rhodes island, Greece | September 2005 | 6 | | 21 | |
| | LyP | Lygaria Port, Crete, Greece | $7^{th}$ May 2012 | 8 | | 20 | |
| | Lyg | Lygaria, Crete, Greece | $7^{th}$ May 2012 | 10 | | 9 | |
| | Gou | Gouves, Crete, Greece | $3^{rd}$ May 2012 | | | 80 | |
| | Elu | Elounda, Crete, Greece | July 2008 | 4 | | 18 | |
| | AgN | Agios Nikolaos, Crete, Greece | $23^{th}$ May 2012 | 36 | | 1 | |
| | AgP | Agios Pavlos, Crete, Greece | $14^{th}$ May 2012 | 20 | | 21 | |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | Ier | Ierapetra, Crete, Greece | 29th May 2012 | 5 | | 35 | |
| | BoA | Baths of Aphrodite, Cyprus | May 2005 | 13 | | | 39 |
| | Per | Pernera, Cyprus | May 2005 | 1 | | | 23 |
| | Ram | Ramkine island, Lebanon | July 2003 | 1 | | | 31 |
| | Han | Hannouch / Batroun, Lebanon | | 1 | | | 32 |
| | Bei | Beirut / Raoucheh / Selaata, Lebanon | July 2005 | 3 | | | 1 |
| | Nak | Nakoura, Lebanon | September 2002 | | | | 1 |
| Total | | | | 544 | 12 | 212 | 127 |

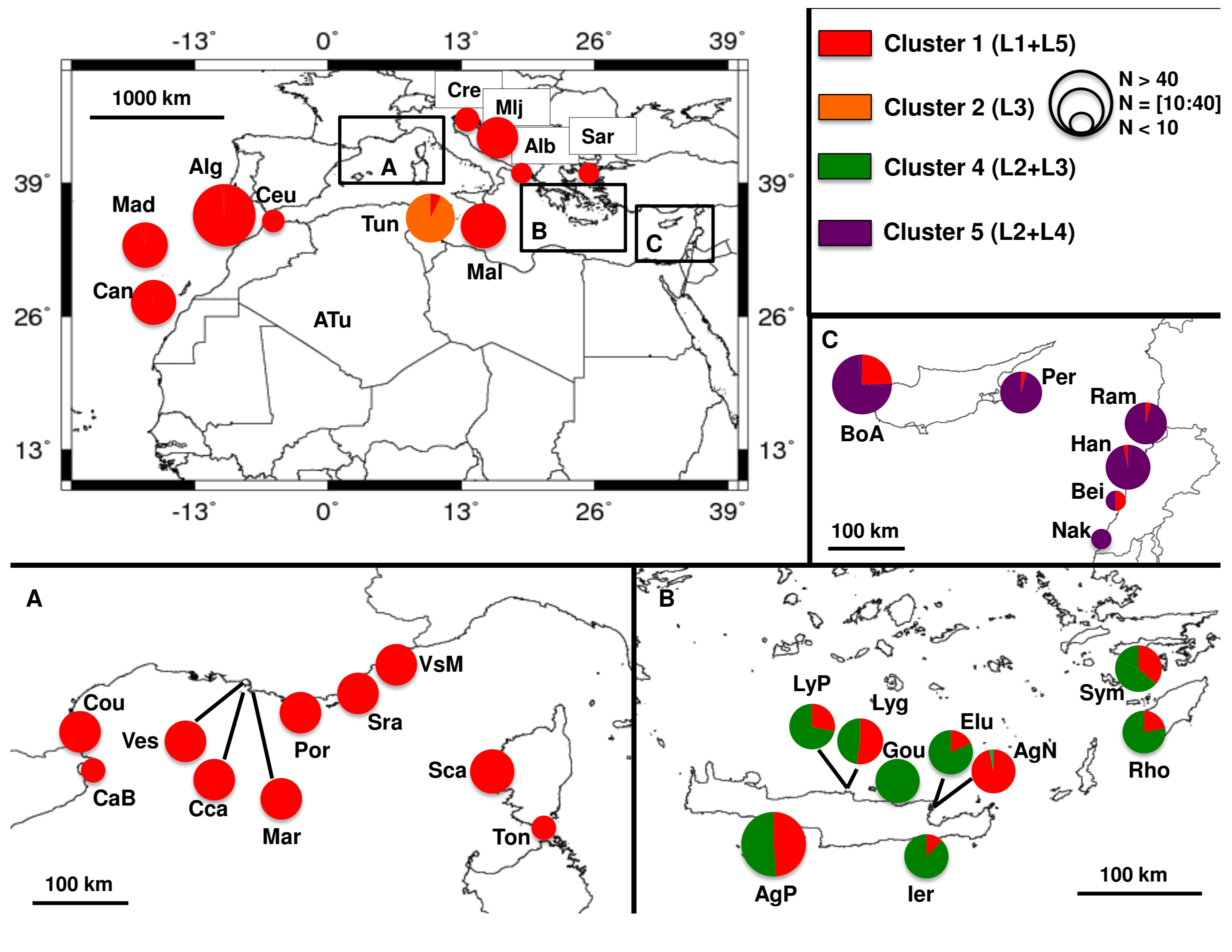



**Figure 5.1**: Map of sampling locations in the Mediterranean Sea and the Atlantic, with COI lineages and numbers of individuals collected. Details of sampling are displayed in Table 1. Can = Canaria; Mad = Madeira; Alg = Algarve; Ceu = Ceuta; Mon = Monastir; Mal = Malta; Cre = Cres; Mlj = Mljet island; Alb = Albania; Sar = Saronic Gulf. A: northwestern Mediterranean Sea. CaB = Cap de Bol; Cou = Cap Oulestrell; Ves = La Vesse ; Cca = Cap Caveau ; Mar = Marseilles ; Por = Porquerolles ; Sra = Saint Raphaël ; VsM = Villefrance sur Mer ; Sca = Scandola ; Ton = Tonnara. B : Greek Islands. AgP = Agios Pavlos ; Ier = Ierapetra ; LyP = Lygaria Port ; Lyg = Lygaria ; Gou = Gouves ; Elu = Elounda ; AgN = Agios Nikolaos ; Sym = Symi Islands ; Rho = Rhodes. C : Levant basin. BoA = Baths of Aphrodite; Per = Pernera ; Ram = Ramkine Islands ; Han = Hannouch ; Bei = Beirut; Nak = Nakoura.

### 5.3.2 *Transcriptome based nuclear markers*

Twelve primer pairs for PCR amplification of nuclear markers were newly developed on the basis of transcriptome sequences from C1 and C4 (see chapter 6 for details about transcriptome generation; Weber et al., in prep). They were developed with the following criteria: they had to be polymorphic in C1 and C4, with shared polymorphism and no diagnostic SNPs between lineages (among the 5 C1 and 2 C4 individual transcriptomes available for C1 and C4, respectively). Their length was between 200 and 400 bp. Two of the primer pairs did not amplify and two contained introns and were thus too long to be sequenced by Illumina. For the eight remaining loci, 192 individuals of C1 (n= 83) and C4 (n= 109) from Crete and Rhodes were amplified. The populations used were Lygaria, Lygaria Port, Elounda, Agios Nikolaos, Ierapetra, Agios Pavlos and Rhodes (Fig. 5.1). Amplicons were paired-end sequenced (2 x 250 bp) with a Miseq sequencer (Illumina technology) by the genomic platform Genotoul. Reads were cleaned, assembled and demultiplexed with the program MOTHUR (Schloss *et al*. 2009). For each primer pair, the total number of reads of five individuals was first manually checked, to infer a read number threshold, since the variance in read numbers among markers was high. Sequences with the highest number of reads were kept, including more than two sequences per individual when this occurred (i.e. allowing paralogs). On the eight amplified primer pairs, one was monomorphic, one displayed very low quality reads, two displayed one or two sequences per individual, and four displayed paralogous genes. For three primer pairs, paralogs could unambiguously be separated, whereas it was not possible for the last pair. In total, 10 markers were available for further analysis. Sequences were converted to genotypes with the program PGDspider (Lischer & Excoffier 2012) and analyzed in Genetix (Belkhir *et al.* 2004) to first assess if the markers displayed linkage disequilibrium, since some paralogs were used. Then, multilocus global and pairwise $F_{ST}$, and $F_{IS}$ were calculated in Genetix. Calculations of isolation by distance were performed as previously described.

### 5.3.3 *Comparative connectivity between L1 and L3 in Crete and Rhodes*

Connectivity of C1 and C4 was assessed using the program MIGRATE-N v. 3.2.1 (Beerli & Felsenstein 2001), which uses Bayesian methods to estimate the posterior probability of the parameters Θ (effective size of each population) and M (=m/μ, the ratio between the migration rate and mutation rate, for each pair of population, allowing

asymmetric migration). A random genealogy and parameter settings inferred by an $F_{ST}$-based method were used for the starting condition. For C1 and C4, the prior distribution for the parameters was uniform. The priors for Θ were bounded between 0 and 5.0 for L1, and between 0 and 0.001 for L3. The priors for M were bounded between 0 and 30,000 for L1, and between 0 and 10,000 for L3. For L1, the first 100,000 steps were discarded, then 2.5 million steps were completed using parallel runs of five replicates. To improve searching, a heating scheme was applied using four different temperatures (1, 1.5, 3 and 1,000,000). For L3, the first 50,000 steps were discarded, then 10 million steps were completed using parallel runs of ten replicates. A heating scheme was used as described above. To ensure the stability of results, three replicate analyses were performed using different initial random seeds.

## 5.4 Results

### 5.4.1 *Large scale analyses: COI and i51*

By combining previously published data and 702 newly analyzed individuals, we obtained a total of 801 COI sequences of 491 bp corresponding to 194 different haplotypes (Table 5.2, Figure 5.2). In the clusters C1, C4 and C5, the haplotype diversity within population was high, ranging from 0.570 in C4 to 0.930 in C1. In contrast, it was low in the cluster C2 (0.196). The nucleotide diversity ranged from $0.46 \times 10^{-3}$ in C2 to $6.32 \times 10^{-3}$ in C5. The different neutrality tests were all significant for clusters C1, C2 and C4, suggesting a recent population expansion event (in C1, illustrated by the star-like pattern of the haplotype network, Figure 5.2A) or a selective sweep on mitochondrial DNA. In contrast, the different neutrality tests were non-significant for C5. Finally, isolation by distance tests were not significant in the cluster C1 across the entire geographic range.

All individuals were genotyped for the marker i51, which displayed 14 alleles in the cluster C1 but was monomorphic in the brooding clusters C4 and C5, preventing a comparison between brooders and broadcasters. Interestingly, the C2 populations from Tunisia (red haplotypes in Figure 5.2B) were polymorphic at the i51 locus and shared alleles with C1.

#### 5.4.1.1 Broadcast spawning cluster C1

We analyzed COI sequences for a total of 23 populations of C1 sampled across the entire geographic range and performed pairwise $F_{ST}$ and $\Phi_{ST}$ among populations. Most of the population comparisons were significant for $F_{ST}$ analyses ($F_{ST}$: 0 - 0.14) (Table 5.3) whereas the majority was non-significant for $\Phi_{ST}$ analyses (values ranging from 0 to 0.15) (Table 5.3). The haplotype network showed no conspicuous phylogeographic structure (Figure 5.2A), as each dominant haplotype was found in the three Mediterranean basins and in the Atlantic. In contrast, the nuclear marker i51 displayed very low genetic structure, as most $F_{ST}$ analyses were non-significant ($F_{ST}$: 0 – 0.21) (Table 5.4).

The AMOVA analyses based on haplotype frequency revealed no significant differences in global $F_{ST}$ among populations ($F_{ST}$; p-value = 0.31), in $F_{ST}$ among populations within group ($F_{SC}$; p-value = 0.21) nor among groups (South Crete, North Crete, Symi Island/ Rhodes) ($F_{CT}$; p-value = 0.73). 99% of the molecular variance was found within populations, 1% was explained by the among-populations within groups level and 0% by the among-groups level. The AMOVA analyses taking into account differences among haplotypes, rather than only haplotype frequencies, yielded the same results. Finally, there was no correlation between $F_{ST}$ and geographic distance or between $\Phi_{ST}$ and geographic distance, either in the whole geographic range of C1 or when restricting the analysis to the western or the eastern Mediterranean basin.

#### 5.4.1.2 Brooding clusters C4 and C5

We compared eight populations of cluster C4, all found in Greece (Symi Island, Rhodes and Crete) (Figure 5.1). They displayed very high genetic structure (values ranging from 0.02 to 0.57). All pairwise $F_{ST}$ comparisons were significant, except the one involving populations separated by only 1 km (Lygaria and Lygaria Port) (Table 5.5A), and the majority of $\Phi_{ST}$ comparisons (values ranging from 0.03 to 0.85) were significant (Table 5.5A). The COI haplotype network also displayed strong phylogeographic structure and displayed three main haplogroups (Figure 5.2B). There was no correlation between $F_{ST}$ and geographic distance, or between $\Phi_{ST}$ and geographic distance in the whole geographic range of cluster C4. The AMOVA analyses were significant for the $F_{ST}$ and $\Phi_{ST}$ (among populations) comparisons (p-value < 0.001) and the $F_{SC}$ and $\Phi_{SC}$ (among

populations within groups) comparisons (p-value < 0.001), but not for the $F_{CT}$ and $\Phi_{CT}$ (among groups; South Crete, North Crete, Symi Island/Rhodes) comparisons (p-value > 0.05). In the $F_{ST}$ analysis, 67% of the molecular variance was explained within populations, 28% by the among-populations within groups comparison and 5% by the among groups comparison. In the $\Phi_{ST}$ analysis, 46% of the molecular variance was explained by within-populations comparison, 54% by the among-populations within groups comparison and 0% by the among groups comparison. Finally, we compared seven populations of cluster C5, a cluster exclusively found in Cyprus and Lebanon (Figure 5.1). The majority of pairwise $F_{ST}$ comparisons were significant and $F_{ST}$ values were high, ranging from 0.01 to 0.34 (Table 5.5B). $\Phi_{ST}$ analyses displayed the same pattern, the majority of comparisons being significant (values ranging from 0.05 to 0.91) (Table 5.5B).

**Table 5.2:** Number of individuals of each cluster (C1, C2, C4, C5) used for genetics in this study. Number of haplotypes per lineage, average per population haplotype (Hd) and nucleotide (π) diversities are also indicated. Neutrality tests D* Fu & Li's, F* Fu & Li's, Tajima's D and F's Fu were performed on the pooled population samples. Significant values are in bold. *: 0.01 < p-value < 0.05; **: 0.001 < p-value < 0.01; ***: p-value < 0.001.

| | C1 | C2 | C4 | C5 | Total |
|---|---|---|---|---|---|
| N sequences | 504 | 12 | 158 | 127 | 801 |
| N haplotypes | 123 | 3 | 28 | 37 | 194 |
| Av.pop Hd | 0,93 | 0,19 | 0,57 | 0,76 | |
| Av.pop .π [x10$^{-3}$] | 5,17 | 0,46 | 3,22 | 6,32 | |
| D* Fu & Li's | **-3,097*** | **-3,246*** | **-3,090*** | -1,367 | |
| F* Fu & Li's | **-3,363**** | **-3,349**** | **-3,058**** | -1,135 | |
| Tajima's D | **-2,522***** | **-2,014*** | **-1,787*** | -0,36 | |
| F's Fu | -260,01 | -5,849 | -13,519 | -7,222 | |

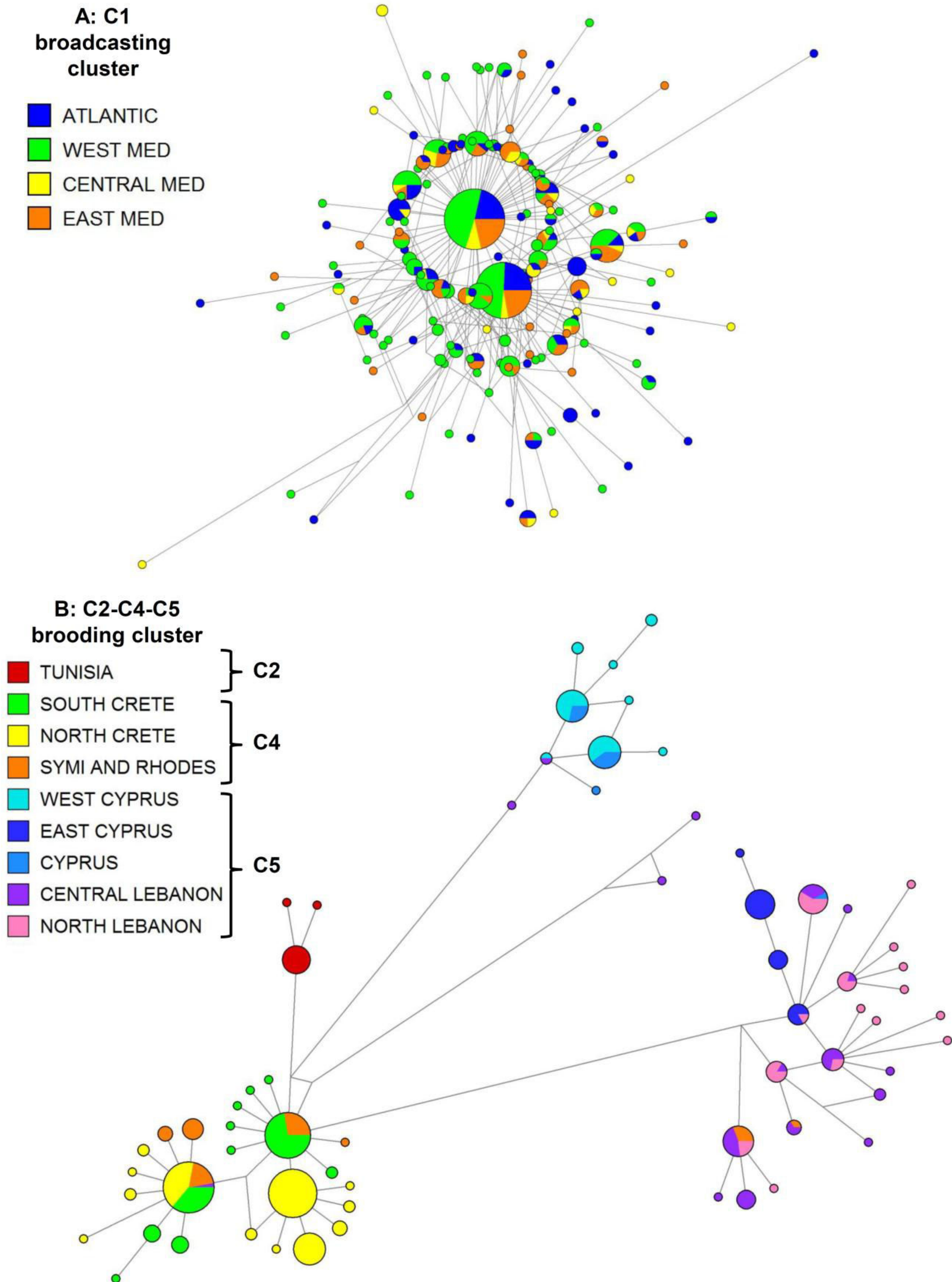



**Figure 5.2**: Mitochondrial COI haplotype network. A: 123 *Ophioderma longicauda* haplotypes from the broadcasting cluster C1. B: 3, 28 and 37 haplotypes of *O. longicauda* brooding clusters C2, C4 and C5, respectively. Reconstructions were performed using the Median Joining calculation of the Network software, version 4.6.1.1. Distances are proportional to mutation events and circle size is proportional to number of sampled individuals.

**Table 5.3**: *O. longicauda* broadcasting cluster C1 pairwise $F_{ST}$ values (under diagonal) and $\Phi_{ST}$ values (upper diagonal) and their significance, based on the mitochondrial COI gene. Significant values are in bold. *: 0.01 < p-value < 0.05; **: 0.001 < p-value < 0.01; ***: p-value < 0.001

| Cluster C1 | Gar (21) | PdM (25) | Sag (68) | CaB (7) | Cou (20) | Ves (27) | Cca (29) | ImL (21) | Por (31) | Sra (31) | VsM (32) | Sca (25) | Ton (9) | Mal (17) | Cre (8) | Mlj (17) | Lyg (10) | LyP (8) | Elu (4) | AgN (31) | AgP (20) | Ier (5) | BoA (14) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Gar (21) | 0 | -0.007 | 0.002 | 0.014 | 0.021 | 0.027 | 0.006 | 0.003 | 0.029* | 0.035** | 0.030* | 0.024 | 0.044 | 0.022 | -0.016 | 0.043* | 0.064* | -0.009 | -0.038 | 0.006 | 0.014 | 0.038 | 0.004 |
| PdM (25) | 0.099*** | 0 | -0.005 | 0.025 | 0.011 | 0.018 | 0.010 | -0.002 | 0.019 | 0.028* | 0.018 | 0.012 | 0.034 | 0.031 | 0.007 | 0.042* | 0.035 | -0.016 | -0.016 | -0.002 | 0.002 | 0.054 | -0.008 |
| Sag (68) | 0.065*** | 0.082*** | 0 | 0.007 | 0.001 | 0.003 | -0.008 | -0.013 | 0.011 | 0.015* | 0.010 | 0.007 | 0.002 | 0.005 | 0.016 | 0.016 | 0.018 | -0.051 | -0.064 | -0.008 | -0.007 | 0.048 | -0.005 |
| CaB (7) | 0.066* | 0.087** | 0.050* | 0 | 0.027 | 0.059 | 0.008 | 0.008 | 0.002 | 0.017 | 0.021 | 0.001 | 0.046 | 0.001 | -0.015 | 0.026 | 0.036 | -0.014 | -0.019 | 0.020 | 0.008 | -0.011 | 0.001 |
| Cou (20) | 0.067*** | 0.085*** | 0.052*** | 0.050* | 0 | 0.006 | 0.016 | -0.005 | 0.014 | -0.001 | 0.010 | 0.003 | 0.001 | 0.004 | 0.033 | 0.004 | 0.005 | -0.035 | -0.008 | 0.014 | -0.004 | 0.035 | -0.013 |
| Ves (27) | 0.108*** | 0.125*** | 0.090*** | 0.098*** | 0.094*** | 0 | 0.011 | -0.006 | 0.012 | -0.008 | -0.001 | -0.001 | -0.002 | 0.006 | 0.086* | 0.010 | 0.020 | 0.011 | 0.031 | 0.013 | -0.002 | 0.153*** | 0.023* |
| Cca (29) | 0.066*** | 0.084*** | 0.052*** | 0.050* | 0.052*** | 0.093*** | 0 | -0.013 | 0.003 | 0.015 | -0.001 | 0.005 | 0.013 | 0.001 | 0.039 | 0.016 | 0.041 | -0.040 | -0.041 | -0.005 | -0.002 | 0.091 | -0.001 |
| ImL (21) | 0.098*** | 0.116*** | 0.081*** | 0.086*** | 0.084*** | 0.124*** | 0.082*** | 0 | 0.002 | -0.002 | -0.005 | -0.011 | 0.004 | 0.002 | 0.017 | 0.009 | 0.022 | -0.044 | -0.030 | -0.018 | -0.014 | 0.069 | -0.012 |
| Por (31) | 0.050*** | 0.068*** | 0.037*** | 0.033* | 0.037*** | 0.077*** | 0.037*** | 0.067*** | 0 | 0.015 | 0.007 | -0.004 | 0.004 | 0.017 | 0.026 | 0.024 | 0.013 | -0.027 | -0.014 | 0.016 | 0.009 | 0.036 | 0.004 |
| Sra (31) | 0.048*** | 0.066*** | 0.035*** | 0.030* | 0.034*** | 0.075*** | 0.034*** | 0.064*** | 0.019** | 0 | 0.010 | -0.017 | -0.022 | 0.004 | 0.054** | 0.006 | -0.013 | -0.009 | 0.000 | 0.023** | 0.003 | 0.074* | 0.015 |
| VsM (32) | 0.071*** | 0.089*** | 0.057*** | 0.056* | 0.058*** | 0.098*** | 0.057*** | 0.088*** | 0.042*** | 0.040*** | 0 | 0.006 | -0.016 | 0.011 | 0.068* | 0.002 | 0.004 | -0.011 | 0.023 | 0.014 | -0.001 | 0.118* | 0.004 |
| Sca (25) | 0.048*** | 0.067*** | 0.035*** | 0.030* | 0.034** | 0.076*** | 0.034*** | 0.065*** | 0.019*** | 0.017*** | 0.040*** | 0 | -0.001 | 0.003 | 0.035 | 0.008 | 0.009 | -0.018 | -0.021 | 0.008 | 0.003 | 0.049 | 0.006 |
| Ton (9) | 0.095*** | 0.114*** | 0.077*** | 0.081 | 0.079*** | 0.124*** | 0.078** | 0.113*** | 0.061*** | 0.059*** | 0.084*** | 0.059*** | 0 | 0.014 | 0.071 | -0.009 | -0.015 | 0.037 | 0.121 | 0.022 | 0.004 | 0.138** | 0.007 |
| Mal (17) | 0.063*** | 0.082*** | 0.048*** | 0.046* | 0.048*** | 0.091*** | 0.048*** | 0.080*** | 0.032** | 0.030*** | 0.054*** | 0.030*** | 0.075** | 0 | 0.030 | -0.009 | 0.008 | -0.041 | -0.046 | 0.021 | 0.004 | 0.051 | 0.001 |
| Cre (8) | 0.044* | 0.065 | 0.029 | 0.023 | 0.029 | 0.075* | 0.029 | 0.063* | 0.012 | 0.009 | 0.035 | 0.009 | 0.057 | 0.024 | 0 | 0.047* | 0.059 | -0.008 | -0.055 | 0.011 | 0.026 | -0.070 | -0.015 |
| Mlj (17) | 0.070*** | 0.089*** | 0.055*** | 0.054* | 0.056** | 0.098*** | 0.055*** | 0.087*** | 0.039*** | 0.037*** | 0.061*** | 0.037*** | 0.083*** | 0.051** | 0.032 | 0 | -0.011 | -0.031 | 0.015 | 0.014 | -0.002 | 0.077* | -0.007 |
| Lyg (10) | 0.105** | 0.124*** | 0.087*** | 0.093* | 0.090** | 0.134*** | 0.088*** | 0.123*** | 0.072** | 0.069*** | 0.094*** | 0.070*** | 0.122*** | 0.086*** | 0.069* | 0.094*** | 0 | 0.024 | 0.113 | 0.039 | -0.001 | 0.071 | -0.007 |
| LyP (8) | 0.061* | 0.081* | 0.045** | 0.042 | 0.045* | 0.091* | 0.045** | 0.079*** | 0.028** | 0.025* | 0.051*** | 0.025 | 0.074 | 0.040* | 0.018 | 0.048* | 0.086 | 0 | -0.057 | -0.039 | -0.027 | 0.044 | -0.040 |
| Elu (4) | 0.112 | 0.134* | 0.092*** | 0.099 | 0.095* | 0.145** | 0.093* | 0.133** | 0.075* | 0.072* | 0.100** | 0.072* | 0.134* | 0.090* | 0.070 | 0.099* | 0.146 | 0.091 | 0 | -0.056 | 0.001 | -0.019 | -0.047 |
| AgN (31) | 0.053*** | 0.071*** | 0.039*** | 0.035** | 0.039*** | 0.079*** | 0.039*** | 0.069*** | 0.024*** | 0.022*** | 0.044*** | 0.021*** | 0.064*** | 0.035*** | 0.014 | 0.042*** | 0.074*** | 0.030* | 0.077* | 0 | -0.003 | 0.065 | -0.007 |
| AgP (20) | 0.059*** | 0.077*** | 0.045*** | 0.042 | 0.045*** | 0.087** | 0.044*** | 0.076*** | 0.029*** | 0.027*** | 0.050*** | 0.027*** | 0.071*** | 0.040*** | 0.020 | 0.048*** | 0.081*** | 0.036* | 0.085* | 0.031*** | 0 | 0.063 | -0.009 |
| Ier (5) | 0.089 | 0.110* | 0.070** | 0.072 | 0.072* | 0.121* | 0.071* | 0.109* | 0.053** | 0.050* | 0.077** | 0.050* | 0.106 | 0.067* | 0.045 | 0.076 | 0.119* | 0.065 | 0.131 | 0.055* | 0.072 | 0 | -0.010 |
| BoA (14) | 0.084*** | 0.103*** | 0.068*** | 0.070** | 0.070*** | 0.112*** | 0.069*** | 0.102*** | 0.053*** | 0.051*** | 0.074*** | 0.051*** | 0.099** | 0.066*** | 0.047* | 0.073*** | 0.110*** | 0.064* | 0.118* | 0.055*** | 0.062** | 0.093* | 0 |

.

**Table 5.4:** *O. longicauda* broadcasting cluster C1 pairwise $F_{ST}$ values and their significance, based on the nuclear i51 marker. Significant values are in bold. *: 0.01 < p-value < 0.05; **: 0.001 < p-value < 0.01; ***: p-value < 0.001.

| Cluster C1 | Lyg (10) | LyP (8) | AgP (20) | AgN (36) | Elu (4) | Ier (5) | BoA (12) | Mal (4) | Ves (29) | VsM (32) | Por (31) | Sca (29) | Cca (30) | Ton (9) | Sra (31) | ImL (35) | Cou (19) | CaB (7) | Sag (43) | PdM (24) | Gar (7) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Lyg (10) | 0 | | | | | | | | | | | | | | | | | | | | |
| LyP (8) | 0.036 | 0 | | | | | | | | | | | | | | | | | | | |
| AgP (20) | 0.015 | -0.019 | 0 | | | | | | | | | | | | | | | | | | |
| AgN (36) | 0.0311 | -0.002 | -0.010 | 0 | | | | | | | | | | | | | | | | | |
| Elu (4) | -0.060 | 0.004 | -0.006 | 0.015 | 0 | | | | | | | | | | | | | | | | |
| Ier (5) | 0.076 | -0.067 | -0.027 | -0.034 | 0.033 | 0 | | | | | | | | | | | | | | | |
| BoA (12) | 0.022 | -0.062 | -0.016 | 0.005 | -0.041 | -0.053 | 0 | | | | | | | | | | | | | | |
| Mal (4) | 0.194 | 0.167 | **0.146*** | 0.083 | **0.190*** | 0.049 | 0.146 | 0 | | | | | | | | | | | | | |
| Ves (29) | 0.038 | -0.024 | -0.009 | -0.004 | 0.019 | -0.047 | -0.009 | **0.129*** | 0 | | | | | | | | | | | | |
| VsM (32) | -0.024 | 0.004 | -0.001 | **0.017*** | -0.052 | 0.033 | 0.002 | **0.176***** | 0.015 | 0 | | | | | | | | | | | |
| Por (31) | 0.007 | 0.011 | 0.012 | **0.027*** | -0.070 | 0.035 | -0.004 | **0.183*** | 0.027 | 0.001 | 0 | | | | | | | | | | |
| Sca (29) | -0.006 | -0.020 | 0.001 | 0.026 | -0.052 | 0.019 | -0.018 | **0.214***** | 0.015 | -0.009 | -0.002 | 0 | | | | | | | | | |
| Cca (30) | -0.013 | 0.004 | 0.011 | **0.039*** | -0.067 | 0.052 | -0.004 | **0.216***** | 0.043 | -0.004 | 0.002 | -0.011 | 0 | | | | | | | | |
| Ton (9) | -0.023 | -0.001 | -0.026 | -0.023 | -0.035 | -0.012 | -0.006 | 0.120 | -0.005 | -0.021 | -0.001 | -0.005 | -0.003 | 0 | | | | | | | |
| Sra (31) | 0.001 | -0.007 | -0.011 | -0.005 | -0.009 | 0.001 | -0.004 | **0.135***** | -0.004 | -0.007 | 0.009 | 0.001 | 0.012 | -0.025 | 0 | | | | | | |
| ImL (35) | -0.003 | 0.018 | -0.001 | 0.003 | -0.029 | 0.023 | 0.013 | **0.134***** | 0.008 | -0.006 | 0.001 | 0.010 | 0.017 | -0.023 | -0.008 | 0 | | | | | |
| Cou (19) | 0.002 | -0.017 | -0.019 | -0.007 | -0.003 | -0.011 | -0.015 | **0.147***** | -0.013 | -0.011 | 0.008 | -0.007 | 0.009 | -0.020 | -0.017 | -0.007 | 0 | | | | |
| CaB (7) | -0.018 | 0.012 | -0.001 | 0.006 | -0.040 | 0.008 | 0.002 | **0.118*** | 0.013 | -0.017 | 0.002 | -0.005 | -0.010 | -0.011 | 0.001 | -0.006 | -0.002 | 0 | | | |
| Sag (43) | 0.015 | 0.002 | -0.006 | -0.006 | 0.016 | -0.007 | 0.011 | **0.116*** | -0.006 | 0.005 | **0.028***** | 0.019 | **0.036***** | -0.010 | -0.010 | -0.002 | -0.013 | 0.008 | 0 | | |
| PdM (24) | 0.011 | 0.036 | 0.012 | 0.006 | 0.032 | 0.022 | 0.039 | 0.076 | 0.014 | 0.015 | **0.051***** | 0.045 | **0.055*** | -0.013 | 0.001 | 0.006 | -0.001 | 0.009 | -0.005 | 0 | |
| Gar (7) | 0.107 | 0.079 | 0.044 | 0.005 | 0.107 | -0.013 | **0.074*** | -0.073 | 0.044 | **0.090***** | **0.102*** | **0.122***** | **0.126*** | 0.016 | **0.049*** | 0.049 | **0.058*** | 0.050 | 0.032 | 0.013 | 0 |

**Table 5.5:** *O. longicauda* brooding clusters C4 (A) and C5 (B) pairwise $F_{ST}$ values (under diagonal) and $\Phi_{ST}$ values (upper diagonal) and their significance, based on the mitochondrial COI gene. Significant values are in bold. *: 0.01 < p-value < 0.05; **: 0.001 < p-value < 0.01; ***: p-value < 0.001.

| (A) cluster C4 | Rho (21) | Sym (9) | Ier (30) | AgP (21) | LyP (18) | Lyg (9) | Gou (30) | Elu (18) |
|---|---|---|---|---|---|---|---|---|
| Rho (21) | 0 | **0.468*** ** | 0.046 | **0.471*** ** | **0.499*** ** | **0.594*** ** | **0.655*** ** | **0.201*** ** |
| Sym (9) | **0.114** ** | 0 | **0.523*** ** | **0.546*** ** | **0.526*** ** | **0.492*** ** | **0.660*** ** | **0.599*** ** |
| Ier (30) | **0.065*** ** | **0.143** ** | 0 | **0.456*** ** | **0.491*** ** | **0.567*** ** | **0.632*** ** | **0.107*** ** |
| AgP (21) | **0.226*** ** | **0.296*** ** | **0.302*** ** | 0 | **0.495*** ** | **0.681*** ** | **0.663*** ** | **0.800*** ** |
| LyP (18) | **0.320*** ** | **0.339*** ** | **0.314*** ** | **0.513*** ** | 0 | 0.003 | **0.209*** ** | **0.727*** ** |
| Lyg (9) | **0.365*** ** | **0.388*** ** | **0.362*** ** | **0.575*** ** | -0.022 | 0 | **0.235*** | **0.857*** ** |
| Gou (30) | **0.310*** ** | **0.315*** ** | **0.311*** ** | **0.472*** ** | **0.182** ** | **0.200*** | 0 | **0.827*** ** |
| Elu (18) | **0.203*** ** | **0.330*** ** | **0.084** ** | **0.548*** ** | **0.465*** ** | **0.546*** ** | **0.448*** ** | 0 |

| (B) Cluster C5 | Per (24) | BoA (27) | Cyp (11) | Han (26) | Bat (4) | Ram (24) | Leb (11) |
|---|---|---|---|---|---|---|---|
| Per (24) | 0 | **0.902*** ** | **0.916*** ** | **0.436*** ** | **0.542*** ** | **0.352*** ** | **0.389*** ** |
| BoA (27) | **0.273*** ** | 0 | **0.006*** ** | **0.826*** ** | **0.882*** ** | **0.840*** ** | **0.594*** ** |
| Cyp (11) | **0.340*** ** | -0.015 | 0 | **0.811*** ** | **0.891*** ** | **0.828*** ** | **0.496*** ** |
| Han (26) | **0.209*** ** | **0.175*** ** | **0.229*** ** | 0 | **0.230*** | **0.119*** ** | **0.197*** ** |
| Bat (4) | **0.303*** ** | **0.266** ** | **0.364*** ** | **0.157*** ** | 0 | 0.056 | 0.096 |
| Ram (24) | **0.187*** ** | **0.166*** ** | **0.219*** ** | **0.041*** | 0.066 | 0 | **0.197** ** |
| Leb (11) | **0.178*** ** | **0.136*** ** | **0.200** ** | 0.035 | **0.102*** | 0.005 | 0 |

**Table 5.6**: Comparison of the genetic diversity among 12 genetic markers in *Ophioderma longicauda* C1 and C4. Mitoch = mitochondrial; Seq = sequence; N = number; Hd = haplotype diversity; π = nucleotide diversity.

| locus | 1972 | | 79905 | | 50183_I | | 50183_II | | 50183_III | | 50183_IV | | 68241_I | | 68241_II | | 11915_I | | 11915_II | | i51 | | COI | | Average | | ratio C1/C4 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| genome | nuclear | | nuclear | | nuclear | | nuclear | | nuclear | | nuclear | | nuclear | | nuclear | | nuclear | | nuclear | | nuclear | | mitoch | | | | |
| type | seq | | seq | | seq | | seq | | seq | | seq | | seq | | seq | | seq | | seq | | µsat | | seq | | | | |
| cluster | C1 | C4 | C1 | C4 | C1 | C4 | C1 | C4 | C1 | C4 | C1 | C4 | C1 | C4 | C1 | C4 | C1 | C4 | C1 | C4 | C1 | C4 | C1 | C4 | C1 | C4 | |
| N alleles (total) | 13 | 3 | 16 | 5 | 17 | 3 | 16 | 7 | 23 | 4 | 22 | 5 | 19 | 2 | 23 | 4 | 7 | 1 | 7 | 1 | 13 | 1 | 29 | 20 | 17.1 | 4.7 | 3.6 |
| N alleles (mean/ pop) | 6.1 | 2 | 6.7 | 2 | 4.4 | 1.8 | 6.4 | 2.5 | 7.1 | 2.2 | 6.1 | 1.7 | 5.9 | 1 | 9.4 | 2.2 | 2.6 | 1 | 3.6 | 1 | 6.4 | 1 | 8.6 | 5 | 6.1 | 1.9 | 3.2 |
| Hd (mean/pop) | 0.83 | 0.34 | 0.82 | 0.25 | 0.48 | 0.19 | 0.78 | 0.31 | 0.67 | 0.25 | 0.57 | 0.05 | 0.71 | 0.02 | 0.89 | 0.25 | 0.44 | 0 | 0.32 | 0 | - | - | 0.92 | 0.55 | 0.68 | 0.20 | 3.4 |
| π [x$10^{-3}$] (mean/ pop) | 16.5 | 3.5 | 18.5 | 5.8 | 1.8 | 0.8 | 4.9 | 1.1 | 3.5 | 0.9 | 2.5 | 0.2 | 7.6 | 0.08 | 13.5 | 5.7 | 2.6 | 0 | 2.8 | 0 | - | - | 5.1 | 2.1 | 7.2 | 1.8 | 4 |

### 5.4.2 *Small scale analyses: comparison of genetic structure and migrant estimation in common locations of C1 and C4 using multilocus genotypes*

Sympatric C1 and C4 individuals were sampled in six locations of Greece (Figure 5.1B), and 10 newly developed nuclear markers were available for analysis (Table 5.6). Two markers (11915-I and 11915-II) were monomorphic in C4 but not in C1, and were thus excluded from the genetic structure analysis. In addition, of the four paralogs found for the marker 50183, three (I-III) displayed linkage disequilibrium, therefore loci 50183-I and 50183-III were excluded from the genetic structure analysis. Finally, six loci were kept for multilocus analyses.

As seen in the large scale analysis, genetic structure was different between brooders and broadcasters, since no genetic structure was detected for C1 (global $F_{ST}$: 0.00054) whereas it was strong in C4 (global $F_{ST}$: 0.49). The pairwise $F_{ST}$ analysis yielded the same pattern of contrasting genetic structure, with $F_{ST}$ values ranging from -0.063 to 0.049 for C1, and from -0.007 to 0.646 for C4 (Table 5.7). The single non-significant comparison in C4 was between the populations separated by 1 km (Lyg and LyP, Figure 5.1B). Most of the studied loci displayed a heterozygote deficiency in C1 (Table 5.8A) as well as in C4 (Table 5.8B), which resulted in heterozygote deficiencies in most populations for the two species. No pattern of isolation by distance (IBD) was detected in C1 at a small scale, whereas IBD was detected for the C4 populations ($p = 0.002$).

Effective size and migrant number estimations also differed between brooders and broadcasters. Indeed, the effective size of broadcasters was about 280 times higher on average than the effective size of brooders (Table 5.9). In addition, the average migrant number was about 50 times higher in broadcasters than in brooders (Table 5.9). Yet, these numbers should be taken with caution since standard deviations were very high, especially in migrant number (sd = 65 and 43 for $\Theta$ and M, respectively; Table 5.9).

### 5.4.3 *Comparison of genetic diversity between brooders and broadcasters*

The genetic diversity comparison was performed on the 10 nuclear loci, as well as using COI and i51 data (Table 5.6). The total allele number, the average per population allele number, haplotype and nucleotide diversities were investigated. Out of

the 12 loci analyzed, three were monomorphic in C4, whereas all were polymorphic in C1. In addition, each summary statistic was higher in C1 compared to C4. The average ratio between C1 and C4 ranged between 3.2 and 4 for each summary statistic. Therefore, the level of genetic diversity is about 3-4 times higher in broadcasters than in brooders.

**Table 5.7**: (A) Multilocus (using 6 nuclear loci) pairwise $F_{ST}$ values (W&C) of *O. longicauda* C1 populations. (B) Multilocus (using 6 nuclear loci) pairwise $F_{ST}$ values (W&C) of *O. longicauda* C4 populations. Significant values are in bold. *: 0.01 < p-value < 0.05; **: 0.001 < p-value < 0.01; ***: p-value < 0.001.

| (A) cluster C1 | AgN | AgP | Elu | Ier | Lyg | LyP |
|---|---|---|---|---|---|---|
| AgP | 0,003 | | | | | |
| Elu | -0,063 | -0,086 | | | | |
| Ier | 0,019 | -0,020 | -0,007 | | | |
| Lyg | 0,012 | 0,029 | 0,036 | 0,104 | | |
| LyP | 0,011 | -0,001 | -0,110 | 0,035 | 0,049 | |
| Rho | -0,033 | -0,006 | -0,085 | -0,017 | 0,022 | -0,031 |

| (B) cluster C4 | AgP | Elu | Ier | Lyg | LyP |
|---|---|---|---|---|---|
| Elu | **0,598*** ** | | | | |
| Ier | **0,363*** ** | **0,543*** ** | | | |
| Lyg | **0,646*** ** | **0,361** ** | **0,581*** ** | | |
| LyP | **0,610*** ** | **0,399*** ** | **0,580*** ** | -0,007 | |
| Rho | **0,617*** ** | **0,244*** ** | **0,472*** ** | **0,422*** ** | **0,482*** ** |

**Table 5.8**: Single locus and multilocus $F_{IS}$ values in *O. longicauda* C1 populations (A) and in C4 populations (B). Significant values are in bold. *: 0.01 < p-value < 0.05; **: 0.001 < p-value < 0.01; ***: p-value < 0.001. M: monomorphic locus in a population. NA: Missing data for a population.

| (A) | 1972 | 79905 | 50183_II | 50183_IV | 68241_I | 68241_II | Multilocus |
|---|---|---|---|---|---|---|---|
| AgN | 0.116 | **0.753*** ** | 0.113 | **0.253** ** | **0.750*** ** | 0.037 | **0.343*** ** |
| AgP | **0.260* ** | **0.427*** ** | 0.084 | 0.155 | **0.926*** ** | 0.045 | **0.349*** ** |
| Elu | 0.000 | 0.428 | -0.200 | **0.400*** ** | NA | -0.142 | 0.092 |
| Ier | **0.529*** ** | **0.225* ** | -0.200 | -0.200 | NA | -0.111 | 0.058 |
| Lyg | 0.160 | 0.066 | **0.296** ** | -0.086 | **0.636*** ** | 0.070 | **0.177* ** |
| LyP | **0.416* ** | **0.440*** ** | **0.200* ** | **1.000*** ** | **0.666*** ** | -0.108 | **0.360*** ** |
| Rho | -0.041 | **0.411** ** | -0.200 | **0.360* ** | **1.000*** ** | -0.132 | **0.252** ** |

| (B) | 1972 | 79905 | 50183_II | 50183_IV | 68241_I | 68241_II | Multilocus |
|---|---|---|---|---|---|---|---|
| AgP | **0.414*** ** | M | M | M | M | M | **0.320** ** |
| Elu | M | **1.000*** ** | -0.398 | M | -0.03226 | **1.000*** ** | **0.380** ** |
| Ier | M | -0.025 | -0.031 | M | M | **0.868*** ** | **0.244* ** |
| Lyg | **0.407** ** | M | -0.066 | M | M | **1.000*** ** | **0.466** ** |
| LyP | -0.259 | M | 0.113 | -0.032 | M | M | -0.049 |
| Rho | 0.090 | **0.588*** ** | -0.032 | M | NA | **1.000*** ** | **0.422*** ** |

**Table 5.9:** Multilocus estimation (using six nuclear loci and COI) of ϴ and M=m/μ performed with Migrate-n. Very close populations (distance between pop < 5 Km) were pooled to increase sample size. Average ratios of ϴ C1/C4 and M C1/C4 are displayed.

| | | C1 | | | C4 | | | |
|---|---|---|---|---|---|---|---|---|
| **ϴ** | | mode | IC 2,5% | IC 97,5% | mode | IC 2,5% | IC 97,5% | ratio C1/C4 |
| AgN+Elu | | 0,01833 | 0 | 0,1 | 0,00008 | 0,00002 | 0,00021 | 229,1 |
| AgP | | 0,03167 | 0 | 0,11333 | 0,00008 | 0,00003 | 0,00018 | 395,9 |
| Ier | | 0,035 | 0 | 0,12 | 0,00014 | 0,00009 | 0,00025 | 250,0 |
| Lyg+LyP | | 0,01833 | 0 | 0,1 | 0,00007 | 0,00003 | 0,00022 | 261,9 |
| Rho | | 0,06167 | 0 | 1,15333 | 0,00022 | 0,00008 | 0,00026 | 280,3 |
| mean ratio and sd | | | | | | | | **283 ± 65** |
| **M=m/μ** | | mode | IC 2,5% | IC 97,5% | mode | IC 2,5% | IC 97,5% | ratio C1/C4 |
| AgN+Elu → | AgP | 10250 | 2960 | 21300 | 176,7 | 0 | 2500 | 58,0 |
| | Ier | 1270 | 0 | 13180 | 196,7 | 0 | 2740 | 6,5 |
| | Lyg+LyP | 4590 | 2800 | 16440 | 196,7 | 0 | 3940 | 23,3 |
| | Rho | 5350 | 320 | 7220 | 183,3 | 0 | 6180 | 29,2 |
| AgP → | AgN+Elu | 9490 | 3700 | 14300 | 170 | 0 | 1933,3 | 55,8 |
| | Ier | 3830 | 1240 | 7620 | 190 | 0 | 3793,3 | 20,2 |
| | Lyg+LyP | 10590 | 5400 | 28640 | 190 | 0 | 2206,7 | 55,7 |
| | Rho | 10370 | 4020 | 19100 | 176,7 | 0 | 1773,3 | 58,7 |
| Ier → | AgN+Elu | 16070 | 8600 | 30000 | 183,3 | 0 | 6540 | 87,7 |
| | AgP | 5930 | 2500 | 23300 | 1296,7 | 0 | 8466,7 | 4,6 |
| | Lyg+LyP | 1770 | 320 | 17140 | 196,7 | 0 | 2273,3 | 9,0 |
| | Rho | 12010 | 860 | 20720 | 196,7 | 0 | 4706,7 | 61,1 |
| Lyg+LyP → | AgN+Elu | 9450 | 1700 | 14560 | 183,3 | 0 | 2826,7 | 51,6 |
| | AgP | 1030 | 260 | 2860 | 183,3 | 0 | 1753,3 | 5,6 |
| | Ier | 530 | 0 | 5980 | 190 | 0 | 2886,7 | 2,8 |
| | Rho | 2090 | 700 | 17060 | 190 | 0 | 2733,3 | 11,0 |
| Rho → | AgN+Elu | 28110 | 14180 | 30000 | 210 | 0 | 4186,7 | 133,9 |
| | AgP | 15850 | 960 | 26480 | 163,3 | 0 | 1333,3 | 97,1 |
| | Ier | 17570 | 7340 | 29840 | 210 | 0 | 3326,7 | 83,7 |
| | Lyg+LyP | 28350 | 16240 | 30000 | 190 | 0 | 1780 | 149,2 |
| mean ratio and sd | | | | | | | | **50 ± 43** |

**Table 5.10**: Comparison of within population percentage of molecular variance (AMOVA) in closely-related species with different dispersal abilities

| Organism | Echinoderms | Echinoderms | Echinoderms | Molluscs | Gastropods |
|---|---|---|---|---|---|
| Species | *Ophioderma longicauda* C1 (L) / *O. longicauda* C4 (B) | *Cucumaria miniata* (L) / *C. pseudocurata* (B) | *Meridiastra calcar* (L) / *Parvulastra exigua* (B) | *Littorina scutulata* (P) / *L. plena* (P) / *L. subrotundata* (B) / *L. sitkana* (B) | *Crepidula fornicata* (P) / *C. convexa* (B)&(L)/ species complex *C. depressa* (P)/ *C. atrasolea* (B) |
| Genetic marker | COI (MtDNA) | tRNA gene (MtDNA) | Allozyme electrophoresis | HSC70; APN 54 (nuclear) Cyt b (MtDNA) | COI (MtDNA) |
| $F_{ST}$ / $\Phi_{ST}$ | $F_{ST}$ | $\Phi_{ST}$ | $F_{ST}$ | $F_{ST}$ | $F_{ST}$ |
| Same populations | yes | yes | partially | yes | partially |
| Distance between pop | 1-200 km | 50-2000 km | 500-2600 km | 10-50 km | 50 -4300 km |
| Geographic range | Greece, eastern Mediterranean | US west coast, overlapping | SO Australia, overlapping | Vancouver island | US east coast |
| Planktotrophic larvae (P) | | | | 95-98.8 | 78-100 |
| Lecithotrophic larvae (L) | 99 | 97.5 | 99 | | 12.8 |
| Brooder or egg masses (B) | 64 | 3.4 | 40 | 72.1-95.7 | 23.9-45.7 |
| References | This study | Arndt and Smith 1998 | Shermann et al. 2008 | Lee and Boulding 2009 | Collin 2001 |

## 5.5 Discussion

### 5.5.1 *Within population genetic structure*

Heterozygote deficiency, displayed by both species C1 and C4, is not rare in marine invertebrates (Addison & Hart 2005) and can be explained by (i) reproduction occurring preferentially between related individuals, (ii) selection against heterozygotes, (iii) the presence of null alleles, (iv) a technical problem when retrieving reads from Miseq data or (v) a Wahlund effect, i.e. the fact that a given sample corresponds to a mixture of differentiated populations. Null alleles are unlikely given that primers were designed within highly conserved coding regions. A technical problem may have occurred when retrieving the alleles due to the use of threshold to keep a sequence, especially if one allele was preferably amplified by PCR. Yet, heterozygote deficiency was also detected in i51, a locus that was not sequenced with the Miseq. Therefore, the heterozygote deficiency does not seem to arise (only) from technical problems. Any selection against heterozygotes should be detected locally (e.g. one or few loci), yet most of the loci displayed high $F_{IS}$ values (5/7 loci with value > 0.2), ruling out this explanation.

Two phenomena that can explain heterozygote deficiency at the genomic level remain to be explored: inbreeding and a Wahlund effect. Interestingly, a study comparing $F_{IS}$ values in marine invertebrate species with different dispersal and reproductive modes found that $F_{IS}$ was generally higher in species where males spawned sperm in comparison with males that directly copulate (Addison & Hart 2005). Dispersal ability (in itself) does not seem to influence $F_{IS}$ values. The authors point out the possibility of within-population genetic structure (e.g. Wahlund effect), but also use the occurrence of higher levels of local inbreeding due to the stochastic nature of broadcasting to explain higher levels of heterozygote deficiency in broadcasters compared with species displaying internal fertilization. In addition, the mating system may also influence the inbreeding coefficient, such as selfing in the sea stars *Parvulastra vivipara* and *P. parvivipara* (Keever *et al.* 2013), as well as in the species *Cryptasterina hystera* (Puritz *et al.* 2012), where heterozygosity was very low or nonexistent.

### 5.5.2 *Influence of life history traits on genetic diversity*

Our results showed that the genetic diversity, investigated using four diversity statistics in 12 loci in C1 and C4, was about 3 to 4 times higher in the broadcasting

species compared to the brooding one. Genetic diversity is expected to be higher in broadcasters than brooders, as it is an increasing function of effective size (Kimura 1969). Indeed, effective sizes theoretically increase with census sizes, everything else being equal, and are thus expected to be higher (i) in most fecund species, which is most often the case in species producing small eggs or species that do not brood (Strathmann 1985), and (ii) in species with wide geographic distributions, which is most often the case in highly dispersing species (Palmer & Strathmann 1981). However, highly fecund species probably display higher variance in reproductive success, which reduces their effective sizes (Hedgecock & Pudovkin 2011). When populations are structured, in ecologically realistic cases (i.e. when there is variance in reproductive success among demes), the effective size at the scale of the species (pooling all demes) is theoretically expected to be reduced (though the simplest unrealistic model predicts the opposite Whitlock 2004). Effective size per deme is always expected to be lower in more structured species. Our finding of a higher diversity within broadcaster populations thus follows theoretical expectations and is congruent with previous surveys comparing species with distinct dispersal abilities due to different developmental modes (Hunt 1993; Hoskin 1997; Foltz 2003; Foltz *et al.* 2004; Lee & Boulding 2009).

Recently, an extensive study investigated the levels of genetic diversity in the transcriptomes of 76 metazoan species (Annex 2; Romiguier *et al.* 2014), including transcriptomic data of *O. longicauda* C1 and C4. These results, at the transcriptomic level, confirmed the signal detected for 12 loci, namely that the genetic diversity was higher in the broadcasting species, though the difference was smaller (1.5 to 2 instead of 3 to 4). Interestingly, the authors showed that levels of genetic diversity were predictable, and were highly correlated to parental investment. Thus, species with high parental investment, brooding abilities and low fecundity displayed lower levels of genetic diversity than species with low parental investment and high fecundity. In addition, this study highlighted the low influence of environmental and historical factors on genetic diversity (Romiguier *et al.* 2014).

### 5.5.3 *Influence of life history traits on genetic structure and connectivity*

Planktotrophic larvae are expected to disperse more than lecithotrophic larvae, because they live longer in the plankton, and in some species it has been shown that they can delay settlement when the environment is not suitable (Hendler 1991).

Similarly, brooders or species laying egg masses are expected to disperse even less than species with lecithotrophic larvae. In a comparison including four species from the snail genus *Littorina*, collected in the same localities, Lee & Boulding (2009) found that the ratio of mean population haplotype diversity in species with planktotrophic larvae relative to brooding species ranged from 1.33 to 8 according to species pairs (using cytochrome b mitochondrial DNA). As shown by the AMOVA analyses (Table 5.10), the proportion of the within populations genetic diversity was lower in the brooding lineage of *Ophioderma longidauda* (64% of the molecular variance) than in the broadcaster lineage (99%), which is in agreement with previous studies in marine taxa (Table 5.10), although relative proportions vary among studies: e.g. 40% in brooder versus 99% in lecithotrophic starfish (Sherman *et al.* 2008), 3.4% in brooders versus 95% in planktotrophic *Cucumaria* sea cucumbers (Arndt & Smith 1998), 23.9-45.7% in direct-developer versus 78-100% in planktotrophic *Crepidula* gastropods (Collin 2001) or 75-95% in *Littorina* species with lecithotrophic larvae versus 97-98% in *Littorina* with planktotrophic larvae in (Lee & Boulding 2009). Thus, the results we obtained for lecithotrophic larvae in *Ophioderma* were similar to the results observed with planktotrophic larvae in other species, where in all cases the percentage of molecular variance within population was higher than 90%, highlighting the fact that even a short PLD can have a great impact on the distribution of genetic diversity.

We observed significant genetic differentiation among populations for both C1 and C4. Finding significant structure for a species with a planktonic larva (in the case of C1) at a moderate geographic scale does not often occur, but it is likely that numerous previously published studies of marine invertebrates could not detect genetic structure due to low sample sizes or exclusive use of $\Phi_{ST}$ statistics (Penant *et al.* 2013). The $F_{ST}$ values however were very distinct among development modes, with almost a 1000-fold difference between brooders and broadcasters. We found lower ratios than Hoffman *et al.* (2011), Barbosa *et al.* (2013), and Puritz *et al.* (2012) who compared species with contrasting developmental modes and/or mating systems. Ratios similar to the ones observed in *O. longicauda* have been reported in marine invertebrates (Hellberg 1996; Teske *et al.* 2007). The brooding species C4 displayed an extremely elevated global $F_{ST}$ of 0.49, at relatively small scale (1-200 km). Other echinoderms which brood or lay egg-masses (with benthic larval development) like the sea stars *Parvulastra exigua* (Lamarck, 1816) and *Asterina gibbosa* displayed similarly high values (Hunt 1993; Baus

*et al.* 2005; Sherman *et al.* 2008; Barbosa *et al.* 2013) and differentiation was revealed among samples separated by a few meters for *P. exigua* (Barbosa *et al.* 2013) and for a brooding sea urchin (Ledoux *et al.* 2012).

Yet, the mode of reproduction is not the single factor influencing genetic structure. Indeed, previous research showed that demographic and biogeographic histories (Ayre & Hughes 2000; Crandall *et al.* 2008; Keever *et al.* 2009; Hart & Marko 2010), larval behavior or survival, selection or habitat specificity (Ayre & Hughes 2000; Miller & Ayre 2008; Ayre *et al.* 2009; Leese *et al.* 2010; Damerau *et al.* 2014) have stronger effects on genetic structure than do larval type and dispersal ability. Nevertheless, population genetic differences among the "*O. longicauda"* species of this study are unlikely to result from something other than life history trait differences (e.g. environmental differences or different evolutionary histories) since we compared two closely related species (C1 and C4) sampled in the exact same locations during a single dive.

The ratio of migrant numbers between the broadcaster (C1) and brooder (C4) (for comparisons ranging between one and 300 km) had a mean value of 50. This ratio must be considered as a very rough theoretical estimate since the standard deviation was 43. Yet, it is the first numerical estimate the influence of larval type on genetic diversity and connectivity. This ratio is sufficient to generate very divergent patterns of geographical distribution of haplotype differences: in brooders, few haplotypes are shared among populations and haplogroups correspond to regions, whereas in the widespread broadcaster C1 most haplotypes are found in each region. Hellberg (1996) compared two coral species with contrasting modes of development and plotted the inferred number of migrants per generation (based on $G_{ST}$ estimates) in both species as a function of population distances. The slope of the linear regression was not significantly different from zero for the species with planktotrophic larvae but for the brooding species, the linear correlation was significant ($r^2 = 0.603$). The estimated number of migrants per generation in the species with planktotrophic larvae was about 20 times that in the brooding lineage for distances of one kilometer, 33 times for 10 km, and 100 times for 3000 km. Although migrant numbers were not explicitly given in most studies, the fact that we obtained similar ratios suggests that the migrant ratios obtained among *Ophioderma longicauda* clusters are not particularly deviant from other species.

When considering the other brooding cluster, C5, we observed the same pattern of extremely high genetic structure, also highlighted by the haplotype networks which displayed a strong phylogeographic structure, completely different from the structure in the broadcaster C1, where most haplotypes are shared among basins. Such a contrast in phylogeographic patterns was found in a comparison between sea urchins from two distinct genera (McMillan *et al.* 1992).

In this study, we compared the connectivity of closely related species with divergent dispersal modes, using samples from the same five locations, for the first time, and found expected trends in relative connectivity and diversity patterns. In addition, we compared the phylogeographic distribution of mitochondrial COI DNA sequences across the whole geographic distribution of the same two species, and for another brooding cluster C5. The large scale comparison and the replicate brooder cluster in the phylogeographic study make this study a reference for understanding the contribution of lecithotrophic larvae to genetic diversity and connectivity.

# Chapitre 6

-

# Transcriptomique comparée entre incubants et dispersants :

# Description des transcriptomes, diversité génétique et gènes sous sélection positive

# Chapitre 6: Transcriptomique comparée entre incubants et dispersants : description des transcriptomes, diversité génétique et gènes sous sélection positive

## 6.1 Résumé

Dans le chapitre précédent, j'ai montré qu'il existait de fortes différences de structure génétique et de connectivité entre les espèces sympatriques, l'incubante et la dispersante, d'*Ophioderma longicauda*. De plus, les 12 marqueurs étudiés ont montré de fortes différences de diversité génétique (nombre d'allèles ; diversités haplotypique et nucléotidique), ceci étant vraisemblablement la conséquence d'une taille efficace réduite chez les incubantes, qui cause une dérive génétique plus forte chez ces dernières. Dans ce chapitre, j'ai réalisé une comparaison des transcriptomes des lignées L1 et L3, afin de décrire les gènes communs chez ces espèces proches, de comparer la diversité génétique (en particulier la proportion de polymorphisme non synonyme) de ces deux espèces au niveau du transcriptome entier et finalement afin de détecter et décrire les gènes évoluant sous sélection positive, en particulier les gènes impliqués dans la reproduction, dont une des espèces présente un caractère fortement dérivé. Ces gènes liés à la reproduction ont été particulièrement étudiés car : 1) l'une des espèces a divergé très récemment des autres et son mode de reproduction est profondément différent du mode de reproduction ancestral prévalent dans le genre *Ophioderma*, et 2) ces gènes peuvent évoluer de façon rapide, sous l'action d'une sélection positive, car les nouveaux (rares) allèles peuvent être avantagés quand différentes espèces dont l'isolement pré-zygotique est incomplet sont en sympatrie (renforcement).

Pour ce faire, des individus des lignées L1 et L3 ont été récoltés en mai 2012 ; les individus L3 incubaient des juvéniles, tandis que les individus L1 avaient les gonades pleines. D'autres individus de la lignée L1 ont également été récoltés à Madère (Portugal) en février 2012, à Sagrès (Portugal) et Marseille (France) en juillet 2012. L'ARN total a été extrait des gonades et d'une partie de bras pour un maximum d'individus. Au total, douze librairies Illumina ont été construites, neuf avec des individus L1 (cinq librairies individuelles et quatre librairies constituées de pools d'individus correspondant à chaque localité) et trois avec des individus L3 (un mâle, une femelle et un pool d'individus).

Le séquençage a été réalisé avec la technologie Illumina (Hi-seq 2000). Les premières étapes ont été réalisées grâce au pipeline utilisé pour le projet Pophyl. Après un tri des reads selon leur longueur et leur qualité, ils ont été assemblés avec les programmes Abyss et Cap3. Les contigs ont ensuite été filtrés selon leur longueur et leur couverture. Les contigs avec un cadre de lecture (ORFs) supérieur à 60 acides aminés ont été filtrés pour la présence de paralogues. La diversité génétique ainsi que l'hérézygotie moyenne ont été calculées. Ensuite, les orthologues entre L1 et L3 ont été trouvés par blast réciproque, sur la base des ORFs trouvés (13'990 L1 et 20'790 L3). Etant donné que le nombre d'orthologues retrouvé était assez faible (4740), de « nouveaux » orthologues ont été reconstruits par mapping des reads L3 sur les ORFs « uniques » de la lignée L1, et vice-versa. Cette étape a permis la récupération de nombreux orthologues, augmentant à 12'229 leur nombre. Au final, 22'123 ORFs L3 et 20'146 ORFs L1 ont été décrits. L'annotation des orthologues a ensuite été réalisée avec Blast2GO, et les gènes sous sélection positive ont été cherchés avec le programme PSG finder.

Sur 12'229 orthologues trouvés, 5925 ont pu être annotés. La majorité des gènes de la catégorie processus biologique étaient impliqués dans les processus métaboliques. Dans la catégorie fonction moléculaire, la majorité des gènes étaient impliqués dans plusieurs fonctions de liaison. Finalement, la majorité de ces gènes étaient localisés, d'après leurs termes de l'ontologie des gènes (Gene Ontology), dans les parties cellulaires ou dans la membrane. La divergence nucléotidique moyenne entre ces deux espèces était de 1%. La diversité génétique (diversité nucléotidique synonyme moyenne ($\pi_S$) et hétérozygotie moyenne) était 1.5 à deux fois plus élevée chez L1 comparé à L3, confirmant la tendance observée sur un petit nombre de marqueurs au chapitre précédent. Cependant, cette mesure est peut-être sous estimée et il serait intéressant de refaire l'analyse de diversité génétique avec la totalité des contigs reconstruits.

L'analyse des gènes évoluant sous sélection positive par PSG finder a détecté 187 fenêtres, correspondant à 41 gènes évoluant sous sélection positive. Parmi ces 41 gènes, 11 ont pu être annotés avec des termes de la Gene Ontology. Près de la moitié de ces gènes étaient impliqués dans le transport ionique, ce dernier faisant partie de la catégorie « établissement de la localisation » au niveau 3 de la Gene Ontology. Cette catégorie ne représente pourtant que 4% de la totalité des orthologues annotés. Parmi

les gènes évoluant sous sélection positive, deux d'entre eux étaient également impliqués dans la reproduction ; plus précisément, il s'agissait de canaux ioniques (TetraKCNG et NHE) spécifiques des spermatozoïdes, impliqués dans la mobilité des spermatozoïdes après activation par des attracteurs chimiques libérés par les oocytes (chimiotaxie). La comparaison de la séquence protéique de ces gènes avec des groupes externes d'échinodermes a indiqué que la sélection positive s'était exercée dans la lignée L3, incubante. Comme la mobilité des spermatozoïdes est un facteur déterminant de la compétition entre spermatozoïdes, ces résultats suggèrent une compétition accrue chez les L3.

Deux hypothèses peuvent expliquer pourquoi ces gènes évoluent sous sélection positive. Premièrement, cela pourrait être du à l'apparition de l'incubation uniquement. En effet, les conditions environnementales autour des oocytes sont alors différentes (oocytes en grande densité, environnement peut-être réduit en oxygène à l'intérieur des bourses). Il est possible que ce canal ionique ait subi plusieurs changements adaptatifs lui permettant de s'adapter pour conserver son efficacité de fécondation dans ce nouvel environnement. Une autre cause théorique d'évolution darwinienne des molécules de reconnaissance entre gamètes est l'évitement de la polyspermie (dont le coût est important dû à la perte de l'ovule), particulièrement plausible dans ces situations de forte densité.

Deuxièmement, L1 et L3 ont pu diverger de manière allopatrique, ne développant pas de mécanismes d'isolement pré-zygotique. Par la suite, les espèces ont pu échanger des gènes par contact secondaire, et à ce moment là les L3 auraient développé des spermatozoïdes plus rapides et donc plus compétitifs afin d'éviter l'hybridation. En effet, Les oocytes L3 devraient être plus gros, ou se trouver en un plus faible nombre que chez L1, afin de permettre un développement des juvéniles dans les bourses. Dès lors, il existe un investissement par oocyte plus élevé chez les L3. Ainsi, il existe une pression de sélection contre l'hybridation entre L3 et L1, ce qui expliquerait pourquoi des spermatozoïdes plus compétitifs auraient été sélectionnés. Cette dernière hypothèse, qui entre dans le cadre de la théorie du « renforcement (de l'isolement reproducteur)», semble très plausible dans le cas d'*Ophioderma*. En effet, j'ai montré dans le chapitre 3 que le scénario historique le plus vraisemblable dans le complexe *Ophioderma* incluait un événement d'hybridation entre les L1 et les L3. Ces populations actuelles, résultant

très probablement d'un événement d'hybridation passé, ont la particularité de montrer un haplotype mitochondrial (hérité de la lignée maternelle) très proches des L3 (3 mutations), et des allèles des marqueurs nucléaires proches de L1, appuyant l'hypothèse de l'hybridation entre spermatozoïdes L1 et oocytes L3.

Plusieurs études sur les invertébrés marins ont révélé une évolution rapide des protéines de reconnaissance gamétique (p. ex. bindine ; lysine ; EBR1 ; suREJ), cependant ces protéines n'ont pas montré de sélection positive chez *Ophioderma longicauda* L1 et L3. Notre étude est la première qui révèle, chez les invertébrés marins, un signal de sélection positive dans la compétition des spermatozoïdes. Ce mécanisme serait donc impliqué dans l'isolement reproducteur complet entre L1 et L3. De plus, j'ai montré qu'il existait des différences de séquence (gap de 4 codons et 9 mutations non-synonymes) entre les récepteurs aux attracteurs chimiques des L1 et des L3 (molécules exprimées sur les ovocytes), suggérant qu'il existerait des attracteurs chimiques différents entre L1 et L3. La chimiotaxie semble être un mécanisme de reconnaissance des espèces spécifique chez les ophiures, puisqu'il a été montré que ces attracteurs chimiques étaient spécifiques au niveau de l'espèce chez 10 espèces d'ophiures (sur 15 étudiées), contrairement aux autres échinodermes où il a été montré que la chimiotaxie n'était pas spécifique au niveau de l'espèce, mais plutôt au niveau de la famille ou de l'ordre, qui représentent des niveaux taxonomiques beaucoup plus élevés. La chimiotaxie chez les ophiures pourrait ainsi être un nouveau mécanisme menant à l'isolement reproducteur chez les échinodermes, et son étude mériterait être développée.

Cet article est intitulé “Comparative transcriptomics in the species complex *Ophioderma longicauda* reveals strong positive selection on genes involved in sperm competition.” Ses auteurs sont Alexandra Anh-Thu Weber, Laurent Abi-Rached, Nicolas Galtier, Aurélien Bernard, Olivier Bouchez, Juan I. Montoya-Burgos et Anne Chenuil. Il est prévu de le soumettre à Molecular Biology and Evolution après relecture et modifications par tous les auteurs. Certaines parties méthodologiques et illustrations ne figureront pas dans l’article à soumettre, mais sont présentées dans cette thèse par souci de clarté.

## 6.2 Introduction

Closely related species are particularly interesting models to study the mechanisms leading to reproductive isolation, since divergence between species is recent. Therefore, it is expected that global genetic divergence between the species is low, whereas genes involved in speciation should be more divergent. There are two categories of mechanism that prevent gene flow between different species. First, there is post-zygotic isolation, occurring after the fusion of gametes, which leads to the death or sterility of newly formed embryos due to genomic incompatibilities. The second category of mechanisms includes pre-zygotic factors, namely each factor involved before the fusion of sperm and egg. Those factors can include temporal pre-zygotic isolation, where reproduction periods are not overlapping. This delay can be as short as 30 minutes, as seen for the species *Ophioderma rubicunda* and *Ophioderma squamosissia* (Hagman & Vize 2003). Furthermore, there can be active avoidance of non-conspecific males and females, via chemical signals, to avoid mating (Hendler 1991; Bickford *et al.* 2007; Mercier & Hamel 2009). Finally, when closely related species occur in sympatry, it has been shown that several genes involved in gamete recognition may evolve under positive selection (see Vacquier & Swanson 2011; Lessios 2011 for reviews).

The most famous examples are the sperm bindin gene and its egg receptor EBR1 mainly studied in sea urchins, and the sperm protein lysin and its egg receptor VERL mostly studied in gastropods. The bindin is a protein involved in recognition between the sperm and the egg vitelline envelope, as well as in the fusion of sperm and egg membranes. In contrast, the lysin is only involved in recognition and dissolution of the vitelline envelope, the fusion of sperm and egg membrane being mediated by another protein (sp18) (Vacquier & Swanson 2011; Lessios 2011). Positive selection on bindin was found between sea urchin species of the genus *Strongylocentrotus, Paracentrotus*,

*Heliocidaris* and *Echinometra* (see Lessios 2011 for a review). In addition, positive selection on bindin was found in populations of the sea star *Patiria miniata* (Sunday & Hart 2013) as well as on the bindin receptor EBR1 (Hart 2013).

Furthermore, positive selection was also found on proteins involved at the level of egg-jelly. Thus, the carbohydrate-recognition domains of sperm receptor for egg-jelly (suREJ) were found to evolve under positive selection in six species of sea urchins (Mah *et al.* 2005). This receptor interacts with sulfated polysaccharides to induce the acrosome reaction, polysaccharides which also display structural changes between closely related species of *Strongylocentrotus* sea urchins (Biermann *et al.* 2004). Therefore, it appears that several mechanisms occur in sea urchins that lead to the same outcome, reproductive isolation.

The study of closely related species is particularly interesting when they display divergent life history strategies. For instance, species with contrasting levels of parental investment should display differences in effective sizes. Since genetic drift is stronger in populations with small effective size, it is expected, everything else being equal, that species with higher parental investment and fewer offspring show lower genetic diversity, a prediction which was confirmed in metazoans (Romiguier *et al.* 2014).

The brittle star species complex *Ophioderma longicauda* is composed of six divergent mitochondrial lineages (L1-L6) displaying contrasted reproductive strategies. The lineages L1-L5-L6 are broadcast spawners and reproduce via a lecithotrophic larva, whereas the lineages L2-L3-L4 are internal brooders (Stöhr *et al.* 2009; Boissin *et al.* 2011). Previous studies showed that L1 and L3 are different biological species, since they do not reproduce at the same time of the year, and display genetic, ecologic and morphological differences (Chapter 2; Weber *et al.* 2014). In addition, differences in thermotolerance were also detected (Chapter 4; Weber *et al.* 2013). Those closely related species with contrasted reproductive strategies provide an interesting model to carry out transcriptomic comparative investigations. In this study, we performed L1 and L3 de novo transcriptome sequencing and assembly, in order to infer the level of genetic divergence and genetic diversity between those two species. In addition, we looked for genes evolving under positive selection, in particular genes involved in reproduction that may have played a role in the reproductive isolation of L1 and L3.

## 6.3 Material and methods

### 6.3.1 *Individual sampling, lineage typing and RNA extraction*

Sampling was performed in four localities across the Mediterranean and the Atlantic. Individuals were sampled in Madeira (February 2012; N 32° 45' 5.774'' / O 17° 13' 26.065''); in Crete, Greece (May 2012; N 35° 24' 8.37''/ E 25° 2' 2.62''); in Sagres, Portugal (July 2012; N 37° 0' 28.837'' / O 8° 55' 27.864'') and sampled in Marseilles, France (July 2012; N 43° 12' 54.765'' / E 5° 19' 35.253''). For each individual, gonads or embryos (when present) and a piece of arm were preserved in RNA later and kept at -20°C. In addition, a piece of arm was preserved in EtOH 95% for DNA extraction. As no morphological character allows lineage identification except brooding, lineage typing was performed by extracting DNA following a Chelex protocol (Walsh *et al.* 1991), amplifying by PCR a fragment of COI gene (forward primer: 5' CTT GCA GGA GGA GGA GAY CC 3'; reverse primer: 5' AGT ATA AGC GTC TGG GTA GTC 3') and sequencing it with an automated sequencer (LGC Genomics, Germany). Individuals were assigned to a specific lineage by comparing their sequence to the ones used by Boissin *et al*. (2011) to describe the species complex.

Each individual from Madeira, Sagres and Marseilles belonged to the broadcasting lineage L1, whereas the individuals of Crete belonged either to the lineage L1 or to the lineage L3. RNA extractions were performed all together using a combination of standard guanidium thiocyanate-phenol-chloroform (GTPC) RNA extraction followed by an extra purification step with a RNeasy kit for RNA extractions (Qiagen). Both methods were used in combination to maximize the quantity and the purity of RNA, as shown efficient for non-model species by Gayral *et al.* (2011). RNA quality and quantity were assessed using a spectrophotometer cell and RNA integrity was measured with a Bioanalyzer (Agilent). Given the quality of the different samples, pools of individuals were performed using the same quantity of RNA per individual.

Six individuals were used for the L1 population of Madeira, ten individuals of each lineage L1 and L3 were used for the L1 and L3 populations of Crete, eight individuals were used for the L1 population of Sagres and six individuals were used for the L1 population of Marseilles. In addition, unique individuals were kept for individual libraries, namely one individual of L1 per population in Madeira, Sagres and Marseilles;

and two individuals of L1 and two individuals of L3 from Crete. Extracted RNAs were stored at -80°C until further manipulation.

### 6.3.2 *Illumina sequencing, assembly and SNP calling*

Librairies preparation and Illumina sequencing were performed by the genomic platform Genotoul (Toulouse). Briefly, quantity, purity and integrity of RNA were checked at the arrival of samples at the platform using a Nanodrop (Thermoscientific) and a Bioanalyzer (Agilent). Librairies were constructed using TruSeq library preparation kit (Illumina) following the manufacturer's instructions. Sequencing was performed with a Hiseq 2000 Illumina sequencer (Illumina). The complete bioinformatics pipeline used in this study is described in Gayral *et al.* (2013). Briefly, raw reads were filtered for length (min 60) and quality (min Phred score: 30). De novo assembly was performed using Abyss and Cap3, according to the method B described in Cahais *et al.* (2012). Illumina reads were mapped to the contigs using BWA. Raw contigs were filtered for length (minimum 200 nt) and coverage (2.5x per individual). ORFs were predicted from the filtered contigs, looking for the longest ORF and applying a filter for paralogous genes as described in Gayral *et al.* (2013). Genotypes were called for each position of each contig. For each species, the per-site synonymous ($\pi_S$) and non-synonymous ($\pi_N$) diversities were calculated, as well as the per-individual heterozygosity.

### 6.3.3 *Research and annotation of orthologous genes*

Orthologous genes were retrieved using a reciprocal best Blast hit approach using the ORFs per species (see Extended Methods for details). Annotation step was performed with a local blastp against the NR (non-redundant) database of NCBI using an e-value cutoff of $10e^{-5}$. The top 20 hits were extracted and loaded in the Blast2GO program for annotation (Conesa *et al.* 2005). Mapping, annotation and slim ontology were performed with Blast2GO using default parameters, except for the annotation cutoff parameter that was set to 45. GO categories were described using the level 3 of slim ontology.

### 6.3.4 *Positively selected genes and gamete recognition genes*

Genes evolving under positive selection were investigated using the new program PSG finder (Tuberosa *et al*. in prep), which allows pairwise comparison of orthologous genes. Sliding windows of different sizes are used to accurately detect parts of genes displaying a signal of positive selection (dN/dS ratio > 1).

In addition, several genes involved in reproduction (in sperm chemotaxis or in sperm-egg interaction) and known to evolve under positive selection in other species were specifically searched. Genes cited hereafter were retrieved from the *Strongylocentrotus purpuratus* genomic database, and blasted against the database L1 (contigs L1) and database L3 (contigs L3). Those genes were the bindin and its receptor EBR1, as well as the sperm receptor for egg jelly, suREJ, involved in activation of the acrosome reaction. In addition, homologs to *Strongylocentrotus purpuratus* sperm activating peptides (speract) and its receptor (speract receptor), known to induce sperm chemotaxis, were searched by BLAST in the L1 and L3 transcriptomes. Finally, sperm-specific calcium channels (Catsper 1-4) known to evolve under positive selection in rodents (Podlaha *et al.* 2005) and primates (Podlaha & Zhang 2003) were also searched by BLAST in L1 and L3 transcriptomes.

## 6.4 Results

### 6.4.1 *Assembly statistics*

The Illumina sequencing provided about 250 millions of raw reads, which were trimmed for length and quality. Assembly yielded 523,565 and 259,916 raw contigs for L1 and L3, respectively (Table 6.1). After filtering for length and coverage, 110,225 L1 and 118,043 L3 filtered contigs were obtained. Open reading frame prediction yielded 13,990 and 20,800 ORFs for L1 and L3, respectively. After ORF reconstruction based on “unique” genes of each species (see Extended Methods), we obtained 20,146 and 22,123 ORFs for L1 and L3, respectively, of which 12,229 were clearly orthologous genes. The average nucleotide divergence of those orthologs at the nucleotide level was about 1%.

**Table 6.1**: Transcriptome assembly statistics generated in study.

| | L1 (5 individuals) | L3 (2 individuals) |
|---|---|---|
| Raw reads | 164,464,891 | 87,959,550 |
| Raw contigs | 523,565 | 259,916 |
| filtered contigs | 110,225 | 118,043 |
| ORFs | 13,990 | 20,790 |
| reconstructed ORFs | 20,146 | 22,123 |
| orthologous genes | 12,229 | 12,229 |
| annotated orthologous genes | 5,925 | 5,925 |

### 6.4.2 *Annotation of orthologous genes*

On the 12,229 identified orthologs, 5,925 (48%) could be successfully annotated with GO terms. For biological process, primary metabolic process was the most abundant Go term (10.8%), followed by organic-substance metabolic process (10.6%) and cellular metabolic process (8.1%) (Fig. 6.1). For molecular function, genes coding for binding were highly represented (67.8%) as well as transferase (10.5%) and hydrolase (12.9%) activities (Fig. 6.1). For cellular component, major categories were cell part (38.2%), membrane-bounded organelle (27.2%) and non-membrane-bounded organelle (11.2%).

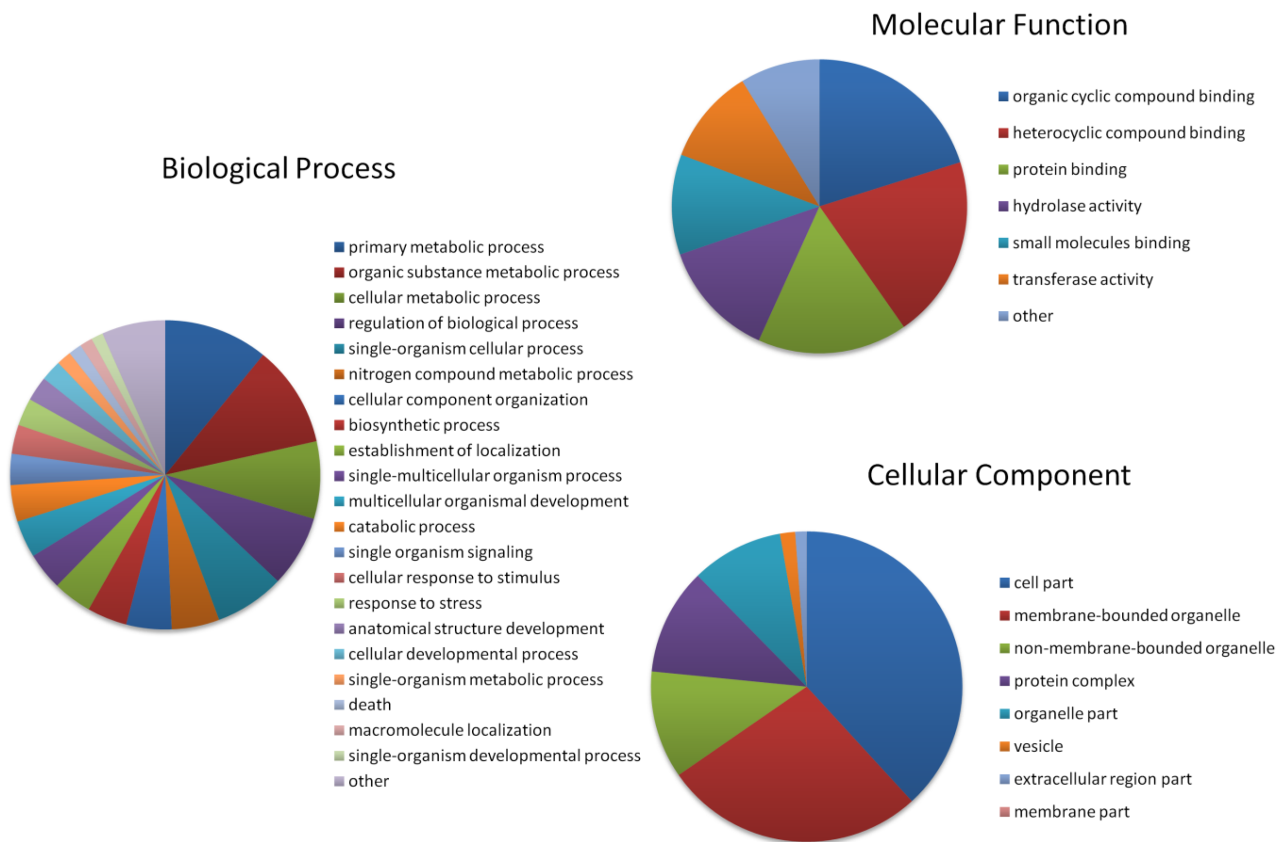



**Figure 6.1**: Results of Blast2Go annotation (Level 3) of 5925 orthologs of *Ophioderma longicauda* L1 and L3 species. The different categories molecular function, biological process and cellular component are displayed.

### 6.4.3 *Comparison of genetic diversity between brooders and broadcasters*

1,388 L1 and 6,700 L3 genes were used to investigate genetic diversity (Table 6.2). 7,607 and 18,408 SNPs were studied in L1 and L3, respectively. The average per-site synonymous diversity ($\pi_S$) was about twice higher in L1 compared to L3, resulting in a higher $\pi_N/\pi_S$ ratio in L3 (Table 6.2). In addition, the average heterozygosity was about 1.5 higher in L1 compared to L3.

**Table 6.2**: Comparison of transcriptomic genetic diversity between broadcasting L1 and brooding L3 species.

| | L1 | L3 | diversity ratio L1/L3 |
|---|---|---|---|
| Number of contigs used | 1,388 | 6,700 | |
| Number of SNPs | 7,607 | 18,408 | |
| average $\pi_S$ | 0,017 | 0,008 | 2,12 |
| average $\pi_N/\pi_S$ ratio | 0,181 | 0,269 | 0,67 |
| average heterozygosity | $6{,}27\text{x}10^{-3}$ | $4{,}16\text{x}10^{-3}$ | 1,51 |

### 6.4.4 *Positive selection detected in proteins involved in sperm competition*

When comparing 12,229 orthologous genes common between L1 and L3, 187 windows evolving under positive selection were found, corresponding to 41 distinct genes. Among them, 11 could be successfully annotated with GO terms (Table 6.3). Interestingly, about half of the genes (5/11) were involved in ion transport. Among them, two genes were particularly well identified, the sperm-specific sodium proton exchanger (NHE) and the tetrameric potassium selective cyclic nucleotide gated channel (TetraKCNG) (Table 6.3). To infer in which species positive selection occurred, specific primer pairs targeting positively selected regions were designed to PCR amplify those genes in closely related species. Unfortunately, numerous trials performed on *Ophioderma* outgroups (*O. teres, O. phoenium* and *O. cinereum*) failed. Nevertheless, orthologs of these genes were available in genomic databases for the echinoderm species *Strongylocentrotus purpuratus* (sea urchin), *Patiria miniata* (sea star) and *Ophionotus victoriae* (brittle star). In the case of TetraKCNG, only the sequence of *S. purpuratus* was available. On the 14 non-synonymous mutations found between L1 and L3, one was only in L1, 5 were only in L3 and 8 were different in the three species. In the case of NHE, two regions of the gene evolved under positive selection. In the first small region, all 5 mutations occurred newly in L3 compared to all the other species, *O. longicauda* L1, *O. victoriae, P. miniata* and *S. purpuratus* (Fig. 6.2). In the second part of

the gene, on the 12 non-synonymous mutations found between L1 and L3, 7 occurred only in L3, 5 were different in each species and none were present only in L1. These results strongly suggest that positive selection occurred in the brooding species L3.

NHE and TetraKCNG are well studied in the sea urchin *Strongylocentrotus purpuratus*, and are involved in sperm mobility after activation of sperm with chemoattractants released by the eggs (Fig 6.3). Since different chemoattractants may differentially attract sperm to the egg, we looked for speract homologs in *Ophioderma*. This research was unsuccessful, most likely due to the small size (10 amino acids) of this peptide (Hirohashi et al 2008). As ligand and receptor usually coevolve, the sequences of the “speract receptor” orthologs in *Ophioderma* were compared. We found the corresponding sequences in L1 and L3, but they did not display a significant signal of positive selection (Table 6.4). Interestingly, however, on the 10 non-synonymous mutations found in this gene, 9 occurred in the extracellular (ligand-binding) part, whereas none were detected in the transmembrane part and only one was found in the intracellular part. In addition, a four-codon gap was found in the ligand-binding part of L3, suggesting that L1 and L3 might have different chemoattractants. Finally, we found homologs of gamete recognition proteins usually described in the literature (Bindin, EBR1, suREJ, Catsper 1-4), yet none of these proteins was found to evolve under positive selection (Table 6.4).

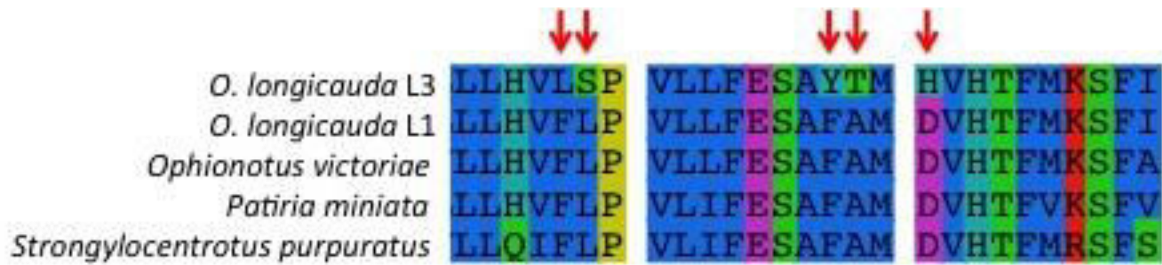



**Figure 6.3**: Positively selected region of the gene NHE in *O. longicauda* L1, L3 and the brittle star *Ophionotus victoriae*, the sea star *Patiria miniata* and the sea urchin *Strongylocentrotus purpuratus.* The positively selected amino-acids are displayed with red arrows, showing that positive selection occurred in the brooding species L3.

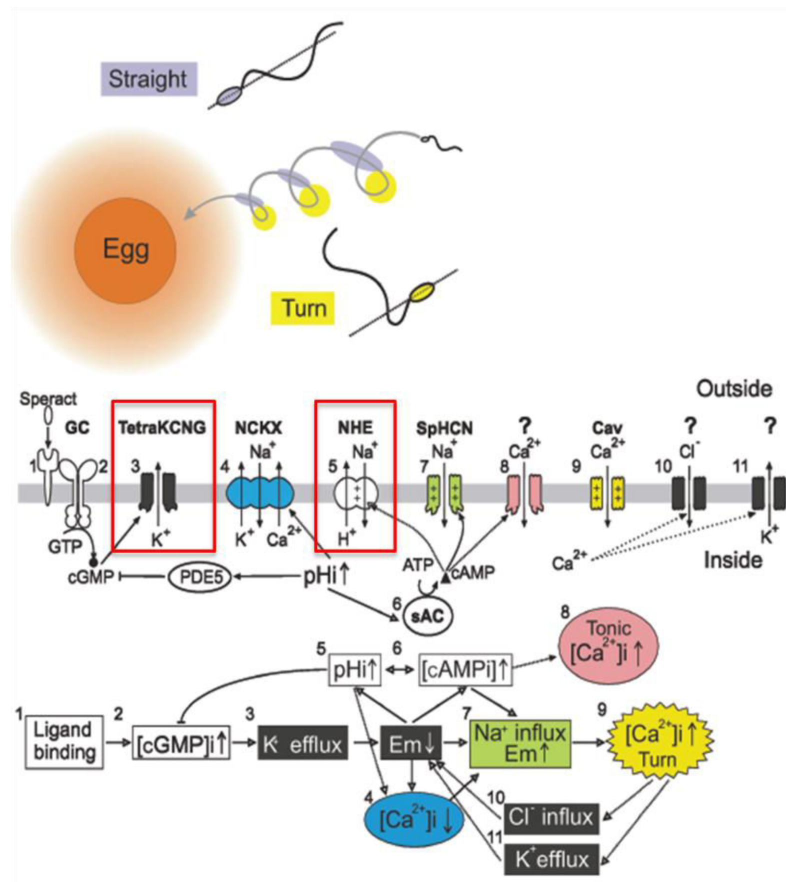



**Figure 6.4**: Protein involved in speract signalling model (modified from Darszon et al, 2008). The genes evolving under positive selection in *O. longicauda,* NHE and TetraKCNG, are highlighted. The sperm activating peptides (SAP) speract, released by the eggs of *Strongylocentrotus purpuratus,* bind to their receptor (speract receptor) and activates sperm motility through changes in intracellular cGMP, intracelluar pH, intracellular cAMP and intracellular calcium concentration. (see Darszon et al, 2008 for description of the complete signalling model).

**Table 6.3:** Positively selected genes between *O. longicauda* L1 and L3 that display GO term annotations. Gene name, matching organism after blast search and available GO terms are displayed.

| Contig L1 | Contig L3 | Gene name | Organism | Full gene name | Biological Process | Molecular Function | Cellular component |
|---|---|---|---|---|---|---|---|
| Contig 157183 | Contig 10800 | **NHE** | *Strongylocentrotus purpuratus* | sperm specific sodium-proton exchanger | **ion transport** | ion channel activity | cell |
| Contig 5262 | Contig 45826 | **TetraKCNG** | *Strongylocentrotus purpuratus* | tetrameric potassium-selective cyclic nucleotide gated channel | **ion transport** | ion channel activity | cell |
| Contig 132964 | Contig751 58 | LOC576432 | *Strongylocentrotus purpuratus* | PREDICTED: excitatory amino acid transporter 3-like | **ion transport**; cellular component organization | transporter activity | cell; plasma membrane |
| Contig 22861 | Contig 23039 | LOC584316 | *Strongylocentrotus purpuratus* | PREDICTED: disrupted in renal carcinoma protein 2 homolog | **ion transport** | ion channel activity | cell |
| Contig 141406 | Contig 7832 | LOC585510 | *Strongylocentrotus purpuratus* | PREDICTED: solute carrier family 13 member 5-like | **ion transport** | transporter activity | plasma membrane |
| Contig 17797 | Contig 45476 | / | *Crassostrea gigas* | Thymidine phosphorylase | nucleobase containing compound metabolic process | transferase activity | no |
| Contig 168194 | Contig 2036 | LOC100376819 | *Saccoglossus kowalevskii* | PREDICTED: arylacetamide deacetylase (esterase)-like | metabolic process | hydrolase activity | no |
| Contig_L1 26484 | Contig 26484 | / | *Crassostrea gigas* | Tripartite motif-containing protein 59 | no | binding; peptidase activity | no |
| Contig_L1 21879 | Contig 21879 | / | *Homo sapiens* | AChain A, Structural Studies Of Protein Tyrosine Phosphatase Beta Catalytic Domain In Complex With Inhibitors | response to external stimulus; signal transduction | phosphoprotei n phosphate activity | plasma membrane |
| Contig_L1 32532 | Contig 32532 | LOC100566073 | *Anolis carolinensis* | PREDICTED: zinc finger protein 167-like | transcription DNA dependant | DNA binding | nucleus |
| Contig_L1 2308 | Contig 2308 | LOC581462 | *Strongylocentrotus purpuratus* | PREDICTED: nephrin-like | multicellular organismal developpement; cell differentiation | protein binding | plasma membrane |

**Table 6.4**: List of known genes involved in sperm chemotaxis and gamete recognition in *Strongylocentrotus purpuratus*, and their occurrence in *Ophioderma longicauda* L1 and L3 transcriptomes. Sp = *Strongylocentrotus purpuratus*. *: but gene small (10 AA) may explain why research was unsuccessful (see text for details)

| Gene | function | hit blast against Sp | BLAST matching Sp [%ID and %coverage] | divergence L1-L3 [nt/protein %] | positively selected |
|---|---|---|---|---|---|
| **TetraKCNG** | sperm-specific potassium channel involved in sperm motility | L1 = yes<br>L3 = yes | 50 - 74 | **[2/4.5]** | **yes** |
| **NHE** | sperm-specific sodium-proton exchanger involved in sperm motility | L1 = yes<br>L3 = yes | 57 - 95 | **[2.1/4.7]** | **yes** |
| speract | egg chemoattractant | L1 = no*<br>L3 = no* | no hits | - | - |
| speract receptor | receptor to egg chemoattractant and activation of sperm motility | L1 = yes<br>L3 = yes | 57 - 93 | [1/1.4] | No |
| Catsper1-4 | sperm specific calcium channel involved in sperm motility | L1 = yes<br>L3 = yes | 1: 56 – 96<br>2: 58 - 97<br>3: 61 - 96<br>4: 41 – 74 (4: against *Ciona intestinalis*) | 1: [0.8/1.1]<br>2: [1.2/0.8]<br>3: [0.1/0.3]<br>4: [0.7/1.0] | No |
| suREJ | sperm receptor for egg jelly, involved in activation of acrosome reaction | L1 = yes<br>L3 = yes | 42 - 83 | [1.2/2] | No |
| bindin | sperm protein involved in attachment with egg vitelline envelope and fusion of sperm-egg membranes | L1 = yes<br>L3 = yes | 62 -20 (against *Lytechnius variegatus*) | [0.5/1] | No |
| EBR1 | egg bindin receptor | L1 = no<br>L3 = yes | - | - | - |

## 6.5 Discussion

### 6.5.1 *Genetic diversity is lower in brooders*

In this study, we sequenced and assembled de novo transcriptomes of two brittle star species belonging to the species complex *Ophioderma longicauda*. Until recently, only few transcriptome researches have focused on brittle stars (e.g. Vaughn *et al.* 2012; Burns *et al.* 2013) except in the field of regeneration (Burns *et al.* 2012; Czarkwiani *et al.* 2013; Purushothaman *et al.* 2015). Yet, increasing genomic or transcriptomic data are being available thanks to the advances of next generation sequencing technologies. For instance, O'Hara *et al.* (2014) published a phylogenomic study of brittle stars using the transcriptomes of 61 species, and two brittle star genome projects are ongoing for the species *Ophiothrix spiculata* and *Amphiura filiformis* (www.echinobase.org).

Our results showed that genetic diversity, a central metric in species response to environmental disturbance, was about 1.5 to 2 times higher in L1 than in L3 at the transcriptomic level. The level of genetic diversity is probably underestimated in L1, since the most polymorphic alleles might have been discarded at the mapping step as stated earlier. In addition, a recent comparative study between sympatric populations of *O. longicauda* L1 and L3 based on 12 genetic markers showed that the genetic diversity was 3-4 times higher in broadcasters than in brooders (see Chapter 5), suggesting that the genetic diversity might actually be underestimated in L1.

Since the two species are recently diverged sister species (about 500,000 years ago, see chapter 3) and were collected in sympatry, historical or environmental factors are unlikely to explain these differences in genetic diversity. The only difference between the two species is the reproductive strategy, with a brooding and a broadcasting species. In fact, a recent study on genetic diversity including transcriptomic data of 76 metazoan species (including *O. longicauda* L1 and L3), showed that the main determinant of genetic diversity is parental investment (Romiguier *et al.* 2014). In particular, the species called K-strategists (high longevity, high parental investment, low number of large offspring) display lower genetic diversity than r-strategists (low longevity, low parental investment, high number of small offspring). Those strategies highlight different responses of species to environmental disturbances. In general, K-strategists are more adapted to stable environments, in which high

investment in few number of offspring is evolutionary advantageous, as risks of strong population bottlenecks are low (Poulin & Feral 1996). In contrast, r-strategists are more adapted to disturbed environments, because the probability that some of the offspring do not fall in the deleterious conditions is higher when numerous offspring are produced. In addition, they can recover from strong bottlenecks thanks to their high level of genetic diversity. In this context, it is likely that the brooding *O. longicauda* L3 would be more impacted by rapid environmental changes than the broadcasting *O. longicauda* L1 in the future.

### 6.5.2 *Mechanisms of prezygotic isolation*

In this study, we were interested in genes displaying positive Darwinian selection involved in reproduction, because they may have played a role in reproductive isolation of *O. longicauda* brooding and broadcasting species. There are several steps leading to fertilization in free-spawning invertebrates: (1) sperm chemotaxis, which is the activation and attraction of sperm mediated by chemoattractants released by the egg, (2) interaction of sperm with the egg jelly, which induces the acrosome reaction, (3) binding of sperm to egg vitelline envelope, (4) penetration of sperm through the egg envelope and (5) fusion of gamete membranes (Vacquier 1998).

So far, positively selected proteins contributing to prezygotic isolation were involved in sperm-egg interactions, whether at the level of egg jelly (2) with the sperm receptors suREJ (Mah *et al.* 2005) or at the level of vitelline layer binding and penetration (3-4) with the sperm protein bindin (see Vacquier & Swanson 2011; Lessios 2011; Evans & Sherman 2013 for reviews). Surprisingly, none of these proteins were found to evolve under positive selection in *Ophioderma longicauda* (Table 6.4). Rather, positive selection was found acting on two ion channels involved in sperm activation and motility (chemotaxis). Interestingly, those genes have never been described as being positively selected between closely related species so far.

Those two sperm-specific channels (TetraKCNG and NHE) and their involvement in sperm activation have been extensively studied in sea urchins (Galindo *et al.* 2007; Beltrán *et al.* 2007; Darszon *et al.* 2008). Briefly, the binding of chemoattractants released by the egg to guanylate cyclase receptors induces a signaling pathway that includes changes in polarization of membrane potential, several ion fluxes, changes in

intracellular pH and finally an increase in intracellular calcium which activates the movement of sperm (see Darszon *et al*. 2008 for the complete signaling model). In addition, the Na+/H+ exchanger (NHE) is also involved in the signaling of the acrosome reaction (Beltrán *et al.* 2007). NHE is essential for sperm motility, as shown with NHE-null male mice which displayed severely diminished sperm motility (Wang *et al.* 2003). The exact mechanism explaining how mutations at the proteic level influence the passage of ions remains unknown, but we hypothesize that ions should be passing through the channels faster, and in this way increasing the movement and thus speed of sperm.

Sperm motility is an important component of sperm competition (Birkhead *et al.* 1999; Levitan 2000; Gage *et al.* 2004; Snook 2005). In fact, positive selection was also found to occur in other sperm-specific ion channels, Catsper (1-4), which are involved in sperm motility in primates and rodents (Podlaha & Zhang 2003; Podlaha *et al.* 2005). These four channels, also found in *O. longicauda*, did not display signals of positive selection (Table 6.4), contrary to tetra-KCNG and NHE.

### 6.5.3 *Adaptation to brooding and/or reinforcement explains the positive selection observed in Ophioderma longicauda reproduction genes?*

Two hypotheses may explain why those genes evolved under positive selection. First, since those genes are not actually involved in sperm-egg interaction, they may have evolved under positive selection as an adaptation to incubation only. Indeed, changes in the environmental conditions surrounding the eggs occurred along with the evolution of brooding. The eggs are gathered into pouches (the bursae), and are therefore present at high density in an environment possibly reduced in oxygen (Strathmann & Chaffee 1984). According to Strathmann & Chaffee (1984), oxygen is a limiting factor in many situations of brooding or in marine species providing brood protection, and the smaller size of such species with respect to their close relatives could be explained as an adaptation to oxygen availability (Strathmann *et al.* 1984). If oxygen levels are lower (or limiting) in the bursae, sperm with enhanced motility might have been positively selected to reach the eggs faster, or despite those conditions, when entering the bursae.

Alternatively, but still related to brooding, it is possible that sperm might have been positively selected to avoid polypsermy, a lethal event occurring if several sperm fertilize the same oocyte, which might occur more often if sperm are present at high density (e.g. Dale & Monroy 1981; Wong & Wessel 2004).

Our second hypothesis, which we believe more likely, is that more competitive sperm may have been selected in the brooding species to avoid hybridization with the broadcasting species and loss of gametes, which corresponds to the reinforcement hypothesis (Lessios 2011). According to this hypothesis, when two species diverge in allopatry, their genomes become incompatible whereas their fertilization systems may remain compatible. If hybridization occurs after secondary contact, hybrid embryos may have reduced fitness or may not develop at all. Therefore, to avoid wastage of gametes, proteins that prevent heterospecific fertilization would be positively selected.

In the case of *Ophioderma*, interestingly, hybridization did not seem to lead to sterility or even lethality, since populations resulting from putative past hybridization events are currently present in Tunisia (chapter 3). Those populations display mitochondrial haplotypes similar to L3 (only 3 mutations), whereas other nuclear markers are closer from the alleles observed in L1, suggesting that L1 sperm fertilized L3 oocytes. Therefore, the mechanism explaining why positive selection was detected in in L3 but not in L1 could be the highest cost of an oocyte in the brooding species (higher parental investment). This mechanism in L3 (maybe also in L1 though undetected) would most likely have completed reproductive isolation between L1 and L3.

### 6.5.4 *Sperm chemotaxis and sperm competition: a common mechanism of species recognition in brittle stars?*

Our results showed for the first time in Echinoderms that sperm competition potentially acts as a mechanism of prezygotic isolation between closely related species. Although we cannot discriminate between the hypothesis of positive selection being only due to the appearance of incubation, and the hypothesis of reinforcement, we believe that the latter one is the most likely. In fact, positive selection may also act at the very beginning of chemotaxis, since different chemoattractants and related receptors might occur between the brooding and broadcasting species. Indeed, we found that the guanylate cyclase receptor of *O. longicauda* (i.e. the receptor, expressed in sperm, of the

egg-expressed chemotaxis molecules) L3 displayed 9 non-synonymous mutations and a four-codons gap in its extracellular (ligand-binding) domain, whereas the transmembrane and intracellular domains remained conserved between L1 and L3. Although not displaying a significant signal of positive selection, those receptors may reflect the existence of different chemoattractants between L1 and L3. It was shown that sperm chemotaxis (i.e. different chemoattractants) plays an important role in sexual selection in the mussel *Mytilus galloprovincialis* (Evans *et al.* 2012), yet its role in species recognition is less obvious. Interestingly, sperm chemotaxis seems very species specific in brittle stars compared to other echinoderm classes. Indeed, sperm of the sea urchin *Lytechinus pictus* could successfully be activated by speract chemoattractants of the species *Strongyloncentrotus purpuratus* (Guerrero *et al.* 2010). More generally, sperm chemotaxis was shown to be order-specific in 17 species of sea urchins, among five tested orders (Suzuki *et al.* 1982). It appeared to be mostly family-specific in 19 species of holothurians (Miller 1997), as well as in six sea star species (Nakachi *et al.* 2006). In contrast, sperm chemotaxis was species-specific in 10 out of 15 species of brittle stars (Miller 1997), suggesting that chemoattractants are mostly species-specific in brittle stars. In particular, sperm chemotaxis specificity allowed separating two unidentified species belonging to the *Macrophiothrix* genus. Therefore, mechanisms leading to reproductive isolation most likely occur during chemotaxis in brittle stars.

In this study, we compared the transcriptomes of two closely related species of brittle stars. We showed that despite very low average genetic divergence, differences in genetic diversity were striking, those differences being due to the contrasted reproductive strategies of the two species. In addition, we showed for the first time in echinoderms positive selection on genes involved in sperm competition, and not on gamete recognition proteins as previously described in other marine invertebrates. These results describe a putative new type of pre-zygotic isolation, occurring before the contact of sperm with egg jelly or vitelline envelope. Research on brittle star sperm-activating peptides, associated receptor and, more generally, genes involved in sperm activation should be pursued, as they might be involved in numerous events of speciation in this particular class of echinoderms.

## 6.6 Extended Methods

Orthologous genes were retrieved using a reciprocal best BLAST hit approach using the ORFs per species. First, ORFs of each lineage were translated in protein to create a protein database. Then, blastp were performed using the 13,990 L1 ORFs as query and the 20,790 L3 ORFs as target, and vice versa with an e-value cutoff of $10e^{-5}$. Results with no hits, with an overlap lower than 60% or with an identity percentage lower than 90% were removed. In addition, results with a ratio of the score of the second best blast hit and the score of the first best blast hit higher than 0.8 were removed, to avoid grouping paralogous genes together. 4740 putative orthologs were found after this first reciprocal best blast hit analysis.

These results pointed out two questions. First, we were surprised to observe that the number of retrieved ORFs was about a third higher in L3 compared to L1 (20,790 ORFs vs 13,990, respectively), although sampling preparation as well as sequencing were performed using the same methodology. In addition, we retrieved "only" 4,740 orthologous genes, which seemed to be low for closely related species. To explain this pattern, we hypothesize that a combination of high polymorphism in L1 and stringency in mapping analysis led to the low number of retrieved ORFs in L1.

To infer if this difference in retrieved ORFs and orthologs was due to an actual difference in expressed genes or due to technical artifacts, we intended to recover some L1 contigs using L3 contigs as reference for mapping. As a control, we performed the same analysis using L1 contigs to recover L3 contigs. To do so, we used not only the reads of single individuals but also the reads of pooled individuals to reconstruct potential "missed" contigs in L1, allowing to obtain a higher coverage for each contig. In addition, we raised to 5% the maximal level of divergence between the L1 reads and the reference L3 contig, keeping in mind that the reads and the reference belonged to two different species.

L1 reads from unique and pool of individuals were mapped to 11405 L3 contigs (displaying lower than 80% identity with L1 contigs after the first blast search) using Bowtie 2, and allowing a maximum of 5 differences with the reference contig. Contigs with average coverage lower than 10x were discarded. L1 contigs were reconstructed with L1 reads using MIRA and allowing no missing data within the contig. Finally,

contigs smaller than 250 bp were discarded. Control reciprocal best blast hit was performed blasting the new reconstructed L1 contig against the 20,790 L3 contigs, to check whether the newly reconstructed L1 contig corresponded to the original L3 reference. Using this method, 6,156 new L1 contigs could be reconstructed, raising the number of L1 contigs to 20,146 (13,990+6,156). The same procedure was performed to reconstruct L3 contigs as a control, using 5,260 L1 contigs (displaying lower than 80% identity with L1 contigs after the first blast search) as reference and L3 reads from unique and pooled individuals. Using the same method as previously described, 1,333 new L3 contigs were reconstructed, raising the number of L3 contigs to 22,123 (20,790+1,333). Using the same approach with reciprocal best BLAST hit as previously described, we obtained a total of 12,229 orthologous genes between L1 and L3.

We paid particular attention that the reconstructed L1 and L3 contigs included no paralogous genes, using the ratio between the first and the second best BLAST hit score. If this ratio was above 0.8, we excluded the contig as there might be some paralogous genes (e.g. the contig corresponding to the first score and the contig corresponding to the second score). If this ratio was lower than 0.8, we could safely argue that there was most likely no paralogous genes which could have been included in the ortholog pairs. This approach being conservative, we might have missed some other "true" orthologous genes, but in this way we ensured that the reconstructed orthologs were accurate.

## 6.7 Extended methodological discussion: issues in transcriptome generation and polymorphism analysis

This study highlights a phenomenon that arose with advances in NGS technologies, which is the compromise between accuracy of results and speed of analysis. Treating such amount of data requires the automation of several steps in data analysis, which requires developers to make choices at each step of the analysis. As discarding valid data seems preferable with respect to taking into account wrong/inaccurate data, pipelines are usually developed to be very stringent. In our case, the pipeline seemed to work accurately with the low polymorphic species L3, retrieving a high number of ORFs and genes for diversity analyses, whereas it seemed too stringent with the highly polymorphic species L1. When comparing datasets that are susceptible to display, a priori, distinct distributions of the parameters used as threshold for bioinformatic treatments, biases may be introduced by these arbitrary thresholds.

Analyses based on likelihood or Bayesian approaches/criteria should be more accurate. In their absence, a critical a posteriori analysis of the comparison of results should be performed trying, when possible, to interpret in which direction the results should be biased by the arbitrary thresholds.

In our study, the critical part was the mapping, where reads are mapped on reference contigs to validate a contig or a SNP based on the coverage. The maximal number of mismatches between the read and the reference contig is fixed by the user, keeping in mind that if this number is too low, “correct” reads with polymorphism will be discarded, decreasing the overall coverage, whereas if this number is too high, reads potentially belonging to paralogous genes will be wrongly mapped to a contig, artificially raising the polymorphism. In our case, the maximal number of mismatch was set to 3 (on a 100 bp read, meaning a maximal divergence of 3%), which seemed to be accurate for the low polymorphic L3 species (20’790 ORFs, 18’408 validated SNPs) but was obviously too low for the highly polymorphic L1 species (13’990 ORFs, 7’607 validated SNPs), since the number of retrieved ORFs was a third higher in L3 species. Using a reconstruction method based on each species’ “unique” genes as a mapping reference, we could retrieve more than 6000 ORFs in L1 and 1300 ORFs in L3, confirming that the low number of ORFs initially found in L1 was due to technical issues.

Nevertheless, our L1 samples were actually not very suitable for the developed pipeline (Gayral *et al.* 2013) which was initially developed to analyze the genetic diversity of single individuals. Therefore, to validate a contig and a SNP, a minimal coverage per individual was necessary. Yet, the five L1 individuals were sampled in four different localities (Madeira, Crete, Portugal and France), therefore at different times in the year (February, May and July). As gene expression might obviously differ between localities and seasons, only a few very important genes (e.g. housekeeping genes) might be highly expressed in each locality, which might partly explain the low number of analyzed genes in the polymorphism analysis. In contrast, the two individuals from L3 were sampled at the same locality, during the same dive. This case study emphasizes the importance of adapting, within the realms of possibility, the analysis pipeline to the biological reality of the studied species, but also the importance of proper sampling.

# Chapitre 7

-

# Discussion et perspectives

# Chapitre 7: Discussion et perspectives

## 7.1 Discussion générale

### 7.1.1 *Les espèces d'Ophioderma longicauda*

Grâce aux travaux menés au cours de cette thèse, j'ai pu clairement déterminer qu'il existait au moins deux, voire cinq espèces biologiques dans le complexe *Ophioderma longicauda.* Tout d'abord, grâce à une combinaison d'observations de la période de reproduction, d'observations morphologiques, génétiques et écologiques, j'ai établi que les lignées sympatriques dispersante L1 et incubante L3 étaient des espèces biologiques différentes. Ensuite, J'ai montré qu'il existait potentiellement des différences de thermotolérance entre les espèces L1 et L3, renforçant l'idée d'une séparation des niches écologiques. Finalement, grâce à l'étude de 31 marqueurs génétiques, j'ai pu déterminer qu'il existait six groupes génétiques distincts dans le complexe *Ophioderma longicauda* (C1-C6).

Il est alors nécessaire de décrire la nouvelle espèce incubante (C4 et/ou C5) selon les critères de taxonomie classique. En effet, trop souvent les espèces cryptiques découvertes restent sans statut taxonomique et désignées comme « clade » ou « lignée » après leur étude, et restent donc taxonomiquement cryptiques (Schlick-Steiner *et al.* 2007; Carstens *et al.* 2013). Il est ainsi prévu, en collaboration avec Sabine Stöhr (Chercheuse au museum d'histoire naturelle de Stockholm, Suède), de décrire une nouvelle espèce d'*Ophioderma*, *Ophioderma vivipara*, comme l'espèce incubante qui inclut les lignées mitochondriales L2-L3-L4-L4b, en s'appuyant sur les données génétiques ainsi que sur des données morphologiques (notamment des parties squelettiques). Depuis quelques années, la communauté des taxonomistes tente de réunir les approches de taxonomie traditionnelle et moléculaire afin de former la taxonomie intégrative (Schlick-Steiner *et al.* 2010). Bien que la tâche de décrire une espèce cryptique ne soit pas forcément facile, il s'agit d'un phénomène en expansion et les études sur la taxonomie moléculaire prennent de l'importance (Jörger & Schrödl 2013).

En ce qui concerne les trois autres espèces potentielles d'*Ophioderma* (C2, C3, C6), des informations supplémentaires sont nécessaires afin de déterminer si nous

possédons assez de données pour décrire ces nouvelles espèces. Cependant, les données suggèrent fortement que le clade C2 de Tunisie (potentiellement incubant) est une espèce à part entière, puisque sa distribution est sympatrique avec l'espèce à larves C1, mais pas syntopique, une différence de profondeur ayant été observée entre les sites d'échantillonnages des C1 et C2. Les deux clusters ne sont pas trouvés dans les mêmes plongées mais étaient intercalés le long de la côte, suggérant fortement que leur différence génétique n'est pas une conséquence de limite au flux génique causée par des barrières locales physiques à la dispersion mais la conséquence de barrières interspécifiques (isolement pré ou post-zygotique). La différence de niche écologique ajoute un critère écologique potentiel, justifiant, s'il est confirmé, la description d'une nouvelle espèce. Par ailleurs, si la différence de mode de reproduction est confirmée, et l'hypothèse de poecilogonie (occurrence de plusieurs modes de reproduction et/ou de développement des œufs au sein d'une même espèce) rejetée, cela confirmera l'existence d'espèces séparées.

Bien que ces résultats doivent être considérés avec prudence étant donné le faible nombre d'individus analysés (10 pour C3 et 5 pour C6), les clusters C3 et C6 pourraient également être des espèces biologiques différentes. En effet, le cluster C3 uniquement trouvé à Dakar se détache assez clairement du cluster C1 sur l'analyse DAPC. Il est difficile de déterminer si les clusters C1 et C3 sont des espèces biologiques différentes, ou si ces clusters nettement différenciés par leurs génotypes multilocus sont dus uniquement à de la structure géographique. Cependant, les individus du cluster C1 provenant des Canaries (donc de l'Atlantique) ainsi que de la Méditerranée orientale (plus de 5000 km de distance, presque les plus éloignés de l'aire de distribution) se groupaient ensemble sans ambiguïté, ce qui suggère fortement que les différences génétiques observées entre C1 et C3 ne sont pas dues à de la structure géographique dans ce clade à larve.

Concernant le clade C6, il semble relativement clair qu'il s'agit d'une nouvelle espèce d'*Ophioderma* d'Atlantique. En effet, les distances génétiques avec les autres clusters d'Ophioderma sont très élevées. De plus, il s'agit du seul groupe monophylétique bien soutenu avec les données des marqueurs concaténés. En outre, Il faut noter qu'une « forme» *Ophioderma longicauda* f. *guineense* (Greeff, 1881) ou encore *Ophioderma guineense* avait été décrite en 1882 par Greeff. Comme la distribution des

caractères morphologiques était largement chevauchante avec *Ophioderma longicauda*, cette « forme » a été incluse à nouveau dans l'espèce *Ophioderma longicauda*. Cette « forme » provenant de Guinée équatoriale pourrait être l'espèce C6 décrite ici. Afin de confirmer cette hypothèse, idéalement, de nouveaux échantillons devraient être récoltés le long des côtes sénégalaises jusqu'en Guinée équatoriale, étude génétique devrait être tentée sur l'individu type d'*Ophioderma guineense*.

### 7.1.2 *Biogéographie, divergence, incubation, hybridation et renforcement : une histoire évolutive complexe chez Ophioderma longicauda*

Dans le chapitre 3, j'ai montré qu'il existait un gradient de spéciation dans le genre *Ophioderma* allant d'Ouest en Est. En effet, le genre *Ophioderma* est apparu dans les environs de la mer des Caraïbes et du Golfe du Mexique actuels ; la majorité des espèces de ce genre y étant trouvés actuellement. De par ces observations, un scénario biogéographique peut être décrit. L'ancêtre commun des *Ophioderma* de l'Est de l'Atlantique a vraisemblablement migré a travers l'océan Atlantique par un moyen inconnu. Arrivé sur les côtes africaines, des événements de diversification se sont produits. Ainsi, l'espèce *Ophioderma wahlbergii* (Müller & Troschel, 1842), présente actuellement en Afrique du Sud, s'est formée, vraisemblablement par séparation allopatrique. Ensuite, l'espèce C6 et aussi vraisemblablement l'espèce C3 ont pu apparaître aux alentours du Sénégal. Puis, en colonisant les côtes Atlantiques plus au Nord, l'espèce *Ophioderma longicauda* C1 s'est séparée. Cette dernière a également colonisé un vaste territoire, des Canaries à l'Espagne, ainsi que la totalité de la Méditerranée. Puis, vraisemblablement en allopatrie, l'incubation est apparue dans une population d'*Ophioderma longicauda* probablement en Méditerranée orientale (ancêtre commun des clusters C4 et C5). L'espèce incubante, suivant toujours un gradient de diversification Ouest-Est, s'est alors dispersée jusqu'aux côtes libanaises donnant le cluster C5.

La comparaison des scénarios historiques par les méthodes ABC a montré que le scénario le plus probable incluait un événement d'hybridation entre les C1 et les C4 (ou l'ancêtre commun de C4-C5) (Chapitre 3). Sachant cela, un scénario évolutif probable peut être décrit, même s'il ne s'agit que d'une hypothèse. L'événement d'hybridation a vraisemblablement eu lieu aux alentours du détroit siculo-tunisien pour former le cluster C2, et est probablement le résultat d'une hybridation entre des spermatozoïdes

C1 « dispersants » et des ovocytes C4 « incubants ». En effet, les individus C2 possèdent un haplotype mitochondrial (hérité de la lignée maternelle) très proche des individus du cluster incubant C4 (3 mutations seulement), mais montrent des allèles nucléaires partagés avec C1 (p.ex. i51) (Chapitre 5).

Il devait donc exister à ce moment là une forte compétition entre les spermatozoïdes C1 et C4. Pour l'espèce C4, les ovocytes fertilisés par les spermatozoïdes C1 représentaient vraisemblablement une perte coûteuse de gamètes, d'où l'existence d'une pression de sélection contre l'hybridation avec les C1. En réponse à cette pression de sélection, des protéines impliquées dans la compétition des spermatozoïdes (plus précisément des canaux ioniques impliqués dans la mobilité des spermatozoïdes) ont alors été sélectionnées positivement chez les spermatozoïdes C4 afin de féconder plus rapidement les ovocytes C4 (Chapitre 6). L'évolution rapide de ces protéines a vraisemblablement permis la mise en place d'un isolement reproducteur entre les C1 et les C4. Par la suite, un décalage temporel des périodes de reproduction des C1 et C4 a fini de compléter l'isolement reproducteur entre ces deux espèces, les C4 se reproduisant en avril et les C1 en juillet (Chapitre 2). Cet isolement temporel, cependant, reste à tester pour les clusters C2 et C5.

### 7.1.3 *Avantages et désavantages évolutifs de l'incubation*

Dans le complexe d'espèces *Ophioderma*, l'incubation est un caractère dérivé et semble être apparue une seule fois au cours de la divergence du complexe (Figure 7.1). En effet, toutes les espèces d'*Ophioderma* d'Atlantique Ouest possèdent une larve planctonique pour leur reproduction. Comme l'espèce C1 possède également une larve pour la reproduction, il est vraisemblable que les clusters C3 et C6 possèdent également une larve pour leur reproduction. Par ailleurs, l'estimation des tailles efficaces de ces clusters a montré qu'elles étaient du même ordre de grandeur que la taille efficace du cluster C1 (Chapitre 3), ce qui renforce

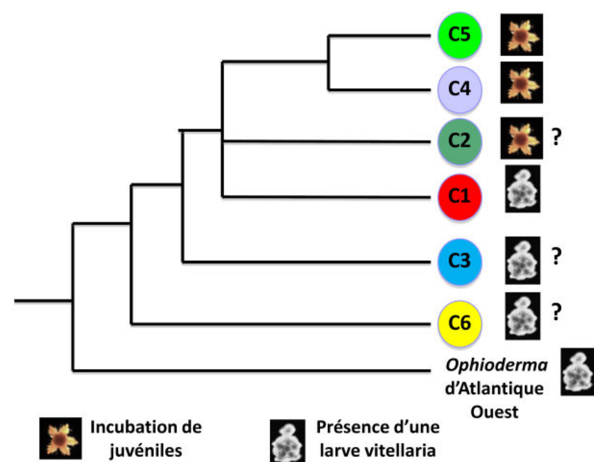



**Figure 7.1**: Résumé des traits d'histoire de vie dans le complexe d'espèces *Ophioderma longicauda*. L'incubation est un caractère dérivé qui semble n'être apparu qu'une seule fois au cours de la divergence du complexe

l'hypothèse que C3 et C6 sont des espèces « à larves ». Ensuite, l'incubation est vraisemblablement apparue une seule fois chez l'ancêtre commun des C4 et C5.

L'incubation ou l'absence de dispersion dans le milieu marin est apparue de nombreuses fois indépendamment chez les invertébrés marins, et semble être un caractère dérivé dans la majorité des cas, la présence de stade larvaire étant généralement admise comme étant ancestrale (revu dans Pechenik (1999). Par exemple, l'étude de l'évolution des traits d'histoire de vie de 12 espèces d'étoiles de mer de la famille des Asterinidae indique que la viviparité est vraisemblablement apparue deux fois indépendamment, et récemment (<2 Mya) chez cette famille (Hart *et al.* 1997). De plus, cette stratégie semble n'être que rarement être réversible vers la dispersion, étant que les taxons incubants se retrouvent toujours dans des branches récentes lorsque l'on observe des reconstructions phylogénétiques incluant la stratégie de reproduction (Hart *et al.* 1997). Cependant, il a été montré que dans certains cas, une forme ancestrale (à larve planctotrophe) pouvait être réacquise par réversion (revu dans Hart (2000). Finalement, il est important de considérer que la reconstruction de caractères ancestraux n'est pas triviale (Cunningham 1999) et ne fournit que des hypothèses plus ou moins vraisemblables ou parcimonieuses selon les données dont on dispose.

Il est connu que la probabilité de survie d'une propagule est plus élevée chez les incubants (Pechenik 1999). En effet, la descendance une fois relâchée dans l'environnement a déjà atteint le stade juvénile, et évite donc l'étape coûteuse du développement larvaire. Cependant, l'incubation est plus risquée sur le long terme, car comme montré dans les chapitres 5 et 6, elle induit une dérive génétique plus forte et donc une diversité génétique réduite. Pour cela, et également en raison de leur aire de distribution réduite, les espèces incubantes ou non-dispersantes sont plus sujettes à l'extinction lorsqu'elles subissent des modifications environnementales. Leclère *et al.* (2009) étudièrent la phylogénie et l'évolution des traits d'histoire de vie chez les hydrozoaires. Chez ces organismes, le cycle ancestral est constitué d'une phase fixée (le polype) et d'une phase dispersive (la méduse). Or, de nombreuses espèces ont montré une perte secondaire récente de la phase méduse, et donc la perte de la capacité de dispersion. Les auteurs proposent l'hypothèse d'un plus fort taux d'extinction chez ces espèces sans phase dispersive pour expliquer la présence de cette stratégie uniquement dans les branches récentes de la phylogénie. A contrario, il est également attendu que

les espèces sans phase dispersive montrent un plus fort taux de diversification (Hart 2000).

En réalité, la stratégie optimale de dispersion et d'investissement parental semble dépendre de la stabilité de l'environnement. Dans un environnement stable, les stratèges K (grande longévité, fort investissement parental, fort ratio d'énergie par gamète) sont favorisés, alors que dans un environnement perturbé, les stratèges r (faible longévité, faible investissement parental, faible ratio d'énergie par gamète) sont plutôt favorisés (Stearns 1977). Dans le cas d'*Ophioderma longicauda*, l'espèce incubante (stratège K) semble donc plus soumise au risque d'extinction.

#### 7.1.4 *Apports des NGS à l'étude des espèces non-modèles : nouvelles données, nouvelles compétences et nouveaux biais*

Au cours de ma thèse, j'ai eu l'opportunité d'obtenir des données provenant des nouvelles technologies de séquençage, à savoir la technologie Illumina qui a été utilisée pour séquencer des transcriptomes de novo, mais aussi pour séquencer des amplicons de PCR, et ce à moindre coût. Les transcriptomes ont non seulement été utiles dans l'étude des gènes communs aux espèces étudiées, l'étude de la diversité génétique de chaque espèce et la détection de gènes évoluant sous sélection positive entre les deux espèces, mais ils ont également servi comme base de données pour développer des nouveaux marqueurs. Ces derniers ont servi à comparer la connectivité entre les espèces dispersante et incubante, mais ils ont également servi à délimiter clairement les espèces dans le complexe d'espèces cryptiques *Ophioderma longicauda*.

Cependant, bien que très puissantes, les NGS amènent de nouvelles contraintes et de nouveaux biais. Premièrement, la quantité de données fournies étant astronomique, il est nécessaire d'avoir les ressources matérielles pour stocker et analyser ces données, ainsi que des compétences bioinformatiques pour les traiter. A moins d'avoir une formation solide en informatique, ces compétences sont souvent acquises de manière empirique. D'autant plus qu'il est vraisemblable qu'à l'avenir, le temps passé au laboratoire de biologie moléculaire sera considérablement réduit au profit de l'analyse des données, tout du moins dans le domaine de génomique au sens large (génétique des populations, spéciation, phylogéographie, etc).

Pour traiter des données de NGS, il est possible d'utiliser des modèles globaux d'évolution, en utilisant des méthodes probabilistes (p.ex. Chenuil (2012). Cependant, la majorité des scripts informatiques actuels, qu'ils soient simples ou complexes, requiert la prise de choix à chaque étape de l'analyse, qui ne sont pas forcément triviaux. En effet, ces choix peuvent dépendre de variables biologiques que seul un biologiste ou un généticien est capable d'appréhender correctement, étant donné la nécessité d'une connaissance à priori des données et des espèces étudiées. Par exemple, lorsque l'on désire reconstruire des relations phylogénétiques entre espèces et que l'on a à disposition un grand nombre de séquences, il est difficile de différencier les allèles des paralogues. Un seuil fixe de divergence n'est probablement pas envisageable étant donné qu'il n'est pas forcément fixe entre taxons et / ou entre gènes, qui peuvent montrer des vitesses d'évolution différentes. Dans mon cas, j'ai reconstruit des arbres de gènes et je me suis aidée des espèces appartenant aux groupes externes afin de définir les groupes de paralogues. Cette méthode, bien qu'étant fructueuse pour certains marqueurs, était quand même considérablement chronophage et n'a pas permis la distinction entre allèles et paralogues pour certains marqueurs. Afin d'être applicable à plus large échelle, cette méthode devrait pouvoir être transcrite en termes de modèle d'évolution et permettre que la décision soit prise au terme d'une analyse de vraisemblance ou bayésienne.

Une autre question survenant lors de l'analyse d'amplicons PCR et de leur tri, est de savoir comment différencier les erreurs de séquençage des véritables allèles. Mais aussi, comment déterminer le seuil qui différencie si un individu est homozygote ou hétérozygote. En général, le nombre de séquences lues (=reads) est informatif pour répondre à ces questions, cependant les cas ne sont pas toujours simples. En effet, un marqueur peut avoir été énormément « lu » (>1000 reads), et un autre moins bien (environ 200 reads). Si le seuil est trop élevé pour tous les marqueurs (p. ex. 500 reads), les risques de jeter des bonnes séquences sont élevés. A contrario, si le seuil est trop bas (p.ex. 100 reads), le risque d'inclure des erreurs de séquençage pour les marqueurs très lus est très élevé. Pour pallier à ce problème, j'ai estimé manuellement un seuil de reads à garder pour chacun des 41 marqueurs, dépendant du nombre maximal de lecture pour chaque marqueur. Simplement prendre la première ou les deux premières séquences les plus lues n'aurait pas été adéquat non plus, étant donné la présence de paralogues dans un grand nombre de marqueurs.

Un aspect non négligeable des biais possibles avec le séquençage d'amplicons est la possibilité de contaminations entre échantillons (cross-contamination). En effet, une étape consiste à manipuler et mélanger les amplicons de différents marqueurs dans une seule plaque, rendant les cross-contaminations entre puits (et donc entre individus) possibles. Etant donnée la très forte profondeur de séquençage de la technologie Illumina, les contaminations peuvent être séquencées, obtenir un grand nombre de lectures et être considérées à tort comme des séquences « vraies ». Ces problèmes sont d'autant plus probables si des ADN de qualités différentes sont utilisés ; en effet les ADN de bonne qualité sont facilement amplifiés contrairement aux ADN de faible qualité. Dès lors, la contamination des échantillons avec ADN de faible qualité (peu ou pas d'amplicons) avec les échantillons avec ADN de bonne qualité (beaucoup d'amplicons) est plus que probable. Un tri des séquences provenant de contaminations est possible à posteriori, si l'on utilise des séquences d'espèces bien différentes par exemple, facilement différenciables. Cependant, cela devient plus compliqué voire impossible si l'on travaille au niveau intraspécifique. Une solution à ce problème serait de n'utiliser, si possible, que des ADN ayant des qualités similaires. Malgré la présence de nombreux biais possibles, les NGS offrent des avancées majeures pour le séquençage de génomes, de transcriptomes, de marqueurs et le génotypage de SNP. Ainsi, de nouvelles méthodes de traitement de ces données sont actuellement en développement, afin de réduire au maximum l'impact de ces biais (p. ex. Lighten *et al.* (2014)).

## 7.2 Perspectives

### 7.2.1 *Etude des gènes sous sélection positive dans chaque cluster*

Une perspective réalisable à court terme pour ce travail serait d'amplifier par PCR et séquencer les gènes codant pour les deux canaux sélectionnés positivement chez tous les clusters d'*Ophioderma longicauda* présentés au cours de cette thèse, afin de savoir si la sélection positive a eu lieu chez tous les clusters incubants ou uniquement chez le cluster C4. De plus, il serait intéressant d'amplifier et séquencer le gène codant pour le récepteur aux chemoattractants (speract receptor) chez tous les clusters, afin de déterminer s'il existe des changements dans la séquence correspondant à la partie extracellulaire (= liaison au ligand). De nombreux tests sur les différentes lignées d'*Ophioderma longicauda* et sur les groupes externes d'*Ophioderma* (*O. teres, O. cinereum, O. phoenium*) ont été réalisés dans ce but, mais l'amplification de ces

marqueurs a toujours été infructueuse. Cependant, grâce à la récente publication de 61 transcriptomes d'ophiures dont une *Ophioderma* (O'Hara *et al.* 2014), de nouvelles amorces dessinées sur les parties conservées de ces gènes devraient pouvoir permettre l'amplification de ces gènes chez *Ophioderma longicauda*.

D'un point de vue fonctionnel, il serait intéressant d'étudier le mécanisme exact menant à une compétition des spermatozoïdes accrue chez les C4. En effet, on ignore encore comment les acides aminés mutés agissent exactement au niveau de la fonction du canal ionique. Grâce à l'étude des domaines fonctionnels du gène tetraKCNG chez *Strongylocentrotus purpuratus* (Galindo *et al.* 2007; Darszon *et al.* 2008), j'ai pu déterminer que les acides aminés mutés se trouvaient dans les quatrième (voltage sensor) et cinquième domaines transmembranaires du quatrième domaine CNG (cyclic nucleotide gated ; KCNG IV) de la protéine. Pour le gène NHE en revanche, je n'ai pas pu trouver d'étude fonctionnelle décrivant précisément les domaines de ce gène. Une perspective plus globale pour ce travail serait d'obtenir les transcriptomes de chacun des six clusters, afin de déterminer si d'autres gènes potentiellement impliqués dans l'isolement reproducteur ont pu être sélectionnés positivement.

Finalement, il serait également intéressant d'étudier ces gènes (le récepteur aux chemoattractants, tetraKCNG et NHE) chez d'autres espèces proches d'ophiures, étant donné que la chimiotaxie des spermatozoïdes semble être un phénomène spécifique à la reconnaissance des espèces chez les ophiures (d'après les données expérimentales testant l'attraction chimique des gamètes), contrairement aux oursins et aux étoiles de mer, où ce sont plutôt les protéines de reconnaissance gamétique qui sont impliquées dans la reconnaissance entre espèces. Le modèle d'ophiure *Astrotoma agassizii* (Heimeier *et al.* 2010) est un candidat prometteur, car les trois lignées mitochondriales décrites dans ce complexe d'espèces semblent également montrer des différences de mode de reproduction (incubant vs larve planctonique). De plus, le modèle d'espèces cryptiques *Ophiothrix fragilis* – *Ophiothrix quinquemaculata* (Baric & Sturmbauer 1999; Muths *et al.* 2009; Pérez-Portela *et al.* 2013) serait intéressant, étant donné que les deux clades existant possèdent tous deux une reproduction via des larves planctoniques. Ainsi, l'étude des gènes précédemment cités permettrait de donner des arguments éventuels pour soutenir la possibilité, ou non, que la sélection positive ait eu lieu indépendamment de l'incubation. Il serait également envisageable de caractériser en

séquences ces gènes de chimiotaxie dans tous les complexes d'espèces cryptiques d'ophiures connus, qu'ils soient sympatriques ou non. Une large approche de ce type permettrait d'établir le degré de généralité de l'implication de ces gènes dans l'isolement prézygotique chez les ophiures.

### 7.2.2 *Génome scans d'adaptation locale*

#### 7.2.2.1 $F_{ST}$ outliers sur les données de fréquence des transcriptomes

Un aspect qui n'a pas pu être abordé au cours de ma thèse pour des raisons de temps était l'étude des fréquences des SNPs entre différentes populations du cluster C1. En effet, en plus des transcriptomes d'individus uniques, des transcriptomes de pools d'individus ont été séquencés. Les ARNm de dix individus (cinq mâles et cinq femelles), huit individus (quatre mâles et quatre femelles) et six individus ont été mélangés et séquencés pour les localités de Crète (10), Marseille (8), Portugal (8) et Madère (6), respectivement. En se basant sur le nombre de reads par gène pour extraire les fréquences des SNPs par population, il est possible de calculer des $F_{ST}$ entre populations. Si l'on calcule des $F_{ST}$ pour un millier de SNPs par exemple, il est possible de déterminer un $F_{ST}$ moyen entre populations. Basé sur cette moyenne, on peut déterminer s'il existe des SNPs dont les fréquences diffèrent significativement entre populations. Ces derniers, appelés $F_{ST}$ outliers, représentent des SNPs appartenant potentiellement à des gènes candidats d'adaptation locale, ou liés physiquement à ces derniers sur le chromosome. Il serait intéressant de déterminer si de tels SNPs sont présents entre les populations du cluster C1, et de trouver s'ils appartiennent à gènes particuliers. Si ces SNPs sont intéressants, ils devraient être étudiés sur un plus grand nombre d'individus et sur un grand nombre de localités.

Il faut cependant être prudent avec l'interprétation de ces données de SNPs outliers, étant donné que les méthodes de détection de ces SNPs sont sujettes à un grand nombre de faux positifs. En effet, il existe plusieurs facteurs (en particulier intrinsèques) qui peuvent aboutir au même signal de fréquences de SNPs significativement différentes entre populations, alors même qu'il ne s'agit pas d'adaptation locale (Bierne *et al.* 2011).

#### 7.2.2.2 Autres techniques

La méthode expliquée précédemment est une méthode en deux étapes, où les gènes candidats d'adaptation locale sont d'abord détectés chez un petit nombre d'individus, puis ces SNPs sont étudiés sur un grand nombre d'individus et de populations (candidate gene approach). Cependant, il existe d'autres méthodes pour déterminer les SNPs ayant des fréquences différentes entre populations (Davey *et al.* 2011; Kumar *et al.* 2012). Par exemple, la méthode Kaspar (Bradbury *et al.* 2010) ou BeadXpress (Thomson *et al.* 2012) proposées par des plateformes génomiques offrent de développer des SNPs et obtenir les génotypes pour environ un millier d'individus. Par ailleurs, la méthode de RAD séquençage permet l'obtention de nombreux SNPs dispersés dans le génome, pour de très nombreux individus (Hohenlohe *et al.* 2011; Baxter *et al.* 2011). L'identification de gènes potentiels d'adaptation locale peut ensuite être mise en relation avec les variables environnementales des localités étudiées.

Lorsque des espèces sont phylogénétiquement très proches et qu'on les trouve en sympatrie, les polymorphismes impliqués dans l'adaptation locale tendent à être conservés entre elles par la sélection naturelle, et ce d'autant plus que leurs niches écologiques sont similaires. Le modèle *O. longicauda* offre donc l'opportunité de traiter finement les questions d'adaptation locale, et ce grâce à l'utilisation d'espèces « réplicas ». Ainsi, la génétique des paysages peut être étudiée, en corrélant les génotypes observés avec des variables environnementales. Dans le cas où ces différences de fréquences entre locus d'adaptation locale sont maintenues, cela pourrait aboutir à un phénomène dit de spéciation écologique, où la sélection disruptive favorise deux génotypes extrêmes, qui peuvent aboutir à la formation de deux nouvelles espèces.

## 7.3 Conclusion

Au commencement de ma thèse, seulement deux marqueurs génétiques étaient disponibles ; le marqueur mitochondrial COI et le marqueur nucléaire i51. Grâce à la récente popularisation des nouvelles technologies de séquençage, j'ai pu obtenir des transcriptomes de deux espèces très proches, les comparer et trouver quels gènes montraient une évolution sous sélection positive. De plus, j'ai pu utiliser ces transcriptomes comme base pour développer de nouveaux marqueurs génétiques. Ces derniers m'ont non seulement permis de clairement définir le nombre d'espèces dans le

complexe *Ophioderma longicauda*, mais aussi d'inférer finement la connectivité entre les espèces incubante et dispersante, ainsi que de déterminer les niveaux de diversité génétique entre ces deux espèces.

Bien qu'incluant une grande partie de génétique et de génomique, j'ai eu l'opportunité d'aborder plusieurs aspects complémentaires au cours de ma thèse, notamment l'échantillonnage et la connaissance de la niche écologique d'*Ophioderma longicauda*; la fixation et la préparation de coupes histologiques ; l'utilisation du microscope électronique à balayage ainsi que l'expérience de thermotolérance et de régénération. De par l'étude d'un seul modèle, j'ai pu aborder de nombreuses questions scientifiques, et tenter d'y répondre le plus précisément possible. D'autres aspects de recherche sont encore à traiter dans ce complexe d'espèces, notamment la recherche de gènes candidats d'adaptation locale, qui devrait être poursuivie. Avec cette thèse, j'ai pu montrer que l'étude d'espèces non-modèles est prometteuse à tous les points de vue, et dans l'optique d'une meilleure connaissance des mécanismes d'adaptation, de sélection et de spéciation, l'étude d'espèces non-modèles devrait être poursuivie.

# Chapitre 8

-

# Annexes

# Chapitre 8: Echinoderm phylogeny finally solved: a review of past pitfalls, from biological to methodological ones (Annexe 1)

Les auteurs de cet article sont Alexandra Anh-Thu Weber, Laurent Abi-Rached et Anne Chenuil. Il sera soumis à PLoS One ou Molecular Phylogenetics and Evolution après relecture et corrections par tous les auteurs.

## 8.1 Abstract

Echinodermata is the largest phylum exclusively constituted of marine species and is the second deuterostome phylum, after chordates, in species number. This phylum encompasses five extant Classes, the crinoids (sea lilies), the echinoids (sea urchins), the holothurians (sea cucumbers), the asteroids (sea stars) and the ophiuroids (brittle stars). Although studied for a long time due to the high number of extent species and the good fossil records, the phylogenetic relationships among Classes remained, until recently, unresolved, the position of ophiuroids being controversial. In this study, we present a bibliographic review of the echinoderm phylogeny history, and show the different pitfalls (biological and methodological) that impaired accurate phylogenetic reconstruction. In particular, we focus on the issue of ancient rapid speciation, long-branch attraction, incongruence among genes, comparison of phylogenetic reconstruction methods, removal of fast-evolving sites and the impact of missing data. In addition, we performed de novo transcriptome sequencing of two brittle stars species, and used available genomic or transcriptomic data of one sea star, one sea cucumber, two sea urchins and one hemichordate to reconstruct the position of ophiuroids in the Echinoderm tree using a complete dataset. After careful selection of highly reliable orthologs shared among those taxa, we obtained an alignment of 138 shared genes without any missing data (17,318 amino-acids). Since taxon sampling was minimal, we checked for compositional biases. All phylogenetic methods applied on this data set yielded the same topology as recent phylogenomic studies (with clades Asterozoa and Echinozoa), with strong support, and no sign an influence of compositional bias or LBA. We also show, reanalyzing another published dataset, that missing data may have a strong negative impact, which is often over-looked in phylogenomics approaches. This study highlights the power of phylogenomic to reconstruct ancient and difficult speciation events, but warns against the impact of missing data in phylogenetic reconstructions.

## 8.2 Introduction

Echinodermata is the largest phylum exclusively constituted of marine species and is the second deuterostome phylum, after chordates, in species number. Its development and morphology are particularly original among animals, with a particular water vascular system, adult pentaradial symmetry and larval bilateral symmetry (Mooi & David 2008). This phylum contains five extant classes: crinoids (sea lilies), asteroids (sea stars), ophiuroids (brittle stars), holothurians (sea cucumbers) and echinoids (sea urchins).

The monophyly of each class was unambiguously established by both morphological and molecular data (e.g. Janies *et al.* 2011) but the phylogenetic relationships among them was, until recently, still a matter of debate. Littlewood *et al.* (1997) resolved most branching and from the 105 possible topologies among those five classes, three remained impossible to disentangle. Whereas both morphological and molecular phylogenetic studies agree to consider crinoids as branching first (Smith 1997a; b; Littlewood *et al.* 1997; Janies *et al.* 2011), and the holothurians and echinoids as being sister classes, which together form the Echinozoa (noted Ez), there was contradictory support from morphology as well as from molecules for the three possible topologies among Asteroidea (A), Ophiuroidea (O) and Echinozoa (Ez).

A first possibility opposes ophiuroids to an asteroids-Echinozoa clade (topology noted [O], where ophiuroids diverge first). In a second topology, asteroids diverged from an ophiuroid-Echinozoa clade (topology noted [A], where asteroids diverge first), and in the third case, asteroids and ophiuroids form a clade named Asterozoa (noted Az), brother clade to the Echinozoa (topology noted [Az]). In most cases, node support was low. When maximum parsimony was used, the most parsimony tree was in general only one or a few steps shorter than the next ones or alternative topologies were equally parsimonious. An important variety of data, e.g. adult and larval morphology (Littlewood *et al.* 1997; Janies 2001), mitochondrial gene order (Smith *et al.* 1993; Perseke *et al.* 2010), amino-acids (e.g. Perseke *et al.* 2010; Pisani *et al.* 2012), ribosomal DNA (e.g. Mallatt & Winchell 2007), even small RNAs (which Pisani *et al.* (2012) proved to be inefficient), was combined with various evolutionary models and phylogenetic methods (reviewed in Table 1). Missing and chimeric data were used in several of these studies to increase the number of characters used (in general molecular characters)

(reported in Table 1). In these studies, mostly based on morphological data or relatively few genes, a consensus about the position of ophiuroids failed to emerge.

With the striking increase in genomic resources available since a few years, phylogenomics generates strong hopes for resolving complex phylogenetic questions. In particular, phylogenomics successfully resolved phylogenetic relationships among, for instance, mammals (Murphy *et al.* 2001), birds (Hackett *et al.* 2008); turtles (Chiari *et al.* 2012) and Rhizaria (Sierra *et al.* 2013). However, non phylogenetic signal and systematic biases also affect such data, resulting in false, but in some cases strongly supported, trees (Philippe *et al.* 2005, 2011; Nishihara *et al.* 2007). Another drawback of molecular studies involving such data sets (large DNA alignments, numerous genes, super-trees or super-matrices) is the level of missing data (i.e. species missing for one or several genetic regions) which have been tolerated, in the absence of significant theoretical or empirical knowledge on the acceptable thresholds.

A recent study (Telford *et al.* 2014) investigated the phylogenetic position of ophuroids in the echinoderm tree, using a matrix of 48,818 positions representing a total of 219 genes. They performed an extensive taxon sampling, which included four echinoids (*Strongylocentrotus purpuratus, Hemicentrotus pulcherrimus, Paracentrotus Eucidaris tribuloides*) two holothurians (*Apostichopus japonicus, Holothuria glaberrima*), two asteroids (*Solaster stimpsoni, Patiria pectinifera*), two ophiuroids (*Amphiura filiformis, Amphipholis squamata*) and one crinoid (*Florometra serratissima*). In addition, they used three hemichordates (*Saccoglossus kowalvskii, Ptychodera flava, Balanoglossus clavigerus*) and one cephalochordate (*Branchiostoma floridae*) as outgroups. They obtained a so-called fully resolved tree (posterior probabilities of 1 for each node of the tree) after Bayesian Inference (BI), supporting the Asterozoa topology. Yet, no results were shown concerning other phylogenetic reconstruction methods, namely Maximum Likelihood (ML), Maximum Parsimony (MP) and Neighbor-Joining (NJ). In addition, the amount of missing data per species was particularly high. Indeed, it ranged between 2 and 90% (with respect to the full alignment length) in echinoids, between 75 and 80% in holothurians, between 64 and 73% in ophiuroids and between 45 and 65% in asteroids. In contrast, the crinoid displayed a low level of missing data (3.7%). The outgroups *S. kowalevskii* and *B. floridae* were well represented (1.8 and 0.6% of missing

data, respectively), whereas the two remaining outgroups *P. flava* and *B. clavigerus* displayed also high level of missing data (79 and 81%, respectively).

An even more recent study (O'Hara *et al.* 2014) focused on the phylogeny of ophiuroids, providing new transcriptomic datasets for 52 brittle star species. They build an echinoderm phylogeny also including transcriptomic or proteomic datasets of two sea stars (*Patiria miniata, Luidia senegalensis*), three sea urchins (*Strongylocentrotus purpuratus, Lytechinus variegatus, Eucidaris tribuloides*), one sea cucumber (*Leptosynapta tenuis*), one crinoid (*Aporometra wilsoni*) and two outgroups (the hemichordate *Saccoglossus kowalevskii* and the vertebrate *Danio rerio*). A 85% complete data matrix of 102,143 amino acids representing 425 genes was used to reconstruct echinoderm phylogeny, using Maximum Likelihood and Bayesian Inference. The authors found full support for the Asterozoa topology with both reconstruction methods.

In this study, we performed a phylogenomic analysis of echinoderm classes using a complete matrix of 138 genes (without missing data), using data from two sea urchins (*S. purpuratus, Lytechinus variegatus*), one sea cucumber (*A. japonicus*), one sea star (*Patiria miniata*), two closely related brittle stars (*Ophioderma longicauda* L1 and L3) and one hemichordate (*S. kowalevskii*) as outgroup. We had three main objectives: first, we inferred the phylogenetic position of ophiuroids in the echinoderm tree, to confirm the results of Telford *et al.* (2014) and O'Hara *et al.* (2014) using a complete dataset without any missing position. Then, we re-analyzed the dataset of Telford *et al.* (2014) with two aims: (i) to check the impact of missing data on traditional methods and estimate whether the results obtained with Bayesian (known to display overestimated supports) were also recovered by Maximum Likelihood methods and traditional methods, and second, (ii) to check whether the results they obtained could be recovered after removal of sites including missing data. Finally, we present a bibliographical review of echinoderm phylogeny, from the use of few morphological characters to hundreds of genetic markers, and emphasize on the different factors that could explain why echinoderm phylogeny was only recently resolved.

## 8.3 Material & Methods

### 8.3.1 *Transcriptome generation and retrieving from databases*

*O. longicauda* samples were collected in Crete (Greece), Marseille (France), Sagres (Portugal) and Madeira (Portugal). Individuals of the broadcasting lineage L1 were collected in each location while individuals of the brooding lineage L3 were collected in Crete. RNA was extracted from gonads and pieces of arm, using a method combining guanidinium thiocyanate–phenol–chloroform extraction (using Qiazol, Qiagen) followed by a silica matrix extraction (using RNeasy columns, Qiagen), as described in Gayral *et al.* (2011). Illumina libraries were constructed after quality controls of purity and integrity of RNA, and sequencing was performed with a Hi-seq 2000 sequencer (Illumina) by the genomic platform Genotoul (Toulouse). Raw reads were filtered for length and quality, and de novo assembly was performed with the programs Abyss and Cap3, following the method B in Gayral *et al.* (2013). Contigs were filtered for length and quality, and ORFs were predicted following the method described in Gayral *et al*. (2013). 20,146 and 22,123 CDS were found for the *O. longicauda* lineages L1 and L3, respectively.

Protein datasets were retrieved from the database www.echinobase.org for the sea urchin *Strongylocentrotus purpuratus* (genome assembly v3.1: 29,890 annotated genes), *Lytechinus variegatus* (genome assembly v0.4: 28,094 annotated genes), *Eucidaris tribuloides* (transcriptomic data) and the sea star *Patiria miniata* (genome assembly v1: 29,697 annotated genes). The transcriptome of the sea cucumber *Apostichopus japonicus* (36,005 CDS) was retrieved from the supplementary materials of Zhou *et al.* (2014). The protein dataset predicted from the genome assembly v3 of the hemicordate *Saccoglossus kowalevskii* (34,239 annotated genes) was retrieved from the genomic database www.metazome.net/skowalevskii.php.

### 8.3.2 *Identification of orthologous gene sets*

The program OrthoMCL (Li *et al.* 2003) was used to retrieve single-copy orthologous genes through reciprocal best-hit blast searches. A first analysis yielded only 47 1:1 orthologous groups shared among the eight studied species: detailed analysis of the results revealed that this low number was due to the incompleteness of the *E. tribuloides* dataset and so this species was discarded. A second analysis was

performed and identified 153 1:1 orthologous groups shared among each of the seven remaining species. Each set of orthologous gene sequences was then aligned using MAFFT (Katoh *et al.* 2009). The resulting alignments were visually inspected to detect problems such as erroneous gene predictions (as several the datasets for several species include gene prediction) and 15 of the 153 genes were removed. The remaining 138 genes were used for further analysis, representing a total of 57,796 amino acids positions.

### 8.3.3 *Gene-by-gene phylogenetic analysis*

A Bayesian concordance analysis was carried out with BUCKy 1.4.0 (Ané *et al.* 2007; Larget *et al.* 2010) using Bayesian phylogenetic trees generated with MrBayes, as described below. This Bayesian concordance analysis used two runs of 100,000 generations with one chain and a discordance parameter $\alpha$ set to 1. For each of the 138 genes of the dataset, Bayesian phylogenetic analyses were performed using MrBayes 3.2.1 (Ronquist & Huelsenbeck 2003) with a WAG+G model of protein substitution. Sampling was performed through three independent runs (each having one cold chain and three heated chains), which were run for 1,000,000 generations or until the average standard deviation of split frequencies for the three runs was <0.01. Trees were sampled every 200 generations and the first one-fourth of the trees was discarded as burnin before a consensus tree was generated.

### 8.3.4 *Concatenated phylogenetic analysis*

The 138 genes were concatenated to create a supermatrix of 57,796 amino acids. To avoid bias in our analyses, all positions with gaps were removed from the alignment so that the final dataset included a complete matrix of 17,318 amino acids for each species. To check whether a base composition bias could wrongly cluster taxa together, we used the posterior predictive analysis implemented in Phylobayes to detect whether the taxa were compositionally homogeneous or heterogeneous. In addition, a Principal Component Analysis (PCA) was performed based on the contingency table describing the amino acid composition of each taxon, to visualize similarity among taxa regarding amino-acid composition.

To analyze this dataset both parametric (Bayesian) and non-parametric (Neighbor-Joining; Parsimony and Maximum-likelihood) bootstrap analyses were

conducted. Neighbor-Joining analyses were performed using MEGA 6.0.5 (Tamura *et al.* 2013) and 1000 replicates. Parsimony analyses were performed using PAUP 4.0 b10 (Swofford 2003) and the tree bisection-reconnection branch swapping algorithm with 500 replicates and a heuristic search. Maximum Likelihood analyses were performed using RaxML 8.0.19 (Stamatakis 2014), with 500 replicates (fast bootstrapping) and the WAG+G model of protein evolution. Bayesian analyses were performed using MrBayes 3.2.1, as described in the gene-by-gene section.

### 8.3.5 *Comparison with previous studies*

A bibliographic search on the previous studies published on echinoderm phylogeny was done. Type of dataset used, level of missing data, reconstruction method and topologies found were reported. In addition, complementary analyses (ML, MP and NJ) were performed using the dataset of Telford *et al*. (2014), available in the supplementary material. The three analyses were carried out with the programs and conditions previously described.

## 8.4 Results

### 8.4.1 *Gene-by-gene analysis cannot resolve class-level Echinoderm phylogeny*

Using a clustering approach with genome-wide data, we could isolate 138 orthologs in seven species. On those 138 analyzed genes, 121 were unresolved (displayed a topology posterior probability <0.9) and 17 displayed posterior probabilities >0.9 for a particular topology. Six genes were fully resolved for the [Az] topology, three genes were fully resolved for the [A] topology and the eight remaining genes indicated spurious topologies. The concordance analysis combines the information of multiple loci, taking into account the number of different topologies per gene and their respective posterior probabilities. The concordance analysis indicated that the most frequent topology was [Az], the topology including two monophyletic groups, the Asteroid-Ophiuroid (A-O) group and the Echinoid-Holothurian (E-H) group, with the hemichordate *S. kowalevskii* as an outgroup. Yet, this topology was not strongly supported, with a concordance factor of 0.48 for the (E-H) group and a concordance factor of 0.45 for the (A-O) group, meaning that 48% of the genes displayed the E-H node and 45% of the genes displayed the A-O node. Concordance factors of the *O. longicauda* L1-L3 and *S. purpuratus* – *L. variegatus* were both 100, indicated that each

gene supported those nodes. Individual genes are thus insufficient to resolve the splits between Ophiuroids, Asteroids and Echinozoa. One possible explanation for this is that these splits occurred over a short evolutionary period of time and to test this possibility we assembled a concatenated dataset.

### 8.4.2 *Concatenated dataset analysis fully resolves Echinoderm phylogeny*

As the concordance analysis did not highlight two or more strongly supported discordant topologies, all genes were concatenated to encompass more information for phylogenetic reconstruction. The 138 genes were concatenated to obtain a complete matrix of 17318 amino acids, without gaps and without missing data. The composition test performed by Phylobayes indicated that *P. miniata, A. japonicus* and *S. kowalevskii* were compositionally heterogeneous, whereas *O. longicauda* L1 and L3, *S. purpuratus* and *L. variegatus* were compositionally homogeneous. Those results were confirmed by the PCA, which showed that *O. longicauda* L1 and L3, *S. purpuratus* and *L. variegatus* formed together a close group, whereas *P. miniata, A. japonicus* and *S. kowalevskii* were more dispersed on the graph (Figure 1).

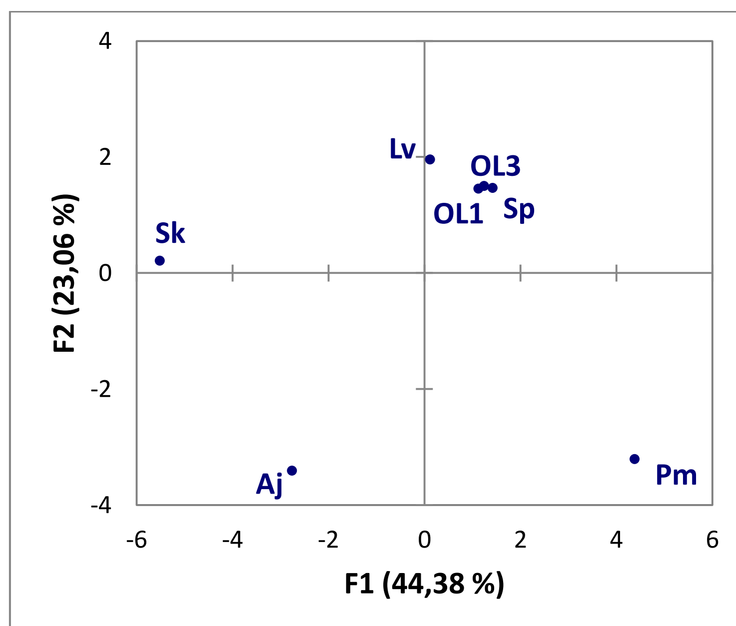



**Figure 1:** Principal component analysis (PCA) on amino-acid composition in each species used in this study. Sk=*Saccoglossus kowalevskii*; Aj=*Apostichopus japonicus*; Lv=*Lytechinus variegatus*; OL1=*Ophioderma longicauda* L1; OL3=*Ophioderma longicauda* L3; Sp=*Strongylocentrotus purpuratus*; Pm=*Patiria miniata*.

The analysis of concatenated dataset provided strong support for the same topology in each phylogenetic method used (Figure 2). NJ indicated maximal support for the A-O group (100) and strong support for the E-H group (93). MP displayed lower support for the A-O group (93) and the E-H group (77). Finally, ML and BI analyses indicated maximal support (100) for both group A-O and E-H.

**Figure 2**: Echinoderm phylogeny based on the concatenation of 138 genes shows unambiguously the monophyly of Asterozoa and Echinozoa. The four reconstruction methods (NJ, MP, ML, BI) were used. OL1=*Ophioderma longicauda* L1; OL3=*Ophioderma longicauda* L3; Pm=*Patiria miniata*; Aj=*Apostichopus japonicus*; Lv=*Lytechinus variegatus*; Sp=*Strongylocentrotus purpuratus*; Sk=*Saccoglossus kowalevskii*.

### 8.4.3 *Comparison with previous studies*

The summary of different studies performed on echinoderm phylogeny is shown in Table 1. Most the studies found the Asterozoa topology, but often with a low support. Interestingly, the early studies of Mooi *et al.* (1994) and David & Mooi (1999) had already found the Asterozoa topology using skeletal homologies, in particular using the axial-extraxial skeletons. Pisani *et al.* (2012) also found the Asterozoa topology using seven housekeeping nuclear genes and three rDNA genes, with high Bayesian support (0.97). Because they thought their full dataset was unreliable, they decided to remove the fast evolving sites as well as the constant sites for the analyses, using a "rate-homogeneous" dataset. They found another topology [A], with the Asteroids more basal, with a lower Bayesian support (0.93). They concluded that this latter topology was the accurate one, arguing that their "rate-homogeneous" dataset was more reliable than the

complete one. This study demonstrates the risk of selecting and removing arbitrarily sites for analyses, which can lead to erroneous conclusions.

The most recent studies of Telford *et al.* (2014) and O'Hara *et al*. (2014) both found maximal support for the Asterozoa topology, using phylogenomic dataset (using 219 and 425 genes, respectively). To compare the results of Telford *et al.* (2014) obtained with BI, we performed ML, MP and NJ analyses on their dataset. We first wanted to perform the same type of analysis we did for this study (without gaps and missing data), but we found that this dataset only contained 194 common amino-acid positions on the 45,818 positions of the initial dataset, which was too small to perform meaningful analyses. Therefore, we performed the analyses on the original dataset, with all methods. We obtained the same topology as they found, supporting the Asterozoa, for the ML and MP analyses (Figure 3). Yet, this node was not well supported in each method (bootstrap estimates were 0.76 and 0.45 in ML and MP, respectively) (Figure 3, Table 1). The NJ analysis produced a spurious topology, with a clade grouping ophiuroids with holothurians (supported with 0.81 of bootstrap support), and a clade joining them with asteroids (bootstrap support 0.72).

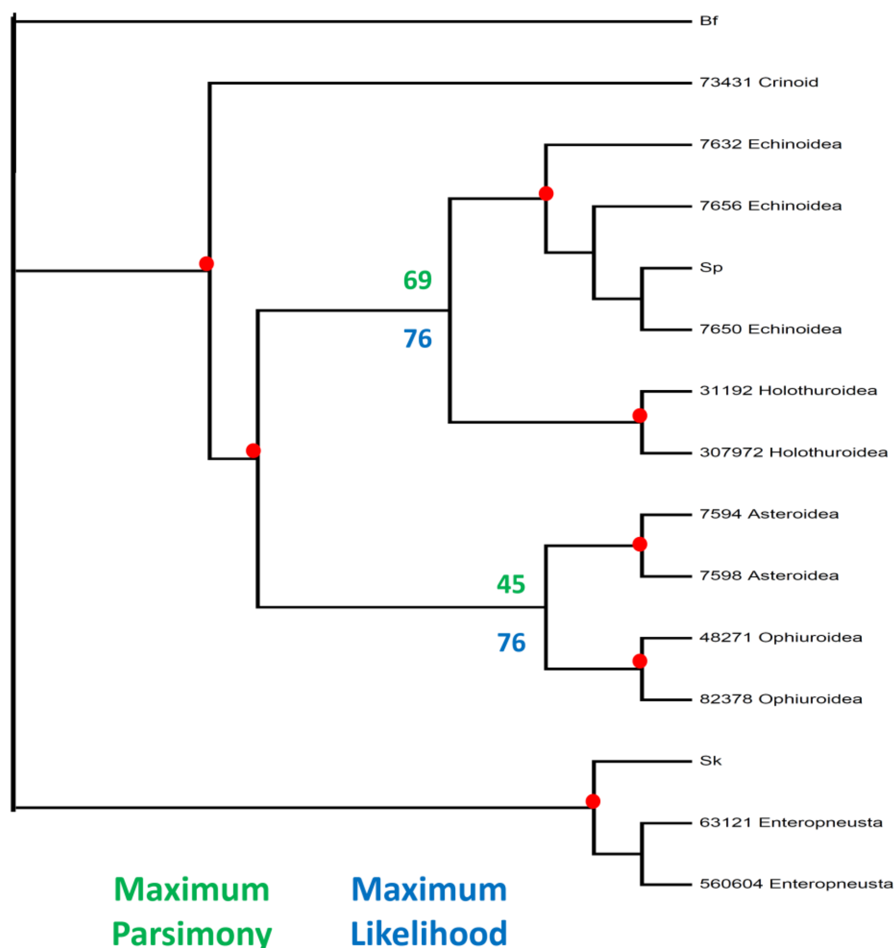



**Figure 3**: Echinoderm phylogeny, obtained with Telford *et al.* (2014) dataset, using Maximum Parsimony and Maximum Likelihood methods. Red circles represent a bootstrap value of 100 in both methods.

**Table 1**: Comparison of the studies that have investigated the phylogenetic relationships among echinoderm classes, in particular the relationships between Asteroidea (A), Ophiuroidea (O) and Echinozoa (Ez). [A] = Topology with Asteroidea most basal; [O] = Topology with Ophiuroidea most basal; [Az] = Topology with Asteroidea and Ophiuroidea grouped together forming the Asterozoa. C = Crinoids. NJ = Neighbour-Joining; MP = Maximum Parsimony; ML = Maximum Likelihood; BI = Bayesian Inference. When all methods gave the same results they are displayed on the same line. * for the two nodes involving (A), (O) and (Ez). ° inconsistent grouping with vertebrate outgroups with those methods. @ Numerous Missing species, varying according to molecular data set 8 (18S)- 32 (5' 28S)- 25 (3' 28S) / 48 spp. & Only unambiguously aligned positions were counted here.

| **Study** | **Data analyzed** | **Level of chimera or Missing data** | **Method(s)** | **Taxon nb in : E – H – A - O** | **Topology** | **Support and posterior probability * (ML – MP - BI)** |
|---|---|---|---|---|---|---|
| Smith (1988) | Morphology | No | Cladistics (parsimony) | not indicated | [A] | ND |
| Smith *et al.* (1993) | Mt gene order | No | Cladistics (parsimony) | 1 – 4 – 2 - 1 | [Az] | ND |
| Wada & Satoh (1994) | 18S rDNA (1720 nt) | No | NJ and ML | 1 – 1 – 1 – 1 | [O] NJ<br>[Az] ML | 0.34<br>ND |
| Mooi *et al.* (1994); David & Mooi (1999) | Morphology (skeletal homologies – axial extraxial skeletons) | No | Cladistics (parsimony) | not indicated | [Az] | ND |
| Littlewood *et al.* (1997) | Morphology (21 larval + 50 adult char.) | No | MP | 1 – 1 – 1 – 1 (consensus) | [A] (MPT/1 step !) | 0.68- *ND* |
| | Partial 28S (313[&] bp, 37 taxa), complete 18S (1,723 bp, 39 taxa) | No | MP-ML | (3-23) –(3-5) –(3-9) – (3-12) | No clear winner among 3 topologies | *0.48 (18S unstable /coding PuPy)* |

| | | | | | | |
|---|---|---|---|---|---|---|
| | Total evidence (cf above) | No | MP | 1 – 1 – 1 – 1 | [A] | 0.62 / *nd* |
| Janies (2001) | - 62 morphology and “gene order” char. | | MP (POY alignment optimization & dynamic homology) | 10 - 4 -15 – 7 | [A] | Other methods |
| | rDNA full 18S (and less species: 28S) | missing data for 28S | | 10 - 4 -15 – 7 | [O] | |
| | both combined, Morphology and genetic data | | | | [Az] | |
| Mallatt & Winchell (2007) | rDNA 3,925 bp | No? | ML – MP – BI | 2 - 1 - 1 – 1 | [Az] | 0.82 - <0.50 – 0.70 |
| | rDNA 4,275 bp | | ML - MP | 2 - 1 - 1 - 1 | [Az] | 0.69 - <0.50 - ND |
| Perseke *et al.* (2010) | Mt genomes Gene order | | MP (min nb of rearrangements) | 6 - 3 – 7 - 6 | (C,O) (A, Ez) | NA |
| | 13 mt-proteins 3,653–3,733 AA | No | ML, BI (CAT & MtRev), NJ°, MP°, | 6 - 3 – 7 - 6 | [O] | 0.74 - *1* |
| Letsch & Kjer (2011) | 18S rDNA (1706 bp) | No | ML (Model-specific-RNA covarion) | 35 – 12 – 55 – 12 | [Az] (or aberrant topology) | 0.41 - *nd* |
| Pisani *et al.* (2012) | Full dataset: 7 genes=2,049AA + 4,682 bp rDNA | Genus lev. chim- 12% (AA)-24% (rDNA) | Heterog. CAT-GTR & GTR+G BI, ML, NJ, MP | 5 - 0- 1- 1 | [Az] | ? - *0.97* |
| | Rate heterogeneous (fast | yes | Heterog. CAT-GTR & | 5 - 0- 1- 1 | [Az] | 0.79 - *1* |

| | | | | | | |
|---|---|---|---|---|---|---|
| | and constant): 1,247 AA and 3,206 bp positions | | GTR+G BI, ML, NJ, MP | | | |
| | rate homogeneous: 811 AA & 1,476 bp | yes | Heterog. CAT-GTR & GTR+G BI, ML, NJ, MP | 5 - 0- 1- 1 | [A] | ? - *0.93* |
| Telford *et al.* (2014) | 219 genes = 45,818 AA (only 194 positions shared by all taxa) | Echinoids: 2-90% Holothurians: 75-80% Asteroids: 45-65% Ophiuroids: 64-73% | BI | 4 – 2 – 2 - 2 | [Az] BI | 1 |
| | | | Analyses performed in this study: ML, MP, NJ | | [Az] ML, MP<br>[((H,O)A)E] NJ | 0.76 – 0.49<br>0.81 (H,O) |
| O'Hara *et al*. (2014) | 425 genes = 102,143 AA | Global: 15% | BI, ML | 3 – 1 – 2 - 5 | [Az] | 1 - 1 |
| This study | 138 genes = 17,318 AA | No | BI, ML, MP, NJ | 2 - 1- 1- 2 | [Az] | 1 - 0.93 – 1 – (1 NJ) |

## 8.5 Discussion

### 8.5.1 *Asterozoa strongly supported*

In this study, we used a complete matrix of 17,318 amino acids to confirm the phylogenetic position of ophiuroids in the echinoderm tree. We obtained a fully resolved tree supporting Asterozoa in each one of the reconstruction method used (BI, ML, MP, NJ), pointing out the unambiguous grouping of ophiuroids and asteroids. The fact that this topology is strongly supported with each reconstruction method highlights its robustness. Although heterogeneous base composition was detected, it should not have erroneously produced the observed topology because the taxa displaying the same base composition did not group together (e. g. echinoids and ophiuroids). This topology is congruent with the axial-extraxial theory, which is based on skeletal homologies (Mooi *et al.* 1994). Furthermore, the topology supporting Asterozoa is also the most frequently observed in the bibliographic review performed for this study, even if in most studies, especially in the least recent ones, support was low. Finally, the most recent study of O'Hara *et al.* (2014) showed maximal support for Asterozoa, using datasets including more than 400 concatenated genes.

### 8.5.2 *Difficulties in phylogenetic reconstruction*

In the light of these results, one could ask why inferring the position of ophiuroids in the echinoderm tree was a matter of debate. There are several reasons for this. The difficulty of phylogenetic reconstruction between a set of taxa increases with the age of their cladogeneses and the rapidity of cladogenesis events, because of the low signal to noise ratio (i.e. the ratio of internode lengths / terminal branch lengths) (e.g. Philippe *et al.* 2011). Another class of factors impairing reliable phylogeny inference is the presence of strong differences in mode of evolution among lineages, such as heterogeneity in rate of evolution (e.g. long branch artefact) or non stationarity in the probability of changes among character states (e.g. compositional differences among taxa). A way to overcome this problem could be the use of different reconstruction methods, namely the distance-based NJ, or character-based MP, ML or BI, to infer the existence of long branch artefact (LBA). Therefore, the congruence or incongruence of the topologies found with the different methods may reveal the presence of LBA. Another solution to overcome such heterogeneity in the mode of evolution among taxa

is to increase taxon sampling (e.g. Lecointre *et al.* 1993; Zwickl & Hillis 2002; Pollock *et al.* 2002; Heath *et al.* 2008) in order to obtain a set of homogeneously evolving species representing all the taxa that need to be compared, or to have a large enough diversity of models within each taxon, to disentangle the noise from the signal. Wide taxon sampling cannot provide an efficient solution when the common ancestors of each clade are relatively recent compared to the times of divergence among clades, since all members of a clade share a long history and thus represent very correlated samples of modes of evolution; if differences in modes of evolution arose before the split from the common ancestor, they cannot be "corrected" by large taxon sampling.

Both problems affect the reconstruction of phylogenetic relationships among echinoderm classes. The divergence of echinoderms dates back to the Cambrian explosion (540 MYA), and the earliest known fossil of extant classes date from the early Ordovician with primitive crinoids (485 MYA), asteroids (475-480 MYA) and late Ordovician for echinoids (450 Ma) (Bottjer *et al.* 2006). In addition, recent results on molecular dating (O'Hara *et al.* 2014) showed that divergence among eleutherozoa occurred very rapidly (in about 30 MYA). During these fast evolutionary radiations, single genes did not accumulate enough information to provide a strongly supported topology. Indeed, on the 138 studied genes taken individually, 121 displayed an unresolved topology. Concatenation of genes allowed to gather the information contained in each gene, raising the global resolution of the tree.

### 8.5.3 *Removal of fast-evolving and constant sites: Not a good idea?*

The example of echinoderm phylogeny highlights the possible bias that may occur when removing sites for analyses. Indeed, Pisani *et al*. (2012) used seven nuclear housekeeping genes and three rDNA genes to resolve the position of ophiuroids in the echinoderm tree. They found strong support (0.97) for the Asterozoa topology, yet the authors decided to remove some positions of the dataset (the fast evolving and invariant ones) and found a lower support (0.93) for another topology [A], with the asteroids more basal. The authors concluded that the [A] topology is the most probable one as their truncated dataset is supposed to be the best, which was not a bad reasoning. Indeed, it is acknowledged that spurious topologies can obtain strong support (in particular) when models of evolution are not adequate, even with large data sets (Philippe *et al.* 2011). This result points out that caution should be taken when

removing, rather arbitrarily, some parts of the dataset as inaccurate results could arise from analyses, as it is the case in echinoderm phylogeny. Indeed, this method of removing heterogeneously evolving sites was previously used on Archaea dataset, and led to the conclusion that Archaea and Bacteria were more closely related to each other compared to Eukarya (Brinkmann & Philippe 1999). Yet, it is now well established that Archaea and Eukarya are more closely related to each other compared to Bacteria (e.g. Ciccarelli *et al.* 2006). We therefore recommend not removing any sites for phylogenetic reconstruction, even if it theoretically makes sense, as one can never know which sites are important or not for phylogenetic reconstructions.

### 8.5.4 *Impact of missing data on phylogenetic reconstruction*

Telford *et al.* (2014) used one to four echinoderm species per class and a matrix of 45,818 positions, based on BI analyses. In the present study, using their dataset, we showed that the Asterozoa clade was not well supported in ML and MP, whereas the NJ analysis yielded a different topology. This low robustness of results certainly arose from the striking high level of missing data in their dataset (>60% in nine species out of fifteen species used, and whole echinoderm classes being represented only by a very partial data set). The level of missing data is a recurrent issue in phylogenomic studies, where tens of species are used with a few hundreds of genes. Increasing taxon sampling is tempting to display numerous species, yet it is done at the cost of a higher level of missing data and thus leading to loss of resolution with methods sensitive to missing data (NJ, MP, ML). A way of dealing with high levels of missing data could be to use Bayesian inference as reconstruction method, since it is less sensitive to missing data, yet BI can display very strong supported topologies that are actually inaccurate. Therefore, validation using another method (e.g. ML) is recommended (Douady *et al.* 2003). In contrast, O'Hara *et al.* (2014) paid particular attention in dataset completeness by removing positions including too many missing data. Using a 85% complete dataset, they obtained a fully resolved topology using ML and BI. Although displaying some missing data, their dataset robustly supported Asterozoa topology.

In this study, we focused on the completeness of the dataset (i.e. absence of missing positions) to infer echinoderm phylogeny, and compare our results with previous studies on the topic. Using few species displaying complete dataset appears particularly justified for the question of Echinoderm class relationships, since the within

class-radiations are very recent compared to their divergence, so increasing taxon sampling does not improve (or much less than generally expected) the power and reliability of the analyses. Indeed, all four phylogenetic methods are congruent and display strong to maximal support for Asterozoa topology.

To conclude, our study demonstrated that ancient yet rapid divergence events could be successfully resolved using phylogenomics. Yet, we also showed that the impact of missing data could greatly influence the support of results. We also emphasize that taxon sampling may not be a solution for such situations, and that in such cases, reduced taxon sampling is acceptable, provided that factors such as composition and rate heterogeneity are carefully examined and taken into account as far as possible. Finally, morphological-based analyses should not be put aside for phylogenetic reconstructions, but rather used in combination with molecular data, since some authors (Mooi *et al.* 1994) already recovered the true topology classes 20 years ago using skeletal homologies, which appeared, at least for some morphologists, definitely unambiguous (David B., pers. com.).

# Annexe 2

LETTER

doi:10.1038/nature13685

# Comparative population genomics in animals uncovers the determinants of genetic diversity

J. Romiguier[1,2], P. Gayral[1,3], M. Ballenghien[1], A. Bernard[1], V. Cahais[1], A. Chenuil[4], Y. Chiari[5], R. Dernat[1], L. Duret[6], N. Faivre[1], E. Loire[1], J. M. Lourenco[1], B. Nabholz[1], C. Roux[1,2], G. Tsagkogeorga[1,7], A. A.-T. Weber[5], L. A. Weinert[1,8], K. Belkhir[1], N. Bierne[1], S. Glémin[1] & N. Galtier[1]

**Genetic diversity is the amount of variation observed between DNA sequences from distinct individuals of a given species. This pivotal concept of population genetics has implications for species health, domestication, management and conservation. Levels of genetic diversity seem to vary greatly in natural populations and species, but the determinants of this variation, and particularly the relative influences of species biology and ecology versus population history, are still largely mysterious[1,2]. Here we show that the diversity of a species is predictable, and is determined in the first place by its ecological strategy. We investigated the genome-wide diversity of 76 non-model animal species by sequencing the transcriptome of two to ten individuals in each species. The distribution of genetic diversity between species revealed no detectable influence of geographic range or invasive status but was accurately predicted by key species traits related to parental investment: long-lived or low-fecundity species with brooding ability were genetically less diverse than short-lived or highly fecund ones. Our analysis demonstrates the influence of long-term life-history strategies on species response to short-term environmental perturbations, a result with immediate implications for conservation policies.**

Since the early studies of evolutionary genetics, there has been no understanding of how and why genetic diversity levels vary between species. This old puzzle, considered four decades ago as 'the central problem in population genetics'[1], is still essentially unsolved in the genomic era[2]. Meanwhile, there is increasing evidence that genetic diversity is central to many conservation challenges, such as species response to environmental changes, ecosystem recovery, and the viability of recently endangered populations[3–7]. In this context, our ability to understand and predict this key aspect of biodiversity seems critical. But is it possible to quantify the contributory ecological and genetic factors? How predictable is the level of genetic diversity of a given species?

Population genetic theory states that neutral genetic polymorphism (that is, diversity) increases with effective population size, $N_e$, which in a panmictic population is equal to the number of individuals contributing to reproduction. One would therefore expect the genetic diversity of a species to be linked to biological traits associated with abundance, such as body size or fecundity. However, this intuitive prediction has not yet been clearly confirmed by empirical data[2,8–10]. This is typically explained by invoking the many confounding factors potentially affecting genetic polymorphism, such as mutation rate, population structure, population bottlenecks, selective sweeps, and, more generally, ecological disturbances[11,12]. Whether demographic or adaptive, historical contingency is often considered to be the main driver of genetic diversity[11]. According to this viewpoint, polymorphism levels would be expected to fluctuate in time more or less randomly, irrespective of life-history traits.

In the absence of compelling empirical evidence, the relative importance of species biology and ecology (on the one hand) and historical, contingent factors (on the other) in shaping the genetic diversity of species is still highly contentious. Indeed, current knowledge on species genetic diversity is based on just a handful of model organisms, or small sets of molecular data[2,8,13]. Various animal taxa and lifestyles, particularly across the invertebrates, have yet to be explored. Here we fill this gap and present the first distribution of genome-wide polymorphism levels across the metazoan tree of life.

We focused on 31 families of animals spread across eight major animal phyla. In each family we produced high-coverage transcriptomic data (RNAseq) for about ten individuals of a particular species. In 25 of these families, we sampled one to three additional species of similar biology and ecology (two to seven individuals each), thus producing taxonomic replicates. The total data set consisted of 374 individual transcriptomes from 76 non-model species (Fig. 1, Extended Data Fig. 1 and Supplementary Tables 1 and 2), from which we predicted protein coding sequences[14] and identified diploid genotypes and single nucleotide polymorphisms[15,16] (Methods). Across species the number of analysed loci varied from 804 to 20,222 (median 5,347) and the number of polymorphic sites from 1,759 to ~230,000 (median 17,924).

Estimates of the synonymous nucleotide diversity ($\pi_s$) spanned two orders of magnitude across species, a range far wider than is usually observed in surveys restricted to a single taxonomic group. The extreme values of $\pi_s$ were observed in two invertebrate species: 0.1% in the subterranean termite *Reticulitermes grassei*; 8.3% in the slipper shell *Bostrycapulus aculeatus*. Figure 1 illustrates the patchy distribution of low-diversity (green) and high-diversity (red) species across the metazoan phylogeny. It also shows that species $\pi_s$ values tend to be similar within families, but distinct between families (analysis of variance; $P < 10^{-12}$). Such a strong taxonomic effect would be unexpected if stochastic disturbances and contingent effects were the main drivers of genetic diversity, because species from a given family are not particularly expected to share a common demographic history. Testing this hypothesis more thoroughly, we detected no strong relationship between $\pi_s$ and any variable related to geography, such as the average distance between GPS records (regression test, $P = 0.19$, $r^2 = 0.02$), maximum distance between GPS records ($P = 0.02$, $r^2 = 0.07$), average distance to Equator ($P = 0.87$, $r^2 = 0.0003$), population structure (measured as $F_{it}$, $P = 0.22$, $r^2 = 0.02$), invasive status (Student's $t$-test, $P = 0.14$) and marine versus continental environment (Student's $t$-test, $P = 0.52$).

To test whether species biology can explain variations in $\pi_s$, we collected data for six life-history traits potentially related to $N_e$: adult size, body mass, maximum longevity, adult dispersion ability, fecundity and propagule size (Supplementary Table 3). In contrast to the geographic variables, all these traits were significantly correlated with the nucleotide diversity (Extended Data Fig. 2) and collectively explained 73% of the variance in $\pi_s$ in a multiple linear regression test ($P < 10^{-10}$). Propagule size, here defined as the size of the stage that leaves its parents and disperses (egg or juvenile depending on species), is by far the most predictive

[1]UMR 5554, Institute of Evolutionary Sciences, University Montpellier 2, Centre national de la recherche scientifique, Place E. Bataillon, 34095 Montpellier, France. [2]Department of Ecology and Evolution, Biophore, University of Lausanne, 1015 Lausanne, Switzerland. [3]UMR 7261, Institut de Recherches sur la Biologie de l'Insecte, Centre national de la recherche scientifique, Université François-Rabelais, 37200 Tours, France. [4]Aix-Marseille Université, Institut Méditerranéen de Biodiversité et d'Écologie marine et continentale (IMBE) – CNRS – IRD – UAPV, 13007 Marseille, France. [5]Department of Biology, University of South Alabama, Mobile, Alabama 36688-0002, USA. [6]UMR 5558, Laboratoire de Biométrie et Biologie Evolutive, Université Lyon 1, CNRS, 69622 Lyon, France. [7]The School of Biological and Chemical Sciences, Queen Mary University of London, Mile End Road, London E1 4NS, UK. [8]Department of Veterinary Medicine, University of Cambridge, Madingley Road, Cambridge CB3 0ES, UK.



©2014 Macmillan Publishers Limited. All rights reserved

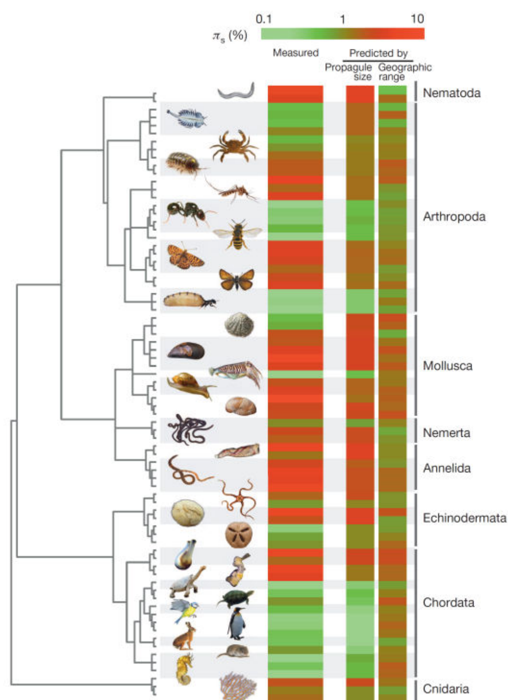



**Figure 1 | Genome-wide genetic diversity across the metazoan tree of life.** Each branch of the tree represents a species ($n = 76$). The leftmost vertical coloured bar is the estimated genome-wide genetic diversity ($\pi_s$), the central bar is the prediction of $\pi_s$ based on a linear model with propagule size as the explanatory variable ($P < 10^{-14}$, $r^2 = 0.56$), and the rightmost bar is the prediction of $\pi_s$ based on a linear model with average distance between GPS records, maximal distance between GPS records, average distance to Equator and invasive status as explanatory variables ($P = 0.16$). Each thumbnail corresponds to one metazoan family. Species are in the same order as in Supplementary Table 2 (from top to bottom).

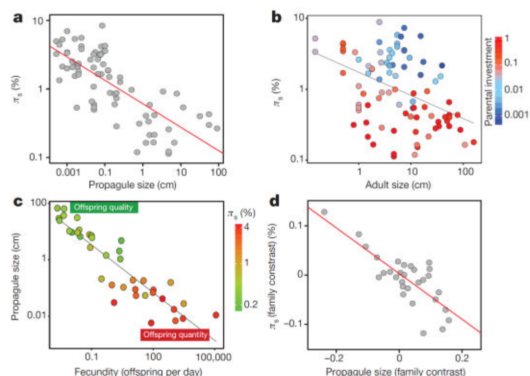



**Figure 2 | Life-history traits and genetic diversity relationships.** **a**, Relationship between propagule size and $\pi_s$ ($P < 10^{-14}$, $r^2 = 0.56$, 76 species included; see Fig. 1). **b**, Relationship between adult size and $\pi_s$ ($P < 0.05$, $r^2 = 0.07$, 76 species included). The colour scale represents the degree of parental investment, here defined as the ratio of propagule size to adult size. **c**, Effect of fecundity per day (*x* axis) and propagule size (*y* axis) on genetic diversity (colour scale; $P < 10^{-6}$, $r^2 = 0.69$, 29 family-averaged data points). **d**, Phylogenetic contrasts of family-averaged $\pi_s$ versus family-averaged propagule size ($P < 10^{-6}$, $r^2 = 0.62$).

of these variables (linear regression test, $r^2 = 0.56$; Fig. 2a). This is illustrated in Fig. 1 by the good agreement between the observed distribution of $\pi_s$ (leftmost coloured vertical bar) and the $\pi_s$ value predicted from propagule size (central bar). The predicted $\pi_s$ based on four demographic metrics is plotted alongside (rightmost bar) for visual comparison.

We explored in more detail the relative impact on $\pi_s$ of the various life-history traits of interest here (Extended Data Fig. 2). Figure 2b plots the relationship between $\pi_s$ and species adult size, a variable typically taken as a proxy for population size in some taxa[9]. Although significant, the correlation is not particularly strong ($P = 0.018$, $r^2 = 0.07$). In particular, species with low genetic diversity cover a large range of body sizes, from less than 1 cm to more than 1 m. Low-polymorphism species include amniotes (turtles, mammals and birds), but also brooding marine species (seahorses, brooding urchins, nemerteans and brittle-stars), eusocial insects (ants, bees and termites) and cuttlefish. These phylogenetically unrelated species have in common a large parental investment in their offspring, as represented in Fig. 2b by the ratio of propagule size to adult size (red). In contrast, species with minimal parental investment (blue) tend to carry high genetic diversity given their size. This is typically the case of highly fecund, broadcast spawning sessile species (such as mussels, non-brooding urchins, nemerteans and brittle-stars, sea squirts and gorgonians). The trade-off between offspring quantity (fecundity) and quality (propagule size) seems to be the most relevant factor explaining variations in polymorphism between species in the animal kingdom (Fig. 2c). We shall for simplicity hereafter categorize as *K*-strategists the species that tend to invest in the quality of their progeny, and as *r*-strategists those that favour quantity[17].

The correlation we report between life-history traits and $\pi_s$ is not due to phylogenetic non-independence of the sampled species: taking family averages from Fig. 1 increased the correlation coefficients (from $r^2 = 0.56$ to $r^2 = 0.66$ with propagule size alone, from $r^2 = 0.73$ to $r^2 = 0.79$ with the six life-history traits). When we took into account the between-family phylogenetic tree using independent contrasts, this still resulted in highly significant correlations between $\pi_s$ and life-history traits ($r^2 = 0.62$ for propagule size; Fig. 2d and Extended Data Fig. 3). These relationships were also unaffected by sampling strategy, sequencing depth, gene expression levels or contaminants (Methods, Supplementary Table 4 and Extended Data Figs 4–6). Finally, our conclusions were unchanged when we included 14 previously published species of mammals[10] or when we restricted the analysis to a subset of common orthologous genes (Supplementary Table 4).

The relationship between $\pi_s$ and life-history traits, however strong, could in principle be mediated by causative variables that were not included in the analysis. One of these potential confounding factors is the mutation rate: a higher average per-generation mutation rate in *r*-strategists than in *K*-strategists could explain our results irrespective of the population size effect. However, theoretical models and empirical measurements actually support the opposite; that is, an increased per-generation mutation rate in large, long-lived organisms due to a larger number of germ-line cell divisions per generation and a reduced efficacy of natural selection on the fidelity of polymerases[18]. Therefore, as far as we can tell, across-species variations of mutation rate are likely to oppose, not strengthen, the main effect we are reporting here.

We computed the non-synonymous nucleotide diversity, $\pi_n$, and this was also found to be correlated with species life-history traits (Extended Data Fig. 2). We found substantial variation in $\pi_n/\pi_s$ across metazoan species, and significant correlations with life-history traits, the best predictor in this case being longevity (Extended Data Fig. 7). This positive correlation is predicted by the nearly neutral theory of molecular evolution[19]: in small populations (long-lived species), the enhanced genetic drift counteracts purifying selection and promotes the segregation of weakly


©2014 Macmillan Publishers Limited. All rights reserved

deleterious, non-synonymous mutations at high allele frequency. These results also confirm that the relationships we uncovered between life-history traits and diversity patterns are mediated in the first place by an effect of $N_e$, not of the mutation rate; synonymous and non-synonymous positions being physically interspersed, the $\pi_n/\pi_s$ ratio is unaffected by the mutation rate.

Our analysis reveals that polymorphism levels are well predicted by species biology, whereas historical and contingent factors are only minor determinants of the genetic diversity of a species. This unexpected result opens new questions. How can life-history traits be so predictive of $\pi_s$ in spite of the overwhelming evidence for the impact of ecological perturbations on patterns of genetic variation[11,12]? Why does the '$r/K$ gradient' affect genetic polymorphism so strongly?

In an attempt to resolve these paradoxes, we suggest that life-history strategies might influence the response of species to environmental perturbations. Because $K$-strategy species have been selected for survival and the optimization of offspring quality in complex, stable environments[17], we speculate that they might experience fewer occasional disturbances (or be less sensitive to them), thus ensuring the long-term viability of even small populations. In contrast, only species with a large population-carrying capacity could sustain the 'riskier' $r$-strategy in the long term, thus buffering the frequent bottlenecks experienced in the context of high environmental sensitivity (see Supplementary Equations for a model formalizing these arguments). According to this hypothesis, environmental perturbations would be a common factor affecting every species, but their demographic impact would depend on the life-history strategy of each species.

This study highlights the importance of species life-history strategy when it comes to turning genetic diversity measures into conservation policy. So far, conservation efforts have mainly been focused on large-sized vertebrates. Here we show that these popular animals represent only a subset of the existing low-diversity, $K$-strategists. Invertebrate species with strong parental investment are probably equally vulnerable to genetic risks. Our results also indicate that $r$-strategists will typically show elevated amounts of genetic diversity irrespective of their current demography, which suggests that species of this kind might face significant extinction risks[20] without any warning genetic signal.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 3 March; accepted 17 July 2014.**
**Published online 20 August 2014.**

1. Lewontin, R. *The Genetic Basis of Evolutionary Change* (Columbia Univ. Press, 1974).
2. Leffler, E. M. *et al.* Revisiting an old riddle: what determines genetic diversity levels within species? *PLoS Biol.* **10,** e1001388 (2012).
3. Keller, L. F. & Waller, D. M. Inbreeding effects in wild populations. *Trends Ecol. Evol.* **17,** 19–23 (2002).
4. Reusch, T. B. H., Ehlers, A., Hämmerli, A. & Worm, B. Ecosystem recovery after climatic extremes enhanced by genotypic diversity. *Proc. Natl Acad. Sci. USA* **102,** 2826–2831 (2005).
5. Hughes, A. R., Inouye, B. D., Johnson, M. T. J., Underwood, N. & Vellend, M. Ecological consequences of genetic diversity. *Ecol. Lett.* **11,** 609–623 (2008).
6. Johnson, W. E. *et al.* Genetic restoration of the Florida panther. *Science* **329,** 1641–1645 (2010).
7. Nair, P. Conservation genomics. *Proc. Natl Acad. Sci. USA* **111,** 569 (2014).
8. Bazin, E., Glémin, S. & Galtier, N. Population size does not influence mitochondrial genetic diversity in animals. *Science* **312,** 570–572 (2006).
9. Nabholz, B., Mauffrey, J.-F., Bazin, E., Galtier, N. & Glemin, S. Determination of mitochondrial genetic diversity in mammals. *Genetics* **178,** 351–361 (2008).
10. Perry, G. *et al.* Comparative RNA sequencing reveals substantial genetic variation in endangered primates. *Genome Res.* **22,** 602–610 (2012).
11. Banks, S. C. *et al.* How does ecological disturbance influence genetic diversity? *Trends Ecol. Evol.* **28,** 670–679 (2013).
12. Hewitt, G. The genetic legacy of the Quaternary ice ages. *Nature* **405,** 907–913 (2000).
13. Ellegren, H. Genome sequencing and population genomics in non-model organisms. *Trends Ecol. Evol.* **29,** 51–63 (2014).
14. Cahais, V. *et al.* Reference-free transcriptome assembly in non-model animals from next-generation sequencing data. *Mol. Ecol. Resources* **12,** 834–845 (2012).
15. Tsagkogeorga, G., Cahais, V. & Galtier, N. The population genomics of a fast evolver: high levels of diversity, functional constraint, and molecular adaptation in the tunicate *Ciona intestinalis. Genome Biol. Evol.* **4,** 740–749 (2012).
16. Gayral, P. *et al.* Reference-free population genomics from next-generation transcriptome data and the vertebrate-invertebrate gap. *PLoS Genet.* **9,** e1003457 (2013).
17. MacArthur, R. H. & Wilson, E. O. *The Theory of Island Biogeography* (Princeton Univ. Press, 1967).
18. Lynch, M. Evolution of the mutation rate. *Trends Genet.* **26,** 345–352 (2010).
19. Ohta, T. Slightly deleterious mutant substitutions in evolution. *Nature* **246,** 96–98 (1973).
20. Pinsky, M. L. *et al.* Unexpected patterns of fisheries collapse in the world's oceans. *Proc. Natl Acad. Sci. USA* **108,** 8317–8322 (2011).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank the following for providing samples: F. Delsuc, E. Douzery, M. Tilak, G. Dugas, S. Harispe, C. Benoist, D. Bouchon (woodlice), J. Bierne, M. Bierne, B. Houseaux, M. Strand, C. Lemaire, D. Lallias, Service Modèle Biologique Station Marine Roscoff (nemertines), X. Turon, S. Lopez-Legentil (Cystodytes), P. Jarne, P. David, R. Dillon, J. Auld, R. Relyea, C. Lively, J. Jokela, V. Poullain, T. Stewart (snails), S. Lapègue, V. Boulo, F. Batista, D. Lallias, L. Fast Jensen, M. Cantou (oysters), J. Do Nascimento, C. Daguin-Thiébaut, M. Cantou (crabs), L. Bonnaud (cuttlefish), D. Aurelle (gorgonians), F. Viard, Y. Pechenik, A. Cahill, R. Colins (slipper limpets), L. Dupont (earthworms), D. Jollivet (trumpet worms), M. A. Felix, I. Nuez (nematodes), N. Rodes, T. Lenormand, E. Flaven (brine shrimps), Rotterdam Zoo, Zurich Zoo, C. Libert, Montpellier Zoo, S. Martin, la Ferme aux Crocodiles, O. Verneau, C. Ayres, M. Carretero, M. Vanberger, K. Pobolsaj, M. Zuffi, C. Palacios, L. du Preez, B. Halpern, Budapest Zoo (turtles), P. Peret, C. Doutrelant, B. Halpern, B. Rosivall (tits), M. de Dinechin, B. Rey (penguins), Z. Melo-Fereira, P. Alves (hares), N. Brand, M. Chapuisat (bees), R. Blatrix, A. Lenoir, I. Nodet, A. Lugagne, S. Blanquart, L. Serres-Giardi, V. Roustang, N. François, G. Ballantyne, A. Carbonnel, Y. Samuel, G. James, G. Kalytta, F. Guerrini, S. Stenzel, J. Beekman, X. Cerda, S. Ikoen (ants), I. Hanski, S. Ikonen, J. Kullberg, Z. Kolev (fritillary butterflies), F. Viard, X. Turon, Di Jiang, D. Chourrout, B. Vercaemer, E. Newman-Smith, Ascidian Stock Center, Service Modèle Biologique Station Marine Roscoff (ciona), L. Excoffier, G. Heckel (voles), F. Dedeine (termites), C. Atyame, O. Duron, M. Weill (mosquitoes), M. Cantou, H. Violette, F. Batista, J. Hondeville (seahorses), C. Fraïsse, G. Pogson, N. Saarman, J. Normand (mussels), E. Poulin, C. Gonzalez-Weivar, and J. P. Feral (sea urchins). This work was supported by European Research Council advanced grant 232971 (PopPhyl).

**Author Contributions** N.G. conceived the project. P.G., M.B., N.F., Y.C., L.A.W., G.T., A.C., A.W., J.R., N.G. and N.B. performed sampling and laboratory work. A.B., V.C., E.L., J.R., J.M.L., C.R., P.G., G.T., B.N., R.D., K.B., S.G. and N.G. developed the data analysis pipeline. J.R. collected life-history/geographic variables and produced figures. J.R., A.B., V.C., L.D., E.L. and N.G. analysed the data. S.G., N.B., B.N., J.R. and N.G. provided interpretations and models. J.R., N.B., S.G. and N.G. wrote the paper.

**Author Information** Data sets are freely available from the Sequence Read Archive (SRA) database (http://www.ncbi.nlm.nih.gov/sra) under project ID SRP042651 and from the Datasets section of the PopPhyl website (http://kimura.univ-montp2.fr/PopPhyl), in which predicted single nucleotide polymorphisms and genotypes are provided as .vcf files. Scripts and executable files are freely available from the Tools section of the PopPhyl website. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to N.G. (nicolas.galtier@univ-montp2.fr).



©2014 Macmillan Publishers Limited. All rights reserved

## METHODS

**Sampling and sequencing.** The 76 analysed species were selected based on phylogenetic and ecological criteria with the goal of gathering a panel representative of the metazoan diversity. In each species, from two to eleven individuals were collected in various localities of their natural geographic range (Supplementary Table 1) or from zoos (*Chelonoidis nigra*). Depending on the species, the whole body, body parts, organs or tissues were dissected and preserved in RNA later, flash-frozen, or processed immediately (Supplementary Table 2). For each sample, total RNA was extracted using standard and improved protocols[21], and a non-normalized complementary DNA library was prepared. The libraries were sequenced on a Genome Analyzer II or Hiseq 2000 (Illumina) to produce 100-base-pair (bp) or 50-bp single-end fragments. In 12 species, an additional normalized random-primed cDNA library was prepared and sequenced for half a run using a 454 Genome Sequencer (GS) FLX Titanium Instrument (Roche Diagnostics). Illumina reads from two to four individuals in 14 mammalian species[10] were downloaded from the SRA database. Reads were trimmed of low-quality terminal portions with the SeqClean program (http://compbio.dfci.harvard.edu/tgi/). The fastQ program was applied to Illumina reads and revealed only a limited amount of motif enrichment: the number of motifs in significant excess varied between 0 and 17 across species, its median being 1.

**Transcriptome assembly, read mapping, coding sequence prediction.** *De novo* transcriptome assembly based on the 454 (when available) and Illumina reads was performed by following strategies B and D in ref. 14, using a combination of the programs Abyss and Cap3. Illumina reads were mapped to predicted cDNAs (contigs) with the BWA program. Contigs with a per-individual average coverage below $\times 2.5$ were discarded. Open reading frames (ORFs) were predicted with the Trinity package. Contigs carrying no ORF longer than 200 bp were discarded. In contigs including ORFs longer than 200 bp, 5′ and 3′ flanking non-coding sequences were deleted, thus producing predicted coding sequences that are hereafter referred to as loci.

**Calling single nucleotide polymorphisms (SNPs) and genotypes.** At each position of each locus and for each individual, diploid genotypes were called according to the method described in ref. 15 (model M1) and improved in ref. 16, using the reads2snps program. This method first estimates the sequencing error rate in the maximum-likelihood framework, calculates the posterior probability of each possible genotype, and retains genotypes supported at >95%; otherwise missing data are called. A minimum of ten reads per position and per individual were required to call a genotype. Then polymorphic positions were filtered for possible hidden paralogues (duplicated genes), using a likelihood ratio test based on explicit modelling of paralogy[16]. The across-species average percentage of SNPs discarded for suspicion of paralogy was 7.65%. Positions at which a genotype could not be called in a sufficient number of individuals were discarded. Calling $k$ the number of sampled individuals for a given species ($2 \leq k \leq 11$), the minimum number of genotyped individuals required to retain a position was set to $k/2$ when $k > 5$, to $k - 1$ when $k = 4$ or $k = 5$, and to $k$ when $k < 4$.

**Control for contamination.** Each locus of each species was translated to protein and compared with the non-redundant NR database using BlastP in search for possible contaminants. The percentage of loci for which no significant hit (e-value <0.001) was retrieved varied greatly between species, reflecting the taxonomic representation of sequences in NR. The percentage of no-hits was below 10% in mammals, but reached values above 20% in echinoderms and cnidarians. When one or several hits were found, the GenBank taxonomy of the first hit was recorded. Overall, 98.7% of first hits were assigned to Metazoa. The percentage of non-metazoan first hits was below 2% in 63 species out of 76, reaching its maximum (5.5%) in the trumpet worm *Pectinaria koreni*. Contamination by known microbes therefore seems negligible in our data set. Extended Data Figure 6 displays the taxonomic distribution of hits for four representative species—a mammal, an insect, a mollusc and an annelid. The results of our analyses were qualitatively unchanged when we removed loci that hit a non-metazoan and/or no-hit loci. In all cases, linear regression tests between $\pi_s$ and propagule size yielded the same $r^2$ of 0.55 ($P < 10^{-13}$) as in our main analysis.

**Life-history, ecological and geographical variables.** Species life-history traits (adult size, juvenile size, body mass, longevity, fecundity and adult speed) were retrieved from the literature (Supplementary Table 3). Invasive/non-invasive status was obtained from the Global Invasive Status Database (http://www.issg.org/database/species/List.asp). The Global Biodiversity Information Facility database (http://www.gbif.org/) was queried to retrieve the GPS records corresponding to documented observations of individuals from the species of interest here. These data were merged with the GPS records of our own samples. For each species, the average and maximum distance between two distinct GPS records and the average distance to the Equator were computed (Supplementary Table 2).

**Statistical analyses.** Population genomic statistics $\pi_s$, $\pi_n$ and $F_{it}$ were calculated by using home-made programs that rely on the Bio++ libraries[22]. The $F_{it}$ calculation was corrected for small sample sizes in accordance with ref. 23. Confidence intervals were obtained by bootstrapping loci. Regression analyses were conducted in R. Variables were log-transformed before linear regressions were performed. The linear model including $\pi_s$ and propagule size alone met the required assumptions of normally distributed residual errors (Shapiro's test, $P = 0.19$) and homoscedasticity (Fligner–Killeen's test, $P = 0.48$). The same remark is valid for the multiple linear model including $\pi_s$ and the six life-history traits (Shapiro's test, $P = 0.31$; Fligner–Killeen's test, $P = 0.49$). Family-level phylogenetic independent contrast analysis was performed with the APE package based on the tree shown in Extended Data Fig. 3, in which branch lengths are proportional to time. Divergence time estimates were retrieved from the TimeTree database (expert result, or average value if expert result was missing). When divergence time estimates were not available (Polycitoridae–Cionidae, Hesperiidae–Nymphalidae, Calyptraeoidae–Physidae, Mytilidae–Ostreidae), they were inferred on the basis of the divergence dates of neighbouring nodes.

**SNP calling quality controls.** The main analyses of this study were reproduced in three ways: first, with an increased minimum number of reads per position per individual of 30 instead of 10, second, removing five bases from each end of each read, and third, not using 454 data, thus controlling for a potential effect of insufficient sequencing depth, low-quality base calls near read ends and sequencing technology. In all three cases the results were highly similar to the main analysis (Supplementary Table 4; columns 'depth = 30X', 'clip_ends' and 'no_454', respectively), indicating that the analysis was robust to these technical caveats. No difference in $\pi_s$ was detected between species showing versus not showing a significant excess of certain motifs by fastQC.

**GC content.** In each species, the correlation coefficient between contig GC content and contig $\pi_s$ was calculated. It was significantly positive in 37 species, significantly negative in 18 species, and non-significantly different from zero in 21 species. The squared correlation coefficients ($r^2$) were relatively low (median $r^2$ 0.007; maximum $r^2$ 0.16 in *Physa acuta*), suggesting only a weak effect of GC content on $\pi_s$. For each species, the average contig GC content was calculated and correlated to the average $\pi_s$ or propagule size. No significant relationship was detected, which indicates that the variation in GC content across genes and across species has no substantial impact on the results of this study.

**Individual and locus sampling.** No significant relationship was found between $\pi_s$ and the number of sampled individuals per species, or between $\pi_s$ and the number of sampled loci per species (Extended Data Fig. 4). The robustness to sampling of the relationship between propagule size and $\pi_s$ was further assessed in two ways. First, for each species, loci were randomly subsampled. Extended Data Figure 5 displays the squared coefficient of correlation between propagule size and $\pi_s$ as a function of the per species number of analysed loci. It shows that as few as 50 loci are enough to capture the relationship with a good probability. Second, for each species, only two individuals were randomly selected and the analyses were conducted again. Results were highly similar to the main analysis: the relationship between propagule size and $\pi_s$ was unchanged and highly significant ($P < 10^{-15}$, $r^2 = 0.55$), thus indicating that population sample size is not an issue.

**Orthologous genes.** The coding sequences of 129 genes or gene fragments previously identified as orthologous across metazoans[24] (hereafter called 'core genes') were downloaded. In each of our species, contigs predicted to be orthologous to one of the core genes by reciprocal best BLAST hit were selected (expected e-value 0.0001, hits with a number of identical matches less than 80 and a bitscore of less than 1,200 were discarded). The number of such predicted core gene orthologues varied between 40 and 122 among species. We restricted the data set to the 39 species including at least 21 core gene orthologues, and reproduced the analysis. Colon 'orthologues' from Supplementary Table 4 shows that the correlation between $\pi_s$ and life-history traits was still strong and significant when a subset of common genes was considered.

**Expression level.** In each species, the expression level of each locus was estimated as the average number of bases read per position. Correlating $\pi_n/\pi_s$ to expression level across genes revealed no significant relationship in 33 species, and weak relationships ($r^2 < 0.27$) in 57 species. The relationship between $\pi_n/\pi_s$ and expression level, when detected, was negative, as expected under the hypothesis or a stronger selective pressure acting on high-expressed genes. Then, for each species, loci were grouped into three equal-sized bins of genes with high, medium or low expression. Each of these categories, taken separately, provided a strong correlation between species propagule size and $\pi_s$ ($r^2 = 0.57$, 0.62 and 0.62, respectively).

**Linkage.** The diversity of a neutral locus might be affected by selection at linked sites. This is particularly true of synonymous sites, which reside within coding sequences; that is, targets for natural selection. One would therefore predict a lower genetic diversity in species experiencing a low genomic average recombination rate. We lacked a recombination map in most of the analysed species; we therefore relied on generic taxonomic patterns to approach this issue. Eusocial hymenopterans are known to experience a recombination rate one order of magnitude higher than most animals[25]. In contrast, dipteran Culicidae (mosquitoes) experience relatively small amounts of recombination (median recombination rate in eusocial hymenoptera, 9.7 centimorgans per megabase (cM $\text{Mb}^{-1}$); median recombination rate in Culicidae, 0.3 cM $\text{Mb}^{-1}$)[25]. The linkage effect would therefore predict a decreased $\pi_s$ in Culicidae and an elevated $\pi_s$ in eusocial hymenopterans. We observed the opposite: $\pi_s$ varied from 0.0016 to 0.0058 in our five eusocial hymenopteran species, which is below the average metazoan

©2014 Macmillan Publishers Limited. All rights reserved

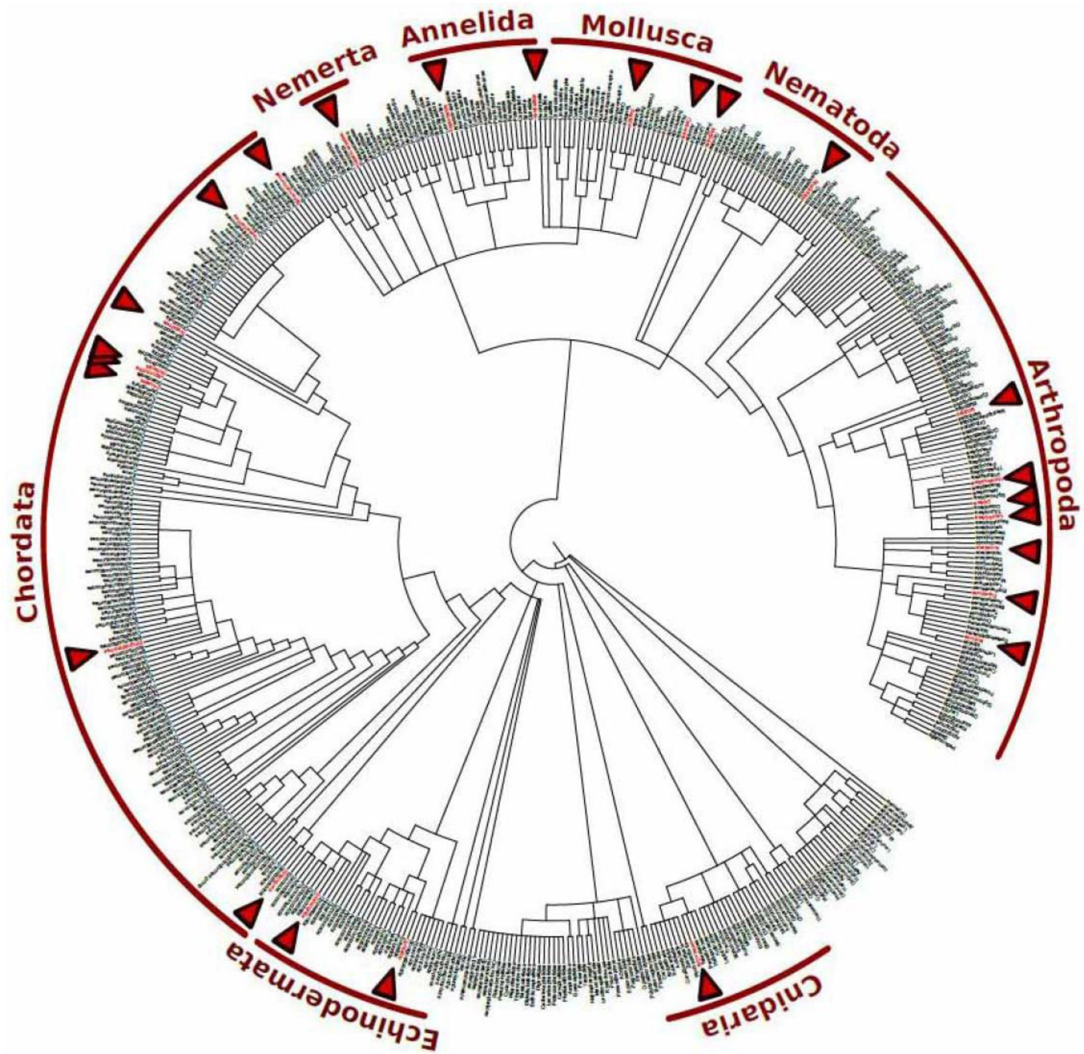
$\pi_s$ (0.015), and one order of magnitude below the $\pi_s$ of our three Culicidae species (0.016–0.041). This result, which is consistent with the propagule-size hypothesis, does not suggest that the between-species variation in genomic average recombination rate strongly influences our results.

**Population structure.** The genetic distance between individuals was defined as $(H_b - H_w)/H_w$, where $H_b$ is the probability of drawing two distinct alleles when sampling one copy from each of the two considered individuals, and $H_w$ is the average heterozygosity of the two considered individuals. In species containing more than four individuals, the genetic distance was calculated for each pair of individuals and correlated to the geographic distance; the squared coefficient correlation, $r^2$, which measures genetic isolation by distance, ranged from 0.0008 to 0.73 (Supplementary Table 5). Consistent with the phylogeographic literature, it was high ($r^2 > 0.35$) and significant in, for example, *Ciona intestinalis* A, *Melitaea cinxia* and *Sepia officinalis*, and low ($r^2 < 0.02$, n.s.) in, for example, *Culex pipiens*, *Lepus granatensis* and *Crepidula fornicata*. The $F_{it}$ statistic was significantly higher, on average, in terrestrial (median $F_{it} = 0.25$) than in marine (median $F_{it} = 0.02$) species ($t$-test, $P = 0.029$) when only species including at least five individuals were considered. No significant relationship was detected between $\pi_s$ and absolute values of $F_{it}$ ($P = 0.22$, $r^2 = 0.05$, only species with more than four individuals included), which does not suggest any confounding effect of population structure in our analysis.

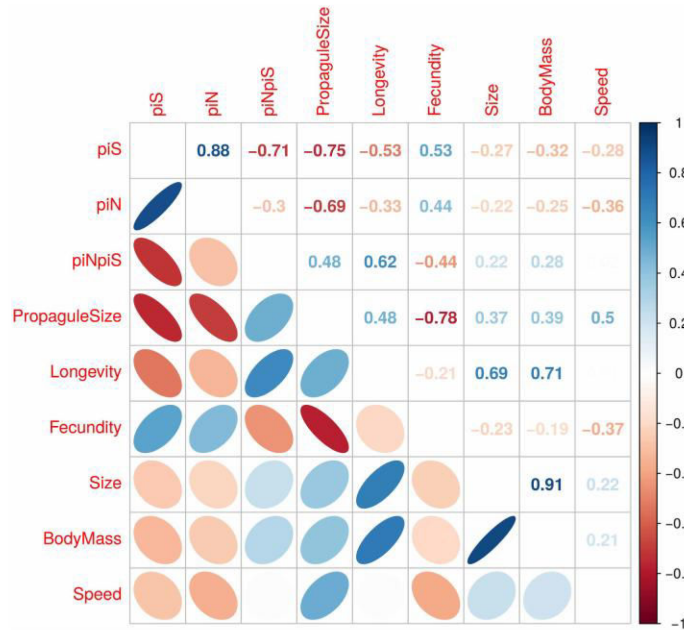
**Ethical statement.** Living animals were manipulated according to the 'Charte Nationale Portant sur l'Ethique de l'Expérimentation Animale'. Sampling of protected species was performed under permits 53/2009 (Galicia, Spain, *Emys orbicularis*), 2009/11/12 (Aude, France, *Emys orbicularis*), 503/05/07/2006 (Pisa, Italy, *Emys orbicularis*), 35601-60/2005-4 (Slovenia, *Emys orbicularis*), and 009-01-1230/a34-455 (France, *Parus caeruleus*). *Aptenodytes patagonicus, Eudyptes moseleyi* and *Eudyptes filholi* individuals were sampled by Institut Polaire Français Paul Emile Victor, program IPEV 131. *Chelonoidis nigra* individuals were handled and sampled by the veterinarians and staff of the Zurich zoo (Switzerland), Rotterdam zoo (the Netherlands), and A Cupulatta zoo (France) in accordance with the Code of Practice and Code of Ethics established by the European Association of Zoos and Aquaria.

21. Gayral, P. *et al.* Next-generation sequencing of transcriptomes: a guide to RNA isolation in non-model animals. *Mol. Ecol. Resources* **11,** 650–661 (2011).
22. Guéguen, L. *et al.* Bio++: efficient, extensible libraries and tools for molecular evolution. *Mol. Biol. Evol.* **30,** 1745–1750 (2013).
23. Weir, B. S. & Cockerham, C. C. Estimating *F*-statistics for the analysis of population structure. *Evolution* **38,** 1358–1370 (1984).
24. Philippe, H. *et al.* Phylogenomics revives traditional views on deep animal relationships. *Curr. Biol.* **19,** 706–712 (2009).
25. Wilfert, L., Gadau, J. & Schmid-Hempel, P. Variation in genomic recombination rates among animal taxa and the case of social insects. *Heredity* **98,** 189–197 (2007).

©2014 Macmillan Publishers Limited. All rights reserved

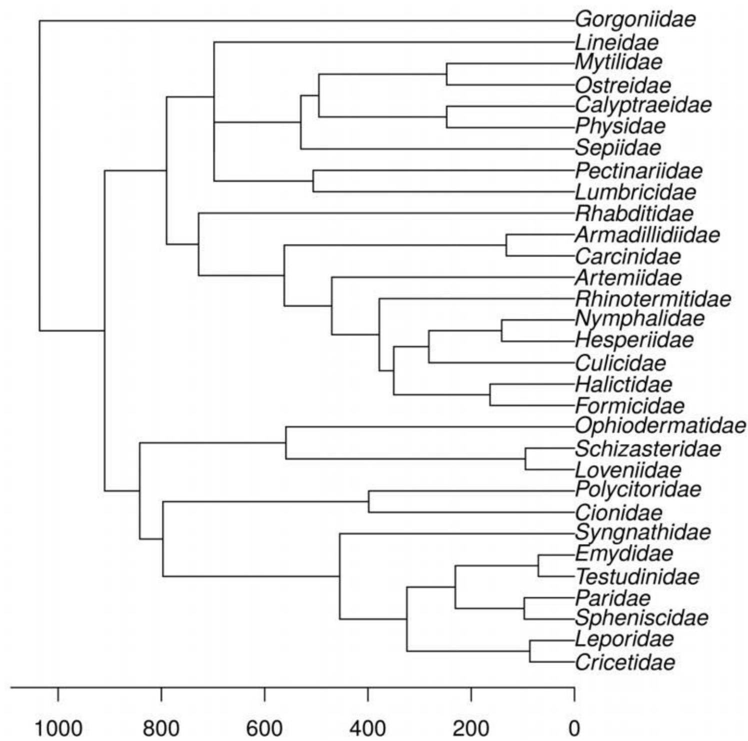



**Extended Data Figure 1 | Phylogenetic tree of metazoan orders and the position of the taxa analysed in this study.** The tree topology is consistent with the NCBI taxonomy. Red arrows identify 25 orders that were sampled. Five gastropod species from two distinct families (Calyptraeidae (*Crepidula fornicata*, *C. plana* and *Bostrycapulus aculeatus*) and Physidae (*Physa acuta* and *P. gyrina*)) are not represented because they lacked any assignment to an order in current taxonomy.

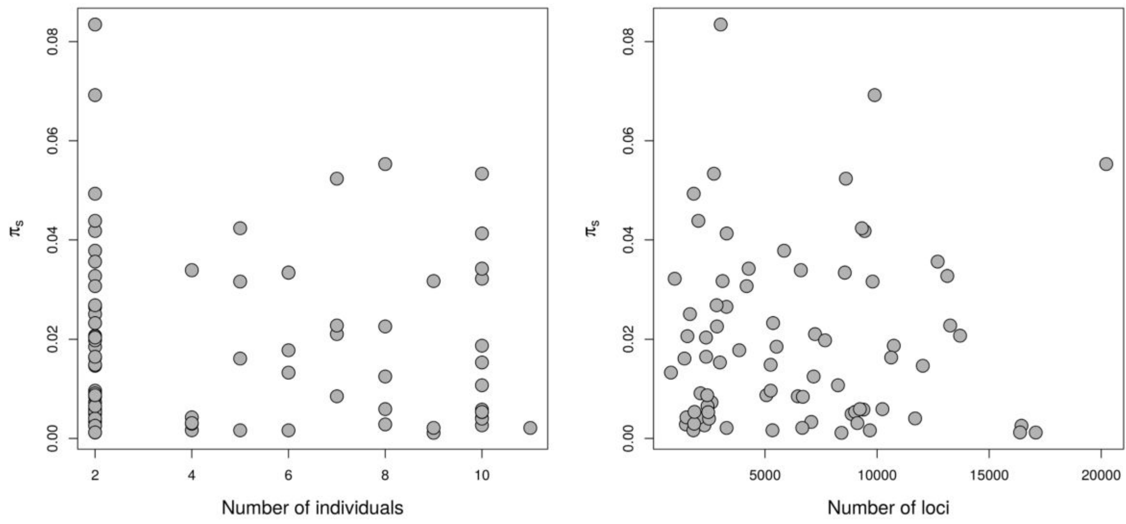



**Extended Data Figure 2 | Correlations between genetic diversity and life history variables.** Blue indicates a positive relationship, red a negative one; colour intensity is proportional to Pearson's correlation coefficient.



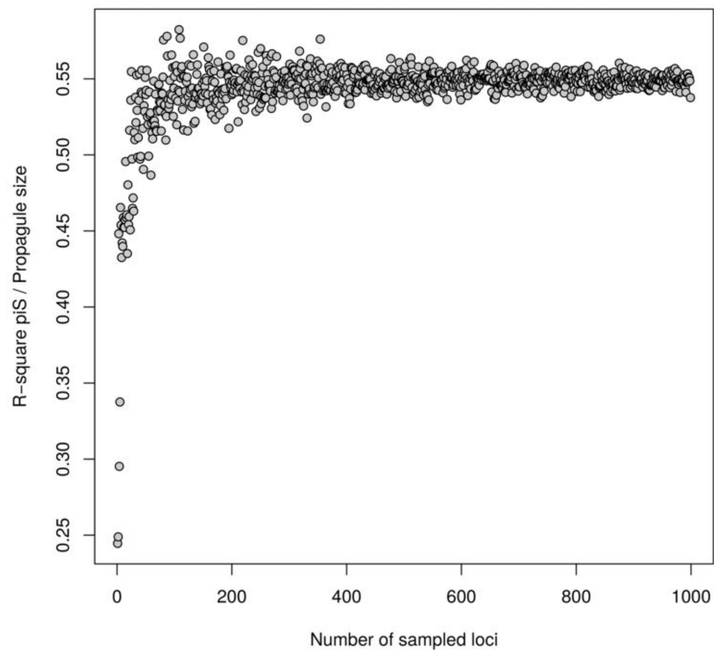



**Extended Data Figure 3 | Family-level phylogenetic tree (31 families included).** The scale is in million years of divergence.

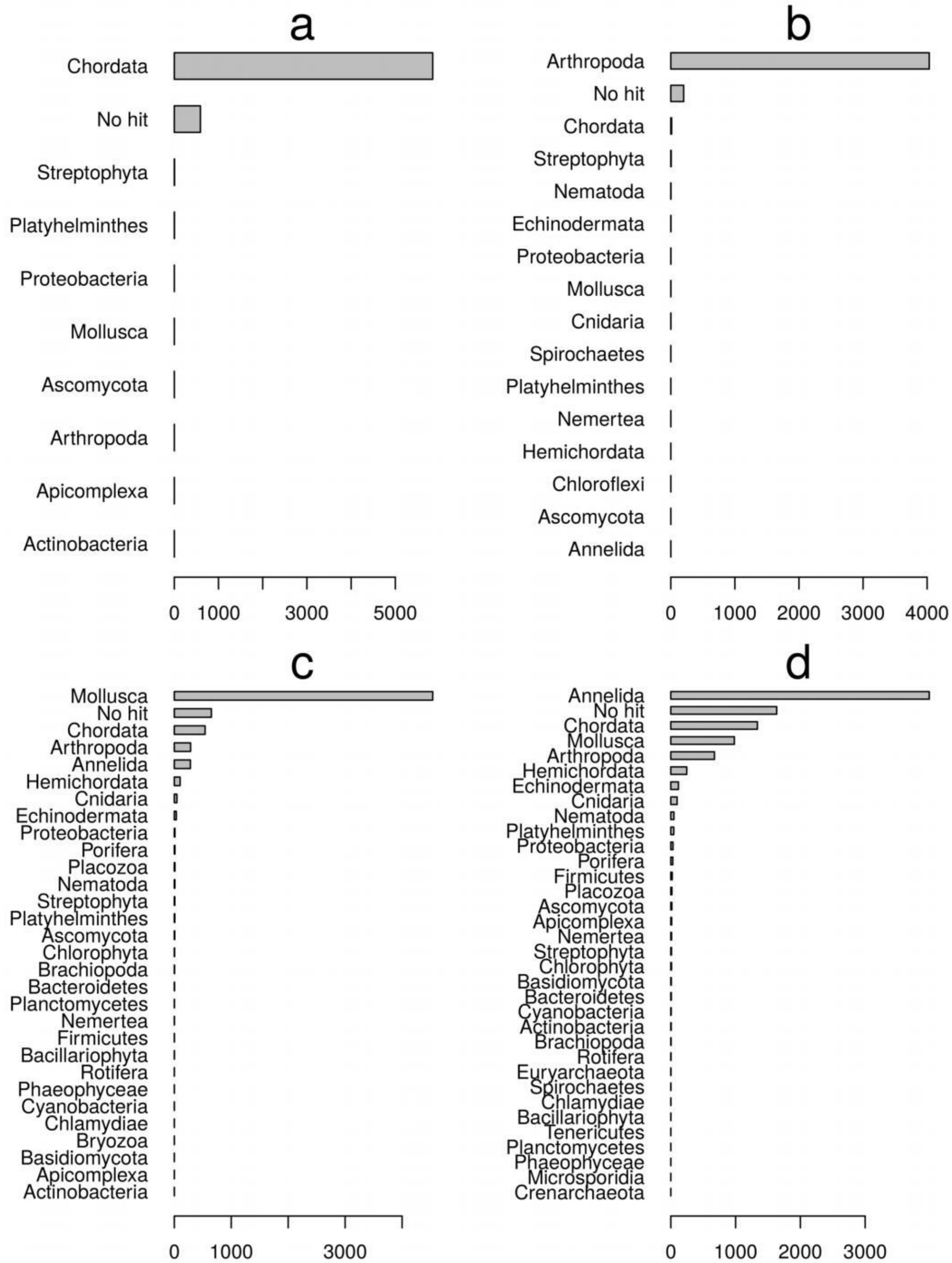



**Extended Data Figure 4 | Absence of significant correlation between species genetic diversity with individual sampling size ($P = 0.47$, $r^2 = 0.007$) and locus sampling size ($P = 0.78$, $r^2 = 0.001$).**

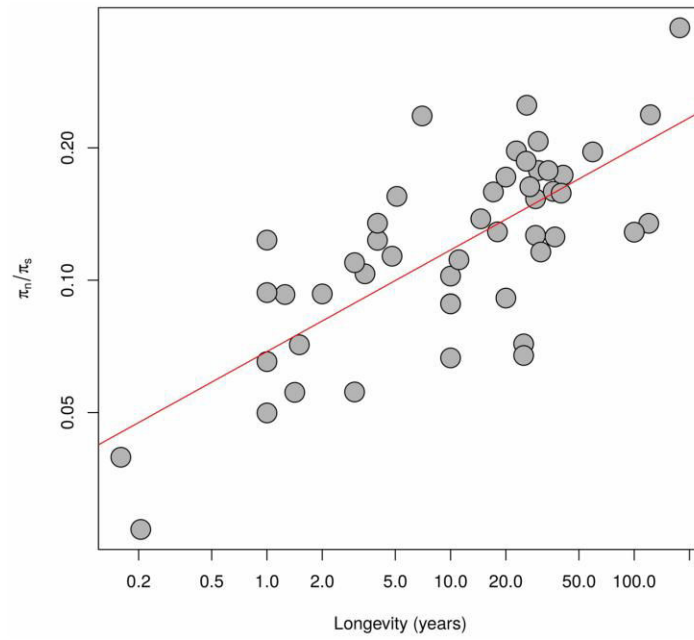





**Extended Data Figure 5 | Relationship between the $\pi_s$/propagule-size $r^2$ and the number of sampled loci.**




**Extended Data Figure 6 | Phylum distribution of the first BLAST hit in four representative species.** **a**, Common vole (*Microtus arvalis*). **b**, Glanville fritillary butterfly (*Melitaea cinxia*). **c**, Blue mussel (*Mytilus edulis*). **d**, Earthworm (*Allolobophora chlorotica*).




**Extended Data Figure 7 | Correlation between $\pi_n/\pi_s$ and maximum longevity ($P < 10^{-8}$, $r^2 = 0.54$).** Only species with at least four individuals are included.

# Annexe 3

C. R. Biologies 337 (2014) 553–560

Contents lists available at ScienceDirect

Comptes Rendus Biologies

www.sciencedirect.com

Development and reproduction biology/Biologie du développement et de la reproduction

## Genetic data, reproduction season and reproductive strategy data support the existence of biological species in *Ophioderma longicauda*

## *Les données génétiques, la saison et la stratégie de reproduction soutiennent l'existence d'espèces biologiques chez* Ophioderma longicauda

Alexandra Anh-Thu Weber [a,*], Sabine Stöhr [b], Anne Chenuil [a]

[a] *Aix-Marseille Université, Institut méditerranéen de biodiversité et d'écologie marine et continentale (IMBE), CNRS - IRD - UAPV, station marine d'Endoume, chemin de la Batterie-des-Lions, 13007 Marseille, France*
[b] *Department of Zoology, Swedish Museum of Natural History, Box 50007, 10405 Stockholm, Sweden*

ARTICLE INFO

*Article history:*
Received 11 June 2014
Accepted after revision 24 July 2014
Available online 22 August 2014

*Keywords:*
Ophiuroidea
Cryptic species
Brooding species
Cytochrome oxidase subunit I
EPIC

ABSTRACT

Cryptic species are numerous in the marine environment. The brittle star *Ophioderma longicauda* is composed of six mitochondrial lineages, encompassing brooders, which form a monophyletic group, and broadcasters, from which the brooders are derived. To clarify the species limits within *O. longicauda*, we compared the reproductive status of the sympatric lineages L1 and L3 (defined after sequencing a portion of the mitochondrial gene *COI*) during the month of May in Greece. In addition, we genotyped a nuclear marker, intron i51. Each L3 female was brooding, whereas all L1 specimens displayed full gonads, suggesting temporal pre-zygotic isolation between brooders and broadcasters. Statistical differences were found among lineages in morphology and bathymetric distribution. Finally, the intron i51 was polymorphic in L1 (60 individuals), but monomorphic in L3 (109 individuals), confirming the absence of gene flow between brooders and broadcasters. In conclusion, the broadcasting lineage L1 and the brooding lineage L3 are different biological species.

© 2014 Académie des sciences. Published by Elsevier Masson SAS. All rights reserved.

*Mots clés :*
Ophiuroidea
Espèces cryptiques
Espèce incubante
Sous-unité I de la cytochrome oxydase
EPIC

RÉSUMÉ

Les espèces cryptiques sont nombreuses dans l'environnement marin. L'ophiure *Ophioderma longicauda* est composée de six lignées mitochondriales, comprenant des incubants, qui forment un groupe monophylétique, et des individus à larves, desquels les incubants sont dérivés. Afin de clarifier les limites d'espèces chez *O. longicauda*, nous avons comparé l'état de reproduction des lignées sympatriques L1 et L3 (définies après séquençage d'une partie du gène mitochondrial *COI*) durant le mois de mai en Grèce. De plus, nous avons génotypé un marqueur nucléaire, l'intron i51. Chaque femelle L3 était incubante tandis que tous les individus L1 avaient les gonades pleines, suggérant un

* Corresponding author.
*E-mail addresses:* alexandra.weber@imbe.fr, alexandra.weber@univ-amu.fr (Alexandra A.-T. Weber).

http://dx.doi.org/10.1016/j.crvi.2014.07.007
1631-0691/© 2014 Académie des sciences. Published by Elsevier Masson SAS. All rights reserved.

# Annexe 4

C. R. Biologies 336 (2013) 572–581

Contents lists available at ScienceDirect

Comptes Rendus Biologies

www.sciencedirect.com

Ecology/Écologie

## Thermotolerance and regeneration in the brittle star species complex *Ophioderma longicauda*: A preliminary study comparing lineages and Mediterranean basins

## *Thermotolérance et régénération dans le complexe d'espèces d'ophiures* Ophioderma longicauda *: une étude préliminaire comparant lignées et bassins méditerranéens*

Alexandra Anh-Thu Weber [*,a], Sam Dupont [b], Anne Chenuil [a]

[a] *Institut méditerranéen de biodiversité et d'écologie marine et continentale (IMBE), Aix-Marseille Université, CNRS, IRD, UAPV, station marine d'Endoume, chemin de la Batterie-des-Lions, 13007 Marseille, France*
[b] *Department for Biological and Environmental Sciences, University of Gothenburg, The Sven Lovén Centre for Marine Sciences, Kristineberg, 45178 Fiskebäckskil, Sweden*

ARTICLE INFO

*Article history:*
Received 23 August 2013
Accepted after revision 4 October 2013
Available online 8 November 2013

*Keywords:*
Thermotolerance
Mediterranean Sea
Regeneration
*Ophioderma longicauda*
Cryptic species

ABSTRACT

Global warming is expected to change marine species distributions; it is thus critical to understand species current thermotolerance. The brittle star species complex *Ophioderma longicauda* comprises a broadcast spawning lineage L1 and a brooding lineage L3. We collected L1 specimens from Marseilles and Crete, and L3 specimens from Crete. We monitored survival, autotomy and arm regeneration at 17, 26 and 30 °C during 14 weeks. Globally *O. longicauda* showed good resistance to elevated temperatures compared to other published studies on ophiuroids. The L3 sample displayed a better thermotolerance than L1 samples. Yet, more research is needed to establish whether these differences are due to lineages, geographic origin, or random effects. We provided for the first time individual regeneration trajectories, and showed that regeneration followed a growth curve and was highly influenced by temperature in both lineages. Our results highlight the importance of taking into account the presence of cryptic species when studying the potential effects of global warming.

© 2013 Académie des sciences. Published by Elsevier Masson SAS. All rights reserved.

*Mots clés:*
Thermotolérance
Mer Méditerranéen
Régénération
*Ophioderma longicauda*
Espèces cryptiques

RÉSUMÉ

Le réchauffement climatique est supposé changer la distribution des espèces marines ; il est donc crucial de comprendre la thermotolérance actuelle des espèces. Le complexe d'espèces d'ophiure *Ophioderma longicauda* comprend une lignée se reproduisant grâce à des larves, L1, et une lignée incubante, L3. Nous avons récolté des spécimens de la lignée L1 provenant de Marseille et de Crète, ainsi que des spécimens de la lignée L3 provenant de Crète. Nous avons mesuré la survie, l'autotomie et la régénération des bras à 17, 26 et 30 °C pendant 14 semaines. Globalement, *O. longicauda* a montré une bonne résistance aux

* Corresponding author.
*E-mail addresses:* alexandra.weber@imbe.fr, alexandra.weber@univ-amu.fr (A.-T. Weber).

1631-0691/$ – see front matter © 2013 Académie des sciences. Published by Elsevier Masson SAS. All rights reserved.
http://dx.doi.org/10.1016/j.crvi.2013.10.004

# Références


Addison JA, Hart MW (2005) Spawning, copulation and inbreeding coefficients in marine invertebrates. *Biology Letters*, **1**, 450–453.

Alonso S, Guacimaro J (2008) Comportamiento de Ophioderma longicaudum (Echinodermata: Ofiuroideos) ante luz de varios colores.

Andersson L (1990) The Driving Force: Species Concepts and Ecology. *Taxon*, **39**, 375–382.

Ané C, Larget B, Baum DA, Smith SD, Rokas A (2007) Bayesian Estimation of Concordance among Gene Trees. *Molecular Biology and Evolution*, **24**, 412–426.

Arndt A, Smith MJ (1998) Genetic diversity and population structure in two species of sea cucumber: differing patterns according to mode of development. *Molecular Ecology*, **7**, 1053–1064.

Avise JC, Ball RM (1990) Principles of genealogical concordance in species concepts and biological taxonomy. *Oxford surveys in evolutionary biology*, **7**, 45–67.

Ayre DJ, Hughes TP (2000) Genotypic Diversity and Gene Flow in Brooding and Spawning Corals Along the Great Barrier Reef, Australia. *Evolution*, **54**, 1590–1605.

Ayre DJ, Minchinton TE, Perrin C (2009) Does life history predict past and current connectivity for rocky intertidal invertebrates across a marine biogeographic barrier? *Molecular Ecology*, **18**, 1887–1903.

Baldauf SL (2003) The Deep Roots of Eukaryotes. *Science*, **300**, 1703–1706.

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

Barbosa SS, Klanten SO, Puritz JB, Toonen RJ, Byrne M (2013) Very fine-scale population genetic structure of sympatric asterinid sea stars with benthic and pelagic larvae: influence of mating system and dispersal potential. *Biological Journal of the Linnean Society*, **108**, 821–833.

Baric S, Sturmbauer C (1999) Ecological Parallelism and Cryptic Species in the GenusOphiothrixDerived from Mitochondrial DNA Sequences. *Molecular Phylogenetics and Evolution*, **11**, 157–162.

Barrett CF, Freudenstein JV (2011) An integrative approach to delimiting species in a rare but widespread mycoheterotrophic orchid. *Molecular Ecology*, **20**, 2771–2786.

Barry JP, Baxter CH, Sagarin RD, Gilman SE (1995) Climate-related, long-term faunal changes in a California rocky intertidal community. *Science(Washington)*, **267**, 672–675.

Baum DA, Shaw KL (1995) Genealogical perspectives on the species problem. *Experimental and molecular approaches to plant biosystematics*, **53**, 289–303.

Baus E, Darrock DJ, Bruford MW (2005) Gene-flow patterns in Atlantic and Mediterranean populations of the Lusitanian sea star Asterina gibbosa. *Molecular Ecology*, **14**, 3373–3382.

Baxter SW, Davey JW, Johnston JS *et al.* (2011) Linkage Mapping and Comparative Genomics Using Next-Generation RAD Sequencing of a Non-Model Organism. *PLoS ONE*, **6**, e19315.

Beerli P, Felsenstein J (2001) Maximum likelihood estimation of a migration matrix and effective population sizes in n subpopulations by using a coalescent approach. *Proceedings of the National Academy of Sciences*, **98**, 4563–4568.

Belkhir K, Borsa P, Chikhi L, Raufaste N, Bonhomme F (2004) GENETIX 4.05, Population genetics software for Windows TM. *Université de Montpellier II. Montpellier*.

Beltrán C, Galindo BE, Rodríguez-Miranda E, Sánchez D (2007) Signal transduction mechanisms regulating ion fluxes in the sea urchin sperm. *Signal Transduction*, **7**, 103–117.

Bensoussan N, Romano J-C, Harmelin J-G, Garrabou J (2010) High resolution characterization of northwest Mediterranean coastal waters thermal regimes: To better understand responses of benthic communities to climate change. *Estuarine, Coastal and Shelf Science*, **87**, 431–441.

Berger EM (1973) GENE-ENZYME VARIATION IN THREE SYMPATRIC SPECIES OF LITTORINA. *The Biological Bulletin*, **145**, 83–90.

Bickford D, Lohman DJ, Sodhi NS *et al.* (2007) Cryptic species as a window on diversity and conservation. *Trends in Ecology & Evolution*, **22**, 148–155.

Biermann CH, Marks JA, Vilela-Silva A-CES, Castro MO, Mourão PAS (2004) Carbohydrate-based species recognition in sea urchin fertilization: another avenue for speciation? *Evolution & Development*, **6**, 353–361.

Bierne N, Welch J, Loire E, Bonhomme F, David P (2011) The coupling hypothesis: why genome scans may fail to map local adaptation genes. *Molecular Ecology*, **20**, 2044–2072.

Biressi ACM, Zou T, Dupont S *et al.* (2010) Wound healing and arm regeneration in Ophioderma longicaudum and Amphiura filiformis (Ophiuroidea, Echinodermata): comparative morphogenesis and histogenesis. *Zoomorphology*, **129**, 1–19.

Birkhead TR, Martinez JG, Burke T, Froman DP (1999) Sperm mobility determines the outcome of sperm competition in the domestic fowl. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, **266**, 1759–1764.

Bohonak AJ (1999) Dispersal, Gene Flow, and Population Structure. *The Quarterly Review of Biology*, **74**, 21–45.

Boissin E, Féral JP, Chenuil A (2008) Defining reproductively isolated units in a cryptic and syntopic species complex using mitochondrial and nuclear markers: the brooding brittle star, Amphipholis squamata (Ophiuroidea). *Molecular Ecology*, **17**, 1732–1744.

Boissin E, Stöhr S, Chenuil A (2011) Did vicariance and adaptation drive cryptic speciation and evolution of brooding in Ophioderma longicauda (Echinodermata: Ophiuroidea), a common Atlanto-Mediterranean ophiuroid? *Molecular Ecology*, **20**, 4737–4755.

Bottjer DJ, Davidson EH, Peterson KJ, Cameron RA (2006) Paleogenomics of Echinoderms. *Science*, **314**, 956–960.

Bowen BW, Nelson WS, Avise JC (1993) A molecular phylogeny for marine turtles: trait mapping, rate assessment, and conservation relevance. *Proceedings of the National Academy of Sciences*, **90**, 5574–5577.

Bradbury IR, Hubert S, Higgins B *et al.* (2010) Parallel adaptive evolution of Atlantic cod on both sides of the Atlantic Ocean in response to temperature. *Proceedings of the Royal Society B: Biological Sciences*, **277**, 3725–3734.

Brinkmann H, Philippe H (1999) Archaea sister group of Bacteria? Indications from tree reconstruction artifacts in ancient phylogenies. *Molecular Biology and Evolution*, **16**, 817–825.

Burns G, Ortega-Martinez O, Dupont S *et al.* (2012) Intrinsic gene expression during regeneration in arm explants of Amphiura filiformis. *Journal of Experimental Marine Biology and Ecology*, **413**, 106–112.

Burns G, Thorndyke MC, Peck LS, Clark MS (2013) Transcriptome pyrosequencing of the Antarctic brittle star Ophionotus victoriae. *Marine Genomics*, **9**, 9–15.

Byrne M (1991) Reproduction, development and population biology of the Caribbean ophiuroidOphionereis olivacea, a protandric hermaphrodite that broods its young. *Marine Biology*, **111**, 387–399.

Byrne M, Cisternas P, O'Hara T (2008) Brooding of pelagic-type larvae in Ophiopeza spinosa: reproduction and development in a tropical ophiodermatid brittlestar. *Invertebrate Biology*, **127**, 98–107.

Byrne M, GonzalezBernat M, Doo S *et al.* (2013) Effects of ocean warming and acidification on embryos and non-calcifying larvae of the invasive sea star Patiriella regularis. *Marine Ecology Progress Series*, **473**, 235–246.

Byrne M, Hart MW, Cerra A, Cisternas P (2003) Reproduction and Larval Morphology of Broadcasting and Viviparous Species in the Cryptasterina Species Complex. *The Biological Bulletin*, **205**, 285–294.

Byrne M, Ho M, Selvakumaraswamy P *et al.* (2009) Temperature, but not pH, compromises sea urchin fertilization and early development under near-future climate change scenarios. *Proceedings of the Royal Society B: Biological Sciences*, **276**, 1883–1888.

Cahais V, Gayral P, Tsagkogeorga G *et al.* (2012) Reference-free transcriptome assembly in non-model animals from next-generation sequencing data. *Molecular Ecology Resources*, **12**, 834–845.

Carstens BC, Pelletier TA, Reid NM, Satler JD (2013) How to fail at species delimitation. *Molecular Ecology*, **22**, 4369–4383.

Chenuil A (2006) Choosing the right molecular genetic markers for studying biodiversity: from molecular evolution to practical aspects. *Genetica*, **127**, 101–120.

Chenuil A (2012) How to infer reliable diploid genotypes from NGS or traditional sequence data: from basic probability to experimental optimization. *Journal of Evolutionary Biology*, **25**, 949–960.

Chenuil A, Féral JP (2003) Sequences of mitochondrial DNA suggest that Echinocardium cordatum is a complex of several sympatric or hybridizing species: a pilot study. In: *Echinoderm Research 2001. Proceedings of the Sixth European Conference on Echinoderm, Banyuls-sur-Mer, France, Swets & Zeitlinger, Lisse* , pp. 15–32.

Chenuil A, Hoareau TB, Egea E *et al.* (2010) An efficient method to find potentially universal population genetic markers, applied to metazoans. *BMC Evolutionary Biology*, **10**, 276.

Chiari Y, Cahais V, Galtier N, Delsuc F (2012) Phylogenomic analyses support the position of turtles as the sister group of birds and crocodiles (Archosauria). *BMC Biology*, **10**, 65.

Christensen AB, Nguyen HD, Byrne M (2011) Thermotolerance and the effects of hypercapnia on the metabolic rate of the ophiuroid Ophionereis schayeri: Inferences for survivorship in a changing ocean. *Journal of Experimental Marine Biology and Ecology*, **403**, 31–38.

Ciccarelli FD, Doerks T, Mering C von *et al.* (2006) Toward Automatic Reconstruction of a Highly Resolved Tree of Life. *Science*, **311**, 1283–1287.

Clark MS, Dupont S, Rossetti H *et al.* (2007) Delayed arm regeneration in the Antarctic brittle star Ophionotus victoriae. *Aquatic Biology*, **1**, 45–53.

Clarke A (2003) Costs and consequences of evolutionary temperature adaptation. *Trends in Ecology & Evolution*, **18**, 573–581.

Clements LAJ, Fielman KT, Stancyk SE (1988) Regeneration by an amphiurid brittlestar exposed to different concentrations of dissolved organic material. *Journal of Experimental Marine Biology and Ecology*, **122**, 47–61.

Collin R (2001) The effects of mode of development on phylogeography and population structure of North Atlantic Crepidula (Gastropoda: Calyptraeidae). *Molecular Ecology*, **10**, 2249–2262.

Coma R, Ribes M, Serrano E *et al.* (2009) Global warming-enhanced stratification and mass mortality events in the Mediterranean. *Proceedings of the National Academy of Sciences*, **106**, 6176–6181.

Conesa A, Götz S, García-Gómez JM *et al.* (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*, **21**, 3674–3676.

Cornuet J-M, Pudlo P, Veyssier J *et al.* (2014) DIYABC v2.0: a software to make approximate Bayesian computation inferences about population history using single nucleotide polymorphism, DNA sequence and microsatellite data. *Bioinformatics*, **30**, 1187–1189.

Cracraft J (1983) Species concepts and speciation analysis. In: *Current ornithology* , pp. 159–187. Springer.

Crandall ED, Frey MA, Grosberg RK, Barber PH (2008) Contrasting demographic history and phylogeographical patterns in two Indo-Pacific gastropods. *Molecular Ecology*, **17**, 611–626.

Csilléry K, Blum MGB, Gaggiotti OE, François O (2010) Approximate Bayesian Computation (ABC) in practice. *Trends in Ecology & Evolution*, **25**, 410–418.

Cunningham CW (1999) Some Limitations of Ancestral Character-State Reconstruction When Testing Evolutionary Hypotheses. *Systematic Biology*, **48**, 665–674.

Currat M, Ruedi M, Petit RJ, Excoffier L (2008) The Hidden Side of Invasions: Massive Introgression by Local Genes. *Evolution*, **62**, 1908–1920.

Czarkwiani A, Dylus DV, Oliveri P (2013) Expression of skeletogenic genes during arm regeneration in the brittle star Amphiura filiformis. *Gene Expression Patterns*, **13**, 464–472.

Dahms H-U, Lee J-S (2010) UV radiation in marine ectotherms: Molecular effects and responses. *Aquatic Toxicology*, **97**, 3–14.

Dale B, Monroy A (1981) How is polyspermy prevented? *Gamete Research*, **4**, 151–169.

Damerau M, Matschiner M, Salzburger W, Hanel R (2014) Population divergences despite long pelagic larval stages: lessons from crocodile icefishes (Channichthyidae). *Molecular Ecology*, **23**, 284–299.

Darszon A, Guerrero A, Galindo BE, Nishigaki T, Wood CD (2008) Sperm-activating peptides in the regulation of ion fluxes, signal transduction and motility. *The International Journal of Developmental Biology*, **52**, 595–606.

Darwin C (1859) On the origins of species by means of natural selection. *London: Murray*.

Daugherty CH, Cree A, Hay JM, Thompson MB (1990) Neglected taxonomy and continuing extinctions of tuatara (Sphenodon). *Nature*, **347**, 177–179.

Davey JW, Hohenlohe PA, Etter PD *et al.* (2011) Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics*, **12**, 499–510.

David B, Mooi R (1999) Comprendre les echinodermes; la contribution du modele extraxial-axial. *Bulletin de la Societe Geologique de France*, **170**, 91–101.

Dawson MN, Louie KD, Barlow M, Jacobs DK, Swift CC (2002) Comparative phylogeography of sympatric sister species, Clevelandia ios and Eucyclogobius newberryi (Teleostei, Gobiidae), across the California Transition Zone. *Molecular Ecology*, **11**, 1065–1075.

DeSalle R, Egan MG, Siddall M (2005) The unholy trinity: taxonomy, species delimitation and DNA barcoding. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **360**, 1905–1916.

Deschuyteneer M, Jangoux M (1978) Comportement alimentaire et structures digestives d'Ophioderma longicauda (Retzius) (Echinodermata, Ophiuroidea). *Annales de l'Institut océanographique*, **54**, 127–138.

Dobzhansky T (1934) Studies on hybrid sterility. *Zeitschrift für Zellforschung und Mikroskopische Anatomie*, **21**, 169–223.

Donoghue MJ (1985) A critique of the biological species concept and recommendations for a phylogenetic alternative. *Bryologist*, 172–181.

D'Ortenzio F, Ribera D'Alcalã M (2009) On the trophic regimes of the Mediterranean Sea: a satellite analysis. *Biogeosciences*, **6**, 139–148.

Douady CJ, Delsuc F, Boucher Y, Doolittle WF, Douzery EJP (2003) Comparison of Bayesian and Maximum Likelihood Bootstrap Measures of Phylogenetic Reliability. *Molecular Biology and Evolution*, **20**, 248–254.

Duffy JE (1993) Genetic population structure in two tropical sponge-dwelling shrimps that differ in dispersal potential. *Marine Biology*, **116**, 459–470.

Dupont S, Thorndyke MC (2006) Growth or differentiation? Adaptive regeneration in the brittlestar Amphiura filiformis. *Journal of Experimental Biology*, **209**, 3873–3881.

Durrieu de Madron X, Guieu C, Sempéré R *et al.* (2011) Marine ecosystems' responses to climatic and anthropogenic forcings in the Mediterranean. *Progress in Oceanography*, **91**, 97–166.

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, **32**, 1792–1797.

Egea E (2011) *Histoire évolutive, structures génétique, morphologique et écologique comparées dans un complexe d'espèces jumelles : Echinocardium cordatum (Echinoidea, Irregularia)*. Aix Marseille 2.

Emlet RB (1995) Developmental Mode and Species Geographic Range in Regular Sea Urchins (Echinodermata: Echinoidea). *Evolution*, **49**, 476–489.

Ence DD, Carstens BC (2011) SpedeSTEM: a rapid and accurate method for species delimitation. *Molecular Ecology Resources*, **11**, 473–480.

Evans JP, Garcia-Gonzalez F, Almbro M, Robinson O, Fitzpatrick JL (2012) Assessing the potential for egg chemoattractants to mediate sexual selection in a broadcast spawning marine invertebrate. *Proceedings of the Royal Society B: Biological Sciences*, **279**, 2855–2861.

Evans JP, Sherman CDH (2013) Sexual Selection and the Evolution of Egg-Sperm Interactions in Broadcast-Spawning Invertebrates. *The Biological Bulletin*, **224**, 166–183.

Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics*, **131**, 479–491.

Fenaux L (1969) Le développement larvaire chez Ophioderma longicauda (Retzius). *Cahiers de Biologie Marine*, **10**.

Fenaux L (1972) Evolution saisonnière des gonades chez l'Ophiure Ophioderma longicauda (RETZIUS), Ophiuroidea. *Internationale Revue der gesamten Hydrobiologie und Hydrographie*, **57**, 257–262.

Foltz DW (2003) Invertebrate Species with Nonpelagic Larvae Have Elevated Levels of Nonsynonymous Substitutions and Reduced Nucleotide Diversities. *Journal of Molecular Evolution*, **57**, 607–612.

Foltz DW, Hrincevich AW, Rocha-Olivares A (2004) Apparent Selection Intensity for the Cytochrome Oxidase Subunit I Gene Varies with Mode of Reproduction in Echinoderms. *Genetica*, **122**, 115–125.

Foltz DW, Stickle WB, Campagnaro EL, Himel AE (1996) Mitochondrial DNA polymorphisms reveal additional genetic heterogeneity within the Leptasterias hexactis (Echinodermata: Asteroidea) species complex. *Marine Biology*, **125**, 569–578.

Fu Y-X (1997) Statistical Tests of Neutrality of Mutations Against Population Growth, Hitchhiking and Background Selection. *Genetics*, **147**, 915–925.

Fu YX, Li WH (1993) Statistical tests of neutrality of mutations. *Genetics*, **133**, 693–709.

Le Gac M, Féral J-P, Poulin E, Veyret M, Chenuil A (2004) Identification of allopatric clades in the cosmopolitan ophiuroid species complex Amphipholis squamata (Echinodermata). The end of a paradox? *Marine Ecology progress series*, **278**, 171–178.

Gage MJG, Macfarlane CP, Yeates S *et al.* (2004) Spermatozoal Traits and Sperm Competition in Atlantic Salmon: Relative Sperm Velocity Is the Primary Determinant of Fertilization Success. *Current Biology*, **14**, 44–47.

Galindo BE, de la Vega-Beltrán JL, Labarca P, Vacquier VD, Darszon A (2007) Sp-tetraKCNG: A novel cyclic nucleotide gated K+ channel. *Biochemical and Biophysical Research Communications*, **354**, 668–675.

Galtier N, Nabholz B, Glémin S, Hurst GDD (2009) Mitochondrial DNA as a marker of molecular diversity: a reappraisal. *Molecular Ecology*, **18**, 4541–4550.

Gavrilets S (2003) Perspective: Models of Speciation: What Have We Learned in 40 Years? *Evolution*, **57**, 2197–2215.

Gayral P, Melo-Ferreira J, Glémin S *et al.* (2013) Reference-Free Population Genomics from Next-Generation Transcriptome Data and the Vertebrate–Invertebrate Gap. *PLoS Genet*, **9**, e1003457.

Gayral P, Weinert L, Chiari Y *et al.* (2011) Next-generation sequencing of transcriptomes: a guide to RNA isolation in nonmodel animals. *Molecular Ecology Resources*, **11**, 650–661.

Geller JB (1999) Decline of a Native Mussel Masked by Sibling Species Invasion. *Conservation Biology*, **13**, 661–664.

Gérard K, Guilloton E, Arnaud-Haond S *et al.* (2013) PCR survey of 50 introns in animals: Cross-amplification of homologous EPIC loci in eight non-bilaterian, protostome and deuterostome phyla. *Marine Genomics*, **12**, 1–8.

Goldstein PZ, Desalle R, Amato G, Vogler AP (2000) Conservation Genetics at the Species Boundary. *Conservation Biology*, **14**, 120–131.

Gouy M, Guindon S, Gascuel O (2010) SeaView Version 4: A Multiplatform Graphical User Interface for Sequence Alignment and Phylogenetic Tree Building. *Molecular Biology and Evolution*, **27**, 221–224.

Guerrero A, Nishigaki T, Carneiro J *et al.* (2010) Tuning sperm chemotaxis by calcium burst timing. *Developmental Biology*, **344**, 52–65.

Guindon S, Dufayard J-F, Lefort V *et al.* (2010) New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. *Systematic Biology*, **59**, 307–321.

Hackett SJ, Kimball RT, Reddy S *et al.* (2008) A Phylogenomic Study of Birds Reveals Their Evolutionary History. *Science*, **320**, 1763–1768.

Häder D-P, Kumar HD, Smith RC, Worrest RC (2007) Effects of solar UV radiation on aquatic ecosystems and interactions with climate change. *Photochemical & Photobiological Sciences*, **6**, 267–285.

Hagman DK, Vize PD (2003) Mass spawning by two brittle star species,Ophioderma rubicundum and O. Squamosissimum (Echinodermata: Ophiuroidea), at the Flower Garden Banks, Gulf of Mexico. *Bulletin of Marine Science*, **72**, 871–876.

Hardin G (1960) The competitive exclusion principle. *Science*, **131**, 1292–1297.

Hart M (2000) Phylogenetic analyses of mode of larval development. *Seminars in Cell & Developmental Biology*, **11**, 411–418.

Hart MW (2013) Structure and evolution of the sea star egg receptor for sperm bindin. *Molecular Ecology*, **22**, 2143–2156.

Hart MW, Byrne M, Johnson SL (2003) Patiriella pseudoexigua (Asteroidea: Asterinidae): a cryptic species complex revealed by molecular and embryological analyses. *Journal of the Marine Biological Association of the United Kingdom*, **83**, 1109–1116.

Hart MW, Byrne M, Smith MJ (1997) Molecular Phylogenetic Analysis of Life-History Evolution in Asterinid Starfish. *Evolution*, **51**, 1848–1861.

Hart MW, Keever CC, Dartnall AJ, Byrne M (2006) Morphological and Genetic Variation Indicate Cryptic Species Within Lamarck's Little Sea Star, Parvulastra (=Patiriella) exigua. *The Biological Bulletin*, **210**, 158–167.

Hart MW, Marko PB (2010) It's About Time: Divergence, Demography, and the Evolution of Developmental Modes in Marine Invertebrates. *Integrative and Comparative Biology*, **50**, 643–661.

Heath TA, Hedtke SM, Hillis DM (2008) Taxon sampling and the accuracy of phylogenetic analyses. *J Syst Evol*, **46**, 239–257.

Hebert PDN, Ratnasingham S, Waard JR de (2003) Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, **270**, S96–S99.

Hedgecock D, Pudovkin AI (2011) Sweepstakes Reproductive Success in Highly Fecund Marine Fish and Shellfish: A Review and Commentary. *Bulletin of Marine Science*, **87**, 971–1002.

Heimeier D, Lavery S, Sewell MA (2010) Molecular Species Identification of Astrotoma agassizii from Planktonic Embryos: Further Evidence for a Cryptic Species Complex. *Journal of Heredity*, **101**, 775–779.

Hellberg ME (1996) Dependence of Gene Flow on Geographic Distance in Two Solitary Corals with Different Larval Dispersal Capabilities. *Evolution*, **50**, 1167–1175.

Hemery LG, Eléaume M, Roussel V *et al.* (2012) Comprehensive sampling reveals circumpolarity and sympatry in seven mitochondrial lineages of the Southern Ocean crinoid species Promachocrinus kerguelensis (Echinodermata). *Molecular Ecology*, **21**, 2502–2518.

Hendler G (1991) *Reproduction of marine invertebrates*. Pacific Grove, California.

Hoareau TB, Boissin E, Paulay G, Bruggemann JH (2013) The Southwestern Indian Ocean as a potential marine evolutionary hotspot: perspectives from comparative phylogeography of reef brittle-stars. *Journal of Biogeography*, **40**, 2167–2179.

Hoffman JI, Clarke A, Linse K, Peck LS (2011) Effects of brooding and broadcasting reproductive modes on the population genetic structure of two Antarctic gastropod molluscs. *Marine Biology*, **158**, 287–296.

Hohenlohe PA, Amish SJ, Catchen JM, Allendorf FW, Luikart G (2011) Next-generation RAD sequencing identifies thousands of SNPs for assessing hybridization between rainbow and westslope cutthroat trout. *Molecular Ecology Resources*, **11**, 117–122.

Hoskin MG (1997) Effects of contrasting modes of larval development on the genetic structures of populations of three species of prosobranch gastropods. *Marine Biology*, **127**, 647–656.

Hrincevich AW, Foltz DW (1996) Mitochondrial DNA Sequence Variation in a Sea Star (Leptasteriasspp.) Species Complex. *Molecular Phylogenetics and Evolution*, **6**, 408–415.

Huelsenbeck JP, Andolfatto P (2007) Inference of Population Structure Under a Dirichlet Process Model. *Genetics*, **175**, 1787–1802.

Huelsenbeck JP, Andolfatto P, Huelsenbeck ET (2011) Structurama: Bayesian Inference of Population Structure. *Evolutionary Bioinformatics Online*, **7**, 55–59.

Huelsenbeck JP, Ronquist F, others (2001) MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics*, **17**, 754–755.

Hughes AR, Inouye BD, Johnson MTJ, Underwood N, Vellend M (2008) Ecological consequences of genetic diversity. *Ecology Letters*, **11**, 609–623.

Hunt R (1982) *Plant growth curves. The functional approach to plant growth analysis.*

Hunt A (1993) Effects of Contrasting Patterns of Larval Dispersal on the Genetic Connectedness of Local-Populations of 2 Intertidal Starfish, Patiriella-Calcar and P-Exigua. *Marine Ecology Progress Series*, **92**, 179–186.

Hunter RL, Halanych KM (2008) Evaluating Connectivity in the Brooding Brittle Star Astrotoma agassizii across the Drake Passage in the Southern Ocean. *Journal of Heredity*, **99**, 137–148.

Janies D (2001) Phylogenetic relationships of extant echinoderm classes. *Canadian Journal of Zoology*, **79**, 1232–1250.

Janies DA, Voight JR, Daly M (2011) Echinoderm Phylogeny Including Xyloplax, a Progenetic Asteroid. *Systematic Biology*, **60**, 420–438.

Jombart T, Devillard S, Balloux F (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics*, **11**, 94.

Jörger KM, Schrödl M (2013) How to describe a cryptic species? Practical challenges of molecular taxonomy. *Frontiers in zoology*, **10**, 1742–9994.

Katoh K, Asimenos G, Toh H (2009) Multiple Alignment of DNA Sequences with MAFFT. In: *Bioinformatics for DNA Sequence Analysis* Methods in Molecular Biology. (ed Posada D), pp. 39–64. Humana Press.

Keeling PJ, Burger G, Durnford DG *et al.* (2005) The tree of eukaryotes. *Trends in Ecology & Evolution*, **20**, 670–676.

Keever CC, Puritz JB, Addison JA *et al.* (2013) Shallow gene pools in the high intertidal: extreme loss of genetic diversity in viviparous sea stars (Parvulastra). *Biology Letters*, **9**.

Keever CC, Sunday J, Puritz JB *et al.* (2009) Discordant Distribution of Populations and Genetic Variation in a Sea Star with High Dispersal Potential. *Evolution*, **63**, 3214–3227.

Kelley AL, de Rivera CE, Buckley BA (2011) Intraspecific variation in thermotolerance and morphology of the invasive European green crab, Carcinus maenas, on the west coast of North America. *Journal of Experimental Marine Biology and Ecology*, **409**, 70–78.

Kimura M (1969) The Number of Heterozygous Nucleotide Sites Maintained in a Finite Population Due to Steady Flux of Mutations. *Genetics*, **61**, 893–903.

Knowles LL, Carstens BC (2007) Delimiting Species without Monophyletic Gene Trees. *Systematic Biology*, **56**, 887–895.

Knowlton N (1993) Sibling Species in the Sea. *Annual Review of Ecology and Systematics*, **24**, 189–216.

Knowlton N (2000) Molecular genetic analyses of species boundaries in the sea. *Hydrobiologia*, **420**, 73–90.

Knowlton N, Jackson JB (1993) Inbreeding and outbreeding in marine invertebrates. *The natural history of inbreeding and outbreeding. University of Chicago Press, Chicago*, 200–249.

Kumar S, Banks TW, Cloutier S (2012) SNP Discovery through Next-Generation Sequencing and Its Applications. *International Journal of Plant Genomics*, **2012**, e831460.

Kuo ESL, Sanford E (2009) Geographic variation in the upper thermal limits of an intertidal snail: implications for climate envelope models. *Marine Ecology, Progress Series*, **388**, 137–146.

Laporte M, Shao Z, Berrebi P *et al.* (2014) Isolation of 12 microsatellite markers following a pyrosequencing procedure and cross-priming in two invasive cryptic species, Alexandrium catenella (group IV) and A. tamarense (group III) (Dinophyceae). *Marine Pollution Bulletin*, **83**, 302–305.

Larget BR, Kotha SK, Dewey CN, Ané C (2010) BUCKy: Gene tree/species tree reconciliation with Bayesian concordance analysis. *Bioinformatics*, **26**, 2910–2911.

Leaché AD, Fujita MK (2010) Bayesian species delimitation in West African forest geckos (Hemidactylus fasciatus). *Proceedings of the Royal Society B: Biological Sciences*, rspb20100662.

Leclère L, Schuchert P, Cruaud C, Couloux A, Manuel M (2009) Molecular Phylogenetics of Thecata (Hydrozoa, Cnidaria) Reveals Long-Term Maintenance of Life History Traits despite High Frequency of Recent Character Changes. *Systematic Biology*, **58**, 509–526.

Lecointre G, Philippe H, Lê HLV, Le Guyader H (1993) Species Sampling Has a Major Impact on Phylogenetic Inference. *Molecular Phylogenetics and Evolution*, **2**, 205–224.

Ledoux J-B, Tarnowska K, Gérard K *et al.* (2012) Fine-scale spatial genetic structure in the brooding sea urchin Abatus cordatus suggests vulnerability of the Southern Ocean marine invertebrates facing global change. *Polar Biology*, **35**, 611–623.

Lee HJ, Boulding EG (2009) Spatial and temporal population genetic structure of four northeastern Pacific littorinid gastropods: the effect of mode of larval development on variation at one mitochondrial and two nuclear DNA markers. *Molecular Ecology*, **18**, 2165–2184.

Leese F, Agrawal S, Held C (2010) Long-distance island hopping without dispersal stages: transportation across major zoogeographic barriers in a Southern Ocean isopod. *Naturwissenschaften*, **97**, 583–594.

Lessios HA (2007) Reproductive isolation between species of sea urchins. *Bulletin of Marine Science*, **81**, 191–208.

Lessios HA (2011) Speciation Genes in Free-Spawning Marine Invertebrates. *Integrative and Comparative Biology*, icr039.

Letsch HO, Kjer KM (2011) Potential pitfalls of modelling ribosomal RNA data in phylogenetic tree reconstruction: Evidence from case studies in the Metazoa. *BMC Evolutionary Biology*, **11**, 146.

Levitan DR (2000) Sperm velocity and longevity trade off each other and influence fertilization in the sea urchin Lytechinus variegatus. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, **267**, 531–534.

Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, **25**, 1451–1452.

Lighten J, van Oosterhout C, Bentzen P (2014) Critical review of NGS analyses for de novo genotyping multigene families. *Molecular Ecology*, **23**, 3957–3972.

Lischer HEL, Excoffier L (2012) PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics*, **28**, 298–299.

Li L, Stoeckert CJ, Roos DS (2003) OrthoMCL: Identification of Ortholog Groups for Eukaryotic Genomes. *Genome Research*, **13**, 2178–2189.

Littlewood DTJ, Smith AB, Clough KA, Emson RH (1997) The interrelationships of the echinoderm classes: morphological and molecular evidence. *Biological Journal of the Linnean Society*, **61**, 409–438.

MacArthur RH, Wilson EO (1967) The theory of island biogeography. *Monographs in Population Biology (Princeton Univ. Press, Princeton)*.

Maddison WP, Knowles LL (2006) Inferring Phylogeny Despite Incomplete Lineage Sorting. *Systematic Biology*, **55**, 21–30.

Madsen FJ (1970) West African Ophiuroids. *Atlantide Report*, **11**, 151–243.

Mah SA, Swanson WJ, Vacquier VD (2005) Positive Selection in the Carbohydrate Recognition Domains of Sea Urchin Sperm Receptor for Egg Jelly (suREJ) Proteins. *Molecular Biology and Evolution*, **22**, 533–541.

Mallatt J, Winchell CJ (2007) Ribosomal RNA genes and deuterostome phylogeny revisited: More cyclostomes, elasmobranchs, reptiles, and a brittle star. *Molecular Phylogenetics and Evolution*, **43**, 1005–1022.

Mayr E (1942) *Systematics and the origin of species, from the viewpoint of a zoologist*. Harvard University Press.

Mayr E (1947) Ecological Factors in Speciation. *Evolution*, **1**, 263–288.

Mayr E (1963) *Animal species and evolution.*

McMillan WO, Raff RA, Palumbi SR (1992) Population Genetic Consequences of Developmental Evolution in Sea Urchins (Genus heliocidaris). *Evolution*, **46**, 1299–1312.

Mercier A, Hamel J-F (2009) *Advances In Marine Biology: Endogenous and Exogenous Control of Gametogenesis and Spawning in Echinoderms*. Academic Press.

Meyer CP, Paulay G (2005) DNA Barcoding: Error Rates Based on Comprehensive Sampling. *PLoS Biol*, **3**, e422.

Michener CD (1970) Diverse approaches to systematics. *Evolutionary Biology*, **4**, 1–38.

Miller RL (1997) Specificity of sperm chemotaxis among great barrier reef shallow-water Holothurians and Ophiuroids. *Journal of Experimental Zoology*, **279**, 189–200.

Miller KJ, Ayre DJ (2008) Population structure is not a simple function of reproductive mode and larval type: insights from tropical corals. *Journal of Animal Ecology*, **77**, 713–724.
Mishler BD (1985) The morphological, developmental, and phylogenetic basis of species concepts in bryophytes. *Bryologist*, 207–214.
Mooi R, David B (2008) Radial Symmetry, the Anterior/Posterior Axis, and Echinoderm Hox Genes. *Annual Review of Ecology, Evolution, and Systematics*, **39**, 43–62.
Mooi R, David B, Marchand D (1994) Echinoderm skeletal homologies: classical morphology meets modern phylogenetics. *Echinoderms through time. Balkema, Rotterdam*, 87–95.
Murphy WJ, Eizirik E, O'Brien SJ *et al.* (2001) Resolution of the Early Placental Mammal Radiation Using Bayesian Phylogenetics. *Science*, **294**, 2348–2351.
Muths D, Davoult D, Gentil F, Jollivet D (2006) Incomplete cryptic speciation between intertidal and subtidal morphs of Acrocnida brachiata (Echinodermata: Ophiuroidea) in the Northeast Atlantic. *Molecular Ecology*, **15**, 3303–3318.
Muths D, Jollivet D, Gentil F, Davoult D (2009) Large-scale genetic patchiness among NE Atlantic populations of the brittle star Ophiothrix fragilis. *Aquatic Biology*, **5**, 117–132.
Nakachi M, Moriyama H, Hoshi M, Matsumoto M (2006) Acrosome reaction is subfamily specific in sea star fertilization. *Developmental Biology*, **298**, 597–604.
Naughton KM, O'Hara TD, Appleton B, Gardner MG (2014) Sympatric cryptic species in the crinoid genus Cenolia (Echinodermata: Crinoidea: Comasteridae) delineated by sequence and microsatellite markers. *Molecular Phylogenetics and Evolution*, **78**, 160–171.
Nelson G, Platnick NI (1981) Systematics and biogeography.
Nguyen HD, Doo SS, Soars NA, Byrne M (2012) Noncalcifying larvae in a changing ocean: warming, not acidification/hypercapnia, is the dominant stressor on development of the sea star Meridiastra calcar. *Global Change Biology*, **18**, 2466–2476.
Nichols R (2001) Gene trees and species trees are not the same. *Trends in Ecology & Evolution*, **16**, 358–364.
Nishihara H, Okada N, Hasegawa M (2007) Rooting the eutherian tree: the power and pitfalls of phylogenomics. *Genome Biol*, **8**, R199.
Nixon KC, Wheeler QD (1990) An amplification of the phylogenetic species concept. *Cladistics*, **6**, 211–223.
Nykjaer L (2009) Mediterranean Sea surface warming 1985-2006. *Climate research*, **39**, 11–17.
O'Hara TD, Byrne M, Cisternas PA (2004) The Ophiocoma erinaceus complex: another case of cryptic speciation in echinoderms. In: *Echinoderms: München. Proceedings of the 11th International Echinoderm Conference* , pp. 537–542.
O'Hara TD, Hugall AF, Thuy B, Moussalli A (2014) Phylogenomic Resolution of the Class Ophiuroidea Unlocks a Global Microfossil Record. *Current Biology*, **24**, 1874–1879.
Palmer AR, Strathmann RR (1981) Scale of dispersal in varying environments and its implications for life histories of marine invertebrates. *Oecologia*, **48**, 308–318.
Palumbi SR (2003) Population genetics, demographic connectivity, and the design of marine reserves. *Ecological Applications*, **13**, 146–158.
Pearson RG, Dawson TP (2003) Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful? *Global Ecology and Biogeography*, **12**, 361–371.
Pechenik JA (1999) On the advantages and disadvantages of larval stages in benthic marine invertebrate life cycles. *Marine Ecology Progress Series*, **177**, 269–297.
Peck LS, Massey A, Thorne MAS, Clark MS (2009) Lack of acclimation in Ophionotus victoriae: brittle stars are not fish. *Polar Biology*, **32**, 399–402.
Penant G, Aurelle D, Feral J, Chenuil A (2013) Planktonic larvae do not ensure gene flow in the edible sea urchin Paracentrotus lividus. *Marine Ecology Progress Series*, **480**, 155–170.
Pérez-Portela R, Almada V, Turon X (2013) Cryptic speciation and genetic structure of widely distributed brittle stars (Ophiuroidea) in Europe. *Zoologica Scripta*, **42**, 151–169.
Perseke M, Bernhard D, Fritzsch G *et al.* (2010) Mitochondrial genome evolution in Ophiuroidea, Echinoidea, and Holothuroidea: Insights in phylogenetic relationships of Echinodermata. *Molecular Phylogenetics and Evolution*, **56**, 201–211.

Pfenninger M, Schwenk K (2007) Cryptic animal species are homogeneously distributed among taxa and biogeographical regions. *BMC Evolutionary Biology*, **7**, 121.

Philippe H, Brinkmann H, Lavrov DV *et al.* (2011) Resolving Difficult Phylogenetic Questions: Why More Sequences Are Not Enough. *PLoS Biol*, **9**, e1000602.

Philippe H, Delsuc F, Brinkmann H, Lartillot N (2005) Phylogenomics. *Annual Review of Ecology, Evolution, and Systematics*, **36**, 541–562.

Pisani D, Feuda R, Peterson KJ, Smith AB (2012) Resolving phylogenetic signal from noise when divergence is rapid: A new look at the old problem of echinoderm class relationships. *Molecular Phylogenetics and Evolution*, **62**, 27–34.

Podlaha O, Webb DM, Tucker PK, Zhang J (2005) Positive Selection for Indel Substitutions in the Rodent Sperm Protein Catsper1. *Molecular Biology and Evolution*, **22**, 1845–1852.

Podlaha O, Zhang J (2003) Positive selection on protein-length in the evolution of a primate sperm ion channel. *Proceedings of the National Academy of Sciences*, **100**, 12241–12246.

Pollock DD, Zwickl DJ, McGuire JA, Hillis DM (2002) Increased Taxon Sampling Is Advantageous for Phylogenetic Inference. *Systematic biology*, **51**, 664–671.

Pons J-M, Sonsthagen S, Dove C, Crochet P-A (2014) Extensive mitochondrial introgression in North American Great Black-backed Gulls (Larus marinus) from the American Herring Gull (Larus smithsonianus) with little nuclear DNA impact. *Heredity*, **112**, 226–239.

Pörtner H. (2002) Climate variations and the physiological basis of temperature dependent biogeography: systemic to molecular hierarchy of thermal tolerance in animals. *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology*, **132**, 739–761.

Poulin E, Feral J-P (1996) Why are There so Many Species of Brooding Antarctic Echinoids? *Evolution*, **50**, 820–830.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of Population Structure Using Multilocus Genotype Data. *Genetics*, **155**, 945–959.

Puritz JB, Keever CC, Addison JA *et al.* (2012) Extraordinarily rapid life-history divergence between Cryptasterina sea star species. *Proceedings of the Royal Society B: Biological Sciences*, **279**, 3914–3922.

Purushothaman S, Saxena S, Meghah V *et al.* (2015) Transcriptomic and proteomic analyses of Amphiura filiformis arm tissue-undergoing regeneration. *Journal of Proteomics*, **112**, 113–124.

De Queiroz K (2007) Species concepts and species delimitation. *Systematic biology*, **56**, 879–886.

Ravaoarimanana IB, Tiedemann R, Montagnon D, Rumpler Y (2004) Molecular and cytogenetic evidence for cryptic speciation within a rare endemic Malagasy lemur, the Northern Sportive Lemur (Lepilemur septentrionalis). *Molecular Phylogenetics and Evolution*, **31**, 440–448.

Rawson PD, Hilbish TJ (1995) Evolutionary relationships among the male and female mitochondrial DNA lineages in the Mytilus edulis species complex. *Molecular Biology and Evolution*, **12**, 893–901.

Reusch TBH, Ehlers A, Hämmerli A, Worm B (2005) Ecosystem recovery after climatic extremes enhanced by genotypic diversity. *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 2826–2831.

Ribout C, Carpentieri C (2013) Automated genomic DNA purification of marine organisms on the epMotion® 5075 VAC from Eppendorf. *Eppendorf Application Note*, **281**, 1–6.

Riginos C, Douglas KE, Jin Y, Shanahan DF, Treml EA (2011) Effects of geography and life history traits on genetic differentiation in benthic marine fishes. *Ecography*, **34**, 566–575.

Romiguier J, Gayral P, Ballenghien M *et al.* (2014) Comparative population genomics in animals uncovers the determinants of genetic diversity. *Nature*, **515**, 261–263.

Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, **19**, 1572–1574.

Rosen DE (1979) Fishes from the uplands and intermontane basins of Guatemala: revisionary studies and comparative geography. Bulletin of the AMNH; v. 162, article 5.

Rosenberg R, Lundberg L (2004) Photoperiodic activity pattern in the brittle star Amphiura filiformis. *Marine Biology*, **145**, 651–656.

Rubinoff D, Cameron S, Will K (2006) A Genomic Perspective on the Shortcomings of Mitochondrial DNA for "Barcoding" Identification. *Journal of Heredity*, **97**, 581–594.

Russello MA, Glaberman S, Gibbs JP *et al.* (2005) A cryptic taxon of Galápagos tortoise in conservation peril. *Biology Letters*, **1**, 287–290.

Sanford E, Kelly MW (2011) Local Adaptation in Marine Invertebrates. *Annual Review of Marine Science*, **3**, 509–535.

Schlick-Steiner BC, Seifert B, Stauffer C *et al.* (2007) Without morphology, cryptic species stay in taxonomic crypsis following discovery. *Trends in Ecology & Evolution*, **22**, 391–392.

Schlick-Steiner BC, Steiner FM, Seifert B *et al.* (2010) Integrative Taxonomy: A Multisource Approach to Exploring Biodiversity. *Annual Review of Entomology*, **55**, 421–438.

Schloss PD, Westcott SL, Ryabin T *et al.* (2009) Introducing mothur: Open-Source, Platform-Independent, Community-Supported Software for Describing and Comparing Microbial Communities. *Applied and Environmental Microbiology*, **75**, 7537–7541.

Schönrogge K, Barr B, Wardlaw JC *et al.* (2002) When rare species become endangered: cryptic speciation in myrmecophilous hoverflies. *Biological Journal of the Linnean Society*, **75**, 291–300.

Selkoe KA, Toonen RJ (2011) Marine connectivity: a new look at pelagic larval duration and genetic metrics of dispersal. *Marine Ecology Progress Series*, **436**, 291–305.

Selvakumaraswamy P, Byrne M (2000) Reproduction, spawning, and development of 5 ophiuroids from Australia and New Zealand. *Invertebrate Biology*, **119**, 394–402.

Shanks AL (2009) Pelagic Larval Duration and Dispersal Distance Revisited. *The Biological Bulletin*, **216**, 373–385.

Shanks AL, Grantham BA, Carr MH (2003) Propagule dispersal distance and the size and spacing of marine reserves. *Ecological Applications*, **13**, 159–169.

Sherman CDH, Hunt A, Ayre DJ (2008) Is life history a barrier to dispersal? Contrasting patterns of genetic differentiation along an oceanographically complex coast. *Biological Journal of the Linnean Society*, **95**, 106–116.

Siegel DA, Kinlan BP, Gaylord B, Gaines SD (2003) Lagrangian descriptions of marine larval dispersion. *Marine Ecology Progress Series*, **260**, 83–96.

Sierra R, Matz MV, Aglyamova G *et al.* (2013) Deep relationships of Rhizaria revealed by phylogenomics: A farewell to Haeckel's Radiolaria. *Molecular Phylogenetics and Evolution*, **67**, 53–59.

Singletary RL (1971) Thermal Tolerance of Ten Shallow-Water Ophiuroids in Biscayne Bay, Florida. *Bulletin of Marine Science*, **21**, 938–943.

Smith AB (1988) Fossil evidence for the relationships of extant echinoderm classes and their times of divergence. *Echinoderm phylogeny and evolutionary biology*, **1**, 85.

Smith AB (1997a) Echinoderm phylogeny: how congruent are morphological and molecular estimates. *Geobiology of echinoderms (ed. JA Waters & CG Maples) Pal. Soc. Pap*, **3**, 337–355.

Smith AB (1997b) Echinoderm Larvae and Phylogeny. *Annual Review of Ecology and Systematics*, **28**, 219–241.

Smith MJ, Arndt A, Gorski S, Fajber E (1993) The phylogeny of echinoderm classes based on mitochondrial gene arrangements. *Journal of Molecular Evolution*, **36**, 545–554.

Sneath PH, Sokal RR, others (1973) *Numerical taxonomy. The principles and practice of numerical classification.*

Snook RR (2005) Sperm in competition: not playing by the numbers. *Trends in Ecology & Evolution*, **20**, 46–53.

Sokal RR, Crovello TJ (1970) The biological species concept: a critical evaluation. *American Naturalist*, 127–153.

Solomon S (2007) *Climate Change 2007 - The Physical Science Basis: Working Group I Contribution to the Fourth Assessment Report of the IPCC*. Cambridge University Press.

Somero GN (2002) Thermal Physiology and Vertical Zonation of Intertidal Animals: Optima, Limits, and Costs of Living. *Integrative and Comparative Biology*, **42**, 780–789.

Somero GN (2005) Linking biogeography to physiology: evolutionary and acclimatory adjustments of thermal limits. *Front Zool*, **2**, 1–1.

Somero GN (2010) The physiology of climate change: how potentials for acclimatization and genetic adaptation will determine "winners" and "losers." *The Journal of Experimental Biology*, **213**, 912–920.

Sorte CJB, Jones SJ, Miller LP (2011) Geographic variation in temperature tolerance as an indicator of potential population responses to climate change. *Journal of Experimental Marine Biology and Ecology*, **400**, 209–217.

Sponer R, Roy MS (2002) Phylogeographic Analysis of the Brooding Brittle Star Amphipholis Squamata (echinodermata) Along the Coast of New Zealand Reveals High Cryptic Genetic Variation and Cryptic Dispersal Potential. *Evolution*, **56**, 1954–1967.

Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, **30**, 1312–1313.

Stearns SC (1977) The Evolution of Life History Traits: A Critique of the Theory and a Review of the Data. *Annual Review of Ecology and Systematics*, **8**, 145–171.

Stearns SC (1992) *The evolution of life histories*. Oxford University Press Oxford.

Steele CA, Baumsteiger J, Storfer A (2009) Influence of life-history variation on the genetic structure of two sympatric salamander taxa. *Molecular Ecology*, **18**, 1629–1639.

Stephens GC, Virkar RA (1966) Uptake of Organic Material by Aquatic Invertebrates. Iv. the Influence of Salinity on the Uptake of Amino Acids by the Brittle Star, Ophiactis Arenosa. *The Biological Bulletin*, **131**, 172–185.

Stöhr S, Boissin E, Chenuil A (2009) Potential cryptic speciation in Mediterranean populations of Ophioderma (Echinodermata: Ophiuroidea). *Zootaxa*, 1–20.

Stöhr S, Muths D (2010) Morphological diagnosis of the two genetic lineages of Acrocnida brachiata (Echinodermata: Ophiuroidea), with description of a new species. *Journal of the Marine Biological Association of the United Kingdom*, **90**, 831–843.

Stöhr S, O'Hara TD, Thuy B (2012) Global Diversity of Brittle Stars (Echinodermata: Ophiuroidea). *PLoS ONE*, **7**, e31940.

Strathmann RR (1985) Feeding and Nonfeeding Larval Development and Life-History Evolution in Marine Invertebrates. *Annual Review of Ecology and Systematics*, **16**, 339–361.

Strathmann RR, Chaffee C (1984) Constraints on egg masses. II. Effect of spacing, size, and number of eggs on ventilation of masses of embryos in jelly, adherent groups, or thin-walled capsules. *Journal of Experimental Marine Biology and Ecology*, **84**, 85–93.

Strathmann RR, Strathmann MF (1982) The Relationship Between Adult Size and Brooding in Marine Invertebrates. *The American Naturalist*, **119**, 91–101.

Strathmann RR, Strathmann MF, Emson RH (1984) Does Limited Brood Capacity Link Adult Size, Brooding, and Simultaneous Hermaphroditism? A Test with the Starfish Asterina phylactica. *The American Naturalist*, **123**, 796–818.

Sunday JM, Hart MW (2013) Sea star populations diverge by positive selection at a sperm-egg compatibility locus. *Ecology and Evolution*, **3**, 640–654.

Suzuki N, Hoshi M, Nomura K, Isaka S (1982) Respiratory stimulation of sea urchin spermatozoa by egg extracts, egg jelly extracts and egg jelly peptides from various species of sea urchins: Taxonomical significance. *Comparative Biochemistry and Physiology Part A: Physiology*, **72**, 489–495.

Swofford D (2003) PAUP*: phylogenetic analysis using parsimony, version 4.0b10.

Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, **123**, 585–595.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Molecular Biology and Evolution*, **30**, 2725–2729.

Tarnowska K, Krakau M, Jacobsen S *et al.* (2012) Comparative phylogeography of two sister (congeneric) species of cardiid bivalve: Strong influence of habitat, life history and post-glacial history. *Estuarine, Coastal and Shelf Science*, **107**, 150–158.

Telford MJ, Lowe CJ, Cameron CB *et al.* (2014) Phylogenomic analysis of echinoderm class relationships supports Asterozoa. *Proceedings of the Royal Society B: Biological Sciences*, **281**, 20140479.

Teske PR, Papadopoulos I, Zardi GI *et al.* (2007) Implications of life history for genetic structure and migration rates of southern African coastal invertebrates: planktonic, abbreviated and direct development. *Marine Biology*, **152**, 697–711.

Thomas CD, Cameron A, Green RE *et al.* (2004) Extinction risk from climate change. *Nature*, **427**, 145–148.

Thomson MJ, Zhao K, Wright M *et al.* (2012) High-throughput single nucleotide polymorphism genotyping for breeding applications in rice using the BeadXpress platform. *Molecular Breeding*, **29**, 875–886.

Toews DPL, Brelsford A (2012) The biogeography of mitochondrial and nuclear discordance in animals. *Molecular Ecology*, **21**, 3907–3930.

Torrence KG, Correia MD, Hoffman EA (2012) Divergent sympatric lineages of the Atlantic and Indian Ocean crinoid Tropiometra carinata. *Invertebrate Biology*, **131**, 355–365.

Tortonese E (1983) Remarks on the morphology and taxonomy of Ophioderma longicaudum (Retz.) from the Mediterranean. *Atti della Societa Italiana di Scienze Naturali e del Museo Civico di Storia Naturale in Milano*, **124**, 21–28.

Touratier F, Goyet C (2011) Impact of the Eastern Mediterranean Transient on the distribution of anthropogenic CO2 and first estimate of acidification for the Mediterranean Sea. *Deep Sea Research Part I: Oceanographic Research Papers*, **58**, 1–15.

Trontelj P, Fišer C (2009) Perspectives: Cryptic species diversity should not be trivialised. *Systematics and Biodiversity*, **7**, 1–3.

Vacquier VD (1998) Evolution of gamete recognition proteins. *Science*, **281**, 1995–1998.

Vacquier VD, Swanson WJ (2011) Selection in the Rapid Evolution of Gamete Recognition Proteins in Marine Invertebrates. *Cold Spring Harbor Perspectives in Biology*, **3**, a002931.

Valen LV (1976) Ecological Species, Multispecies, and Oaks. *Taxon*, **25**, 233–239.

Vargas-Yáñez M, Jesús García M, Salat J *et al.* (2008) Warming trends and decadal variability in the Western Mediterranean shelf. *Global and Planetary Change*, **63**, 177–184.

Vargas-Yáñez M, Moya F, García-Martínez MC *et al.* (2010) Climate change in the Western Mediterranean sea 1900–2008. *Journal of Marine Systems*, **82**, 171–176.

Vaughn R, Garnhart N, Garey JR, Thomas WK, Livingston BT (2012) Sequencing and analysis of the gastrula transcriptome of the brittle star Ophiocoma wendtii. *EvoDevo*, **3**, 19–19.

Verling E, Crook A, Barnes D (2002) Covering behaviour in Paracentrotus lividus: is light important? *Marine Biology*, **140**, 391–396.

Wada H, Satoh N (1994) Phylogenetic relationships among extant classes of echinoderms, as inferred from sequences of 18S rDNA, coincide with relationships deduced from the fossil record. *Journal of Molecular Evolution*, **38**, 41–49.

Wake DB, Roth G, Wake MH (1983) On the problem of stasis in organismal evolution. *Journal of Theoretical Biology*, **101**, 211–224.

Walsh P, Metzger D, Higuchi R (1991) Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. *BioTechniques*, **10**, 506–513.

Wang D, King SM, Quill TA, Doolittle LK, Garbers DL (2003) A new sperm-specific Na+/H+ Exchanger required for sperm motility and fertility. *Nature Cell Biology*, **5**, 1117–1122.

Ward RD, Holmes BH, O'hara TD (2008) DNA barcoding discriminates echinoderm species. *Molecular Ecology Resources*, **8**, 1202–1211.

Weber AA-T, Dupont S, Chenuil A (2013) Thermotolerance and regeneration in the brittle star species complex Ophioderma longicauda: A preliminary study comparing lineages and Mediterranean basins. *Comptes Rendus Biologies*, **336**, 572–581.

Weber AA-T, Pawlowski J (2014) Wide Occurrence of SSU rDNA Intragenomic Polymorphism in Foraminifera and its Implications for Molecular Species Identification. *Protist*, **165**, 645–661.

Weber AA-T, Stöhr S, Chenuil A (2014) Genetic data, reproduction season and reproductive strategy data support the existence of biological species in Ophioderma longicauda. *Comptes Rendus Biologies*, **337**, 553–560.

Weersing K, Toonen RJ (2009) Population genetics, larval dispersal, and connectivity in marine systems. *Marine Ecology Progress Series*, **393**, 1–12.

Whitfield AK, Panfili J, Durand J-D (2012) A global review of the cosmopolitan flathead mullet Mugil cephalus Linnaeus 1758 (Teleostei: Mugilidae), with emphasis on the biology, genetics, ecology and fisheries aspects of this apparent species complex. *Reviews in Fish Biology and Fisheries*, **22**, 641–681.

Whitlock MC (2004) Selection and drift in metapopulations. In: *Ecology, genetics and evolution of metapopulations* , pp. 153–173. Hanski I, Gaggiotti O.E.

Wilkins JS (2011) Philosophically speaking, how many species concepts are there. *Zootaxa*, **2765**, 58–60.

Williams ST (2000) Species boundaries in the starfish genus Linckia. *Marine Biology*, **136**, 137–148.

Will KW, Mishler BD, Wheeler QD (2005) The Perils of DNA Barcoding and the Need for Integrative Taxonomy. *Systematic Biology*, **54**, 844–851.

Wilson NG, Hunter RL, Lockhart SJ, Halanych KM (2007) Multiple lineages and absence of panmixia in the “circumpolar” crinoid Promachocrinus kerguelensis from the Atlantic sector of Antarctica. *Marine Biology*, **152**, 895–904.

Wong JL, Wessel GM (2004) Major components of a sea urchin block to polyspermy are structurally and functionally conserved. *Evolution & Development*, **6**, 134–153.

Wood HL, Spicer JI, Lowe DM, Widdicombe S (2010) Interaction of ocean acidification and temperature; the high cost of survival in the brittlestar Ophiura ophiura. *Marine Biology*, **157**, 2001–2013.

Yang Z, Rannala B (2010) Bayesian species delimitation using multilocus sequence data. *Proceedings of the National Academy of Sciences*, **107**, 9264–9269.

Zhou ZC, Dong Y, Sun HJ *et al.* (2014) Transcriptome sequencing of sea cucumber (Apostichopus japonicus) and the identification of gene-associated markers. *Molecular Ecology Resources*, **14**, 127–138.

Zwickl DJ, Hillis DM (2002) Increased Taxon Sampling Greatly Reduces Phylogenetic Error. *Systematic Biology*, **51**, 588–598.

Ecological and genetic study of the cryptic species complex *Ophioderma longicauda* (Ophiuroidea: Echinodermata): comparison between brooding and broadcasting lineages.

Cryptic species (species wrongly described with the same name due to morphological resemblance) are numerous in the marine environment. *Ophioderma longicauda* (Bruzelius, 1805) is a cryptic species complex including six mitochondrial lineages (L1-L6), of which three (L2-L3-L4) brood their juveniles during the reproductive period, whereas the other lineages most likely reproduce using lecithotrophic larvae (confirmed for L1). The first objective of this thesis was to define the species limits in the *O. longicauda* complex. To do this, the reproductive status of lineages L1 and L3 collected in sympatry was studied for a month. The morphological and genetic study of those two lineages showed that they were indeed different biological species, with notably different reproductive periods. Furthermore, a comparison of thermotolerance and regeneration of the two lineages showed that the species complex *O. longicauda* was globally resistant to elevated temperatures, and that the population L3 was potentially more thermotolerant than population L1. Finally, the analysis of 31 genetic markers using DAPC showed that the *O. longicauda* complex included six distinct genetic groups, which were slightly different from the mitochondrial lineages. In addition, a comparison of historical scenarios using ABC revealed that the most likely scenario included a hybridization event between brooders and broadcasters. The second objective of this thesis was to study the influence of life-history traits on connectivity and genetic diversity. To do so, 10 markers were sequenced for six sympatric populations of lineages L1 and L3 from Greece. The genetic structure was high for the brooding species, whereas the broadcasting species did not display any genetic structure at that scale. Furthermore, the average estimated migrant number between populations was about 50 times higher in L1 compared to L3. The analysis of genetic diversity (the average per population nucleotide and haplotype diversity) for these 10 markers showed that diversity was three to four times higher in broadcasters than in brooders. The analysis of genetic diversity (synonymous nucleotide diversity and heterozygosity) in the L1 and L3 transcriptomes showed that diversity was 1.5 to 2 times higher in broadcasters than in brooders, confirming the results obtained using a small number of markers. Finally, genes that were positively selected between L1 and L3 and involved in reproduction were identified. Two ion channels (TetraKCNG and NHE) involved in sperm motility showed an evolution under positive selection. These results suggest that sperm competition might be a mechanism of pre-zygotic isolation in *Ophioderma longicauda*.

Key words: brooding; brittle star; species delimitation ; thermotolerance ; regeneration ; dispersal ; connectivity ; genetic diversity ; comparative transcriptomics ; positive selection.

Etude écologique et génétique du complexe d'espèces cryptiques *Ophioderma longicauda* (Ophiuroidea : Echinodermata) : comparaison entre lignées incubantes et lignées produisant des larves planctoniques.

Les espèces cryptiques (espèces incorrectement décrites sous le même nom à cause d'une ressemblance morphologique) sont nombreuses en milieu marin. *Ophioderma longicauda* (Bruzelius, 1805) est un complexe d'espèces cryptiques incluant six lignées mitochondriales (L1-L6), dont certaines (L2-L3-L4) possèdent la caractéristique d'incuber leur descendance durant la période de reproduction, alors que les autres lignées se reproduisent probablement via des larves lécithotrophes (confirmé pour L1). Cette thèse avait comme premier objectif de définir les limites d'espèces dans le complexe *O. longicauda*. Pour ce faire, le statut reproductif des lignées L1 et L3 se trouvant en sympatrie a été étudié pendant un mois. L'analyse morphologique et génétique de ces deux lignées a montré qu'il s'agissait effectivement d'espèces biologiques différentes, avec notamment différentes périodes de reproduction. De plus, une comparaison de la thermotolérance et de la régénération chez ces deux lignées a montré que le complexe *O. longicauda* était globalement résistant aux températures élevées, et que la population L3 de Crète était potentiellement plus thermotolérante que la lignée L1. Finalement, l'analyse par DAPC de 31 marqueurs génétiques a montré que le complexe *O. longicauda* était constitué de six groupes génétiques distincts, légèrement différents des six lignées mitochondriales. La comparaison de scénarios historiques par la méthode ABC a montré que le scénario le plus probable incluait un événement d'hybridation entre les dispersants et les incubants. Le deuxième objectif de cette thèse consistait à étudier l'influence des traits d'histoire de vie sur la connectivité et la diversité génétique. Pour ce faire, 10 marqueurs ont été séquencés pour six populations sympatriques des lignées L1 et L3 en Grèce. La structure génétique était très marquée pour l'espèce incubante L3, tandis que l'espèce dispersante L1 n'a pas montré de structure génétique à cette échelle. De plus, le nombre de migrants moyen estimé entre populations était environ 50 fois plus élevé chez les L1 comparé aux L3. L'analyse de la diversité génétique (diversités nucléotidique et haplotypique moyenne par population) pour ces 10 marqueurs a montré que celle des dispersantes était trois à quatre fois plus élevée que celle des incubantes. L'analyse de la diversité génétique (diversité nucléotidique synonyme et hétérozygotie) dans les transcriptomes des L1 et L3 a montré qu'elle était 1.5 à 2 fois plus élevée chez les dispersantes que chez les incubantes, confirmant la tendance retrouvée sur un petit nombre de marqueurs. Finalement, les gènes évoluant sous sélection positive entre L1 et L3 et impliqués dans la reproduction ont été identifiés. Deux canaux ioniques (TetraKCNG et NHE) impliqués dans la mobilité des spermatozoïdes ont montré une évolution sous sélection positive. Ces résultats suggèrent que la compétition des spermatozoïdes pourrait être un mécanisme d'isolement pré-zygotique chez *Ophioderma longicauda*.

Mots clé : incubation ; ophiures ; délimitation d'espèces ; thermotolérance ; régénération ; dispersion ; connectivité ; diversité génétique ; transcriptomique comparée ; sélection positive.