



**HAL**  
open science

# Histoire évolutive des remaniements chromosomiques en liaison avec la mobilisation d'éléments transposables chez les téléostéens antarctiques Nototheniidae : la radiation adaptative du groupe " Trematomus "

Juliette Auvinet

## ► To cite this version:

Juliette Auvinet. Histoire évolutive des remaniements chromosomiques en liaison avec la mobilisation d'éléments transposables chez les téléostéens antarctiques Nototheniidae : la radiation adaptative du groupe " Trematomus ". Génétique. Sorbonne Université, 2018. Français. NNT : 2018SORUS371 . tel-02438851

**HAL Id: tel-02438851**

**<https://theses.hal.science/tel-02438851v1>**

Submitted on 14 Jan 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Ecole doctorale Complexité du Vivant (ED 515)

Programme doctoral Interface pour le Vivant

Institut de Biologie Paris-Seine

UMR7138 "Evolution Paris-Seine" SU-CNRS

Equipe : Evolution des génomes Eucaryotes

**Histoire évolutive des remaniements chromosomiques  
en liaison avec la mobilisation d'éléments transposables  
chez les téléostéens antarctiques Nototheniidae :  
la radiation adaptative du groupe « *Trematomus* ».**

Par Juliette Auvinet

Soutenance le **19 octobre 2018** pour l'obtention du titre de Docteur en  
génétique évolutive

Composition du jury :

**Yves Desdevises**, PU, Sorbonne Université (SU)

**Eva Pisano**, PI Université de Gênes

**Filip Volckaert**, PU, Université de Louvain

**Jean-Nicolas Volff**, PU, ENS de Lyon

**Catherine Montchamp-Moreau**, DR, Université Paris-Saclay

**Dominique Higuët**, PU, SU

**Agnès Dettai**, MCF, MNHN

**Catherine Ozouf-Costaz**, IR, CNRS, SU

**Examineur**

**Rapporteur**

**Rapporteur**

**Examineur**

**Examinatrice**

**Directeur de thèse**

**Directrice de thèse**

**Co-encadrante (invitée)**



# Table des matières

|  |           |
|--|-----------|
| <b>REMERCIEMENTS</b> .....   | <b>5</b>  |
| <b>I) INTRODUCTION GENERALE</b> .....  | <b>7</b>  |
| <b>II) CHAPITRE 1 : CARACTERISATION DES REMANIEMENTS CHROMOSOMIQUES AU SEIN DU GROUPE « TREMATOMUS »</b> ..... | <b>17</b> |
| <b>II.1 : NATURE DES REMANIEMENTS CHROMOSOMIQUES</b> .....   | <b>17</b> |
| II.1.A : HISTOIRE DES DIVERSIFICATIONS CHROMOSOMIQUES CHEZ LES NOTOTHENIIDAE .....                             | 17        |
| II.1.B : CONSERVATION DE LA TAILLE DES GENOMES CHEZ LES NOTOTHENIIDAE.....                                     | 21        |
| II.1.B.1 : Introduction.....   | 21        |
| II.1.B.2 : Matériels et Méthodes.....  | 23        |
| II.1.B.3 : Résultats.....  | 26        |
| II.1.B.4 : Discussion.....   | 28        |
| <b>II.2 : RECHERCHE ET IDENTIFICATION DES HOMOLOGIES CHROMOSOMIQUES INTERSPECIFIQUES</b> .....                 | <b>30</b> |
| II.2.A : LES HOMOLOGIES CHROMOSOMIQUES INTERSPECIFIQUES.....   | 30        |
| II.2.A.1 : Homologies interspécifiques entre chromosomes entiers.....  | 35        |
| II.2.A.2 : Homologies entre unités structurales au sein des « <i>Trematomus</i> » .....                        | 37        |
| II.2.B : METHODES.....   | 39        |
| II.2.C : RECHERCHE D’HCI PAR PEINTURE DE LA GRANDE PAIRE DE <i>T. PENNELLII</i> MICRODISSEQUEE .....           | 44        |
| II.2.C.1 : Résultats .....   | 44        |
| II.2.C.2 : Discussion :.....   | 48        |
| II.2.D : RECHERCHE DES HOMOLOGIES DES UNITES STRUCTURALES INTERSPECIFIQUES PAR BAC-FISH .....                  | 53        |
| II.2.D.1 : Résultats.....  | 53        |
| II.2.D.2 : Discussion.....   | 59        |
| II.2.E : DISCUSSION .....  | 64        |
| <b>III) CHAPITRE 2 : PHYLOGENIE DES « TREMATOMUS »</b> .....   | <b>67</b> |
| <b>III.A : INTRODUCTION</b> .....  | <b>67</b> |
| III.A.1 : COMPOSITION DU GROUPE « <i>Trematomus</i> » .....  | 67        |
| III.A.2 : PHYLOGENIE DU GROUPE « <i>Trematomus</i> » .....   | 69        |
| III.A.3 : ANALYSE PHYLOGENOMIQUE.....  | 71        |

|  |                   |
|--|-------------------|
| <b>III.B : MATERIELS ET METHODES .....</b>   | <b>72</b>         |
| <b>III.C : RESULTATS .....</b>   | <b>78</b>         |
| III.C.1 : GENERATION DU JEU DE DONNEES POUR FORMER LES LOCUS .....   | 78                |
| III.C.2 : RECONSTRUCTIONS PHYLOGENETIQUES.....   | 81                |
| <b>III.D : DISCUSSION .....</b>  | <b>86</b>         |
| III.D.1 : JEUX DE DONNEES ET PARAMETRES D'ANALYSE .....  | 86                |
| III.D.2 : INFLUENCE DES DONNEES MANQUANTES.....  | 88                |
| III.D.3 : RESOLUTION DES RELATIONS PHYLOGENETIQUES CHEZ LES « <i>TREMATOMUS</i> » .....  | 91                |
| <br>   |                   |
| <b><u>IV) CHAPITRE 3 : LES ELEMENTS TRANSPOSABLES, FACILITATEURS DE CES<br/>REMANIEMENTS AU COURS DE L'EVOLUTION? .....</u></b>  | <b><u>95</u></b>  |
| <br>   |                   |
| IV.1 : INTRODUCTION .....  | 95                |
| <b>IV.2 : DIVERSITE, QUANTIFICATION, LOCALISATION CHROMOSOMIQUES DES RETROTRANSPOSONS A LTR<br/>GYPSY ET COPIA ET RETROTRANSPOSONS A YR DIRS1 CHEZ LES « <i>TREMATOMUS</i> » .....</b> | <b>100</b>        |
| IV.2.A : RESUME DES PRINCIPAUX RESULTATS .....   | 101               |
| IV.2.B : MOBILIZATION OF RETROTRANSPOSONS AS A CAUSE OF CHROMOSOMAL DIVERSIFICATION AND RAPID<br>SPECIATION: THE CASE FOR THE ANTARCTIC TELEOST GENUS <i>TREMATOMUS</i> (ARTICLE)..... | 103               |
| <b>IV.3 : DIVERSITE DES ELEMENTS TRANSPOSABLES A L'ECHELLE DE LA FAMILLE DES NOTOTHENIIDAE... 145</b>  |                   |
| IV.3.A : RECHERCHE DES FAMILLES DES TROIS RETROTRANSPOSONS CHEZ LES NOTOTHENIIDAE .....  | 145               |
| IV.3.B : DIVERSITE DES ELEMENTS TRANSPOSABLES CHEZ LES « <i>TREMATOMUS</i> ».....  | 148               |
| <br>   |                   |
| <b><u>V) DISCUSSION GENERALE : HISTOIRE EVOLUTIVE DES REMANIEMENTS<br/>CHROMOSOMIQUES DES « <i>TREMATOMUS</i> » .....</u></b>  | <b><u>153</u></b> |
| <br>   |                   |
| V.1 : LES FUSIONS CHROMOSOMIQUES, REMANIEMENTS MAJORITAIRES SURVENUS LORS DE LA<br>DIVERSIFICATION DES « <i>TREMATOMUS</i> ».....  | 153               |
| V.2 : SPECIATIONS CHROMOSOMIQUES MULTIPLES CHEZ LES « <i>TREMATOMUS</i> » .....  | 158               |
| V.3 : VERS UNE RECONSTRUCTION DES SCENARIOS EVOLUTIFS DES DIVERSIFICATIONS CHROMOSOMIQUES<br>.....   | 161               |
| <br>   |                   |
| <b><u>VI) CONCLUSION ET PERSPECTIVES .....</u></b>   | <b><u>169</u></b> |
| <br>   |                   |
| <b><u>VII) BIBLIOGRAPHIE : .....</u></b>   | <b><u>171</u></b> |
| <br>   |                   |
| <b><u>VIII) ANNEXES.....</u></b>   | <b><u>199</u></b> |

## Remerciements

Je souhaiterais remercier tout d'abord mes deux encadrants de thèse. Dominique Higuët, côté Sorbonne Université et Agnès Dettaï côté Muséum. Merci de m'avoir accueillie au sein de l'unité et d'avoir supervisé ces travaux de master suivis de la thèse. Ces dernières années ont été un parcours enrichissant et passionnant à conduire à vos côtés.

Je remercie les membres de mon jury, mes rapporteurs Eva Pisano et Filip Volkaert ainsi que mes examinateurs, Yves Desdevises, Catherine Montchamps Moreau et Jean-Nicolas Volff qui ont accepté de porter leur regard sur mon travail.

Je tiens également à remercier Laure Teyssset, Dider Casane et de nouveau Jean-Nicolas Volff pour avoir accepté de faire partie de mon comité de thèse.

Je tiens à remercier toutes les personnes avec qui j'ai pu nourrir des échanges intéressants et constructifs et qui m'ont fait progresser tout au long de ma thèse. En particulier Pedro Bausero, Laurence Petit, Olivier Coriton, Jawad Abdelkrim, Sophie Brouillet, Guillaume Achaz, Régis Debruyne, Delphine Gey, Josie Lambourdière, Céline Bonnillo, sans oublier Paula Graça pour leur aide précieuse.

Merci à l'équipe pédagogique de la Licence Sciences de la Vie (SDV, SU) de m'avoir accompagnée durant toutes ces années, et de m'avoir permis de m'accomplir en tant que chargée de mission enseignement.

Un grand merci aux membres du Service de Systématique Moléculaire (MNHN) et de l'Unité Evolution Paris Seine (SU) pour leur aide, leur soutien, les moments de partage. Merci à mes collègues doctorants, en particulier Thomas, Gabriel et Anne-Sophie ; ensemble nous avons formé une bien belle équipe.

Je voudrais remercier particulièrement Catherine Ozouf-Costaz qui a joué un rôle fondamental dans mon apprentissage et ma réflexion, ainsi que dans ma vocation à poursuivre dans ce merveilleux monde antarctique. Merci de m'avoir accompagnée durant le Master et la thèse avec un regard toujours bienveillant et attentionné.

Merci à l'équipe de Bill Detrich et de John Postlethwait, ainsi qu'à Thomas Desvignes, qui, en plus d'une collaboration fructueuse, m'ont permis de réaliser une mission sur la base de Palmer en péninsule antarctique durant mes derniers mois de thèse, dans la perspective de poursuivre les recherches au sein de leur laboratoire.

Enfin, un grand merci à mes amis, ma meilleure amie Valentine, et à toute ma famille qui ont été très présents tout au long de ce périple. Une pensée pour zeyde, mon grand-père tant aimé aux yeux pétillants d'intelligence et d'humour, et dont le rire résonnera toujours dans mon cœur. Merci à l'homme que j'aime pour son accompagnement et la force qu'il m'a communiquée. Merci pour ce soutien de tous les jours qui m'a poussée à me dépasser et me permet d'avancer avec combativité et détermination pour réaliser mes rêves.



## I) Introduction générale

Les téléostéens antarctiques de la famille des Nototheniidae dominent l'ichtyofaune australe actuelle (Eastman 2005; Cheng and Detrich 2007; Rutschmann et al. 2011). Ils représentent en effet 46 % des téléostéens de l'Océan Austral, 77 % de la diversité spécifique et 91,2 % de leur biomasse totale (Eastman and McCune 2000; Eastman 2005). Cette prédominance peut être liée à plusieurs séries de diversifications rapides au sein de cette famille dans cet environnement marin « extrême ». Leur diversité est à la fois d'ordre écologique, spécifique et morphologique (Eastman and McCune 2000; Lecointre et al. 2013). Chez certains groupes de cette famille, la diversité spécifique est accompagnée d'une variabilité chromosomique importante (Ozouf-Costaz et al. 1991; Pisano and Ozouf-Costaz 2000; Ghigliotti et al. 2015), rare parmi les autres téléostéens (Molina 2006; Molina and de Freitas Bacurau 2006). Tenter de comprendre comment une telle diversification a pu s'établir au cours de l'évolution de ces organismes est particulièrement intéressant dans le contexte des changements environnementaux qu'a connu le continent antarctique au cours des quarante derniers millions d'années (Ma) (Kennett et al. 1985; Kennett and Barker 1990; Stott et al. 1990). Du fait du nombre limité d'espèce chez les Nototheniidae et des efforts d'échantillonnage menés ces dernières années, une étude comparative est maintenant possible à l'échelle de cette famille.

Le sous-ordre des Notothenioidei est constitué de quatre familles (Duhamel et al. 2014). La première famille à diverger des autres est celle des Bovichtidae, suivie des Pseudaphritidae, et enfin, les Eleginopsidae, groupe frère des Nototheniidae (Balushkin 1992; Lecointre et al. 1997; Ritchie et al. 1997; Balushkin 2000) (Figure I.1). Leurs divergences interviennent dans des périodes concomitantes à celles de l'ouverture du passage de Drake durant l'Eocène (il y a 40 - 34 Ma) (Kennett 1977; Livermore et al. 2005; Scher and Martin 2006) suivie de la mise en place du front polaire et du courant circumpolaire antarctique à la jonction Oligocène - Miocène (il y a 23 - 25 Ma) (Matschiner et al. 2011; Near et al. 2012; Colombo et al. 2015). Les Bovichtidae, Pseudaphritidae, et Eleginopsidae sont actuellement distribués en dehors de la convergence Antarctique (front polaire), principalement dans les eaux côtières de Nouvelle-Zélande, Australie, Amérique du Sud, îles sub-antarctiques des secteurs indiens et atlantiques (Gon and Heemstra 1990; Balushkin 2000; Duhamel et al. 2014). Au contraire, les espèces appartenant aux Nototheniidae sont presque toutes retrouvées à l'intérieur du périmètre antarctique, à l'exception des espèces des genres *Indonotothenia*, *Lepidonotothen*, *Patagonotothen* et de l'espèce *Notothenia angustata*, retrouvées au niveau des îles sub-antarctiques telles que Kerguelen, Malouines ou Heard (Gon and Heemstra 1990; Duhamel et al. 2014). Cette famille a été par conséquent appelée « clade Antarctique » par certains auteurs (Near 2004).

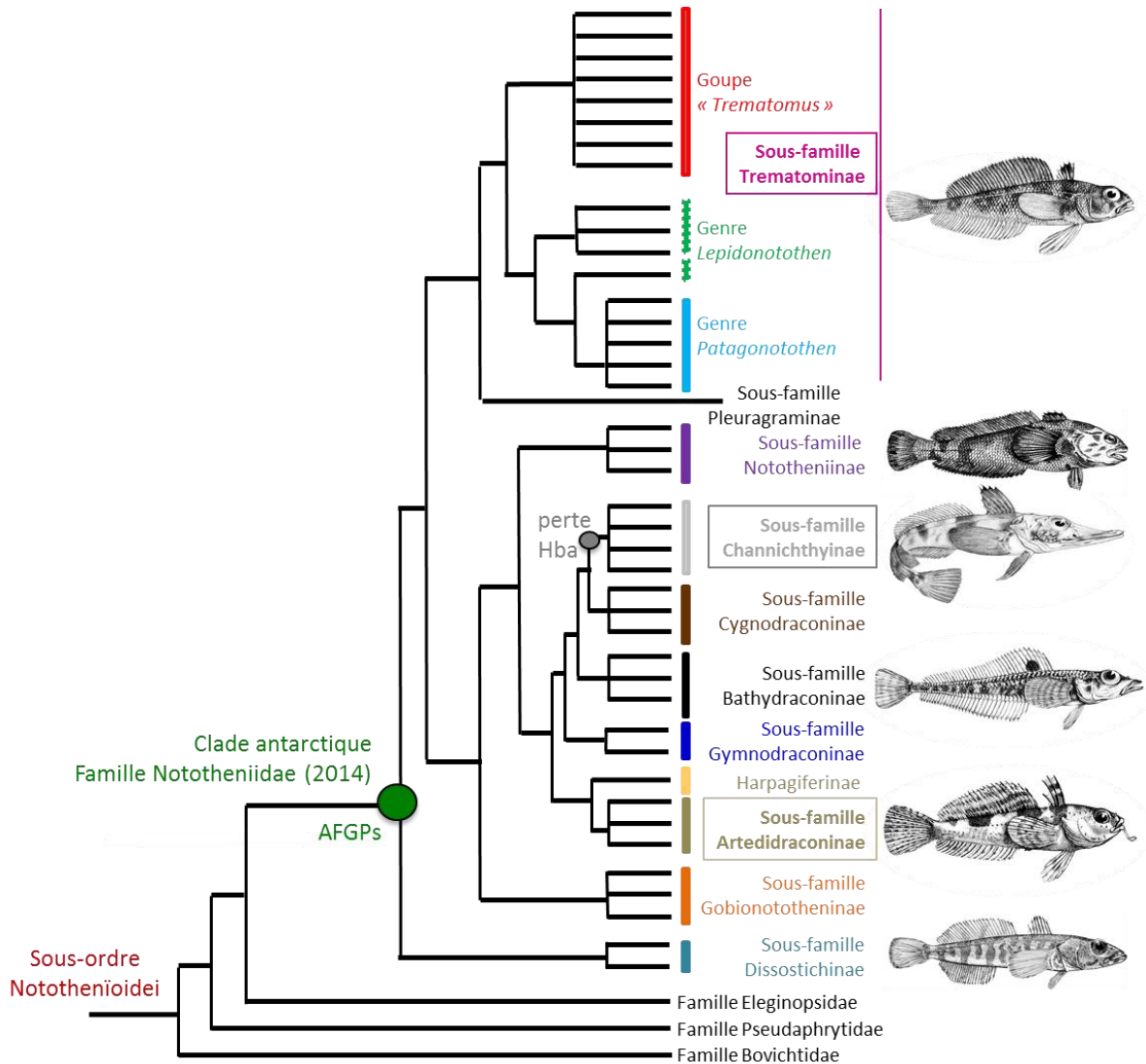


Les événements paléoclimatiques et paléogéographiques du Cénozoïque ont mené à l'isolement géographique et thermique et à un refroidissement consécutif de l'Océan Austral (Kennett 1977; Eastman and McCune 2000). Au sein de ce milieu « extrême », le succès des Nototheniidae est probablement dû à l'apparition d'innovations biochimiques et physiologiques peu communes. La plus importante est certainement l'acquisition d'une combinaison de glyco-protéines antigel (AGFPs pour « anti-freeze glycoproteins ») (De Vries 1971; De Vries 1980), caractéristique du clade antarctique (Figure I.1). D'autres caractères ont été sélectionnés au cours de l'évolution de ce clade dans ce nouvel environnement. L'absence de vessie natatoire, la réduction de leur minéralisation et densité osseuse (Balushkin 2000), associée aux modifications de leur métabolisme basal et lipidique (Eastman and McCune 2000; Logue et al. 2000) facilitent les migrations verticales dans la colonne d'eau et l'adaptation au fort gradient de profondeur. Ces caractères concernent la moitié des espèces de Nototheniidae (Eastman 2005) et ont été acquis indépendamment dans différents groupes (Eastman and Hubold 1999; Near et al. 2007). Ils ont pu contribuer au passage d'un mode de vie benthique à des modes de vie plus diversifiés (semipélagique, épibenthique, cryopélagique, ou pélagique). Enfin, la sténothermie des Nototheniidae à basses températures résulte en partie de l'assemblage particulier de leurs microtubules (Williams et al. 1985; Detrich et al. 2000), de l'optimisation du transport et du catabolisme protéique (Detrich and Amemiya 2010; Near et al. 2012), de la mise en place de cristallins cryostables prévenant la formation de cataracte (Kiss 2004) et de la perte de réponse au stress thermique de type « heat shock » (Hofmann 2000; Buckley 2004).

Lors des refroidissements, une grande partie des espèces tempérées qui peuplaient la région antarctique et chez qui ces caractères n'étaient pas présents ou n'ont pas été fixés se sont éteintes (Eastman 2005; Detrich et al. 2010). Ces extinctions ont libéré un nombre important de niches écologiques. De plus, durant les 25 derniers Ma, l'alternance de périodes glaciaires et interglaciaires a été à l'origine d'une dynamique de la calotte glaciaire (Anderson et al. 2002; Huybrechts 2002; Thatje et al. 2005; Smale et al. 2008). En période glaciaire, la calotte atteint des maxima d'extension, ce qui cause une fragmentation importante des habitats avec présence de micro-niches refuges. En période de réchauffement, elle se rétracte, et les habitats fragmentés sont de nouveau en contact, avec possibilité de recolonisation des niches vacantes (Kennett et al. 1985; Kennett and Barker 1990; Stott et al. 1990; Clarke and Johnston 1996). C'est dans ce contexte que les populations issues de la famille des Nototheniidae se sont diversifiées sur le plateau continental antarctique et les plateaux péri-insulaires (Eastman and McCune 2000; Lecointre et al. 2013). Ce milieu se caractérise aujourd'hui par une grande stabilité thermique (Clarke 1990; Kaus 2005), avec des eaux océaniques dont la température se situe entre -1,9 et -1,5°C (Anderson 2000).

Avec 115 espèces décrites (Duhamel et al. 2014; Ghigliotti et al. 2015), la divergence du clade antarctique a été datée soit vers la transition Oligocène-Miocène aux alentours de 23 Ma (23,9 Ma, Matschiner *et al.* (2011) ; 22,4 Ma, Near *et al.* (2012)), soit vers le milieu du Miocène il y a 15 Ma (14,7 Ma, Colombo *et al.* (2015)). Les spéciations par vagues se caractérisent par des diversifications en successions rapides (radiations) (Near 2004; Matschiner et al. 2011; Near et al. 2012; Colombo et al. 2015) dans des aires géographiques circonscrites, avec des groupes dominants en termes de biomasse (Eastman and McCune 2000; Eastman 2005; Lecointre et al. 2013) conduisant à un fort degré d'endémisme. Elles concernent des espèces apparentées formant un groupe monophylétique et présentant une grande diversité écologique et morphologique, ce qui s'applique bien au clade antarctique à antigels (Eastman and McCune 2000; Near 2004; Near et al. 2012; Lecointre et al. 2013; Colombo et al. 2015). Ces spéciations par vagues sont aussi appelées explosions de diversité, plus communément retrouvées sous le terme de « species flocks » (Martens 1997; Eastman and McCune 2000; Rutschmann et al. 2011; Lecointre et al. 2013).

Les relations de parentés au sein des Nototheniidae ont fait l'objet de nombreux débats. Sur la base de comparaisons de caractères morphologiques (Voskoboinikova 1993; Balushkin 2000), cinq familles avaient été définies (les Channichthyidae, les Artedidraconidae, les Harpagiferidae, les Bathydraconidae et les Nototheniidae). Cependant, l'absence de synapomorphie ne permettait pas de déterminer de manière fiable les relations de parentés entre ces clades, ni même leur monophylie (Sanchez et al. 2007). La remise en question de ces relations d'apparentement s'est accrue avec le développement des phylogénies moléculaires. La multiplication des analyses a permis d'aboutir à une congruence sur de nombreuses parties des topologies obtenues, confirmant la monophylie de certains groupes (les Channichthyidae, les Artedidraconidae, les Harpagiferidae) (Cheng and Detrich 2007; Rutschmann et al. 2011; Dettai et al. 2012; Lautrédou et al. 2012; Near et al. 2012; Colombo et al. 2015), mais a souligné la paraphylie d'autres (Nototheniidae, Bathydraconidae).



**Figure I.1 :** cladogramme présentant les relations de parenté dans le sous ordre des Notothenioidei (adapté de Dettai *et al.* (2012), synthétisant Near et Cheng (2008), Dettai *et al.* (2012), Near *et al.* (2012) et Colombo *et al.* (2015)). Une barre verticale pleine indique les groupes monophylétiques, tandis que les groupes paraphylétiques (genre *Lepidonotothen*) sont indiqués en pointillés. Un représentant par sous-famille est présenté à droite (de haut en bas : *Trematomus pennellii*, *Notothenia coriiceps*, *Chaenocephalus aceratus*, *Prionodraaco evansii*, *Artedidracono oriana*, *Dissostichus mawsoni*). Les tailles relatives entre spécimens ne sont pas respectées. L'acquisition des protéines antigels ou AFGPs (caractéristique du clade antarctique : la famille des Nototheniidae) est indiquée par un point vert et la perte d'hémoglobine (caractéristique de la famille des Channichthyinae ou « icefish à sang blanc ») par un point gris. Les trois radiations récentes (Artedidraconinae, Channichthyinae et Trematominae) sont encadrées. Le groupe monophylétique « *Trematomus* » regroupe les quatre genres : *Cryothenia*, *Indonotothenia*, *Pagothenia* et *Trematomus*.

Sur la base des congruences entre les topologies récemment obtenues, une nouvelle classification a été proposée en 2014 (Duhamel *et al.* 2014) et publiée dans l'Atlas Bibliographique du « Scientific Committee on Antarctic Research » (SCAR). Elle sera employée tout au long de cette étude (Figure I.1, Tableau I.1). Les Nototheniidae deviennent une seule et

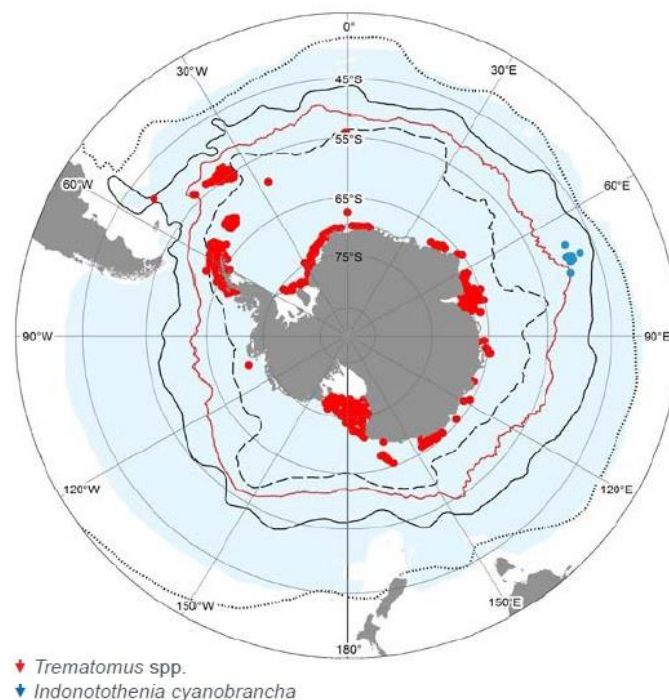
même famille (incluant les quatre autres) correspondant au clade antarctique. Cette famille comporte 36 genres, répartis dans 11 sous-familles (Tableau I.1).

| Sous-famille            | Genres compris dans la sous-famille  |
|-------------------------|--|
| DISSOSTICHINAE          | <i>Dissostichus, Aethotaxis, Gvodarus</i>  |
| GOBIONOTOTHENINAE       | <i>Gobionotothen</i>   |
| NOTOTHENIINAE           | <i>Notothenia, Paranotothenia</i>  |
| PLEURAGRAMMINAE         | <i>Pleuragramma</i>  |
| <b>TREMATOMINAE</b>     | <i>Lepidonotothen, Patagonothen,</i><br><i>Cryothenia, Indonotothenia, Trematomus, Pagothernia</i>   |
| <b>ARTEDIDRACONINAE</b> | <i>Artedidraco, Dolloidraco, Histiodraco, Pogonophryne</i>   |
| HARPAGIFERINAE          | <i>Harpagifer</i>  |
| GYMNODRACONINAE         | <i>Gymnodraco, Acanthodraco, Psilodraco</i>  |
| BATHYDRACONINAE         | <i>Bathyraco, Akarotaxis, Racovitzia, Prionodraco, Vomeridens</i>  |
| CYGNODRACONINAE         | <i>Cygnodraco, Parachaenichthys, Gerlachea</i>   |
| <b>CHANNICHTHYINAE</b>  | <i>Chaenocephalus, Chaenodraco, Champsocephalus,</i><br><i>Channichthys, Chionobathyscus, Chionodraco, Cryodraco,</i><br><i>Dacodraco, Neopagetopsis, Pagetopsis, Pseudochaenichthys</i> |

**Tableau I.1 : récapitulatif taxonomique de la famille des Nototheniidae (clade antarctique) avec les différents genres présents dans chaque sous-famille, adapté de Duhamel *et al.* (2014).** En gras, les trois sous-familles incluant des radiations adaptatives (Eastman and McCune 2000; Near 2004; Matschiner *et al.* 2011; Rutschmann *et al.* 2011; Near *et al.* 2012).

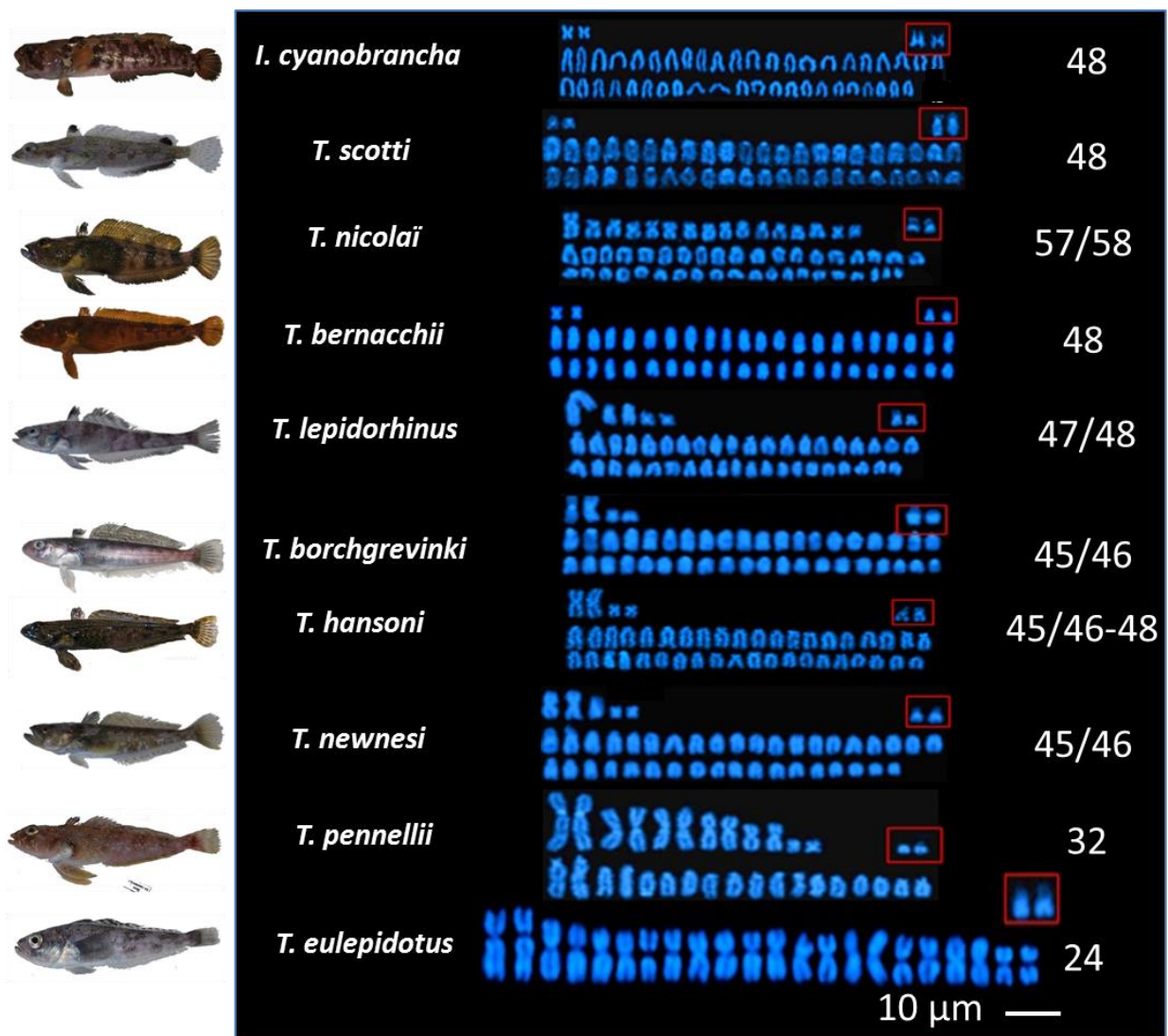
Trois des onze sous-familles contiennent des radiations récentes (Eastman and McCune 2000; Near 2004; Matschiner *et al.* 2011; Rutschmann *et al.* 2011; Near *et al.* 2012) : les Channichthyinae (-6,3 Ma, Near *et al.* (2012), -3,49, Ma Colombo *et al.* (2015)), les Artedidraconinae (-3 Ma, Near *et al.* (2012), -1,24 Ma, Colombo *et al.* (2015)) et les Trematominae (11 Ma, Near *et al.* (2012), 6 Ma, Colombo *et al.* (2015)) (Figure I.1, Tableau I.1). Si les Channichthyinae (aussi connus sous le nom de poissons des glaces ou « icefish ») font l'objet de nombreux travaux de recherche en raison de nombreuses particularités dont la perte de leur hémoglobine (unique chez les vertébrés), les deux autres sous-familles sont beaucoup moins étudiées. Nous nous sommes intéressés à l'une de ces deux autres radiations : celle présente au sein des Trematominae.

Les Trematominae sont composés de six genres (Tableau I.1). Toutes les études phylogénétiques récentes regroupent d'une part les deux genres subantarctiques *Lepidonotothen* (cinq espèces) et *Patagonotothen* (quinze espèces), et d'autre part les quatre genres *Cryothernia* (deux espèces), *Indonotothenia* (une espèce), *Trematomus* (onze espèces) et *Pagothenia* (deux espèces) (Near 2004; Sanchez et al. 2007; Near and Cheng 2008; Dettai et al. 2012; Near et al. 2012; Colombo et al. 2015). C'est ce dernier regroupement, identifié comme une radiation, auquel nous avons choisi de nous intéresser. Nous le désignerons par la suite sous le nom de groupe « *Trematomus* ». Ce groupe est endémique du plateau continental antarctique et des plateaux péri-insulaires (Figure I.2). La monophylie de ce groupe est établie, mais les relations phylogénétiques au sein du groupe restent à éclaircir (Ritchie et al. 1996; Sanchez et al. 2007; Kuhn and Near 2009; Lautrédou et al. 2012). La diversification des « *Trematomus* » regroupe tous les critères d'une radiation dite adaptative (Matschiner et al. 2011; Colombo et al. 2015) à savoir une divergence récente depuis l'ancêtre commun, puis des diversifications écologiques rapides avec des adaptations phénotypiques associées (corrélation « phénotype-environnement ») en lien avec la diversification des ressources et des habitats disponibles (augmentation de la valeur sélective) (Schluter 2000). Les « *Trematomus* » constituent la diversification la plus spectaculaire parmi les Nototheniidae.



**Figure I.2 :** Localités de capture de spécimens du groupe « *Trematomus* » dans l'Océan Austral (repris de Duhamel et al. (2014)). Le courant circumpolaire est représenté en noir et le front polaire par un trait rouge. Les spécimens appartenant aux espèces du genre *Trematomus* (*sensus stricto*) sont indiqués par des points rouges. Les spécimens appartenant au genre *Indonotothenia* sont indiqués par des points bleus (plateaux Kerguelen-Heard).

Cette diversité est écologique, avec la diversification des régimes alimentaires (espèces planctonivores, piscivores), des milieux de vie (espèces benthiques, épibenthiques, pélagiques, semi-pélagiques, cryopélagiques) (Montgomery et al. 1993; Eastman 2005; Rutschmann et al. 2011), et des répartitions bathymétriques (Gon and Heemstra 1990; Eastman 2017). Cette diversité est aussi morphologique et spécifique. Enfin, les « *Trematomus* » se singularisent par la variabilité chromosomique la plus importante au sein des Nototheniidae (Pisano and Ozouf-Costaz 2000; Pisano et al. 2003) et rare chez les autres téléostéens Eupercaria (« Perciformes ») (Molina 2006; Molina and de Freitas Bacurau 2006; Betancur et al. 2017). Le nombre diploïde de chromosomes varie entre 24 et 58 (Figure I.3). Cette diversité comprend non seulement la taille, mais aussi la morphologie des chromosomes (Ozouf-Costaz et al. 1991; Pisano and Ozouf-Costaz 2000; Ghigliotti et al. 2015). Certaines espèces possèdent de multiples petits chromosomes acrocentriques (*I. cyanobrancha*, *T. hansonii*, *T. borchgrevinkii*), tandis que d'autres ont des chromosomes en moins grand nombre ; à la fois acrocentriques, mais aussi sub-méta et/ou métacentriques, et plus grands en taille (*T. pennellii*, *T. eulepidotus*) (Figure I.3 et II.1). Plusieurs des espèces du groupe « *Trematomus* », (*C. amphitreta*, *C. peninsulæ*, *T. tokarevi*, *P. brachysoma*) n'ont jamais fait l'objet d'un caryotypage à ce jour. Si la plupart des caryotypes sont stables au sein d'une espèce, une variabilité interindividuelle a parfois été observée, notamment en fonction de la localité géographique des spécimens collectés (Morescalchi et al. 1992; Ghigliotti et al. 2015). C'est le cas pour *T. hansonii* dont le caryotype est de  $2n = 46$  en Terre Adélie, 45-46 en mer de Ross, 48 dans le secteur dit atlantique de l'Océan Austral (Pisano and Ozouf-Costaz 2000), ou encore de *T. loennbergii*, avec  $2n = 26-27-28-29-30-31-33$  (Pisano and Ozouf-Costaz 2000; Ghigliotti et al. 2015). De plus, certaines espèces ou populations de « *Trematomus* » possèdent des chromosomes sexuels multiples différenciés (mâle hétérogamétique, Ozouf *et al.* (1991), Morescalchi *et al.* (1992)). Ce dimorphisme chromosomique sexuel, de type  $X_1X_1-X_2X_2/X_1YX_2$  (femelle/mâle) est présent chez *T. nicolai*, *T. newnesi*, *T. borchgrevinkii*, *T. hansonii* et *T. lepidorhinus* (Ozouf-Costaz et al. 1991; Morescalchi et al. 1992; Ghigliotti et al. 2015). Cette variabilité caryotypique intra et interspécifique suggère de nombreux remaniements ayant accompagné les spéciations au sein du groupe (Morescalchi et al. 1992; Ozouf-Costaz et al. 1997; Pisano and Ozouf-Costaz 2000; Ozouf-Costaz et al. 2004; Tomaszewicz et al. 2011).



**Figure 1.3:** diversité chromosomique observée chez les « *Trematomus* ». Classement des chromosomes colorés au DAPI en taille décroissante. Les proportions relatives des caryotypes entre les espèces ont été respectées (échelle de 10 µm). Caryotypes mâles présentés pour *T. nicolai* (2n = 57) et *T. lepidorhinus* (2n = 47). La paire chromosomique entourée en rouge dans chaque caryotype correspond à la paire qui porte l'ADN correspondant aux ARN ribosomiques 18 et 28S. Les espèces représentées sur cette figure correspondent aux « *Trematomus* » de notre échantillonnage.

De tels remaniements chromosomiques impliquent des cassures doubles brins de l'ADN suivies de réparations. Dans les génomes eucaryotes, les éléments transposables (ETs) sont connus pour induire des cassures double brins et favoriser la recombinaison ectopique (recombinaisons homologues à des sites non homologues) (Devos et al. 2002; Bonnivard and Higuët 2009). Les ETs sont des constituants majeurs des génomes eucaryotes (Eickbush and Malik 2002; Kidwell 2002; Volf 2005). Par leur activité, ils sont impliqués dans la structuration et la plasticité des génomes au cours de l'évolution (Biéumont and Vieira 2005; Bonnivard and Higuët 2009; Oliver and Greene 2009). Ils génèrent ainsi de la variabilité génétique (modification ou interruption d'expression de gènes par exemple) sur lesquelles différentes forces évolutives peuvent agir

(van de Lagemaat et al. 2003; Sela et al. 2010). Ils sont susceptibles d'intervenir dans des événements de remaniements chromosomiques (Lönnig and Saedler 2002; Raskina et al. 2008; Valente et al. 2011; Chénais et al. 2012), notamment, par le biais de recombinaisons ectopiques. Certains éléments (*élément P*, *DIRS*) sont connus pour s'insérer préférentiellement dans des copies déjà présentes (« homing ») (Capello et al. 1985; Taillebourg and Dura 1999; Eickbush and Jamburuthugoda 2008). Ces accumulations par insertions répétées dans les mêmes régions chromosomiques sont appelées points chauds d'insertion (Caceres 2001; Lönnig and Saedler 2002). Ils impliquent nécessairement des cassures d'ADN double brin suivies de réparations associées pouvant mener aux remaniements (Bonnivard and Higuete 2009). Une fois fixés, ces réarrangements chromosomiques constituent une première étape vers un isolement reproductif complet (Ozouf-Costaz et al. 1997). Leur rôle en tant que possibles médiateurs des réarrangements chromosomiques en lien avec la spéciation est souvent évoqué (Navarro and Barton 2003; Raskina et al. 2008; Kraaijeveld 2010; Rebollo et al. 2010), mais reste toutefois débattu. En temps normal, l'expression des ETs est finement régulée dans les génomes (Bourque 2009). Cependant, dans un contexte de variations environnementales, incluant les stress thermiques, leur mobilisation peut être induite suite à des dérégulations épigénétiques (levée des méthylation de l'ADN, et déméthylation et acétylation des histones) (Mansour 2007; Slotkin and Martienssen 2007; Rebollo et al. 2010). Des transpositions massives, augmentant potentiellement la fréquence des remaniements et donc d'isolement, conduisant à la spéciation, ont ainsi pu être détectées chez la drosophile (Lim and Simmons 1994; Aulard et al. 2004; Ladevèze et al. 2012), le maïs (Steward et al. 2002), ou encore chez les primates (Pace and Feschotte 2007) ou les Salmonidae (Symonová et al. 2013). Cependant, des analyses sur des groupes entiers intégrant des données moléculaires, chromosomiques et environnementales seraient requises pour confirmer ces hypothèses.

Au cours de ma thèse, je me suis intéressée à la mise en place d'une telle diversité chromosomique chez les « *Trematomus* » en reconstituant leur histoire évolutive. J'ai donc cherché à caractériser ces remaniements, puis à les replacer sur une phylogénie résolue du groupe « *Trematomus* ». Enfin, j'ai recherché si les éléments transposables avaient pu jouer un rôle dans ces réarrangements. Ces travaux se décomposent en trois grands chapitres : le premier chapitre traite de la caractérisation des remaniements chromosomiques dans le groupe « *Trematomus* » ; le deuxième chapitre s'intéresse à la phylogénie des « *Trematomus* » ; et le troisième chapitre cherche à évaluer le rôle des éléments transposables comme facilitateurs de ces remaniements au cours de l'évolution.



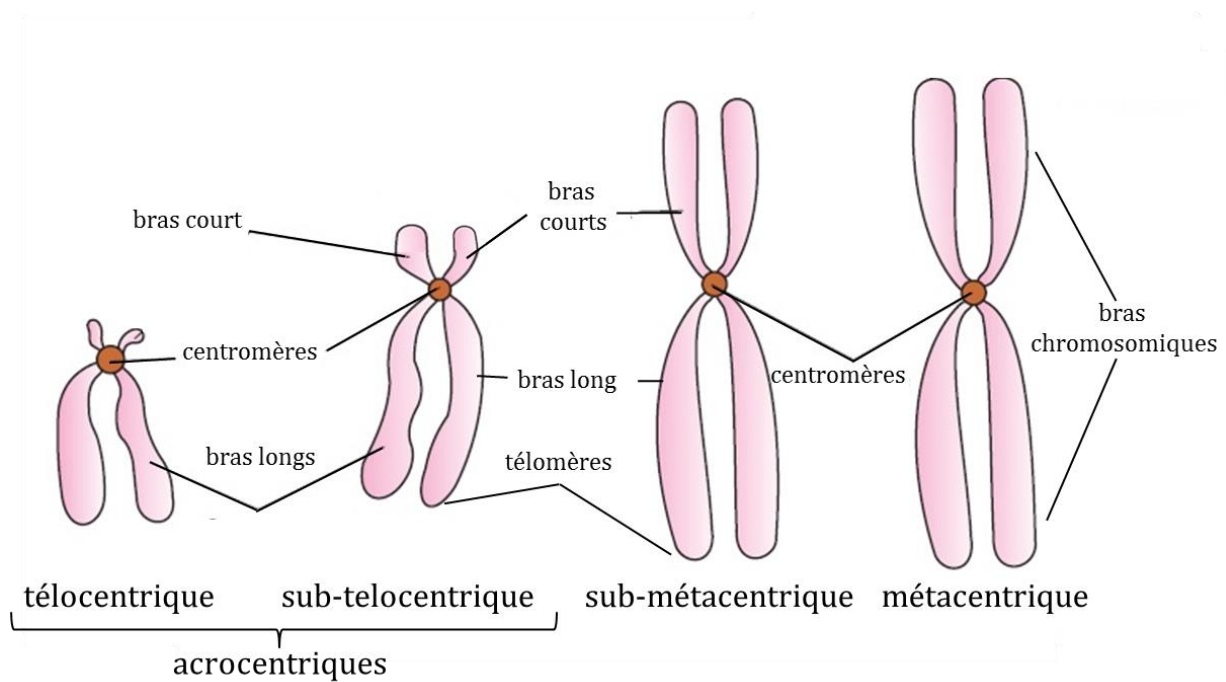


## II) Chapitre 1 : caractérisation des remaniements chromosomiques au sein du groupe « *Trematomus* »

### II.1 : Nature des remaniements chromosomiques

#### II.1.A : Histoire des diversifications chromosomiques chez les Nototheniidae

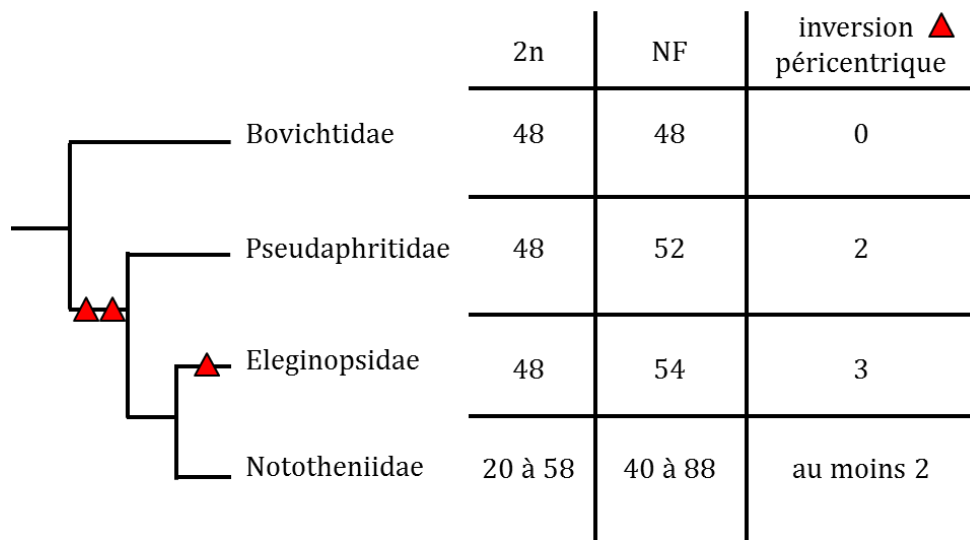
Par reconstruction du caryotype ancestral des vertébrés ( $2n = 20$  Nakatani *et al.* (2007) ou  $2n = 24$ , Jaillon *et al.* (2004)), il a été inféré chez l'ancêtre commun des téléostéens un « proto-caryotype » de  $2n = 48$  chromosomes. Ce caryotype résulte d'une duplication de l'ensemble du génome (TGD pour « teleost genome duplication ») (Ohno 1974; Hartley 1987; Jaillon *et al.* 2004; Naruse *et al.* 2004; Woods 2005; Kohn *et al.* 2006; Nakatani *et al.* 2007) datée entre 316 et 226 Ma (Amores 1998; Taylor 2003; Jaillon *et al.* 2004; Hurley *et al.* 2005; Santini *et al.* 2009) chez l'ancêtre commun des téléostéens, après leur divergence d'avec les Holostéens (Hoegg *et al.* 2004; Amores *et al.* 2011). Chacun des chromosomes de ce « proto-caryotype » est acrocentrique, c'est-à-dire porteur d'un seul bras chromosomique (Figure II.1). Ce caryotype correspond donc à un nombre fondamental (nombre de bras) de 48 (NF = 48). Une grande stabilité caryotypique est observée depuis la TGD (Amores *et al.* 2017). Les téléostéens dont les génomes ont été séquencés ont généralement conservé des nombres de chromosomes proches de  $2n = 48$  : le poisson-zèbre (*Danio rerio*) ( $2n = 50$ ) (Sola and Gornung 2001), le medaka (*Oryzias latipes*) et le platyfish (*Xiphophorus maculatus*) ( $2n = 48$ ) (Ocalewicz 2005; Kasahara *et al.* 2007), la morue (*Gadus morhua*) ( $2n = 46$ ) (Ghigliotti *et al.* 2012), le fugu (*Takifugu rubripes*) ( $2n = 44$ ) (Miyaki *et al.* 1995), l'épinoche (*Gasterosteus aculeatus*), ou encore le « poisson-coffre » ou tétraodon (*Tetraodon nigroviridis*) avec  $2n = 42$  chromosomes (Grutzner *et al.* 1999; Urton *et al.* 2011). Ces fluctuations autour de la valeur 48 sont imputables à certains remaniements structuraux (de type fusions ou translocations chromosomiques) ayant abouti par exemple à des paires chromosomiques submétacentriques ou métacentriques (Miyaki *et al.* 1995; Ghigliotti *et al.* 2012). Ce nombre diploïde de 48 chromosomes est rencontré chez la majorité des autres téléostéens ayant fait l'objet d'un caryotypage à ce jour. Cette constance dans la composition des caryotypes s'applique en particulier aux lignées proches des Nototheniidae, appartenant à l'ordre des Eupercaria (« Perciformes ») (Gold *et al.* 1980; Morescalchi *et al.* 1992; Molina 2006; Betancur *et al.* 2017).



**Figure II.1 :** classification des différents types de chromosomes en fonction de la position relative de leur centromère (adapté de Levan *et al.* (1964)). Par convention, les chromosomes sont orientés avec le bras court au dessus du centromère, et le bras long positionné en-dessous. Les chromosomes télocentriques ont un centromère terminal situé à leur extrémité ne laissant dépasser qu'un bras court réduit *a minima*. Pour les chromosomes subtélocentriques, le bras court reste réduit, mais il est toutefois plus développé. Pour les chromosomes submétacentriques, ce bras court est réellement visible et clairement identifiable. Les chromosomes métacentriques ont un centromère médian et deux bras de taille relativement équivalente.

Ce caryotype de  $2n = 48$  chromosomes acrocentriques a été décrit comme plésiomorphe pour le sous-ordre des Notothenioidei (Pisano *et al.* 2003; Mazzei *et al.* 2006; Mazzei *et al.* 2008). Les Bovichtidae sont les premiers à se séparer des autres familles chez les Notothenioidei (Bargelloni *et al.* 2000; Cheng 2003; Near 2004). Ils présentent l'ensemble de ces caractéristiques caryotypiques : 48 chromosomes acrocentriques télocentriques, de tailles sensiblement égales mais très légèrement décroissantes,  $NF = 48$  (Pisano *et al.* 1995; Mazzei *et al.* 2006). Les deux autres familles : Pseudaphritidae et Eleginopsidae, présentent également des caryotypes de  $2n = 48$  chromosomes, mais avec un  $NF$  supérieur à 48 (respectivement 52 et 54) (Pisano *et al.* 1995; Mazzei *et al.* 2008). Cette augmentation du nombre fondamental, sans changement de ploïdie, est imputable à des inversions péricentriques dans deux (Pisano *et al.* 1995) ou trois (Mazzei *et al.* 2008) paires chromosomiques acrocentriques produisant des paires métacentriques de taille équivalente (Figures I.3 et II.1). Elles sont considérées comme des états dérivés (Pisano *et al.* 2003; Mazzei *et al.* 2008) acquis après la divergence des Bovichtidae d'avec les autres familles de Notothenioidei (Pseudaphritidae, Eleginopsidae, Nototheniidae). Deux de ces inversions sont partagées par l'ensemble des Nototheniidae (Ozouf-Costaz *et al.* 1991; Pisano *et al.* 2000; Ghigliotti *et al.* 2015) (Figures I.3 et II.2). Une des deux paires chromosomiques métacentriques résultant de ces inversions porte l'ADN correspondant

aux ARN ribosomiques 5,8, 18 et 28S, l'autre constitue très probablement la plus petite paire chromosomique métacentrique observée chez la quasi totalité des espèces de Nototheniidae (Figure I.3) (Ozouf-Costaz et al. 1991; Pisano et al. 1995; Ozouf-Costaz et al. 1997).



**Figure II.2:** cladogramme représentant les relations phylogénétiques entre les quatre familles de Notothenioidei (adapté de Near *et al.* (2004) et Mazzei *et al.* (2006)). Leurs caryotypes respectifs sont repositionnés sur la topologie. Les Bovichtidae ( $2n = 48$ ,  $NF = 48$ ) sont les premiers à émerger, suivis des Pseudaphritidae ( $2n = 48$ ,  $NF = 52$ ). Suivent les Eleginopsidae ( $2n = 48$ ,  $NF = 54$ ), groupes frères plus proches parents des Nototheniidae ( $2n = 20$  à  $58$ ,  $NF = 40$  à  $88$ ). La mise en place des deux inversions péricentriques chez deux paires chromosomiques acrocentriques (devenues métacentriques ou submétacentriques tout en conservant leur taille) partagées par l'ensemble des Nototheniidae a donc été datée au moment de la divergence des Bovichtidae d'avec les autres Notothenioidei (Pseudaphritidae, Eleginopsidae, Nototheniidae). Une autre inversion péricentrique supplémentaire caractérise les caryotypes des Eleginopsidae.

Cependant, chez certaines des espèces de Nototheniidae, le nombre fondamental peut varier entre 40 et 88 (Pisano and Ozouf-Costaz 2000), suggérant, des modifications de l'arrangement des bras de type fusions, fissions, inversions, sans polyploïdisation, à l'origine de la grande variabilité des caryotypes observés. Une tendance à la réduction du nombre total de chromosomes (appelée aussi « symétrisation ») a été évoquée à plusieurs reprises dans la littérature pour cette famille (Morescalchi et al. 1992; Morescalchi et al. 1996; Pisano et al. 2000; Pisano and Ozouf-Costaz 2000; Pisano et al. 2003). Elle correspond à des caryotypes qui sont constitués, en plus de petites paires chromosomiques acrocentriques, de plus grandes paires métacentriques (*T. eulepidotus* ( $2n = 24$ ), *T. pennellii* ( $2n = 32$ ), *N. coriiceps* ( $2n = 22$ )), ou bien acrocentriques (*T. pennellii* ( $2n = 32$ )). Dans un seul cas, chez *T. nicolai* ( $2n = 57-58$ ), on trouve au contraire des paires chromosomiques acrocentriques nettement plus petites que les 48 paires acrocentriques télocentriques observées chez les *Bovichtus* (Figure I.3). Les caryotypes des « *Trematomus* » peuvent être partitionnés en trois catégories distinctes (Morescalchi et al. 1992; Pisano and Ozouf-Costaz 2000; Ghigliotti et al. 2015). La première regroupe les

caryotypes à  $2n = 48$  chromosomes (*I. cyanobrancha*, *T. bernacchii*, *T. scotti*). Ils sont constitués de vingt-deux paires chromosomiques acrocentriques, ainsi qu'une paire submétacentrique (portant les gènes ribosomiques) et une paire métacentrique, la plus petite paire en taille. Ces deux dernières paires sont partagées par l'ensemble des « *Trematomus* » (Morescalchi et al. 1992). Le deuxième groupe concerne les caryotypes à  $2n < 48$  chromosomes. Ils sont les plus fréquents dans le genre (*T. borchgrevinki*, *T. eulepidotus*, *T. hansonii*, *T. lepidorhinus*, *T. newnesi*, *T. pennellii*) et contiennent probablement des chromosomes issus de simples ou doubles fusions (Morescalchi et al. 1992; Pisano and Ozouf-Costaz 2000) par rapport à l'état ancestral. La dernière catégorie est celle des caryotypes à  $2n > 48$ . Elle concerne uniquement une espèce (*T. nicolai*) dont cinq paires chromosomiques sont probablement issues de fissions chromosomiques (Morescalchi et al. 1992) par rapport à l'état ancestral.

Afin de caractériser et de retracer de manière fiable l'histoire évolutive des remaniements chromosomiques ayant accompagné la diversification des « *Trematomus* » d'après ces hypothèses, il est nécessaire de 1- tester l'absence de polyploïdisation précédant les remaniements structuraux et 2- établir une correspondance entre les chromosomes des différentes espèces. Il s'agit de l'établissement des homologues chromosomiques interspécifiques (HCI). Pour cela, nous devons définir une unité pour pouvoir décrire ces chromosomes et suivre leurs homologues sans *a priori* et le plus objectivement possible. Le nombre fondamental ne paraît pas le paramètre le mieux adapté pour cette unité chromosomique car il n'existe pas de consensus au sein de la communauté scientifique pour le définir. Un chromosome acrocentrique subtélocentrique pourra être considéré tantôt porteur d'un, tantôt de deux bras chromosomiques. Par simplification, nous confondrons par la suite les chromosomes télocentriques et subtélocentriques sous le terme de chromosomes acrocentriques (Figure II.1). Nous préférons donc plutôt que le terme de « nombre fondamental » utiliser le terme « nombre d'unités structurales ». Pour définir ce que nous entendons par unité structurale chromosomique, nous avons choisi le caryotype d'*I. cyanobrancha* ( $2n = 48$ , NF = 52 ou 54, Doussau De Bazignan et Ozouf-Costaz, 1985) comme référence (Figure I.3). C'est l'espèce présentant une position basale (avec *T. scotti*) chez les « *Trematomus* » (Bargelloni et al. 2000; Near 2004; Dettai et al. 2012). Nous considérons que le nombre d'unités structurales constituant ce caryotype est de 52. Les 22 paires acrocentriques (subtélocentriques) correspondent ainsi à une unité structurale par chromosome, et les deux paires métacentriques ou submétacentriques issues d'inversions péricentriques correspondent à deux unités structurales chromosomiques (Figures I.3 et II.1). Ce caryotype pourrait ressembler à l'état plésiomorphe, ayant subi le moins de modifications depuis l'ancêtre commun des « *Trematomus* ».

## **II.1.B : Conservation de la taille des génomes chez les Nototheniidae**

### **II.1.B.1 : Introduction**

L'inférence de caryotype ancestral et la reconstitution des événements de remaniements génomiques et/ou chromosomiques constituent des étapes clés dans la compréhension de l'émergence des génomes et des caryotypes actuels (Hartley 1987; Jaillon et al. 2004). La variation des tailles de génomes au sein d'une population ou d'une communauté permet de mettre en évidence ces phénomènes de plasticité génomique interindividuelle ou interspécifique au cours de l'évolution (Lipnerova et al. 2013). La taille des génomes rend en effet directement compte du contenu d'ADN dans un noyau haploïde au travers de la valeur C (exprimée usuellement en paires de bases ou en picogrammes) (Greilhuber 2005). Dans le cadre de notre hypothèse, les remaniements intra (inversions, translocations) ou interchromosomiques (translocations, fusions, fissions) ayant accompagné les diversifications rapides des Nototheniidae sont des réarrangements structuraux. Cependant, une hypothèse alternative prenant en compte d'autres types de réarrangements : duplication chromosomique suivie de réduction drastique (pour passer de  $2n = 48$  à  $2n = 24$  par exemple) est également envisageable.

Les variations structurales (« structural variant ») décrites dans la littérature à l'échelle intraindividuelle (Feuk et al. 2006) ou interspécifique (Hurley et al. 2005) impliquent une faible variation des tailles de génomes. Il peut s'agir de variations du nombre de copies (1 à 50 kilobases), de duplications ou de réductions segmentales (1 à 100 kilobases), ou encore d'inversions, de translocations ou de disomie segmentale uniparentale (de l'ordre de plusieurs mégabases). Ces variations peuvent notamment être liées à des différences de nombre de copies en éléments transposables, et en particulier en rétrotransposons (Sanmiguel 1998; Kidwell 2002; Chénais et al. 2012), ou encore à des régions dupliquées responsables d'anomalies génétiques (Emanuel and Shaikh 2001; Bailey 2002).

Pour les cas de duplications de génomes entiers, de plus grandes fluctuations de tailles de génomes sont théoriquement attendues (de facteur deux). Des duplications génomiques ont ainsi pu être identifiées par reconstructions de paralogies, orthologies, et synténies moléculaires entre les deux génomes de levures *S. cerevisiae* et *K. waltii* (Kellis et al. 2004), entre les génomes de mammifères et du tétraodon (Jaillon et al. 2004), ou encore entre les génomes de l'arabette *Arabidopsis thaliana*, du coton *Gossypium* ou du riz *Oriza* (Bowers et al. 2003). Cependant, chez la levure (Kellis et al. 2004; Scannell et al. 2007), les téléostéens (Jaillon et al. 2004), comme chez les dicotylédones (Bowers et al. 2003), les études ont montré que ces duplications chromosomiques (ou « polyploïdisations ») sont souvent suivies de retours à l'état diploïde

(« diploïdisation ») par des délétions rapides et extensives de la plupart des paralogues (Inoue et al. 2015). Seules quelques copies dupliquées (1 078 chez le *Tetraodon* et 995 chez le *Takifugu*) sont conservées et sont sujettes à des néo ou sous-fonctionnalisations (Jaillon et al. 2004; Kellis et al. 2004; Hurley et al. 2005). Cela complique considérablement l'établissement des correspondances entre les segments homologues chez différentes espèces. De plus, la reconstruction du « proto-caryotype vertébré » présent chez l'ancêtre commun mammifères-téléostéens était d'autant plus compliquée à établir qu'une correspondance stricte 1:1 entre chromosomes dupliqués était loin d'être strictement retrouvée. De nombreux réarrangements structuraux à la fois intra et interchromosomiques ont en effet accompagné ces événements de pertes massives. Ils impliquent des événements de types fusions, fissions, inversions, translocations intra ou interchromosomiques (Dobigny et al. 2003; Pinton et al. 2003; Ross et al. 2009; Amores et al. 2017). Du fait de ces remaniements extensifs, et malgré la perte d'un certain nombre de copies de gènes identifiés actuellement, le contenu total d'ADN génomique reste toutefois bien plus élevé qu'avant duplication génomique.

Afin de mieux comprendre la nature des remaniements observés chez les *Nototheniidae*, et en particulier chez les espèces du groupe « *Trematomus* », nous avons donc besoin de connaître les valeurs C de leur génome. La comparaison de ces valeurs avec les caryotypes actuels et le caryotype ancestral inféré à  $2n = 48$  chromosomes acrocentriques, constitue une première étape dans l'identification d'un scénario évolutif de ces réarrangements complexes. Contrairement à la divergence des téléostéens au sein des *Osteichthyens* datée à 350 Ma (Jaillon et al. 2004; Naruse et al. 2004; Woods 2005; Kohn et al. 2006) ou encore celle des monocotylédones et dicotylédones datée à 140-150 Ma (Chaw et al. 2004), la divergence des *Nototheniidae* est très récente (entre 22 et 13 Ma, Near *et al.* (2012), Colombo *et al.* (2015)). L'échelle de diversification des *Trematomus* est encore plus courte puisque la radiation a été datée entre 4 et 7 Ma (Near et al. 2012; Colombo et al. 2015). Des traces de cette duplication et réduction génomique putative chez ce groupe pourraient être détectées d'autant plus facilement du fait de la divergence récente. Elles se manifesteraient par une différence significative des valeurs C mesurées entre les espèces. Si une telle différence n'est pas établie, cela soutiendrait plutôt notre hypothèse de départ : une évolution des génomes à nombre d'unités structurales et à taille de génomes constantes (Lipnerova et al. 2013).

Les valeurs C peuvent être déterminées expérimentalement par coloration de Feulgen (Hardie and Hebert 2003; Hardie and Hebert 2004) ou encore par cytométrie en flux (Detrich et al. 2010; Detrich and Amemiya 2010), et aussi *in silico* sur la base de données génomiques (Volf 2005). Aucun génome entier suffisamment assemblé n'est disponible dans la littérature à l'heure actuelle chez les téléostéens antarctiques. Cependant, certaines valeurs C de *Nototheniidae* ont

déjà été établies expérimentalement par coloration de Feulgen et mesure de densité optique (Hardie and Hebert 2003; Hardie and Hebert 2004). Pour la plupart des données publiées pour ce groupe, elles ont été mesurées par cytométrie en flux (Detrich et al. 2010; Detrich and Amemiya 2010). Nous avons élargi cet ensemble de données existant en nous focalisant surtout sur les espèces du groupe « *Trematomus* », en incluant toutefois un représentant des Nototheniinae (*N. coriiceps*), des Channichthyinae (*C. hamatus*) et des Gymnodraconinae (*G. acuticeps*). Les mesures de tailles de génomes ont été établies par cytométrie en flux, relativement à une référence bien caractérisée : la valeur C de la truite arc-en-ciel (*Oncorhynchus mykiss*).

### ***1.1.B.2 : Matériels et Méthodes***

#### Matériels

Les échantillons de sang hépariné ont été prélevés et préparés pour un total de 52 spécimens de douze espèces, dont huit espèces du groupe « *Trematomus* » (*T. bernacchii*, *T. borchgrevinki*, *T. eulepidotus*, *T. hansonii*, *T. loennbergii*, *T. newnesi*, *T. nicolaï*, *T. pennellii*), une espèce de *Notothenia* (*N. coriiceps*), une espèce de *Dissostichus* (*D. mawsoni*), une espèce de *Gymnodraco* (*G. acuticeps*) et une espèce de *Chionodraco* (*C. hamatus*) lors des campagnes REVOLTA (Radiation EVOLutives en Terre Adélie, 2010-2014). Ils proviennent tous de la même localité géographique (Terre Adélie). Pour chacun des spécimens, les échantillons de sang sont référencés en annexe (annexe 1).

Nous avons suivi les mêmes protocoles pour les érythrocytes de truite arc-en-ciel (*Oncorhynchus mykiss*), récoltés à l'Institut National de la Recherche Agronomique (INRA), Jouy-en-Josas (souche "synthétique INRA", année 2015) avec l'aimable aide de Dimitri Rigau. Ces échantillons nous serviront de référence ou standard pour calibrer nos mesures de valeurs C.

#### Méthodes

##### Préparation des échantillons sanguins

Les poissons ont été anesthésiés à l'aide de MS 222 avant les prélèvements sanguins (minimum de 0,5 ml) réalisés par ponction sur la veine caudale à l'aide de seringues héparinées. Pour tous les Nototheniidae exceptés les *Chionodraco hamatus*, les échantillons ont ensuite été dilués au dixième dans du tampon Phosphate Buffered saline (PBS). Un minimum de cinq à dix aliquots de 50 µl a été distribué goutte par goutte dans des tubes Eppendorf contenant 1 ml d'éthanol à 70 % glacé avant conservation à -20°C. Les cellules de sang de truite arc-en-ciel (*Oncorhynchus*



*mykiss*) ont été fixées en suivant les mêmes protocoles que celles récoltées sur les Nototheniidae. Pour *C. hamatus*, dépourvu d'érythrocytes, les leucocytes prélevés ont d'abord été concentrés par centrifugation 10 min à 2 000 rpm avant d'être resuspendus dans 1ml de PBS. Un minimum de cinq à dix aliquots de 100 µl a été distribué goutte à goutte dans des tubes Eppendorf contenant 1 ml d'éthanol à 70 % glacé avant conservation à -20°C.

#### Mesure de la taille des génomes par cytométrie de flux

Les protocoles de cytométrie de flux ont été adaptés des procédures décrites par Detrich *et al.* (2010). 500 µl de chaque échantillon ont été centrifugés (microcentrifugeuse Eppendorf 5430, 1500 rpm, 4°C, 5 min) à deux reprises. La première centrifugation a eu lieu dans un tampon de 0,01 M de PBS, la deuxième dans ce même tampon supplémenté en RNASE (10 µg/ml, Miltenyi Biotec) et en Iodure de Propidium (IP 50 µg/ml, Sigma Aldrich). Les mesures ont été réalisées au moyen d'un cytomètre de flux MACSQuant 10. Des mesures effectuées sur des cellules sans IP incorporé dans les noyaux ont permis de déterminer un seuil d'autofluorescence résiduel et de calibrer le mode d'acquisition. Avec une taille de génome estimée à 2,695 pg dans la littérature (génome size database, Volff (2005)), la truite arc-en-ciel nous a servi de référence (ou standard) pour calibrer nos mesures. Les cellules de truite arc-en-ciel ont donc fait l'objet de mesures en trois réplicats réalisées au début de chaque nouvelle plaque. Un même échantillon de *N. coriiceps* (Ich 1096) et de *T. bernacchii* (Ich 1100) a également été mesuré à chaque reprise. Cela nous a permis d'éviter tout biais de mesure imputable à un défaut de calibration en vérifiant la répétabilité entre les différentes plaques. Un minimum de 10 000 cellules a été mesuré par échantillon et chaque série de mesures a été réalisée en trois réplicats.

#### Test qualité des échantillons

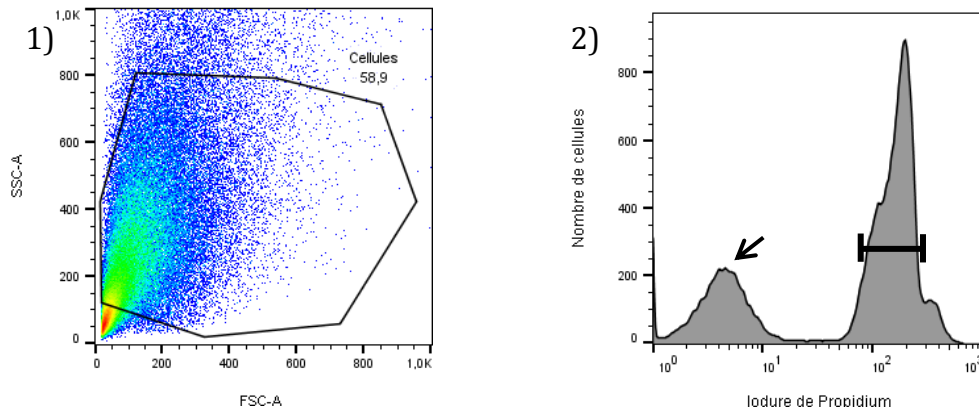
Les données d'intensités de fluorescence ont été analysées en utilisant le logiciel FlowJo v10. Elles reflètent l'incorporation du PI dans l'ADN des cellules préparées et sont directement reliées aux valeurs C recherchées.

Pour chaque échantillon, nous avons sélectionné un ensemble de cellules sur le profil SSC x FSC en évitant de prendre en compte d'éventuels débris (identifiés en rouge sur les Figure II.3.A.1 et II.3.B.1). Nous avons ainsi déterminé les profils correspondants de distribution de fluorescence des cellules en fonction du nombre de cellules analysées (Figures II.3.A.2 et II.3.B.2).

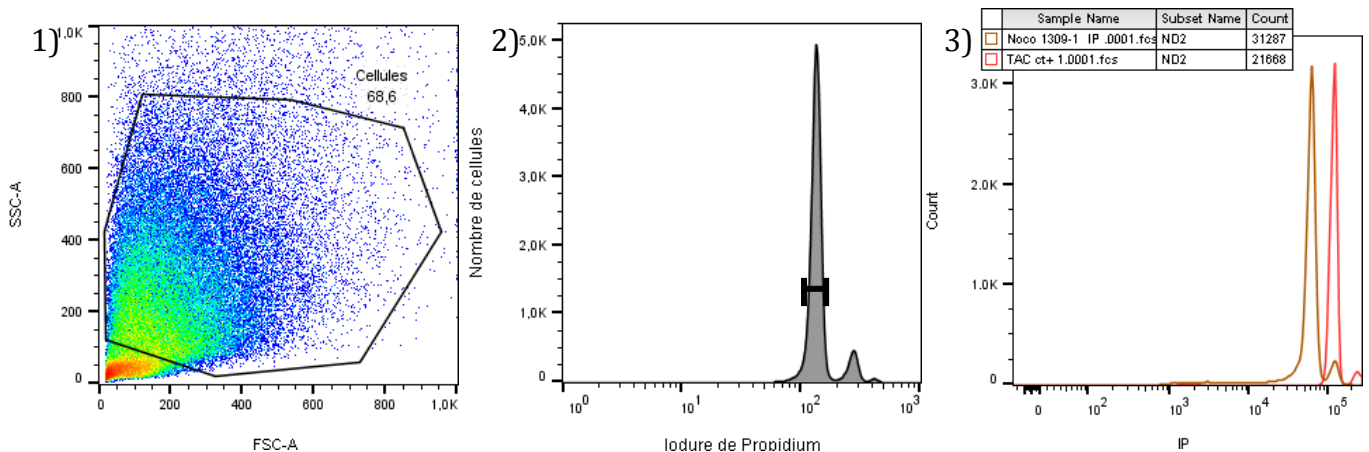
Les deux échantillons (*N. coriiceps* Ich1103 et Ich1104) présentant un profil correspondant à celui de la Figure II.3.A.2 ont été écartés de l'analyse en raison de la présence de débris, d'un

faible nombre de cellules prises en compte dans l'analyse, et de l'étalement en fluorescence du pic obtenu. Au contraire, les échantillons présentant un profil correspondant à celui présenté chez *N. coriiceps* (Ich1096) (Figure II.3.B.2) ont été retenus pour la suite des analyses.

(A)



(B)



**Figure II.3 :** Exemples pour deux échantillons : *N. coriiceps* Ich1103 (A) et *N. coriiceps* Ich1096 (B) de profils SSC x FSC (A.1, B.1) et de profils de distribution des intensités de fluorescence des cellules sélectionnées correspondants (A.2, B.2). Le Side Scatter (SSC) rend compte de la granulosité des cellules et le Forward scatter (FSC) de leur taille. Le profil SSC x FSC représente la taille et la complexité des cellules présentes dans les échantillons. Dans le cas A, l'échantillon contient de nombreux débris visibles en rouge sur le profil A.1 et peu de cellules de bonne qualité. Le profil de distribution des cellules prises en compte (58,9 % des cellules totales) présenté en A.2 montre la persistance de débris, indiqués par la flèche noire, et un étalement des intensités de fluorescence mesurées. Sur la base de ces critères, cet échantillon a donc été jugé de mauvaise qualité et écarté de l'analyse. Au contraire, le faible nombre de débris et le pic d'intensité de fluorescence détecté pour le deuxième cas (B.1, B.2) a permis de déterminer la moyenne d'intensité de fluorescence précise. Les profils de l'individu ont donc été comparés à celui du standard pour estimer sa taille de génome (B.3, superposition des profils de *N. coriiceps* Ich1096 (courbe marron) et de la truite arc-en-ciel (courbe rouge)).

## Mesure de la taille des génomes

Les tailles de génomes des échantillons retenus suite au test qualité ont été estimées en comparant les intensités de fluorescence moyenne (hauteur et étalement des pics obtenus) des échantillons à celles du standard (Truite arc-en-ciel) (Figure II.3.B.3).

Tests statistiques des différences entre les tailles de génomes :

Compte tenu du nombre d'individus analysés très variable entre espèces (de 1 à 17), nous avons choisi de tester les éventuelles différences de taille de génomes par des ANOVA entre individus d'une même espèce, puis par une ANOVA entre espèces. Cette dernière analyse a été suivie par des comparaisons multiples (t de Student) sous la protection de l'ANOVA en utilisant la variance résiduelle.

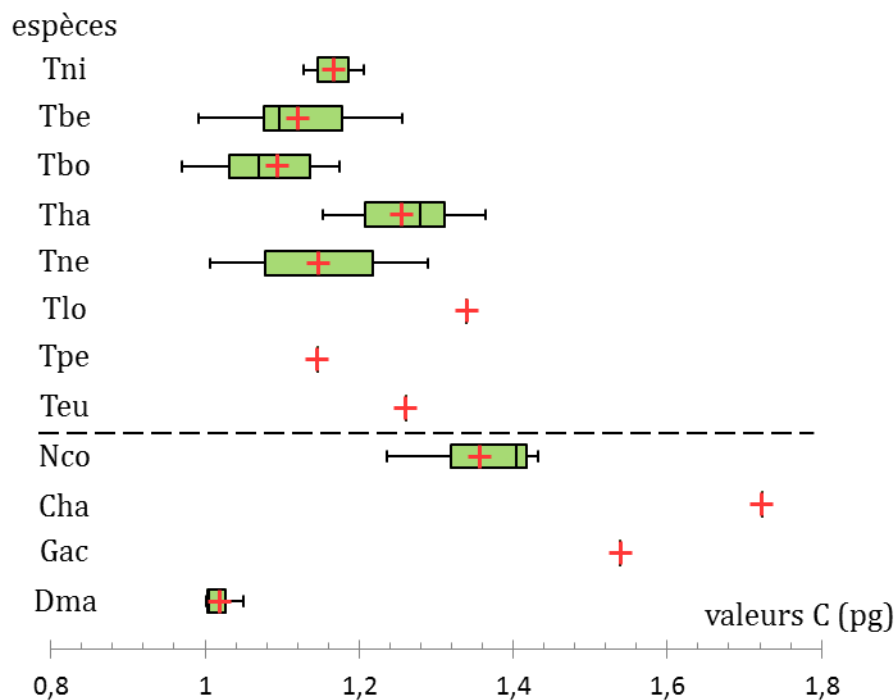
### I.1.B.3 : Résultats

Des différences de tailles de génomes existent tant entre individus de la même espèce (intraspécifique) qu'entre individus d'espèces différentes (interspécifique) (Tableau II.1 et Figure II.4).

| Sous famille    | Genre Espèce                  | Nc         | Ni | Min et Max intra | Valeurs C (pg) ± SE | t multiple à 5 % |
|-----------------|-------------------------------|------------|----|------------------|---------------------|------------------|
| Trematominae    | <i>I. cyanobrancha</i> (Icy)  | 48         | 0  |                  | ND                  |                  |
|                 | <i>T. scotti</i> (Tsc)        | 48         | 0  |                  | ND                  |                  |
|                 | <i>T. nicolai</i> (Tni)       | 57 - 58    | 2  | 1,13 - 1,21 *    | 1,17 ± 0,04         | A B C            |
|                 | <i>T. bernacchii</i> (Tbe)    | 48         | 17 | 0,99 - 1,26 *    | 1,12 ± 0,02         | A B              |
|                 | <i>T. lepidorhinus</i> (Tle)  | 47 - 48    | 0  |                  | ND                  |                  |
|                 | <i>T. borchgrevinki</i> (Tbo) | 45 - 46    | 8  | 0,97 - 1,31 *    | 1,09 ± 0,04         | A B              |
|                 | <i>T. hansonii</i> (Tha)      | 45 - 46/48 | 12 | 1,05 - 1,36 *    | 1,25 ± 0,03         | C                |
|                 | <i>T. newnesi</i> (Tne)       | 45 - 46    | 2  | 1,01 - 1,29 *    | 1,15 ± 0,14         | A B C            |
|                 | <i>T. loennbergii</i> (Tlo)   | 47 - 48    | 1  |                  | 1,34                | C D E            |
|                 | <i>T. pennellii</i> (Tpe)     | 32         | 1  |                  | 1,15                | A B C            |
|                 | <i>T. eulepidotus</i> (Teu)   | 24         | 1  |                  | 1,26                | B C D            |
| Nototheniinae   | <i>N. coriiceps</i> (Nco)     | 22         | 5  | 0,98 - 1,43 *    | 1,36 ± 0,06         | D E              |
| Channichthyinae | <i>C. hamatus</i> (Cha)       | 48         | 1  |                  | 1,72                | F                |
| Bathydraconinae | <i>G. acuticeps</i> (Gac)     | 48         | 1  |                  | 1,54                | E F              |
| Dissostichinae  | <i>D. mawsoni</i> (Dma)       | 48         | 3  | 1,00 - 1,05      | 1,02 ± 0,02         | A                |

**Tableau II.1 : Estimation de la taille des génomes par cytométrie de flux.** Nc : nombre de chromosomes (2n). Ni : nombre d'individus analysés. Min et Max : taille minimale et maximale au sein d'une même espèce. Les différences significatives en intraspécifique sont indiquées par des étoiles. Valeurs C : moyenne de la taille du génome haploïde ± erreur standard. t multiple : résultats des comparaisons des moyennes deux à deux sous protection de l'ANOVA, les espèces ne présentant pas de différence significative de taille de génomes partagent une même lettre.

Les douze espèces examinées se distribuent dans six groupes (notés de A à E) plus ou moins chevauchant et les « *Trematomus* » sont présents dans cinq de ces groupes (tableau II.1). Les deux premiers groupes (A et B) sont relativement proches en taille, avec respectivement une moyenne arithmétique de 1,12 pg et 1,16 pg (pour une erreur standard de 0,02 pg dans les deux cas). Ces deux groupes ont cinq espèces de « *Trematomus* » en commun (*T. nicolai*, *T. bernacchii*, *T. borchgrevinki*, *T. newnesi*, et *T. pennellii*) et se distinguent au niveau de deux espèces : *D. mawsoni* présent uniquement dans le groupe A et *T. eulepidotus* présent uniquement dans le groupe B. Le groupe C a une moyenne de 1,22 pg (erreur standard 0,03) et inclut les deux derniers « *Trematomus* » de notre étude : *T. hansonii* et *T. loennbergii*. La variation des tailles au sein du groupe « *Trematomus* » est donc relativement faible, de l'ordre de 9 %. Ce résultat peut également être visualisé au niveau des intervalles à 95 % de la taille des génomes chez douze espèces de Nototheniidae qui se chevauchent clairement au sein des huit espèces de « *Trematomus* » (Figure II.4).



**Figure II.4 :** variations des tailles de génomes chez douze espèces de Nototheniidae. Les diminutifs des espèces examinées correspondent respectivement du haut vers le bas à *T. nicolai*, *T. bernacchii*, *T. borchgrevinki*, *T. hansonii*, *T. newnesi*, *T. loennbergii*, *T. pennellii*, *T. eulepidotus*, pour le groupe « *Trematomus* », et *N. coriiceps*, *C. hamatus*, *G. acuticeps*, *D. mawsoni*. Ces variations sont représentées sous forme de boîtes à moustaches (ou « boxplot »). Les moyennes des valeurs C sont indiquées par des croix rouges. La médiane est représentée par un trait vertical noir dans la boîte. Les intervalles de confiance à 95 % sont indiqués par les barres horizontales de part et d'autre de chacune des boîtes.

#### ***I.1.B.4 : Discussion***

Pour les échantillons ayant passé avec succès le contrôle qualité, on note une bonne répétabilité des mesures pour chaque individu (écart type moyen des mesures par individu toutes espèces confondues = 0,015 pg). Nos résultats sont du même ordre que ceux précédemment publiés (Tableau IV.3, chapitre IV, article) (Hardie and Hebert 2003; Hardie and Hebert 2004; Chen et al. 2008; Detrich et al. 2010; Detrich and Amemiya 2010). L'ensemble des espèces du groupe « *Trematomus* » présente une faible gamme de variation des tailles de génome, assez comparable à celle estimée par Detrich *et al.* (2010) (0.98 - 1.20 pg). Ils appartiennent aux groupes A, B et C définis précédemment. Cependant, la variabilité du nombre de spécimens par espèce est telle au sein de notre échantillonnage qu'il est difficile de tirer des conclusions, notamment pour trois « *Trematomus* » (*T. loennbergii*, *T. eulepidotus* et *T. pennellii*). L'élargissement et l'homogénéisation de l'échantillonnage d'étude en termes du nombre de spécimens par espèce serait requis pour pouvoir comparer les variations mesurées. De plus, nous avons établi une différence significative des tailles de génomes mesurées en trois réplicats chez un même spécimen, mais lors d'acquisitions différentes. Les différences observées pourraient donc refléter également un effet « moment de la mesure », en plus d'une réalité biologique. Nous considérons donc que les variations de tailles de génomes au sein de la radiation « *Trematomus* » ne sont pas notablement différentes entre les espèces et sont du même ordre de grandeur.

Les variations de tailles de génomes entre individus/populations au sein d'une même espèce ont été montrées à des nombreuses reprises chez l'homme (Redon et al. 2006; Korbelt et al. 2007; The 1000 Genomes Project Consortium 2012) ou encore chez les plantes (Cavallini and Natali 1991) comme par exemple l'orge sauvage *Dasyphyrum villosum* (Caceres et al. 1998). Il peut s'agir de différences allant de 3 kb à plusieurs centaines de Mpb, soit l'équivalent de 0,1 à 0,4 pg : le même ordre de grandeur observé dans notre étude. Les auteurs suggèrent que ces différences observées pourraient être imputables à des différences interindividuelles en contenu génomique appelées aussi variations structurales (variations du nombre de copies en éléments transposables *Alu* et LINEs chez l'homme par exemple) (Redon et al. 2006; Korbelt et al. 2007; The 1000 Genomes Project Consortium 2012), pouvant impliquer des insertions/délétions (Korbelt et al. 2007) dont les répercussions sont visibles à l'échelle chromosomique (Caceres et al. 1998). N'ayant pas toutes les données concernant le sexe des individus examinés, nous ne pouvons pas conclure quant à une éventuelle différence ou conservation des tailles de génomes entre mâles et femelles. Cependant, chez les espèces examinées, il est à noter que l'on n'observe aucune corrélation entre le nombre de chromosomes et l'appartenance à un des six groupes

(Teu, 24 chromosomes, 1,26 pg ; Tpe, 32 chromosomes 1,32 pg ; Tha, 45-46 chromosomes, 1,26±0,025 pg, Tbe, 48 chromosomes, 1,12±0,019 pg).

Aucune donnée de taille de génomes n'est disponible pour les trois genres les plus apparentés à la famille des Nototheniidae (Bovichthidae, Eleginopsidae ou Pseudaphritidae), ni pour le groupe frère récemment proposé pour le sous-ordre Notothenioidei : les Percophidae (Near et al. 2015). Cependant, cette faible variation entre les tailles de génomes correspond à celles mesurées chez les espèces appartenant aux familles les plus apparentées à ce sous-ordre comme les Percidae (*P. flavescens*, 0,92 pg, *P. fluviatilis*, 0,90 pg, *Percina caprodes*, 1,06 pg, *Sander lucioperca*, 1,14 pg, *Sander vitreus*, 1,06 pg) (Hardie and Hebert 2004; Detrich et al. 2010), ou encore les Serranidae (gamme de variation 0,80 - 1,30, moyenne de 1,09 ±0,17 pg) (Dettaï and Lecointre 2004; Hardie and Hebert 2004). Ces résultats suggèrent fortement que cette valeur C de l'ordre de 1 - 1,30 pg pourrait correspondre à un état plésiomorphe, conservé notamment au sein de la radiation « *Trematomus* ».

Compte tenu de nos résultats, nous écartons l'hypothèse de duplication de génome entier suivi de réarrangements ultérieurs au profit de l'hypothèse alternative la plus vraisemblable qui sera retenue dans cette étude. Les réarrangements à l'origine de la diversité chromosomique ayant accompagné la diversification des « *Trematomus* » sont des modifications structurales (fusions centriques ou en tandem, fissions, inversions, translocations) sans polyploïdisation. En prenant uniquement en compte les valeurs de tailles de génomes mesurées ainsi que la variation interspécifique du nombre de chromosomes, cette hypothèse s'applique également à d'autres genres de Nototheniidae : *Dissostichus* (*D. mawsoni*, 1.02 pg), *Lepidonotothen* (*L. nudifrons*, 1.12 pg), et *Gobionotothen* (*G. gibberifrons*, 0.98 pg) (Detrich et al. 2010; Detrich and Amemiya 2010). Le Nototheniinae *N. coriiceps*, appartenant aux groupes D et E semble également être inclus en limite supérieure dans cet intervalle, même si la moyenne calculée est supérieure à celle des groupes A, B et C.

Cette stabilité des tailles de génomes n'a pas été observée dans tous les clades de Nototheniidae. En effet, les valeurs C mesurées chez le Bathydraconinae *G. acuticeps* (1,54pg) et la Channichthyinae *C. hamatus* (1,72pg), appartenant aux deux derniers groupes E et F, paraissent se détacher avec des tailles plus élevées que les autres. Elles sont conformes aux mesures effectuées par Chen *et al.* (2008) et par Detrich *et al.* (2010) chez une espèce de Bathydraconinae (*P. charcoti* : 1.34±0.05 pg, Ni = 2) et surtout chez quatre espèces différentes de Channichthyinae (*C. aceratus*, *C. gunnari*, *C. rastrospinosus*, *P. georgianus*, moyenne des tailles de génomes de 1,77pg, Ni = 8). Ces expansions génomiques pourraient être imputables à des événements de transposition massives dans ces génomes au cours de l'évolution, ou encore à

des délétions ou duplications segmentales (Kidwell 2002; Kraaijeveld 2010; Chénais et al. 2012). Detrich (2010) et Chen (2008) soulèvent l'hypothèse d'une expansion de gènes (concernant 118 familles géniques) dans un but adaptatif, notamment en liaison avec l'acquisition du matériel génique nécessaire à des compensations au niveau de l'appareil cardiovasculaire et du métabolisme, consécutives à la perte de l'expression de l'hémoglobine chez ces espèces.

## **II.2 : Recherche et identification des homologies chromosomiques interspécifiques**

### **II.2.A : les homologies chromosomiques interspécifiques**

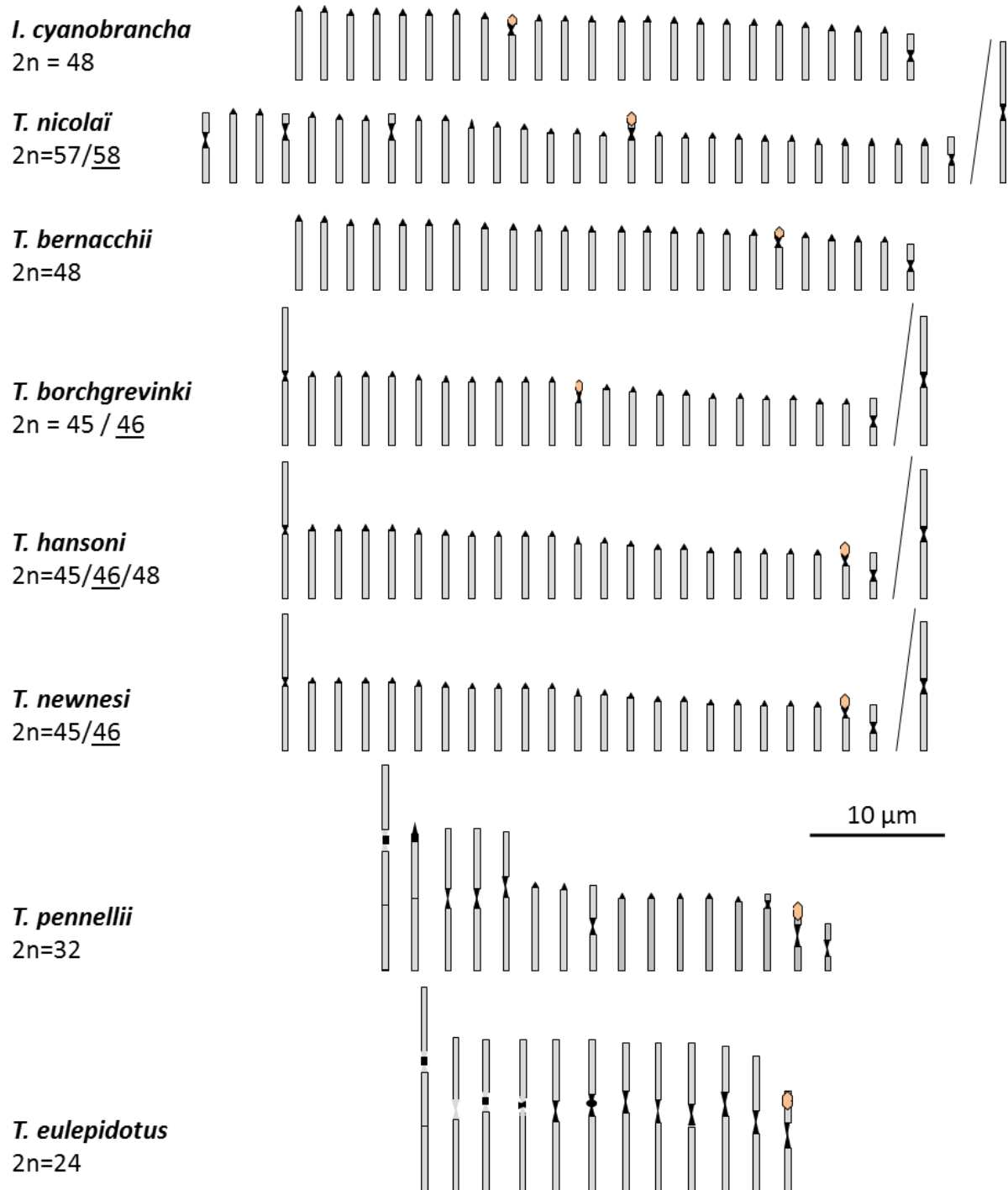
Avec une faible variabilité des tailles de génomes chez les différentes espèces de « *Trematomus* » (et plus largement chez d'autres représentants des Nototheniidae), aucune duplication génomique ou de réduction massive n'ont été mises en évidence. Nous retenons donc comme plus probable l'hypothèse alternative des remaniements structuraux ayant accompagné la radiation « *Trematomus* ». Pour identifier et reconstituer ces remaniements, nous avons défini comme unité de référence les 52 unités structurales présentes chez *I. cyanobrancha*, dont le caryotype pourrait refléter l'état plésiomorphe des « *Trematomus* », et plus largement, des Nototheniidae (Figures I.3 et II.2).

En cytogénétique, l'établissement des caryotypes par classification des chromosomes prend principalement en compte deux critères (Levan et al. 1964). Le premier est la taille et le deuxième, la morphologie propre à chaque paire (notamment par rapport à la position de leur centromère, Levan *et al.* (1964)). Les tailles des chromosomes dépendent essentiellement du degré de condensation de l'ADN, et donc de l'état plutôt précoce ou tardif de la métaphase observée. Ces tailles peuvent ainsi varier d'une métaphase à une autre pour un même individu, et *a fortiori* entre individus et entre espèces. Il est donc difficile d'établir une taille absolue pour chaque paire. Cependant, il est possible de déterminer des tailles relatives au sein d'une même métaphase. Cela permet la comparaison de caryomorphes moyens entre individus et de différents caryotypes entre espèces. Chez les espèces à  $2n = 48$  comme *T. bernacchii* ou encore *T. scotti*, la taille et la morphologie des chromosomes acrocentriques ou métacentriques sont comparables à celle des 48 unités structurales de référence (*I. cyanobrancha*). Par contre, on observe des chromosomes de plus grande taille, équivalente à deux ou trois unités structurales d'*I. cyanobrancha* pour les « *Trematomus* » dont le caryotype est  $2n < 48$ . C'est par exemple le cas pour la plus grande paire chromosomique métacentrique de *T. hansonii*, *T. newnesi* et *T. borchgrevinkii* ; pour la plus grande paire chromosomique submétacentrique de *T. pennellii* et *T. eulepidotus* ; ou encore pour les chromosomes sexuels différenciés Y de *T. lepidorhinus*,

*T. nicolai*, *T. newnesi*, *T. hansonii* et *T. borzhgrevinkii*. A l'inverse, la taille des plus petites paires de *T. nicolai*, dont le caryotype est à  $2n > 48$ , semble au moins deux fois plus petite que les unités structurales définies chez notre espèce de référence. Les comparaisons des tailles des chromosomes composant les caryomorphes des différents « *Trematomus* » avec notre taille de référence constituée par les unités structurales définies chez *I. cyanobrancha* apparaissent comme de bons indicateurs des remaniements chromosomiques survenus lors de la diversification du groupe. Compte tenu des caryotypes observés et comparés à celui de *I. cyanobrancha*, l'hypothèse d'une conservation des 52 unités structurales de référence chez l'ensemble des « *Trematomus* » semble vraisemblable (Figure II.5).

Les remaniements chromosomiques survenus au cours de cette radiation récente seraient donc des combinaisons des 52 unités structurales héritées de l'ancêtre commun des « *Trematomus* ». Comme le montre la figure II.5, les fusions chromosomiques semblent majoritaires dans ce groupe. Elles semblent marquer une tendance préférentielle à la réduction du nombre total de chromosomes (Morescalchi et al. 1992; Morescalchi et al. 1996; Pisano et al. 2000; Pisano and Ozouf-Costaz 2000; Pisano et al. 2003). Cependant, nous ne savons pas si ces événements de réarrangements sont apparus chez l'ancêtre commun et partagés par la descendance, avec possibles pertes ou gains secondaires. Dans le premier cas de figure, les fusions mettraient en jeu les mêmes unités structurales, alors que sous la deuxième hypothèse, les produits de ces réarrangements ne seraient pas homologues et les fusions mettraient en jeu des unités structurales différentes. Pour aller plus loin dans la caractérisation et la reconstitution de l'histoire évolutive de ces réarrangements (fusions, fissions, inversions), et tester notre scénario évolutif, il est nécessaire de pouvoir identifier clairement chacune de ces unités structurales. C'est l'identification de ces dernières, impliquées dans les fusions/fissions qui permettra (1), d'établir les homologies chromosomiques interspécifiques (HCI) (correspondances entre les chromosomes des différentes espèces), et (2), en reconstituant l'histoire des réarrangements mis en évidence, de déterminer s'il y a partage ou évolution indépendante de ces remaniements entre les différentes espèces.





**Figure II.5 :** schéma d'une carte chromosomique haploïde présentant les caryomorphes des différents « *Trematomus* » caryotypés. Seul le lot haploïde est représenté ici. Les unités structurales ont été identifiées pour les caryotypes de chaque espèce relativement à notre référence, *I. cyanobrancha*. On compte un total de 52 unités structurales dans l'ensemble des espèces de « *Trematomus* » caryotypées. *T. nicolai* est la seule espèce à présenter des fissions chromosomiques au sein du groupe. Du fait du choix d'*I. cyanobrancha* comme référence, dans certains cas chez *T. nicolai*, il faut au moins deux paires chromosomiques pour constituer une unité structurale. La localisation correspondant aux ARN ribosomiques 18 et 28S sur une paire ayant fait l'objet d'une inversion péricentrique (héritée de l'ancêtre commun du groupe formé par les genres Pseudaphritidae, Eleginopsidae et Nototheniidae (Figure II.2)) est indiquée au moyen du losange orange.

Les recherches d'homologies chromosomiques (HC) ont souvent été réalisées chez l'Homme (*Homo sapiens*) et ses proches parents (Scherthan et al. 1994; Müller et al. 1999; Dumas et al. 2005; Picone et al. 2011), des espèces modèles (Romanenko et al. 2007; Ventura et al. 2009) ou encore des espèces domestiques ou d'intérêt agronomique (Chaves et al. 2004; Nanda et al. 2011; Mandáková, Hloušková, et al. 2017; Mandáková, Pouch, et al. 2017). Dans des cas moins fréquents, elles ont été appliquées à des groupes d'espèces évoluant en milieu naturel (Graphodatsky et al. 2000; Ráb et al. 2008). Les identifications de ces HC ont divers objectifs, notamment dans la recherche d'anomalies chromosomiques (Breneman et al. 1993; Weimer et al. 1999; Blennow 2004) à des fins thérapeutiques (Guan et al. 1994; Ried 1998; Arens et al. 2004). Dans un cadre plus fondamental, les recherches d'HC à l'intérieur d'une même espèce permettent d'identifier les duplications de gènes paralogues, la mise en place de familles multigéniques, ainsi que d'autres remaniements de grande ampleur tels que les duplications de génomes entiers (Jaillon et al. 2004). Dans la perspective d'études comparatives (au niveau interspécifique), connaître les HCI présente de nombreux intérêts comme l'identification de l'origine de paires sexuelles hétéromorphes (Henning et al. 2008; Machado et al. 2011; Hirai et al. 2012), de chromosomes surnuméraires B (Trifonov et al. 2002; Teruel et al. 2009; Voltolin et al. 2010), ou encore la reconstruction de liens de parentés (Mandáková, Pouch, et al. 2017). Ces HCI sont alors identifiables chez un nombre important d'espèces (Ried 1998; Ventura et al. 2009). Elles peuvent aussi être utilisées afin de retracer des remaniements interchromosomiques structuraux (types fissions, fusions, translocations) (Ried 1998; Vicari et al. 2015) dans une logique de reconstitution de scénario évolutif d'une diversification spécifique (Romanenko et al. 2007; Stanyon and Stone 2008; Picone et al. 2011) et/ou chromosomique (Yang et al. 2004; Ráb et al. 2008).

Différents types de méthodologie ont été employés au cours de ces dernières années pour identifier les HCI. On peut citer le marquage en bandes permettant d'identifier les paires ou des régions chromosomiques d'intérêt sur la base du contraste des bandes obtenues. C'est un outil pratique et peu coûteux offrant rapidement la possibilité d'établir des caryotypes ainsi que des HC, permettant de reconstituer certains remaniements intra ou interchromosomiques intra ou interspécifiques. Cette technique est surtout utilisée chez les Mammifères (Couturier and Dutrillaux 1981; Wienberg et al. 1992; Graphodatsky et al. 2000; Yang et al. 2004; Ventura et al. 2009) en marquant les bandes chromatidiennes pour lesquels il offre une bonne résolution (Wienberg et al. 1992; Dumas et al. 2005). Des fusions massives ont ainsi été détectées dans des populations de souris (Auffray 1993; Castiglia et al. 2002) ainsi que d'autres remaniements structuraux chez d'autres rongeurs comme les gerbilles (Dobigny et al. 2003; Dobigny et al. 2005) ou les musaraignes (Biltueva et al. 2011). De la même manière, on peut citer la mise en évidence des remaniements chromosomiques entre les cercopithèques (Dutrillaux 1979;

Dutrillaux et al. 1980; Muleris et al. 1983; Sineo 1990) et l'homme, ou encore chez les oiseaux et les amphibiens (Stock and Mengden 1975; Bed'Hom 2000; Pokorná et al. 2012). Toutefois, le marquage en bandes ne fonctionne pas systématiquement ou n'est pas toujours informatif, même chez des taxons aux caryotypes hautement remaniés (Chowdhary and Raudsepp 2001). Pour une résolution affinée à l'échelle génique, le génotypage, ainsi que le séquençage de génomes entiers donnent lieu à la reconstruction de synténies moléculaires à large échelle (Jaillon et al. 2004). Les segments chromosomiques étudiés et potentiellement sujets aux remaniements sont ainsi plus finement bornés, même si les données ne concernent en général que les régions géniques, au détriment des séquences hautement et moyennement répétées, difficiles à assembler et à repositionner au sein des contigs. L'essor des technologies de séquençages dits de nouvelle génération (NGS) permet un accès croissant aux données suffisamment bien assemblées et annotées pour des reconstructions fiables et nombreuses (Green 2001). Néanmoins, les difficultés de reconstruction de larges « contigs » et « scaffolds » (Denton et al. 2014; Ekblom and Wolf 2014) compliquent l'assemblage et l'annotation de génomes de grande taille (Alkan et al. 2011).

Dans le cas où les données en séquences suffisamment assemblées et annotées sont indisponibles, comme pour les Nototheniidae, la cytogénomique comparative est un moyen rapide et facile pour identifier les HCI. Chez les téléostéens, le marquage en bandes offre une résolution nettement moins bonne que chez les mammifères, oiseaux et amphibiens (Hartley and Horne 1985; Medrano et al. 1988; Vicari et al. 2015). Des paires ou des bras chromosomiques ont cependant pu être identifiés chez les anguilliformes (Amores et al. 1995), les siluridés (Ráb et al. 2008), ou encore les salmonidés (Hartley 1987). Ces marquages en bandes permettant de mettre en évidence l'hétérochromatine (banding C), les gènes ribosomiques (Ag-NORs), ou bien des sites de restriction (RE banding pour « restriction endonucléase banding »). Ces différents types de marquages en bandes ont été testés au laboratoire sur les chromosomes de Nototheniidae et seul le marquage en bandes C produit un marquage net et reproductible (Ozouf-Costaz et al. 1997). Cependant, l'hétérochromatine étant concentrée principalement dans les régions centrométriques et péricentromériques et rarement en position intercalaire ou terminale, la résolution de ce type de marquage ne permet pas de faire la différence entre les multiples chromosomes acrocentriques composant les caryotypes des « *Trematomus* », et encore moins d'établir les HCI.

Nous avons testé deux stratégies permettant d'établir les HCI au sein des « *Trematomus* », et plus largement entre les différentes espèces de Nototheniidae. D'une part, nous avons choisi de prélever et marquer l'intégralité d'un chromosome d'une espèce, puis de rechercher par hybridation les chromosomes homologues chez les autres espèces (technique de « peinture

chromosomique »). La deuxième approche concerne l'identification des homologies à l'échelle d'une unité structurale (décrites précédemment chez notre espèce de référence *I. cyanobrancha*) par localisation de fragments génomiques clonés dans des chromosomes artificiels bactériens (BACs pour « Bacterial Artificial Chromosome »), utilisés comme marqueurs de ces différentes unités chromosomiques.

### ***II.2.A.1 : Homologies interspécifiques entre chromosomes entiers***

La peinture chromosomique, utilisée en cytogénomique comparative, est susceptible de fournir de nombreuses informations concernant l'identification des HCl. Ainsi, on appelle sonde de peinture chromosome-spécifique un ensemble de séquences d'ADN génomique issues d'un seul chromosome ou d'un segment de chromosome d'une espèce d'intérêt (Laudicina and Mühlmann 2015; Vicari et al. 2015). Après prélèvement, amplification et marquage de l'ADN obtenu, cette sonde est hybridée par hybridation *in situ* en Fluorescence (*FISH*) et peint les chromosomes homologues des différentes espèces étudiées (*Zoo-FISH*). Cette technique est aujourd'hui au point et employée chez une grande diversité de taxons : chien (Graphodatsky et al. 2000), blé (Šafář et al. 2004), cheval (Yang et al. 2004), mammifères (Stanyon and Stone 2008), rongeurs (Ventura et al. 2009). Les chromosomes peuvent être triés et séparés en fonction de leur taille et de leur constitution en bases azotées par électrophorèse en champ pulsé (Schwartz and Cantor 1984) (chromosomes de petite taille < à 10 Mb) ou par tri en flux (« flow sorting ») (Gray, Carrano, Moore, et al. 1975; Gray, Carrano, Steinmetz, et al. 1975; Monard 1998). Des méthodologies ont également été développées pour optimiser l'obtention d'un signal chromosome-spécifique en enrichissant ces produits chromosomiques isolés en séquences uniques ou en utilisant un mélange de BACs choisis pour couvrir l'intégralité du segment voulu (Hirai et al. 2012; Mandáková, Pouch, et al. 2017; Mandáková, Hloušková, et al. 2017). Le nombre de séquences répétées présentes dans la sonde est alors limité, réduisant ainsi la possibilité d'hybridations sur d'autres chromosomes ou segments chromosomiques que ceux qui sont ciblés (Laudicina and Mühlmann 2015). Ces peintures chromosomiques déjà préparées, commercialisées (par exemple [www.cytocell.com](http://www.cytocell.com), cf REF « Painting Probes LPP xxR/G ») et directement prêtes à l'emploi pour la *FISH*, sont des sondes très accessibles et simples d'application. En plus de cet enrichissement en séquences uniques, la pré-hybridation de l'ADN contenu dans la sonde de peinture avec une quantité excédentaire d'ADN génomique non marqué très riche en séquences répétées (type ADN Cot-1, directement préparé et commercialisé ou ADN compétiteur), permet le ré-appariement double brin des séquences répétées avant hybridation des sondes, et donc l'hybridation préférentielle des séquences

uniques sur les chromosomes cibles (Medrano et al. 1988; Vicari et al. 2011; Swennenhuis et al. 2012; Trifonov et al. 2017).

En raison de la proximité génétique entre certains mammifères, il est possible d'hybrider des sondes de peintures hétérologues sur les chromosomes d'espèces apparentées (Chien-Renard polaire/roux (Graphodatsky et al. 2000), homme-antilope (Chaves et al. 2004), homme-éléphant (Stanyon and Stone 2008)). Par contre, chez les taxons dont le degré d'apparentement avec l'homme ou les espèces domestiques pour lesquelles existent des sondes commerciales (poulets, porcs, bovins) est faible, la distance génétique est trop importante pour que les sondes de peinture hétérologues hybrident sur les cibles. La séparation des chromosomes un à un peut également être compromise pour les caryotypes présentant pour plusieurs paires chromosomiques une homogénéité phénotypique (en taille) et en composition en bases, diminuant alors les possibilités d'utilisations du tri en flux (Pinton et al. 2003) pour la constitution des peintures.

La microdissection chromosomique apparaît alors comme une des seules alternatives afin de pouvoir préparer une sonde de peinture « maison » ciblant une région chromosomique d'intérêt chez n'importe quelle espèce (Guan et al. 1994; Laudicina and Mühlmann 2015; Vicari et al. 2015). Cette méthodologie a été développée sur les chromosomes polytènes de la drosophile (Hill and Watt 1977; Scalenghe et al. 1981), puis chez l'homme (Bates et al. 1986) et la souris (Röhme et al. 1984). C'est au début des années 1990 que paraissent les premières utilisations de sondes de peinture chromosomique obtenues à partir d'éléments microdisséqués (Deng et al. 1992; Kao 1993). Cette technique est aujourd'hui régulièrement employée pour un grand nombre d'organismes : poulets (Griffin et al. 1999), tortues (Ulsh et al. 2000), insectes (Teruel et al. 2009), « poissons » (Henning et al. 2008; Vicari et al. 2011), « plantes » (Stein et al. 1998; Matsunaga et al. 1999; Zhou et al. 1999; Kerk 2003). Elle consiste à prélever manuellement à partir d'une préparation chromosomique d'un individu un ou plusieurs chromosomes, bras chromosomiques, ou bandes chromatidiennes de la même paire à l'aide d'un laser ou d'une micro-aiguille en verre (Weimer et al. 1999; Vicari et al. 2015). Une fois amplifié et marqué, ce produit de microdissection est hybridé par *FISH* sur les chromosomes des espèces d'intérêt. Dans de rares cas, un seul chromosome est suffisant pour élaborer la sonde de peinture (Guan et al. 1993; Viersbach et al. 1994; Christian et al. 1999; Griffin et al. 1999; Trifonov et al. 2002). Par contre, dans la majorité des situations, il est nécessaire d'obtenir des lots de plusieurs chromosomes de la même paire, constituant une quantité minimale de produit microdisséqué suffisante pour une bonne amplification. L'identification des HCl par peinture de sondes microdisséquées nécessite cependant que trois conditions soient réunies. Il est en effet nécessaire 1- de pouvoir reconnaître/identifier de manière fiable et répétable la paire de

chromosomes à microdisséquer au sein des différentes métaphases de l'échantillon observé, du fait de la nécessité de récolter plusieurs copies de ce chromosome pour réaliser une sonde (Zhou and Hu 2007; Vicari et al. 2015), 2- d'étudier un groupe d'espèces dont les génomes sont relativement conservés pour garantir une identité de séquences sondes/cibles suffisante et donc assurer leur hybridation, 3- d'avoir à disposition et en quantité de l'ADN génomique de l'espèce dont la sonde est issue ou d'une espèce génétiquement proche afin de pouvoir élaborer un ADN compétiteur ou de l'ADN Cot-1 qui sera pré-hybridé avec la sonde de peinture avant la *FISH* (Vicari et al. 2011; Swennenhuis et al. 2012; Trifonov et al. 2017).

La plupart des paires de chromosomes des espèces de Nototheniidae sont trop semblables pour les identifier clairement sur la base de leur seule taille, morphologie ou contraste DAPI. L'approche par hybridation de sondes de peinture sur les chromosomes de diverses espèces de la famille a été proposée pour la première fois par Ozouf-Costaz *et al.* (1997). Bien qu'appliquée pour la première fois pour la paire sexuelle différenciée Y de *C. hamatus* (Cocca et al. 2015), cette méthode n'avait jamais été réalisée chez les Nototheniidae dans le cadre d'une démarche comparative.

Nous avons réalisé une sonde de peinture, à partir d'un chromosome spécifique obtenu à partir de la microdissection de la plus grande paire chromosomique de *Trematomus pennellii*, probable produit de fusions. C'est une des rares paires facilement identifiable par sa très grande taille et sa morphologie submétacentrique. Le choix des espèces à hybrider s'est porté sur des espèces aux caryotypes très différents dont deux « *Trematomus* » : *T. pennellii* ( $2n = 32$ ) et *T. hansonii* ( $2n = 45-46, 48$ ). D'autres représentants de la famille des Nototheniidae ont été inclus pour comparaison : *Notothenia coriiceps* ( $2n = 22$ ) et *Dissostichus mawsoni* ( $2n = 48$ ).

### **II.2.A.2 : Homologies entre unités structurales au sein des « *Trematomus* »**

Une autre façon de retracer l'histoire des remaniements chromosomiques ayant accompagné la radiation des « *Trematomus* » est d'identifier les unités structurales correspondant à celles identifiées chez notre espèce de référence *I. cyanobranca* pour chacun des caryotypes de « *Trematomus* » et d'établir ainsi les homologies des bras. Nous serons ainsi en mesure de pouvoir reconnaître et de suivre les unités structurales impliquées ou non dans les fusions ou fissions, et donc de reconstituer une histoire commune ou indépendante de ces événements. Pour identifier ces unités structurales chez les « *Trematomus* » en utilisant la cytogénomique comparative, l'idéal serait de pouvoir connaître le devenir de chacune dans son intégralité. Cependant, comme explicité précédemment, l'approche peinture chromosomique ne peut pas être utilisée à cette échelle en raison de la grande ressemblance entre les multiples

chromosomes acrocentriques constituant les caryotypes des « *Trematomus* ». Si l'on n'est pas capable d'identifier les chromosomes, on l'est encore moins à distinguer de manière fiable les différentes unités structurales qui les composent.

Nous avons donc hybridé des BACs issus de la banque génomique ayant servi au séquençage du génome de *N. coriiceps* (Shin et al. 2014) sur les chromosomes des « *Trematomus* ». Nous avons besoin d'identifier une même unité structurale par plusieurs BACs afin de pouvoir délimiter et orienter un segment représentatif de ces unités chromosomiques, et ainsi mettre en évidence des fusions, fissions ou inversions. Nous considérons que le marquage d'une unité structurale sera assuré par une co-localisation d'un minimum de deux BACs, suffisamment espacés dans le bras pour pouvoir distinguer chacun d'entre eux (pas de superposition des signaux détectés).

Les BACs (772 au total) ont été généreusement fournis par l'équipe de John Postelthwait (University of Oregon) dans le cadre d'une collaboration. Etant donné qu'ils n'avaient pas été caractérisés, nous les avons dans un premier temps localisés sur les chromosomes de *N. coriiceps*. Cette première étape nous a permis de fournir des résultats utiles à nos collaborateurs, et d'identifier des combinaisons favorables (deux BACs ou plus) présentes sur un même bras. Les BACs ont tous hybridé sans problème sur les chromosomes de « *Trematomus* » du fait de la divergence entre les Nototheniidae considérée comme récente (23,9 Ma, Matchiner *et al.* (2011), 22,4 Ma, Near *et al.* (2012), 13,4, Colombo *et al.* (2015)) et de la conservation des génomes à l'échelle de la famille (Detrich et al. 2010; Detrich and Amemiya 2010). En raison de sa proximité caryotypique avec notre référence *I. cyanobrancha* et d'une quantité importante de matériel permettant une étape d'exploration, *T. newnesi* ( $2n = 45/46$ ) a été choisi pour identifier en aveugle les unités structurales. Les BACs sélectionnés à l'issue de cette étape ont été utilisés par la suite pour l'étude des caryotypes d'autres espèces appartenant au groupe « *Trematomus* ».

## II.2.B : Méthodes

### Préparations chromosomiques

La sonde de peinture a été hybridée sur les chromosomes de deux « *Trematomus* » (*T. pennellii* et *T. hansonii*), un Nototheniinae (*N. coriiceps*) et un Dissostichinae (*D. mawsoni*). Les spécimens ayant servi pour les étalements chromosomiques sont référencés en annexe (annexe 1).

Nous avons hybridé les BACs sur les chromosomes de *N. coriiceps*, huit « *Trematomus* » (*I. cyanobranca*, *T. bernacchii*, *T. eulepidotus*, *T. newnesi*, *T. pennellii*, *T. hansonii*, *T. nicolai*, *T. borchgrevinkii*), ainsi que d'autres Nototheniidae : un *Lepidonotothen* (*L. larseni*), un *Patagonotothen* (*P. ramsayi*), un Dissostichinae (*D. mawsoni*), un Channichthyinae (*C. hamatus*), un Gymnodraconinae (*G. acuticeps*), un Cygnodraconinae (*C. mawsoni*); et deux Notothenioidei n'appartenant pas aux Nototheniidae : un Eleginopsidae (*E. maclovinus*) et un Bovichtidae (*B. diacanthus*). Les spécimens ayant servi pour les étalements chromosomiques sont référencés en annexe (annexe 1).

Les chromosomes mitotiques des différentes espèces de Nototheniidae ont été obtenus à partir de cultures cellulaires de rein céphalique et rate (Rey et al. 2015), organes hématopoïétiques dans lesquels de nombreuses cellules souches donnent naissance à diverses lignées cellulaires (lymphocytes, précurseurs lymphoïdes...). Brièvement, les cellules de rein céphalique et de rate ont été tamisées puis mises en culture dans du milieu L-15 Leibovitz sans bicarbonate, enrichi en L-glutamine et en sérum de veau fœtal, et supplémenté avec des lectines (les mitogènes concanavaline A et pokeweed) et des antibiotiques (solution antibiotique-antimycotique, Sigma-Aldrich) pour une période de 6,5 jours entre 0° et +2°C. La colchicine a été ajoutée dans les milieux 6h avant le choc hypotonique (KCL 0,075M, 1h à +2°C). Les suspensions cellulaires ainsi obtenues ont été fixées trois fois dans un mélange éthanol absolu/acide acétique glacial 3:1 et congelées sur le terrain.

Ces préparations chromosomiques ont été récoltées durant les campagnes antarctiques des programmes de l'IPEV (Institut Paul Emile Victor) et de la NSF (U.S National Science Foundation) : **REVOLTA** (Radiation EVOLutives en Terre Adélie, 2010-2014, Terre Adélie), **ICOTA** (Ichtyologie côtière en Terre Adélie, 1996-2008, Terre Adélie), et **ICEFISH2** (International Collaborative Expedition to collect and study Fish Indigenous to Sub-antarctic Habitats, 2004, Océan Austral et Atlantique Sud). Les identifications et caryotypages des poissons ont été réalisés par Catherine Ozouf-Costaz, par Guillaume Lecointre, ou par des volontaires aides techniques (VAT).



Pour les étalements, les suspensions cellulaires ont été décongelées et centrifugées 10 minutes à 4° C / 1500 rpm. Le surnageant a été retiré et le culot resuspendu dans du fixateur frais. Les étalements pour la *FISH* (1) et la microdissection (2) ont été réalisés à l'aide d'une pipette Pasteur, par dépôt sur lames superfrost (1) ou lamelles 24 x 60 mm (2) préalablement dégraissées, sous vapeur d'un bain marie à 37° C afin d'optimiser la dispersion des chromosomes métaphasiques.

Pour la *BAC-FISH*, les lames contenant un grand nombre de métaphases (> 10-15) bien étalées peuvent être deshybridées (séries de bains de trois minutes chacun, détergent tampon phosphate (PBD pour « phosphate buffer detergent ») 1X / SSC 4X / SSC 2X, puis de deux minutes chacun, éthanol à 70 % glacé / éthanol à 80 % ; et enfin bain éthanol à 100 % une nuit à 4°C), puis une à trois fois et réhybridées en présence d'autres BACs.

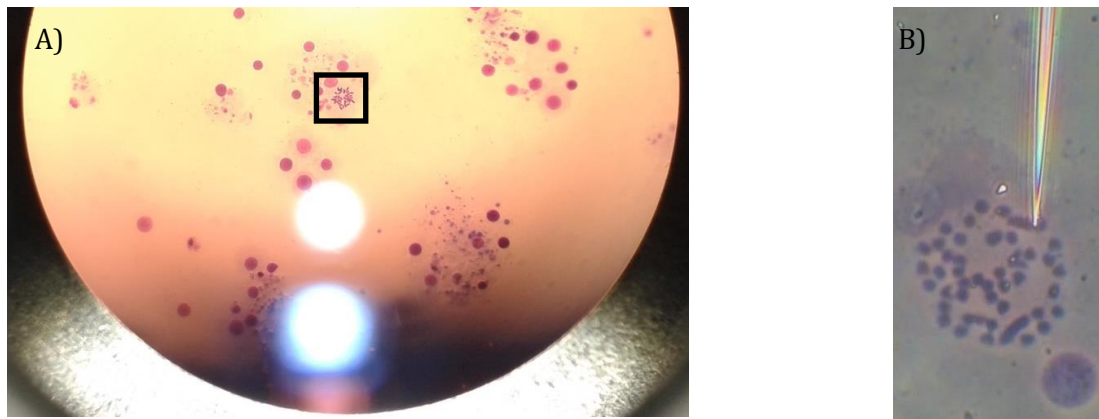
## **Préparations des sondes**

### Microdissection et élaboration de la sonde de peinture

Pour la microdissection, le contrôle de la qualité des étalements a été effectué au microscope (Zeiss, axioplan) en contraste de phase, afin d'évaluer la quantité de métaphases correctement étalées. Les lames sélectionnées ont été colorées au Giemsa 4 % durant 12mn.

### Constitution des lots de paires de chromosomes homologues microdisséqués

La microdissection de la plus grande paire de *T. pennellii* a été réalisée à l'aide d'un microscope inversé (Axiovert 25, Zeiss) équipé d'un « joystick » : micromanipulateur (Narishige). L'isolement des chromosomes a été obtenu à l'aide d'aiguilles de verre fabriquées au moyen d'un étireur (Narishige). Un exemple illustrant une métaphase sous l'objectif du microscope inversé et une microdissection à l'aide de l'aiguille en verre est présenté Figure II.6.A et II.6.B. Des protocoles publiés nous ont guidés dans notre démarche (Seifertova et al. 2013; Laudicina and Mühlmann 2015; Vicari et al. 2015). Une quantité minimum d'ADN est requise pour l'amplification par WGA4 du matériel microdisséqué. Cette amplification (kit WGA4 « single cell », Sigma Aldrich) est en effet optimisée pour une cellule humaine unique (3,5 pg). Nous avons déterminé par calcul un nombre minimum de 30 chromosomes homologues de cette plus grande paire de chromosomes de *T. pennellii* à microdisséquer pour la constitution des lots en prenant pour référence de taille de génome 1,15 pg (valeur C).



**Figure II.6: Observation au microscope inversé de deux préparations chromosomiques pour la microdissection.** A) observation au microscope inversé de l'étalement chromosomique de *T. pennellii* coloré au GIEMSA au grossissement 400. Métaphase encadrée en noir. B) observation d'une métaphase de *O. niloticus* au microscope inversé au grossissement 1 000. Microaiguille de verre placée sous l'objectif à proximité de la plus grande paire de chromosomes.

#### - Amplification, purification et quantification de ces produits

Les amplifications par « Whole Genome Amplification » (WGA) des lots de chromosomes microdisséqués ont été effectuées en utilisant les protocoles décrits précédemment dans la littérature (Gribble et al. 2004; Arneson et al. 2008; Hockner et al. 2009). Les kits GenomePlex de WGA4 et WGA3 (Sigma-Aldrich) ont été utilisés en suivant le protocole du fournisseur. Le kit WGA4 a permis d'amplifier une quantité d'ADN initiale estimée entre 4 et 6 pg selon les lots (Tableau II.2). L'amplification a été vérifiée par dépôt sur gel d'agarose 1,5 %, et l'amplifiat purifié par précipitation à l'éthanol. 10 ng de produit WGA4 purifié ont été réamplifié avec le kit WGA3 (quantité minimum de matériel requis de 1µg pour le marquage).

A chaque étape de l'élaboration de la sonde de peinture, les échantillons et les témoins ont été dosés précisément en utilisant le kit Qubit 3.0 Fluorometer HS et BR (Invitrogen Corporation).

#### Multiplication des BACs

Nous avons disposé d'un total de 792 BACs issus de la banque génomique VMRC-19 de *N. coriiceps* (Detrich et al. 2010; Detrich and Amemiya 2010) ayant servi au séquençage du génome complet (Shin et al. 2014).

Afin de disposer d'une quantité suffisante d'ADN génomique nécessaire à l'élaboration de sonde pour la BAC-FISH, nous avons préparé quarante de ces BACs ; quatre étaient ciblés et intéressaient nos collaborateurs et les autres (trente-six) étaient tirés aléatoirement dans les deux plaques 384 puits fournies. Dans un premier temps, 5 µl de matériel glycérolé ont été prélevés, puis mis en culture sous agitation à 200 rpm durant 5h à 37°C, dans 5 ml de milieu Lysogeny Broth (LB) enrichi de chloramphénicol (12,5 µg/ml). Les cultures ont dans un deuxième temps été transvasées dans 95 ml de milieu enrichi et mises sous agitation à 200 rpm, 18h à 37°C. Après vérification des multiplications bactériennes par mesures de densité optique (DO), les BACs ont été récupérés et purifiés par l'utilisation du kit Plasmid Midi prep (QUIAGEN). L'ADN récolté et précipité à l'éthanol, puis resuspendu dans 30 µl d'eau stérile bi-distillée et enfin dosé au fluoromètre Qubit (quantification Broad Range, Thermo Fisher Scientific).

### Marquages des sondes

**Les sondes de peinture** ont été marquées à la biotine-16-dUTP par random priming (kit Biotin-High-Prime Roche Diagnostics) selon le protocole du fournisseur à partir d'une quantité de matériel initiale de 1 µg. La réaction de marquage est stoppée par ajout 2 µl d'EDTA, 2M à pH 8 et chauffage 10 mins à 65°C.

**Pour les sondes de BACs**, 1 µg d'ADN de chaque BAC préparé a été dilué dans 16 µl d'eau stérile RNase free, puis chauffé 25 mins à 95°C avant d'être placé dans la glace. Pour la détection indirecte, 4 µl de mix contenant de la biotine-16-dUTP ou de la digoxigénine-11-dUTP (kits Biotin ou DIG-High Prime, Sigma Aldrich) ont été ajoutés à l'ADN selon les recommandations du fournisseur (20h à 37°C). La réaction de marquage est stoppée par ajout 2 µl d'EDTA, 2M à pH 8 et chauffage 10 mins à 65°C. Pour la détection indirecte, 2 µl de tampon de marquage et 2 µl de fluorochrome : fluorescéine-ULS ou rhodamine-ULS ou Dyomics415-ULS-dGTP (kits *Platinum Bright™*, ULS 495-Green, 550 Red/Orange et 415 Blue, Nucleic Acid Labeling Kit, Leica Biosystems) ont été ajoutés à l'ADN selon les recommandations du fournisseur (30 mins à 85°C). Dans ce dernier cas, les fluorochromes étant déjà incorporés dans les nucléotides servant au marquage (liés en position N7 à la guanine), les précautions nécessaires ont été prises tout au long de la procédure pour protéger les échantillons de la lumière, et éviter ainsi le déclin prématuré de l'intensité de fluorescence des sondes.

Pour les deux types de marquages, les sondes ont été précipitées à l'éthanol avant d'être resuspendues dans 60 µl de tampon d'hybridation (50 % formamide / 2x SSC / 10 % dextran sulfate) pour la double BAC-FISH et la peinture (cf = 16 ng/µl) ou dans 40 µl de ce tampon d'hybridation pour la triple BAC-FISH (cf = 25 ng/µl).

Le passage de la détection indirecte à la détection directe des BACs a permis non seulement d'améliorer significativement la qualité des signaux (intensité de fluorescence des BACs augmentée et bruit de fond diminué nettement), rendant plus confortable et rapide l'étape d'observation ; mais aussi d'optimiser l'identification des bras chromosomiques en hybridant les BACs par combinaisons de trois et non plus de deux (troisième marquage en utilisant le fluorochrome Dyomics415-ULS qui s'ajoute à ceux par la Fluorescéine et la Rhodamine).

### FISH

Les sondes marquées ont été dénaturées par chauffage à 85°C durant 5 mins en présence d'ADN compétiteur. Elles ont ensuite été préhybridées 2h à 37°C en présence d'ADN compétiteurs avant d'être appliquées sur les étalements chromosomiques. Ces étalements ont préalablement été dénaturés à 72°C durant 10 sec à 1 min dans une solution de dénaturation (70 % formamide / 2x SSC). Les temps de dénaturation sont déterminés en fonction de l'âge des étalements, de la méthode de préparation des chromosomes (culture cellulaire ou méthode directe) ou encore du nombre d'hybridations successives d'une même lame. Les lames hybridées sont ensuite laissées en chambre humide 48h pour une hybridation optimale des sondes sur leurs cibles en présence d'ADN compétiteurs. Les différents lavages des lames, ainsi que le couplage des sondes hybridées à l'avidine-FITC (Fluoresceine, Roche) ou à la Rhodamine anti-DIG (Rhodamine, Roche) dans le cas de la détection indirecte ont été effectués selon Bonillo *et al.* (2015).

### Elaboration d'ADN compétiteur

Afin de saturer les séquences hautement et moyennement répétées présentes dans la sonde de peinture de la grande paire de *T. pennellii* ou dans les inserts des BACs, de l'ADN génomique compétiteur a été préparé à partir de tissu musculaire respectivement de *Trematomus hansonii* ou de *N. coriiceps*. Pour les BACs, l'ADN du cerf *Cervus elaphus* a été choisi pour saturer les sites aspécifiques. Les protocoles qui ont été suivis sont ceux d'une extraction d'ADN : lyse tissulaire avec Cetyl triméthylammonium bromide (CTAB) et protéinase K, récupération d'ADN après traitement au chloroforme Alcool Isoamylique (CIA) et précipitation à l'isopropanol, rinçage à l'éthanol 70 %. L'ADN est fragmenté par autoclavage et choc thermique, précipité et resuspendu dans du tampon d'hybridation (50 % formamide / 2x SSC / 10 % dextran sulfate) à une concentration finale de 8 µg/µl pour les extraits de *T. hansonii* et *N. coriiceps*, et de 10 µg/µl pour les extraits de *Cervus elaphus*.

## Observation, capture et traitement des images

Les lames ont été analysées au microscope (Axioplan, Zeiss) équipé d'une caméra refroidie N&B (Coolsnap Photometrics) et d'une lampe à épifluorescence (XCite), couplé à un logiciel de caryotypage / *FISH*-imaging (CytoVision 3.93.2 / Genus). Un minimum de trente métaphases pour la peinture et de cinq métaphases pour la BAC-*FISH* observées par échantillon a été retenu pour assurer la répétabilité des résultats. Dans la mesure du possible (chromosomes suffisamment étalés présentant peu de chevauchements), les chromosomes d'une métaphase ont été ordonnés selon les critères suivants : taille décroissante, ressemblance morphologique des paires homologues et contraste DAPI. C'est seulement après classement des chromosomes que les signaux de peinture ont été superposés sur l'image de synthèse.

### **II.2.C : Recherche d'HCI par peinture de la grande paire de *T. pennellii* microdisséquée**

#### ***II.2.C.1 : Résultats***

#### **MICRODISSECTION**

Deux lots de trente (n°1 et n°2) et un lot de quarante chromosomes (n°3) ont été récoltés par microdissection chromosomique (Tableau II.2).

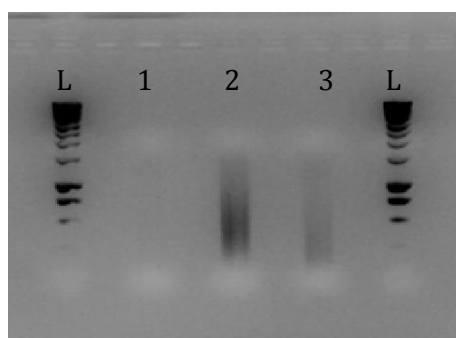
Le lot n°1 a été directement amplifié avec le kit WGA4 sans usage de la protéinase K dans la première étape de lyse et fragmentation du matériel génétique. N'ayant pas été suffisamment amplifié en WGA4, aucune amplification n'a été détectée après dépôt sur gel d'agarose 1,5 %. La quantification au Qubit confirme une faible concentration du lot numéro 1 amplifié en WGA4, inférieure à 4 ng/μl, donc en-dessous de la limite de détection sur gel (Tableau II.2).

Les lots n° 2 et 3 ont été amplifiés en suivant strictement le protocole du fournisseur du kit WGA4. On note une amplification de respectivement neuf et près de seize fois plus importante que le lot numéro 1. Ces deux lots ont donc fait l'objet d'une réamplification en WGA3. Après différents tests, nous avons estimé que les rendements d'amplifications étaient significativement plus importants lorsque les lots étaient systématiquement purifiés après amplification (élimination des tampons, excès d'enzymes, d'adaptateurs issus des kits WGA). La précipitation à l'éthanol a été retenue car, contrairement à d'autres systèmes de purification (colonnes), elle permet de choisir le volume de resuspension du matériel et donc la concentration finale avant réamplification.

| Lots           | Nb chr | Quantité initiale estimée (pg) | Concentration après WGA4p (ng/μl) | Taux d'amplification après WGA4 | Concentration après WGA3p (ng/μl) | Rendement purif (%) |
|----------------|--------|--------------------------------|-----------------------------------|---------------------------------|-----------------------------------|---------------------|
| lot n°1        | 30     | 4.3                            | 3,82                              | 50 000                          | X                                 | X                   |
| témoin positif | X      | 10 000                         | 38,6                              | 270                             | X                                 | X                   |
| lot n°2        | 30     | 4.3                            | 15,70                             | 450 000                         | 75,30                             | 53,8                |
| témoin positif | X      | 10                             | 21,50                             | 370 000                         | 151,00                            | 73,7                |
| lot n°3        | 40     | 5.7                            | 55                                | 783 333                         | 103,5                             | 26,8                |
| témoin positif | X      | 10                             | 153                               | 540 000                         | 170                               | 58,7                |

**Tableau II.2 : quantifications des différentes étapes après amplifications et purifications pour l'obtention de la sonde de peinture.** Le témoin positif est un échantillon d'ADN humain fourni dans le kit WGA4 (Sigma-Aldrich).

Le lot numéro 3 a donné le meilleur résultat concernant la quantité finale de matériel après WGA4 puis WGA3. De plus, aucune contamination n'a été détectée (cf témoin négatif, Figure II.7). Il a donc été retenu pour la suite du processus. Les fragments d'ADN génomique constituant la sonde de peinture sont compris entre 100 pb et 1 kb avec une plus forte concentration de séquences situées entre 200 et 400 pb (Figure II.7), taille optimale pour l'accès des fragments de la sonde à leur cible lors de la *FISH* (Vicari et al. 2015).

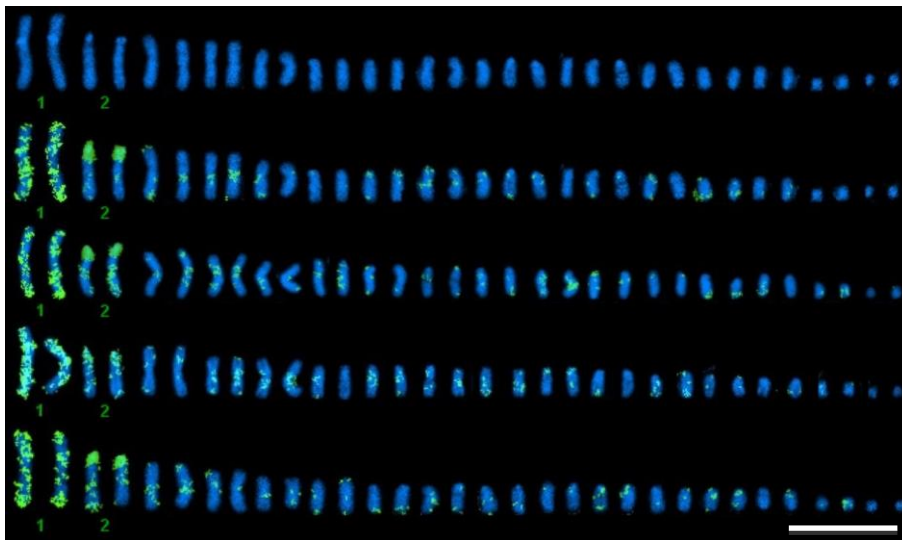


**Figure II.7 : résultats de l'amplification en WGA4 du lot n°3 par dépôt sur gel d'agarose 1,5 %.** Piste 1 : témoin négatif, piste 2 : témoin positif/ADN génomique humain fournit par le kit WGA4 (Sigma-Aldrich) (quantité initiale estimée = 10 pg / quantité finale dosée = 5,4 μg), piste 3 : produit microdisséqué et amplifié contenant 40 chromosomes de la plus grande paire de *T. pennellii* (quantité initiale estimée = 8 pg / quantité finale dosée = 4,7 μg). L : Marqueurs de poids moléculaires BenchTop 100 pb (Proméga).

## PEINTURE

*Hybridations de la sonde de peinture chromosome-spécifique sur les chromosomes du spécimen dont elle est issue, T. pennellii (CE 4313) (témoin positif) :*

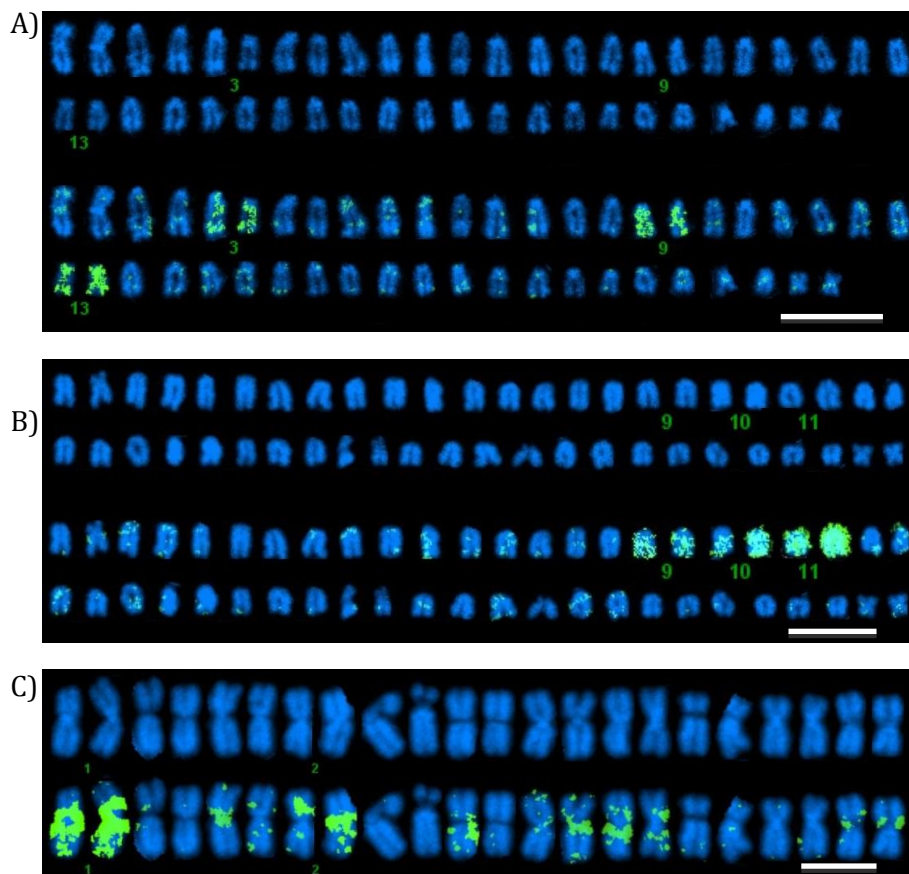
La plus grande paire de chromosomes de *T. pennellii* est peinte sur toute la longueur de manière relativement homogène (paire numérotée 1, Figure II.8). Ce signal chromosome-spécifique est retrouvé dans l'ensemble des métaphases observées. D'autres signaux plus ponctuels et minoritaires et non spécifiques de la plus grande paire sont également observés sur les autres paires chromosomiques. Ces signaux annexes sont variables d'une métaphase à une autre, mais un spot très intense est retrouvé de manière systématique au niveau de la région péracentromérique de la plus grande paire de chromosomes acrocentriques de *T. pennellii* (numérotée paire 2, Figure II.8).



**Figure II.8 : Hybridation *in situ* en Fluorescence de la sonde de peinture microdisséquée sur les chromosomes de *T. pennellii*.** Présentation d'un classement des chromosomes de *T. pennellii* contre-colorés au DAPI suivi des quatre métaphases avec chromosomes ordonnés auxquels est ajouté le signal de la peinture. La sonde de peinture est issue de la microdissection de la plus grande paire en taille de *T. pennellii* (numérotée paire 1). Elle a été amplifiée par la méthode de WGA et marquée indirectement en biotine, puis détectée au moyen de l'avidine-FITC (fluorescéine, signal vert). Sont numérotées les paires peintes ou aisément identifiables mentionnées dans le texte. Barre d'échelle : 10  $\mu$ m.

*Hybridations de la sonde de peinture chromosome-spécifique sur les chromosomes d'autres espèces de Nototheniidae : T. hansonii, N. coriiceps et D. mawsonii.*

Lorsque la sonde de peinture est hybridée sur les chromosomes de *T. hansonii* ( $2n = 46$ ) et *D. mawsonii* ( $2n = 48$ ), trois paires de chromosomes acrocentriques équivalant à trois bras chromosomiques sont peintes (numérotées 3, 9 et 13 Figure II.9.A, et 9, 10 et 11 Figure II.9.B). Ce signal est reproductible et intense sur l'intégralité des paires qu'il recouvre. De manière similaire à l'hybridation sur les chromosomes de *T. pennellii*, on retrouve des signaux annexes présents de manière éparpillée sur d'autres paires de chromosomes chez ces deux espèces.



**Figure II.9 : Hybridation *in situ* en Fluorescence de la sonde de peinture microdisséquée sur les chromosomes de A) *T. hansonii*, B) *D. mawsonii* et C) *N. coriiceps*.** Pour chacune de figure, présentation d'un classement des chromosomes contre-colorés au DAPI suivi de la même image sur laquelle figurent les signaux de peinture. La sonde de peinture est issue de la microdissection de la paire de chromosomes -numérotée paire 1 de *T. pennellii* (Figure II.8). Elle a été amplifiée par méthode de WGA et marquée indirectement en biotine, puis détectée au moyen de l'avidine-FITC (fluorescéine, signal vert sur la figure). Sont numérotées les paires peintes mentionnées dans le texte. Barre d'échelle : 10  $\mu$ m.



Les signaux de la sonde de peinture hybridée sur les chromosomes de *N. coriiceps* sont plus dispersés. Sept des onze paires de chromosomes présentent un signal intense sur un segment au moins. Néanmoins, ce signal couvre une grande partie de la plus grande paire facilement identifiable par la contre-coloration au DAPI (numérotée paire 1, Figure II.9.C) ainsi qu'une autre paire de plus petite taille (numérotée paire 2, Figure II.9.C). Cependant, le signal ne couvre pas les paires dans leur intégralité, des régions télomériques et subtélomériques ne sont pas marquées. Des signaux annexes assez intenses couvrent également plusieurs autres paires. Ils se concentrent dans des régions hétérochromatiques péricentromériques (et dans une moindre mesure dans des régions télomériques) qui sont riches en ADN satellites.

### ***II.2.C.2 : Discussion :***

#### **Aspects techniques mis au point et perspectives d'amélioration du signal de peinture**

La peinture préparée à partir de la microdissection de la plus grande paire de *T. pennellii* a bien produit des signaux spécifiques de cette plus grande paire, mais aussi des signaux annexes détectés sur les autres paires chromosomiques. Même si les métaphases étaient en nombre suffisant sur les préparations chromosomiques et les chromosomes relativement bien étalés au sein d'une métaphase pour permettre leurs prélèvements, il existe toujours des risques de contamination intrinsèque des lots en prélevant des fragments d'autres chromosomes avec l'aiguille (Laudicina and Mühlmann 2015; Vicari et al. 2015). Ce risque est accentué par les multiples prélèvements (30) du chromosome cible sur les différentes métaphases pour constituer les lots (Weimer et al. 1999). De plus, ce risque est d'autant plus marqué que la présence de résidus de cytoplasme agit comme une matrice liant entre eux les chromosomes, ce qui peut entraîner le décrochage synchrone de plusieurs chromosomes différents .

Une des solutions pour minimiser ce risque et maximiser le nombre d'HCI détectables potentielles consiste à préparer des sondes de peinture à partir de la microdissection d'un petit nombre (<5) ou bien d'un unique chromosome (Guan et al. 1993; Viersbach et al. 1994; Christian et al. 1999; Griffin et al. 1999; Trifonov et al. 2002). Nous avons réalisé avec succès une amplification de sonde préparée à partir de 15 chromosomes de la plus grande paire de *T. pennellii* ou de *T. hansonii*. Ce résultat montre qu'il est a priori possible de réduire le nombre de chromosomes pour la constitution de peintures, et donc potentiellement d'éliminer une partie de la contamination intrinsèque. Dans tous ces cas, y compris celui d'une préparation de peinture à partir d'un seul chromosome, il est toutefois nécessaire de pouvoir distinguer clairement ce chromosome dans un souci de répétabilité de l'expérience. Une taille suffisante des chromosomes, et surtout, l'identification fiable de ces paires est alors strictement requise.

Or, si pour *T. pennellii* et *N. coriiceps*, certaines paires sont identifiables sur la base de la taille, de la morphologie et/ou du marquage DAPI (Figure 3 et 4c), il n'est pas possible de reconnaître la plupart des paires de manière non équivoque (surtout les petites paires acrocentriques présentes chez *T. pennellii*, *T. hansonii* et *D. mawsonii*, Figures 4A et 4B). Leur similarité en taille et en forme limite également leur isolement par électrophorèse en champ pulsé comme cela est fait pour d'autres espèces. Cela limite donc le nombre de paires candidates possibles pour la microdissection chromosomique et donc pour la détection d'HCI par peinture dans cette famille.

La technique de « Whole Genome Amplification 4 » employée, de type « single cell » repose sur l'amplification aspécifique de produit récolté au moyen d'hexamères dégénérés jouant le rôle d'amorces (Vicari et al. 2015). Elles sont optimisées pour l'amplification de l'équivalent du contenu ADN d'une cellule humaine unique (7 pg) (Arneson et al. 2008; Hockner et al. 2009). C'est donc une méthode particulièrement adaptée à l'amplification de lots de chromosomes microdisséqués (estimés à 4 pg pour les lots numéros 1 et 2 et à 6 pg pour le lot numéro 3). Cependant, les risques de contamination des échantillons microdisséqués par de l'ADN exogène sont très importants car l'étape d'amplification WGA est extrêmement sensible (Laudicina and Mühlmann 2015). Des contaminations imputables à l'utilisation du WGA des lots numéros 1 et 2 ont été détectées lors de l'élaboration de la sonde de peinture : des témoins négatifs constitués d'eau stérile amplifiés avec le kit WGA4 produisent une faible amplification lors des dépôts sur gel d'agarose et la détection d'une quantité résiduelle d'ADN contaminant lors des quantifications. Après avoir pris les précautions nécessaires pour le lot numéro 3 (stérilisation des outils par exposition aux UVs, pas de confinement sous hotte lors de l'incorporation des produits, ouverture des tubes en dehors de la zone de manipulation...), nous n'avons plus détecté de contamination lors des dépôts sur gel. Cependant, le séquençage de certains de ces produits a révélé une haute proportion de séquences contaminantes, principalement d'origine humaine. Etant donné la distance génétique entre l'homme et les téléostéens antarctiques, ces séquences n'ont pas pu empêcher ou biaiser l'hybridation de la peinture préparée sur les chromosomes de Nototheniidae. Par contre, la quantité réelle de sonde Nototheniidae-spécifique a été largement sur-estimée lors des dosages des produits d'amplifications.

Malgré la présence d'ADN compétiteur lors de l'hybridation, le signal produit par les *FISH* de peinture ne se révèle pas uniquement chromosome-spécifique. Des signaux épars sont en effet présents sur la quasi-totalité des autres paires chromosomiques. Si ce signal annexe est plutôt réduit chez *T. hansonii* et *D. mawsonii*, il est plus marqué chez notre témoin positif *T. pennellii* et chez *N. coriiceps*. Il est alors difficile d'établir clairement les HCI car le signal chromosome-spécifique (correspondant à une majorité d'hybridations des séquences uniques présentes dans la sonde) et les signaux répétés annexes se confondent. Ces derniers correspondent très

probablement aux séquences hautement et moyennement répétées présentes en quantité plus ou moins importantes dans la sonde et dans tous les chromosomes : satellites centromériques, séquences télomériques et certaines familles d'éléments transposables (Auvinet et al. 2018).

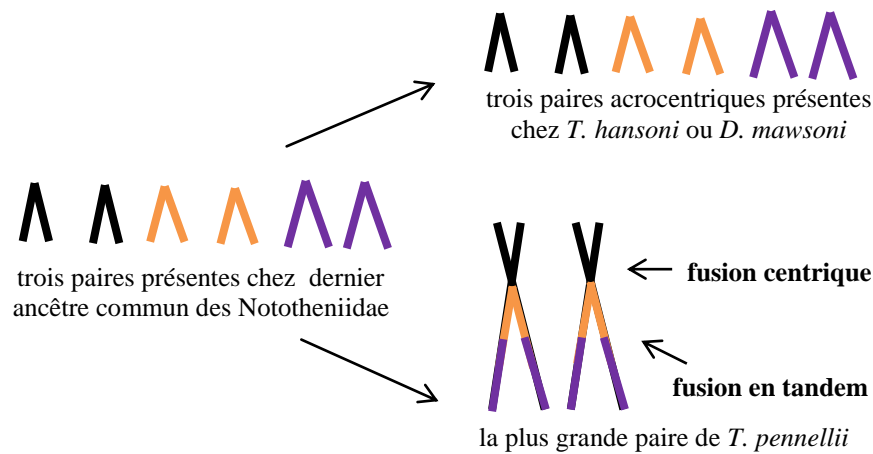
En dépit de ces signaux présents sur les autres paires de chromosomes, on note une couverture homogène sur toute la longueur de la plus grande paire de *T. pennellii*. Si la distribution des signaux annexes est plutôt fluctuante, le signal de peinture est bien reproductible (observé sur plus de 70 métaphases différentes). Les signaux de peinture retrouvés sur les autres paires chromosomiques peuvent être imputables à la méthode de fixation des chromosomes lors de leur préparation (dépuration partielle probable de l'ADN des chromosomes microdisséqués par l'acide acétique contenu dans le fixateur, et donc hybridation de sonde dégradée et moins spécifique (Vicari et al. 2015). Ces signaux aspécifiques peuvent également être dus à un biais d'amplification en faveur des séquences répétées présentes en plus grand nombre de copies par rapport aux séquences uniques. Ce produit microdisséqué est en effet une très grande paire de chromosome riche en hétérochromatine, notamment au niveau de deux constriction (une centromérique et une présente en position médiane dans le grand bras), et donc plus sujette à contenir des accumulations de séquences répétées (Pimpinelli et al. 1995; Biémont and Vieira 2005; Biémont and Vieira 2006).

L'approche par peinture de sondes microdisséquées est intéressante pour identifier ces homologies chromosomiques interspécifiques au sein des Nototheniidae, notamment pour certaines paires facilement identifiables. Cependant, l'utilisation de ces approches nécessiterait des améliorations techniques pour rendre le signal plus facile à interpréter. On peut proposer la préparation des chromosomes avec la dernière fixation dans l'éthanol pur (sans acide acétique) pour éviter la dépuration et donc la dégradation des futures sondes. L'élaboration d'ADN Cot-1 (riche en séquences répétées) pourrait jouer le rôle d'un compétiteur « maison » (Trifonov et al. 2002; Vicari et al. 2011; Trifonov et al. 2017). L'amélioration et l'optimisation des étapes d'amplifications et de réamplifications est à envisager, notamment en réduisant autant que possible le nombre de chromosomes dans chaque lot et en prenant des précautions pour les contaminations d'origine humaine. Enfin, l'emploi d'un marquage avec détection directe des sondes de peinture pourrait nettement augmenter la résolution et la puissance du signal. En plus, l'utilisation de plusieurs peintures différentes (« multicolore », Ried *et al.* (1998), Weimer *et al.* (2001)) et leur cohybridation permettraient de distinguer plus clairement le signal chromosome-spécifique (une seule fluorescence) du signal répété (superposition des fluorescences).

### **La plus grande paire chromosomique de *T. pennellii*, produit de deux fusions chromosomiques (Robertsonienne = centrique et en tandem)**

Chez les Nototheniidae, l'utilisation des sondes de peinture devrait permettre de rechercher des homologies chromosomiques interspécifiques afin de reconstituer les nombreux remaniements survenus au cours de leur diversification (Pisano and Ozouf-Costaz 2000). Avec des signaux d'hybridation abondamment détectés sur les chromosomes de *T. hansonii*, *D. mawsonii* et *N. coriiceps*, les résultats de peinture obtenus chez les différentes espèces de cette famille laissent supposer une bonne conservation des séquences présentes dans la plus grande paire de *T. pennellii* entre les différentes espèces examinées.

La plus grande paire de chromosomes de *T. pennellii* semble être homologue à trois paires peintes de chromosomes acrocentriques présentes chez *T. hansonii* et *D. mawsonii*. Les résultats obtenus par peinture suggèrent donc que cette grande paire de chromosomes de *T. pennellii* résulterait d'une double fusion (une centrique et une en tandem), à partir de trois chromosomes hérités du dernier ancêtre commun des Nototheniidae (Figure II.14). Ce résultat est cohérent avec les données caryotypiques de *Trematomus pennellii* de  $2n = 32$ , inférieur au caryotype ancestral inféré ( $2n = 48$ ). Certaines des paires chromosomiques de *T. pennellii* sont en effet le produit de fusions à nombre constant d'unités structurales (telles que définies chez *I. cyanobrancha*) (Figure II.5).



**Figure II.14 : Scénario évolutif des remaniements chromosomiques ayant conduit à la plus grande paire de chromosome de *T. pennellii*.** L'état plésiomorphe inféré est  $2n = 48$  chromosomes acrocentriques présents chez l'ancêtre commun des Nototheniidae. Trois paires (peinte par la sonde de peinture) ont été conservées dans un premier cas chez de nombreuses espèces dont *T. hansonii* et *D. mawsonii*. Elles ont doublement fusionné (fusion centrique et fusion en tandem) dans un deuxième cas pour donner la plus grande paire de *T. pennellii* (sonde de peinture).

De la même manière, les résultats semblent indiquer une homologie au moins partielle de la plus grande paire de chromosomes de *T. pennellii* avec celle de *N. coriiceps*. Ceci suscite de nombreuses questions concernant par exemple l'homologie des fusions chromosomiques mises en évidence chez *N. coriiceps* (Nototheniinae) et *T. pennellii* (Trematominae). Les mêmes bras chromosomiques présents chez l'ancêtre commun de Nototheniidae sont-ils impliqués dans les mêmes fusions comme cela a été proposé au sein de la sous-famille des Nototheniinae (« one-to-one » massive fusions, (Amores et al. 2017)) ? D'autre part, ce résultat soulève le problème du nombre de remaniements qui se sont produits au cours du temps depuis la diversification de ces espèces. La peinture ne permet pas de savoir s'il y a eu un ou plusieurs remaniements intra-chromosomiques après la fusion des paires. D'autres types de réarrangements pourraient être détectés au moyen de cet outil comme les fissions ou les translocations. Une manière d'affiner ces résultats de peinture et d'augmenter le nombre d'HCI identifiées au sein des « *Trematomus* » est de s'intéresser non plus à l'échelle du chromosome dans sa totalité, mais à celle de l'unité structurale.

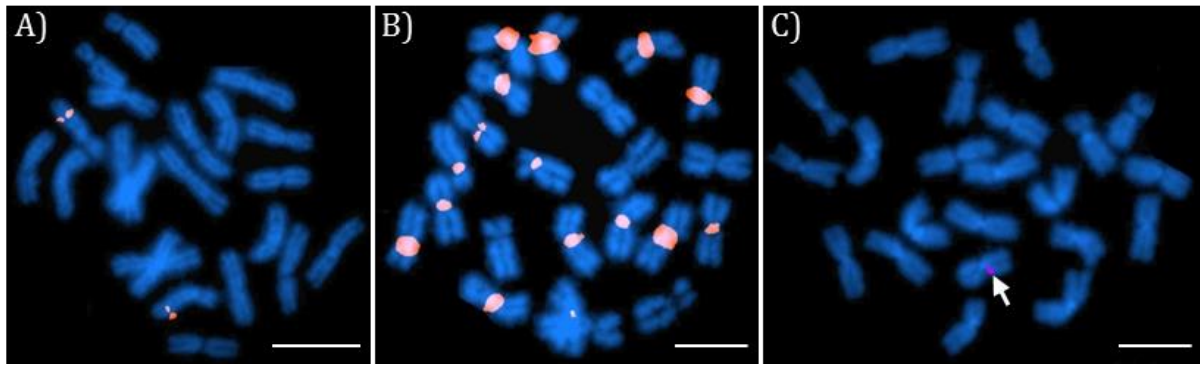
## **II.2.D : Recherche des homologies des unités structurales interspécifiques par BAC-FISH**

### ***II.2.D.1 : Résultats***

#### **Localisation des BACs sur les chromosomes de *N. coriiceps***

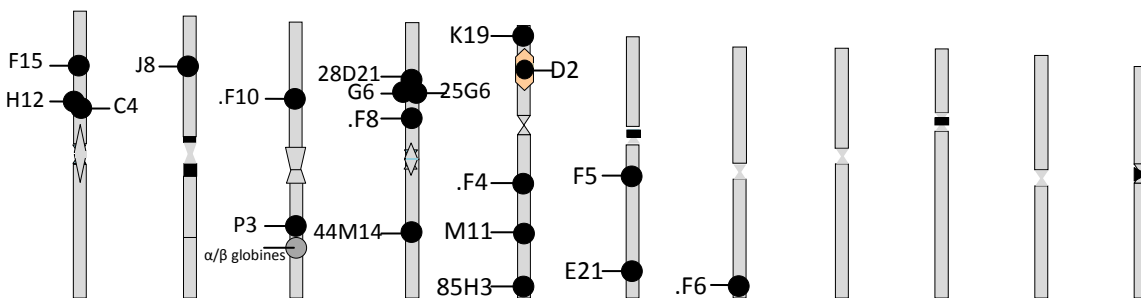
Après préparation des sondes pour la BAC-FISH, nous avons choisi en premier lieu de les hybrider sur l'espèce dont ils sont issus : *N. coriiceps*. Ceci nous a permis d'étudier le contenu de ces différents BACs au moyen des signaux produits lors des FISH et de trier les marqueurs utilisables : produisant un signal clair et unique. De plus, les résultats de localisation chromosomique du premier couple de BACs hybridé : 28D21 - 44M14 (ayant chacun fait l'objet d'un séquençage plus poussé que les autres BACs) sur chacun des bras du quatrième chromosome de *N. coriiceps* a été confirmée lors de cette étape par l'équipe impliquée dans la collaboration.

Trois types de signaux ont été détectés lors des hybridations des quarante BACs préparés, indépendamment des fluorochromes employés pour la détection des sondes. Le premier type de signal observé pour 21 BACs (52,5 %) est un signal unique, net et précis, avec deux doubles spots par paire de chromosomes homologues (Figure II.10.A). Il correspond à l'hybridation d'une majorité de séquences uniques (gènes, introns...) présentes dans la sonde. Le deuxième type de signal, observé pour 7 BACs (17,5 %), est présent sur de multiples chromosomes, souvent dans les régions centromériques ou péricentromériques (Figure II.10.B). Il s'agit probablement de BACs contenant une forte proportion en éléments transposables et/ou en séquences moyennement et hautement répétées. Le troisième et dernier type de signal, observé pour 12 BACs (30 %) produit de très faibles spots dans les noyaux (uniques ou multiples), mais aucun visible clairement et localisable sur les chromosomes métaphasiques (Figure II.10.C). Ce faible signal observé pourrait être dû aux contenus génomiques des BACs eux-mêmes ou bien à un défaut de marquage de ceux-ci.



**Figure II.10** : exemple des signaux observés suite à la localisation des BACs sur les chromosomes de *N. coriiceps* : A) signaux uniques et nets, exemple du BAC M11; B) signaux répétés, exemple du BAC B9; C) signaux trop faibles pour une localisation nette et répétable, exemple du BAC hybridé J2. Barres d'échelle : 10  $\mu$ m. Etant donné le nombre déjà important de BACs utilisables en tant que marqueurs (A), nous avons décidé de ne pas aller plus loin avec ces deux dernières catégories (B et C).

Seuls les 21 BACS produisant un signal clair et précis (correspondants à ce que l'on a considéré être de la séquence unique) ont été conservés pour rechercher des homologues entre unités structurales chez les « *Trematomus* ». Leurs positions relatives sur les bras chromosomiques de *N. coriiceps* ont été réalisées suite à de la double ou triple *FISH* (hybridations combinées) et par leur localisation par rapport à des structures chromosomiques particulières visibles en DAPI (Figure II.11).



**Figure II.11** : positionnement des BACs sur la carte chromosomique haploïde reconstituée de *N. coriiceps* ( $2n = 22$ ). Les chromosomes sont classés selon leur taille décroissante. Certaines paires sont identifiables par des caractéristiques de leur marquage en DAPI indiquées sur la figure. Les bras sont situés de part et d'autre du centromère de chaque paire chromosomique. Les gènes ribosomiques 18 et 28S (colorés en orange) et globines (colorés en gris) ne font pas partie du jeu de BACs fourni. Ils ont été localisés respectivement par Pisano *et al.* (2000) et Pisano *et al.* (2003) (cDNA) et sont indiqués pour permettre d'identifier les paires. Cependant, ils n'ont pas été hybridés lors de cette étude.

Onze bras chromosomiques de *N. coriiceps* ont été marqués par au moins un BAC sélectionné, le maximum étant de quatre BACs par bras (Figure II.11). Au contraire, les bras constituant les

quatre plus petites paires chromosomiques de *N. coriiceps* n'ont fait l'objet d'aucun marquage par les BACs testés au cours de l'étude.

Sous l'hypothèse que deux BACs localisés dans un même bras chez *N. coriiceps* pourraient marquer une même unité structurale chez les « *Trematomus* », nous avons voulu tester cette correspondance. Pour cela, les BACs identifiés par couple ou triplets sur un même bras chromosomique de *N. coriiceps* ont été hybridés sur les chromosomes d'un « *Trematomus* » : *T. newnesi*. Seuls les BACs K19 et D2, positionnés dans le bras chromosomique de *N. coriiceps* portant les gènes ribosomiques 28S n'ont pas fait l'objet d'investigations sur les chromosomes des « *Trematomus* » car les bras homologues portant ces marqueurs sont déjà connus et identifiés.

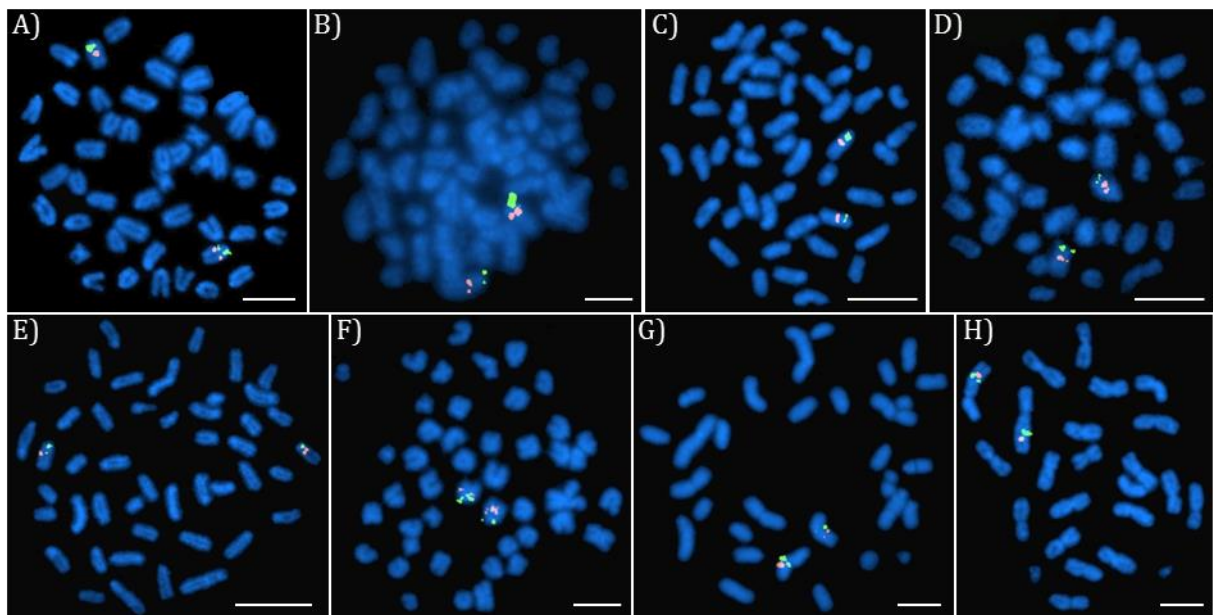
### **Caractérisation des unités structurales chez les « *Trematomus* »**

Les BACs sélectionnés à l'issue de la première étape de tri chez *N. coriiceps* ont hybridé avec succès sur les chromosomes de « *Trematomus* » (Figure II.12). Toutes les combinaisons de BACs ayant marqué un même bras chromosomique de *N. coriiceps* ont été retrouvées dans une même unité structurale chez *T. newnesi*. L'hypothèse de cette correspondance a donc été vérifiée, nous permettant d'identifier ces unités structurales et d'établir leurs correspondances entre espèces de « *Trematomus* ». Au cours de cette étape, nous avons choisi dans un premier temps de caractériser par deux ou trois BACs les unités structurales, de manière à pouvoir suivre leur orientation et leur distance relative entre les espèces (*T. bernacchii*, *T. eulepidotus*, *T. newnesi*, *T. pennellii*, *T. hansonii*, *T. nicolaii*, *T. borchgrevinkii*, *I. cyanobranca*). Compte tenu de cette correspondance systématique entre *N. coriiceps* et *T. newnesi*, nous avons rapidement choisi de définir des combinaisons de BACs directement chez *N. coriiceps*.

De façon globale, la position relative des BACs dans les bras observés chez *N. coriiceps* a été retrouvée sur les unités structurales des espèces de « *Trematomus* » examinées (position centromériques, péricentromériques, milieu du bras, télomérique). Ainsi, les BACs 28D21, C4 et M11 sont toujours retrouvés en position péricentromérique des unités structurales sur lesquels ils ont été localisés, comme 85H3 en position télomérique (Figure II.13). Une exception a été remarquée pour le BAC 44M14 qui semble situé en milieu de bras chez *T. pennellii*, alors qu'il est localisé en position péricentromérique d'un bras chez *T. newnesi*, *T. eulepidotus* et *T. bernacchii* (Figure II.13). Cela montre l'avantage de s'intéresser aux unités structurales plutôt qu'aux bras chromosomiques.

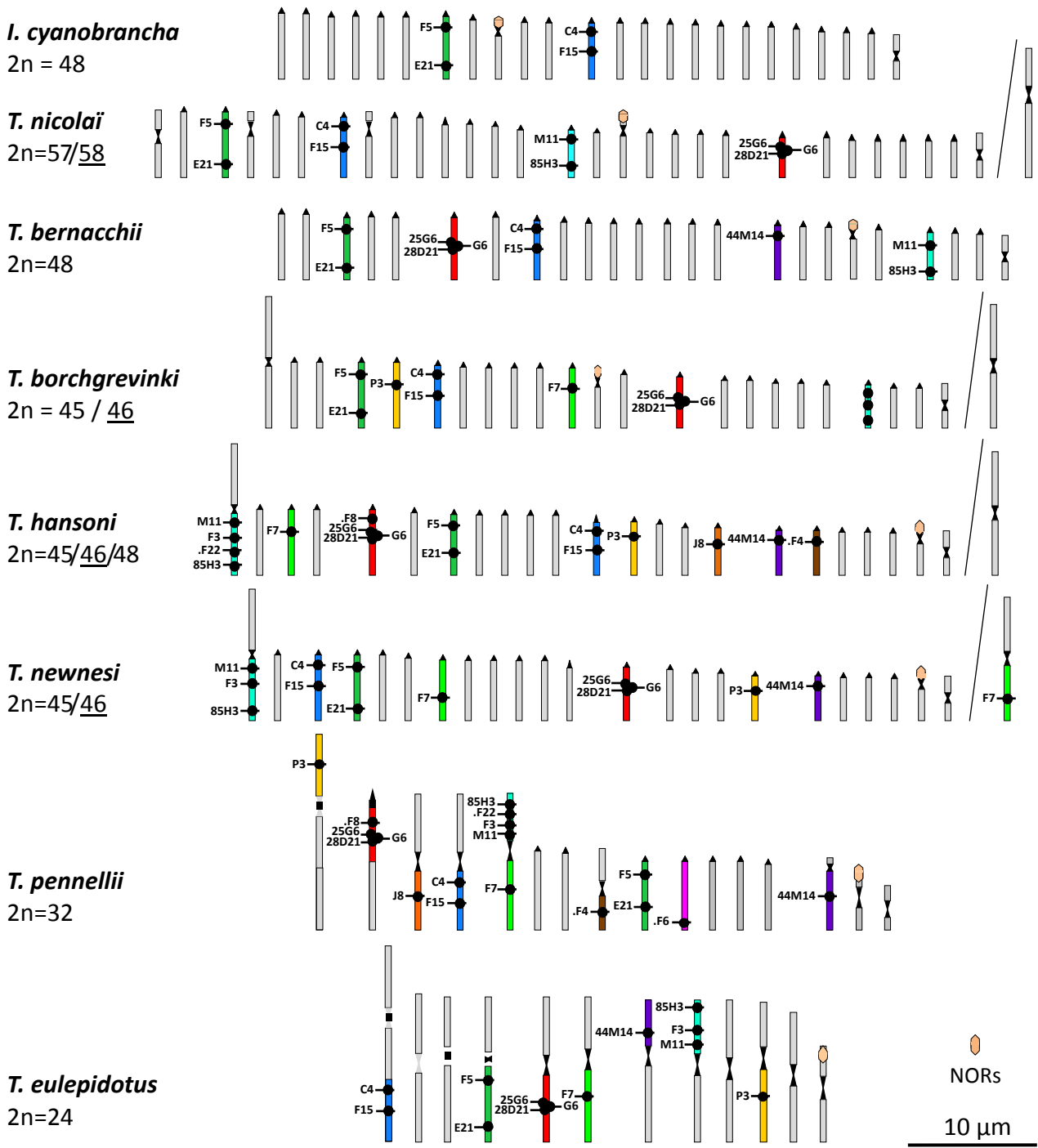


Une synténie systématique semble exister entre les unités structurales des « *Trematomus* ». Cette conservation se caractérise par l'ordre et la distance des BACs sur les unités structurales identifiées chez les différentes espèces. Par exemple, le couple de BACs C4-F15 est retrouvé sur une même unité structurale, et orienté dans le même sens par rapport au centromère chez toutes les espèces examinées: C4 en position péricentromérique et F15 en position médiane, légèrement décalé vers le télomère (Figure II.12 et II.13). De la même façon, on retrouve cette synténie entre *N. coriiceps* et les « *Trematomus* », montrant non seulement la correspondance mais aussi la conservation de l'ordre des marqueurs entre bras chromosomiques chez *N. coriiceps* et les unités structurales chez les « *Trematomus* ». C'est par exemple le cas avec le segment borné par le couple de BACs F5-E21, situés aux deux extrémités de l'unité structurale, avec F5 en position péricentromérique et E21 en position subtélomérique (Figure II.11, II.12 et II.13). Etant donné cette synténie observée, nous avons considéré dans un second temps qu'un seul BAC serait désormais suffisant pour assurer le marquage d'une unité structurale chez les « *Trematomus* ».



**Figure II.12 :** Exemple de double BAC-FISH avec l'hybridation du couple C4 (vert) - F15 (rouge) sur les chromosomes de A) *I. cyanobrancha*, B) *T. nicolai*, C) *T. bernacchii*, D) *T. borchgrevinki*, E) *T. hansonii*, F) *T. newnesi*, G) *T. pennellii*, H) *T. eulepidotus*. Sur chacune de ces doubles hybridations, on retrouve le BAC C4 (détectés directs par la fluoresceine, en vert) en position centromérique et le BAC F15 (détectés directs par la rhodamine, en rouge) en position médiane, localisés sur une même unité structurale. Barres d'échelle : 10  $\mu$ m.

S'il y a conservation des unités structurales entre espèces de « *Trematomus* » (et plus largement chez d'autres espèces de Nototheniidae), l'arrangement entre ces unités structurales pour former les grandes paires chromosomiques acrocentriques, métacentriques ou submétacentriques n'est pas le même suivant les différentes espèces (Figure II.13).



**Figure II.13: positionnements des BACs validés et sélectionnés sur les cartes chromosomiques haploïdes de huit « *Trematomus* ».** Les chromosomes sont classés selon leur taille décroissante. Certaines paires sont identifiables par certaines caractéristiques de leur marquage en DAPI indiquées sur la figure. Pour *T. nicolai*, *T. borchgrevinki*, et *T. newnesi* ayant des chromosomes sexuels multiples différenciés (de type  $X_1X_1X_2X_2/X_1YX_2$  (femelle/mâle)), le chromosome sexuel mâle Y est représenté à droite du caryotype entier (de type femelle). Les positionnements des gènes ribosomiques 18S et 28S (NORs) proviennent de la littérature (Tomaszkiewicz et al. 2011; Ghigliotti et al. 2013). Les noms de chaque BAC hybridé sont indiqués à côté de chaque unité structurale où il a été localisé.

C'est ce que l'on observe à l'échelle du chromosome entier entre *N. coriiceps* et les « *Trematomus* ». Par exemple, les BACs 28D21 et 44M14 qui hybrident sur deux bras différents d'une même paire chromosomique métacentrique chez *N. coriiceps* (Figures II.11) sont retrouvés sur deux unités structurales différentes de deux chromosomes distincts chez *T. eulepidotus*, *T. pennellii*, *T. newnesi* et *T. bernacchii* (Figure II.13). Cette absence d'homologie entre grands chromosomes méta ou submétacentriques est également retrouvée au sein même des « *Trematomus* ». Par exemple, les BACs F3 et F7, identifiés sur les deux unités structurales composant la deuxième plus grande paire chromosomique submétacentrique chez *T. pennellii* sont bien localisés sur deux unités structurales différentes chez *T. eulepidotus*, mais ils ne sont pas retrouvés sur une même paire chromosomique (Figure II.13).

De la même manière, le couple C4 - F15 est localisé dans la plus grande paire chromosomique métacentrique de *N. coriiceps* (Figure II.11), alors qu'il est retrouvé dans une unité structurale contenue dans une paire submétacentrique (4<sup>e</sup> en taille) de *T. pennellii*; dans une unité structurale située à l'extrémité distale de la plus grande paire submétacentrique de *T. eulepidotus*; et dans des unités structurales constituant des paires acrocentriques chez *I. cyanobranca*, *T. borchgrevinki*, *T. nicolai*, *T. newnesi*, *T. hansonii* et *T. bernacchii* (Figure II.12 et II.13).

On note également une différence de localisation chromosomique de l'unité structurale délimitée par le couple M11 - 85H3, retrouvé sur une unité structurale contenue dans la plus grande paire métacentrique chez *T. newnesi* et *T. hansonii*, mais dans une unité structurale constituant une paire acrocentrique chez *T. borchgrevinki*, dont le caryotype ( $2n = 45/46$ ,  $NF = 52$ ) semble pourtant similaire aux deux espèces citées précédemment (Figure II.13) (Morescalchi et al. 1992).

Enfin, le triplet 28D21-25G6-G6 a été localisé dans une unité structurale située à l'extrémité proximale de la plus grande paire acrocentrique de *T. pennellii*, dans une unité structurale d'une moyenne paire métacentrique de *T. eulepidotus*, et dans des unités structurales constituant des petites paires acrocentriques chez les autres « *Trematomus* » (Figures II.13).

Les mêmes combinaisons de BACs sont donc parfois localisées sur des unités structurales constituant à elles seules des petits chromosomes acrocentriques (*T. nicolai*, *I. cyanobranca*, *T. bernacchii*), ou bien des unités structurales associées à d'autres dans des chromosomes de plus grande taille, à la fois acrocentriques (*T. pennellii*), ou métacentriques (*T. pennellii*, *T. eulepidotus*, *T. hansonii* et *T. newnesi*), ou submétacentriques (*T. pennellii*, *T. eulepidotus*) (Figure II.13). Ces dernières ont probablement fait l'objet d'un ou plusieurs remaniements.

## ***II.2.D.2 : Discussion***

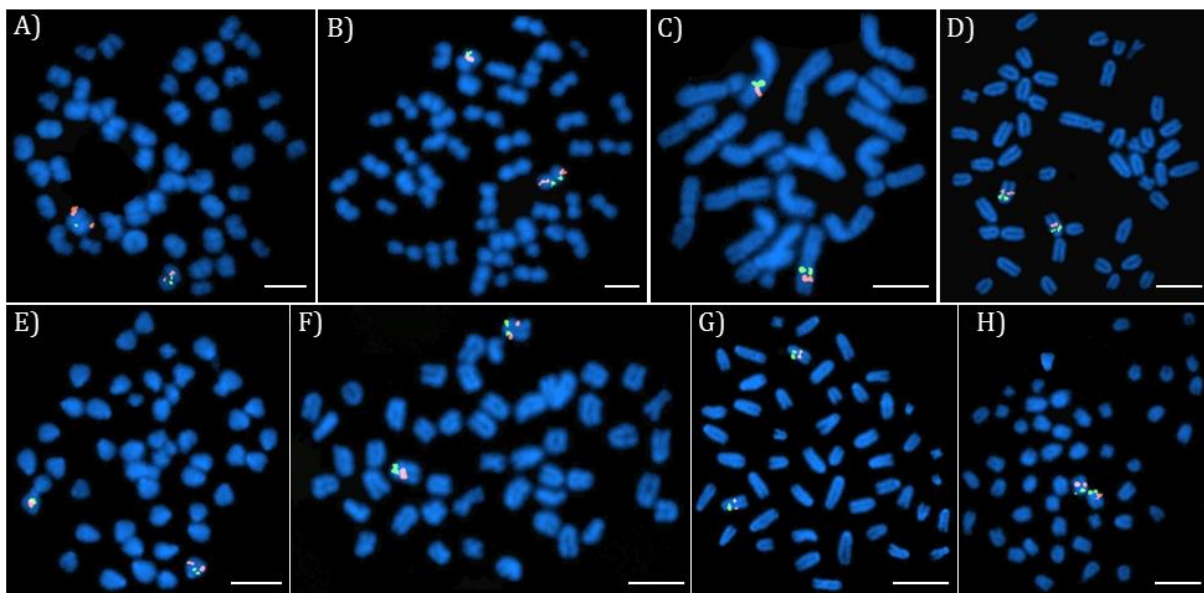
### **Conservation interspécifique des séquences d'ADNg présentes dans les BACs de *N. coriiceps***

Nous avons observé une proportion beaucoup plus importante des BACs localisés sur les plus grandes paires de chromosomes de *N. coriiceps*, au détriment des quatre plus petites paires chromosomiques. Ces dernières ayant un contenu en ADN moins important, la probabilité d'y identifier un BAC tiré aléatoirement est théoriquement plus faible. D'autre part, les deux plaques reçues dans le cadre de la collaboration contiennent peut être préférentiellement des BACs localisés dans les sept plus grandes paires chromosomiques de *N. coriiceps*. Ce biais de localisation a constitué un avantage dans l'étape d'exploration car il a permis de localiser plus facilement deux marqueurs présents sur un même bras. Par contre, les bras présents dans les plus petites paires n'ont pas pu être représentés dans nos recherches d'identifications. La plus petite paire chromosomique métacentrique (probable produit d'une inversion péricentrique (Pisano et al. 1995; Pisano and Ozouf-Costaz 2000; Mazzei et al. 2006; Mazzei et al. 2008) présente chez l'ensemble des « *Trematomus* » étudiés (à l'exception probable de *T. eulepidotus*) n'a donc pas pu être caractérisée avec ce matériel.

Le succès de l'hybridation de ces BACs à large échelle taxinomique au sein de la famille de Nototheniidae suggère une forte conservation des séquences génomiques présentes dans les BACs entre les différentes espèces examinées. Nous avons également montré que ces BACs hybrident au-delà des Nototheniidae, sur une espèce de Notothenioidei (*Eleginops maclovinus*, divergence avec les Nototheniidae estimée entre 37 et 42 Ma, Near et al. (2012), Colombo et al. (2015) (Figure II.15), mais pas chez une espèce plus éloignée du sous-ordre (échec chez *Bovichtus diacanthus*, dont la divergence avec les Nototheniidae est estimée supérieure à 60Ma, Near et al. (2012), Colombo et al. (2015)). Cela confirme une divergence relativement récente de la famille des Nototheniidae (22,4 Ma, Near et al. (2012), 13,35 Ma, Colombo et al. (2015)), et a fortiori de la radiation des « *Trematomus* » (9,1 Ma Near et al. (2012), 4,26 Ma, Colombo et al. (2015)). L'étude des contenus très similaires, voire quasiment identiques en GC, en séquences répétées, en simple-sequence repeats (SSRs), et en ADN satellites entre les génomes du Nototheniinae *N. coriiceps* et du Channichthyinae *C. aceratus* (Detrich et al. 2010; Detrich and Amemiya 2010) confirme ce haut degré d'identité de séquences au sein de la famille.

### **Conservation des unités structurales, différemment réarrangées dans les caryotypes de « *Trematomus* »**

On note une forte conservation des segments délimités par les combinaisons de BACs identifiés sur une même unité structurale. Les BACs localisés dans un même bras chromosomique chez *N. coriiceps* se retrouvent en effet systématiquement localisés dans une même unité structurale, à distance comparable, et dans la même orientation chez les « *Trematomus* ». Nous avons voulu voir jusqu'où l'on pouvait retrouver cette conservation en hybridant le couple de BACs C4-F15 chez d'autres Nototheniidae. Cette synténie a été retrouvée à plus large échelle taxinomique chez d'autres représentants de la sous-famille des Trematominae (genres *Lepidonotothen* et *Patagonotothen*), ainsi que chez les autres sous-familles des Nototheniinae, des Channichtyinae, des Gymnodraconinae, des Cygnodraconinae, et des Dissostichinae (Figure II.15). Nous avons considéré que la présence d'un minimum de deux, puis d'un seul BAC sur une même unité structurale était représentative de cette unité dans son intégralité. Compte-tenu de la conservation retrouvée entre unités structurales, dans le cas où les BACs sont très proches, mais aussi lorsqu'ils sont plus éloignés les uns des autres ; il est possible que cette conservation interspécifique de la nature et de l'ordre des séquences soit extrapolable à l'unité structurale entière.

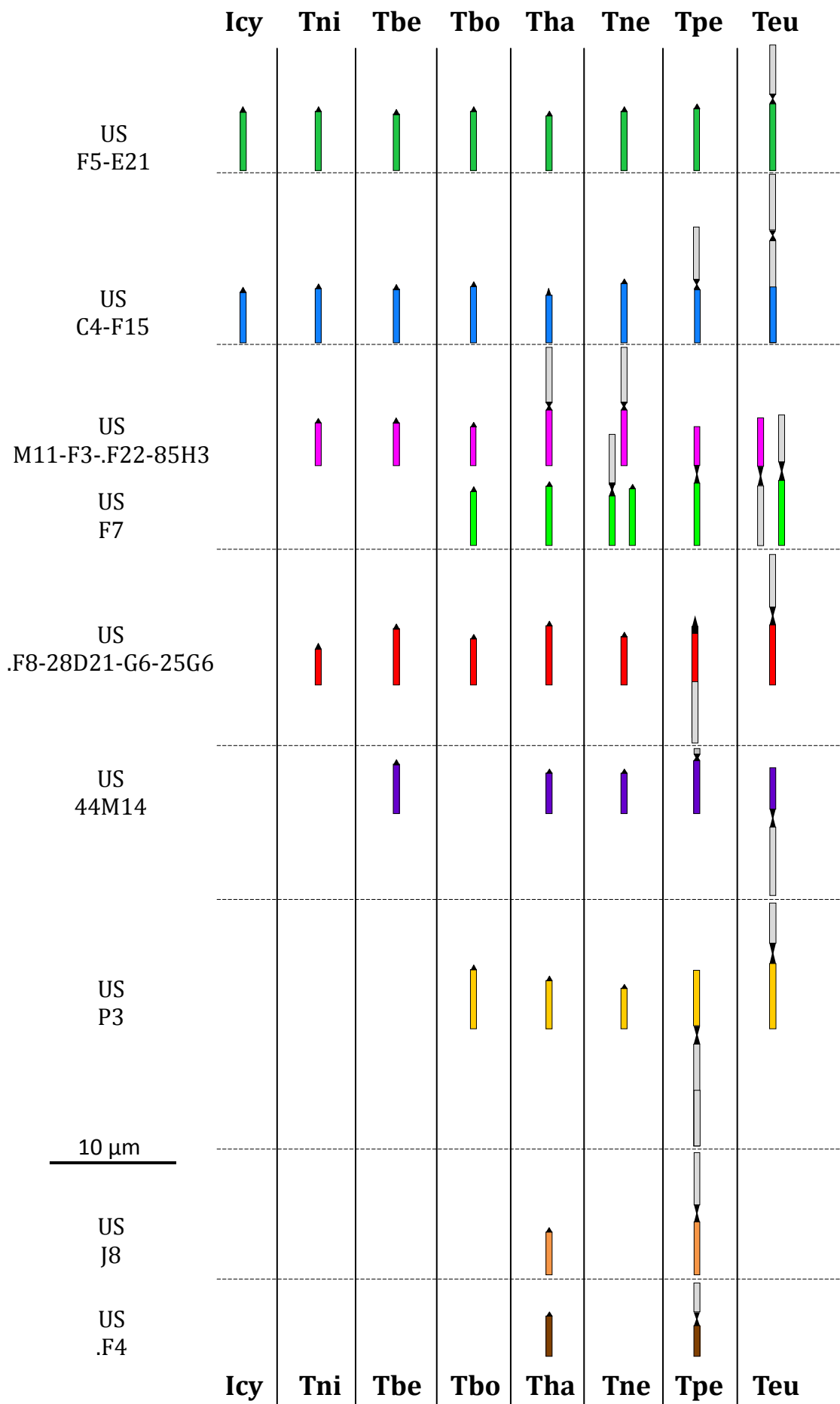


**Figure II.15 :** Exemple de double BAC-FISH avec l'hybridation du couple C4 (vert) - F15 (rouge) sur les chromosomes de A) *L. larseni*, B) *P. ramsayi*, C) *N. coriiceps*, D) *C. hamatus*, E) *C. mawsoni*, F) *D. mawsoni*, G) *G. acuticeps*, H) *E. maclovinus*. Tout comme chez les « *Trematomus* », on retrouve sur chacune de ces doubles hybridations le BAC C4 (détectés directs par la fluorescéine, en vert) en position centromérique et le BAC F15 (détectés directs par la rhodamine en rouge) en position médiane sur une même unité structurale. Barres d'échelle : 10  $\mu$ m.

Cette conservation des unités structurales chez les « *Trematomus* » va de pair avec une faible proportion de remaniements à l'intérieur d'un même bras (appelé ici remaniements secondaires). Nous n'avons mis en évidence aucune inversion ou insertion/délétion chromosomique au sein d'une même unité structurale avec les BACs utilisés au cours de cette

étude, bien que la littérature fasse état des deux inversions péricentriques présentes chez l'ensemble des « *Trematomus* » (Pisano and Ozouf-Costaz 2000; Pisano et al. 2003). Comme mentionnés précédemment, ces deux événements sont apparus après la divergence des Bovichtidae avec les autres groupes (63,2 Ma, Near *et al.* (2012), 61,5 Ma, Colombo *et al.* (2015)). Ils sont donc anciens par rapport à la diversification des Nototheniidae (22,4 Ma, Near *et al.* (2012), 13,4 Ma, Colombo *et al.* (2015)). On note de possibles insertions ou délétions segmentales dans certaines unités structurales suite à l'observation d'une différence de localisation du BAC 44M14 par rapport au centromère dans les unités structurales des différentes espèces examinées (péricentromérique vs dans le milieu du bras). Notre hypothèse est renforcée par l'absence ou le faible nombre de remaniements secondaires suggérés par Amores *et al.* (2017) chez la sous-famille des Nototheniinae (divergence estimée entre 8 et 6 Ma, Near *et al.* (2012), Colombo *et al.* (2015)), dans le cadre d'une étude comparative de l'organisation génomique et de la diversification des caryotypes par fusions massives.

Sur la base de ces observations, les unités structurales définies chez notre référence *I. cyanobrancha* pourraient donc être les unités de remaniements. Ainsi, ces bras constitutifs des caryotypes chez les « *Trematomus* » seraient conservés par rapport au caryotype de l'ancêtre commun des Nototheniidae (Morescalchi et al. 1992; Pisano et al. 2000; Pisano et al. 2003; Ghigliotti et al. 2015). C'est la combinaison de ces unités structurales, plus ou moins réarrangées qui serait à l'origine de la diversité caryotypique actuellement observée dans ce groupe. Les BACs sont donc parfois localisés sur des unités structurales constituant l'intégralité de petits chromosomes acrocentriques (Figures II.13 et II.16). Nous postulons que ces derniers sont directement hérités des paires chromosomiques constituant le caryotype ancestral inféré. Les mêmes combinaisons de BACs ont aussi été retrouvées dans des unités structurales faisant partie de paires chromosomiques de plus grande taille. Ces paires chromosomiques peuvent être acrocentriques (*T. pennellii*), métacentriques (*T. pennellii*, *T. eulepidotus*, *T. hansonii* et *T. newnesi*) ou submétacentriques (*T. pennellii*, *T. eulepidotus*) (Figures II.13 et II.16). Sur la base de la synténie observée entre unités structurales sur les chromosomes des différentes espèces, de la taille de la morphologie de ces paires chromosomiques, l'hypothèse la plus parcimonieuse est qu'elles sont le produit de fusion(s) centrique(s) et/ou en tandem de différentes unités structurales héritées de l'ancêtre commun. Ces fusions sont les événements de remaniements majoritaires dans ce groupe (Morescalchi et al. 1992; Pisano et al. 2000; Pisano et al. 2003). Les positionnements et la comparaison des localisations des BACs sélectionnés sur les unités structurales entre les différentes espèces de « *Trematomus* » nous permettent de retracer l'histoire évolutive de ces fusions.



**Figure II.16:** homologues interspécifiques des unités structurales (US) mises en évidence par BAC-FISH sur les chromosomes de huit espèces de « *Trematomus* ». De gauche à droite, *I. cyanobrancha*, *T. nicolai*, *T. bernacchii*, *T. borchgrevinki*, *T. hansonii*, *T. newnesi*, *T. pennellii*, *T. eulepidotus*. Les échelles de longueur des unités structurales présentées Figure II.13 ont été conservées. Les unités structurales sont nommées en fonction des BACs ayant servi à les identifier.

Si l'unité structurale semble très fortement conservée au sein de « *Trematomus* » (Figure II.12, II.13 et II.16), et même plus largement avec des bras chromosomiques chez d'autres Nototheniidae (Figure II.15), on constate un très fort degré des réarrangements de ces unités structurales survenu lors de la diversification (Figures II.12, II.13, II.16). Les fusions/fissions/inversions péricentriques ayant accompagné la radiation du groupe « *Trematomus* » peuvent être homologues et donc résulter d'une histoire commune. Dans ce premier cas, elles auraient été acquises chez l'ancêtre commun et seraient partagées par la descendance, avec de possibles pertes ou gains secondaires. L'hypothèse alternative serait l'acquisition indépendante de ces remaniements au cours de l'évolution.

Les fusions centriques ou en tandem mises en évidence dans cette étude n'impliquent pas forcément les mêmes unités structurales d'une espèce à une autre. Ainsi, au moins deux des trois unités structurales présentes dans la plus grande paire chromosomique submétacentrique présente chez *T. pennellii* et *T. eulepidotus* ne sont pas homologues. Ces très grandes paires chromosomiques résultent pourtant toutes deux d'une double fusion : centrique et en tandem. Il en est de même pour une des deux unités structurales situées dans la plus grande paire métacentrique présente chez *T. newnesi*, *T. hansonii* d'une part et *T. borchgrevinki* d'autre part, qui résultent chacune d'une simple fusion centrique. Pour pouvoir identifier un maximum de fusions chromosomiques à l'échelle des « *Trematomus* », et étant donné la synténie retrouvée entre les unités structurales au sein des Nototheniidae, nous avons repris les hybridations de BACs, avec pour l'objectif d'identifier un BAC par unité sur des paires chromosomiques constituées de deux ou trois unités structurales. Les BACs F3 et F7 en sont un exemple, avec une hybridation dans chacune des deux unités structurales constituant une même paire chromosomique chez *T. pennellii*, mais présentes sur deux unités structurales différentes de deux paires métacentriques différentes chez *T. eulepidotus* (Figures II.13 et II.16). Dans presque tous les cas (Figures II.16), les fusions mises en évidence au travers de cette étude d'homologies des unités structurales ne semblent donc pas partagées entre les espèces.

Par contre, il est surprenant de noter qu'une même unité structurale est impliquée dans une fusion centrique aboutissant à la plus grande paire chromosomique métacentrique de *T. newnesi* et *T. hansonii* (Figures II.16). Ceci n'est pas sans rappeler l'établissement des fusions chromosomiques à l'origine des caryotypes observés chez *Gasterosteus aculeatus* et *Tetraodon*



*nigroviridis* ( $2n = 42$ ) (Amores et al. 2014). Ces trois fusions indépendantes impliquent cependant deux bras communs, hérités du caryotype ancestral inféré sur le caryotype de *Xiphophorus maculatus* ( $2n = 48$ ). Les auteurs notent que ces bras qu'ils nomment « sticky chromosomes » auraient plus de propension à fusionner que les autres. Ils sont riches en éléments répétés et pauvres en séquences géniques. Cependant, il faudrait pouvoir identifier des BACs dans la seconde unité structurale impliquée dans la fusion pour pouvoir trancher entre un partage ou non de cette fusion centrique entre ces deux espèces de « *Trematomus* ».

## II.2.E : Discussion

### Deux approches de cytogénomique complémentaires pour identifier les homologies chromosomiques interspécifiques

Grâce aux nouvelles techniques de séquençage NGS permettant de générer une importante quantité de données de séquences en très peu de temps, la génomique comparative est en plein essor. Si les séquençages de génomes complets sont couramment pratiqués pour des petits génomes Procaryotes ou Eucaryotes (unicellulaires pour la plupart) ainsi que chez des espèces modèles, cela reste des études beaucoup plus ponctuelles et difficiles à mettre en place sur des lignées entières vivant en milieu naturel. Ainsi, parmi les 115 espèces recensées actuellement chez les Nototheniidae (Duhamel et al. 2014; Ghigliotti et al. 2015), seules trois espèces (*N. coriiceps*, *C. aceratus*, *P. charcoti*) ont fait l'objet d'un séquençage de génome entier (Shin et al. 2014; Ahn et al. 2017). L'assemblage et l'annotation corrects des séquences de ces génomes ne sont pas encore publiés, rendant impossible leur utilisation.

Les approches par cytogénomique comparative (*FISH*) apparaissent comme une alternative efficace pour répondre à ces problématiques. Elles sont relativement rapides à mettre en œuvre (Ried 1998; Vicari et al. 2015), et offrent la possibilité d'examiner un grand nombre d'espèces dans la perspective d'études comparées. Malgré certaines contraintes d'ordre technique inhérentes à l'élaboration des sondes microdisséquées, la peinture chromosomique permet de s'affranchir du manque de données génomiques et de sondes commercialisées chez les espèces non-modèles (Laudicina and Mühlmann 2015). Elle nous a permis de mettre en évidence des homologies chromosomiques interspécifiques à l'échelle d'un chromosome entier choisi pour sa reconnaissance facile : la plus grande paire chromosomique de *T. pennellii* qui combine trois unités structurales doublement fusionnées. Nous avons donc besoin d'identifier ces unités structurales pour déterminer leurs homologies entre espèces et reconstruire les remaniements chromosomiques survenus au cours des spéciations. L'incapacité d'identifier de manière non équivoque des multiples unités structurales constituant les petites paires chromosomiques

acrocentriques des « *Trematomus* » rend impossible l'emploi d'une approche de peinture à l'échelle des unités structurales. C'est l'approche par BAC-*FISH* comparative qui nous a permis de marquer et d'identifier un grand nombre d'unités structurales différentes pour chaque espèce. Ainsi, nous avons pu établir leurs homologies entre un bon nombre d'espèces de « *Trematomus* », et mettre en évidence un nombre non négligeable de réarrangements (majoritairement des fusions).

Nous avons confirmé, par l'emploi de ces deux méthodes, l'hypothèse d'une conservation de ces unités structurales chez les « *Trematomus* ». La peinture a en effet montré une correspondance entre trois unités structurales entières constituant trois petites paires chromosomiques acrocentriques chez *T. hansonii* ( $2n = 45-46/48$ ) et trois unités structurales formant la plus grande paire chromosomique de *T. pennellii* ( $2n = 48$ ). Les multiples positionnements de BACs ont permis de mettre en évidence des synténies de toutes les unités structurales identifiées chez les « *Trematomus* » au cours de l'étude. Ces deux approches ont également révélé l'extension de cette synténie à l'échelle de la sous-famille des Nototheniidae. La correspondance 3 : 1 est en effet retrouvée par hybridation de la sonde de peinture sur les chromosomes de *D. mawsonii*. De plus, l'unité structurale marquée par le couple de BAC C4 - F15 est conservée chez de nombreuses autres espèces représentant d'autres clades de téléostéens antarctiques. Sur la base de ces résultats, les unités structurales ont été définies comme des unités de réarrangement. Premièrement définies chez *I. cyanobranca* en vue d'une étude chez les « *Trematomus* », on pourrait tout à fait envisager de replacer ces unités structurales sur les chromosomes des autres Nototheniidae.

Nous retrouvons donc une cohérence dans les résultats obtenus par ces deux types d'approches complémentaires. En effet, la sonde de peinture issue de la plus grande paire submétacentrique de *T. pennellii* produit un signal fort identifié dans une grande partie d'une grande paire chromosomique de *N. coriiceps* (Figure II.9.C). Il s'agit d'une paire aisément identifiable par son marquage en DAPI positif de part et d'autre du marquage DAPI : en forme de « diabolo » au niveau du centromère. C'est sur cette même paire de *N. coriiceps* qu'a été localisé le seul BAC retrouvé dans la plus grande paire chromosomique de *T. pennellii* au cours de cette étude (P3, Figures II.11 et II.13). De la même manière, ce BAC P3 a été localisé dans une paire chromosomique acrocentrique chez *T. hansonii*, probablement une des trois paires peintes par la sonde de peinture utilisée.

Le séquençage, l'assemblage et surtout l'annotation complète du génome de *N. coriiceps* ou encore *P. charcoti* et *C. aceratus* (Shin et al. 2014; Ahn et al. 2017), ainsi que de nombreux autres Nototheniidae à venir ouvrent de nouvelles perspectives pour une meilleure caractérisation de

ces génomes. Ces données de séquence pourront nous permettre de vérifier ou d'invalider nos hypothèses concernant ces HCl.

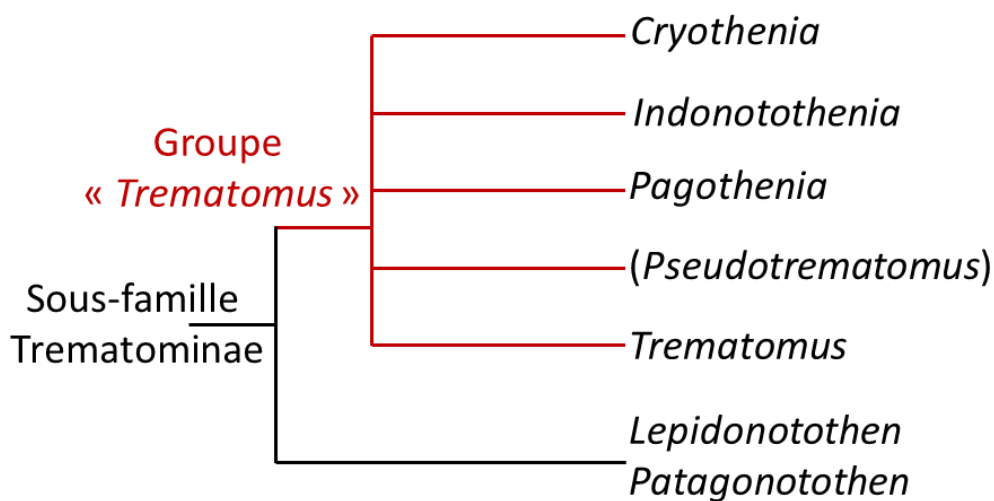
Une très large proportion des fusions étudiées ne serait pas partagées, mais au contraire, plutôt acquises indépendamment au cours de l'évolution, comme suggéré par Morescalchi *et al.* (1992). Cependant, nous n'avons pas été exhaustifs : le nombre de remaniements mis en évidence au cours de cette étude ne représente pas la totalité des réarrangements chromosomiques survenus lors de la diversification des « *Trematomus* ». D'autres événements pourraient avoir été acquis chez l'ancêtre commun et partagés par la descendance (avec possibles pertes ou gains secondaires). Pour pouvoir explorer cette hypothèse et ordonner chronologiquement ces réarrangements en rapport avec les événements de spéciation, il est nécessaire de les positionner sur une phylogénie bien résolue des « *Trematomus* ».

## III) Chapitre 2 : phylogénie des « *Trematomus* »

### III.A : Introduction

#### III.A.1 : Composition du groupe « *Trematomus* »

Quelles sont les espèces qui constituent le groupe « *Trematomus* » ? A cette simple question, il n'est pas si facile de répondre sans faire un petit rappel historique.



**Figure III.1 :** Cladogramme montrant la composition de chaque genre et leurs liens phylogénétiques au sein de la sous-famille des Trematominae. Adapté de Balushkin (2000), Sanchez *et al.* (2007), Lautre dou *et al.* (2012) et Detta  *et al.* (2012). Le groupe monophyl tique « *Trematomus* » indiqu  en rouge rassemble cinq genres (*Cryothenia*, *Indonotothenia*, *Pagothenia*, *Pseudotrematomus* et *Trematomus*) d crits dans la litt rature. Ils sont ici tous consid r s comme appartenant   l'ensemble monophyl tique « *Trematomus* ».

Le nom de genre *Trematomus* appara t dans Boulenger (1902) qui d crit *T. bernacchii*, *T. hansonii*, *T. newnesi*, *T. nicolai* et *T. scotti*. Toutes sont inf od es au plateau continental antarctique,   l'exception de *T. hansonii* (retrouv   galement dans les  les Orcades du Sud et l' le de G orgie du Sud), *T. bernacchii* et *T. scotti* (retrouv s  galement dans les  les Shetland du Sud et les  les Orcades du Sud) (Duhamel *et al.* 2014). En 1982, sur la base de ressemblances morphologiques (principalement des caract res morphologiques en lien avec la forme du squelette) et des modes de vie ( pibenthique vs benthop lagique,  pipelagique ou cryop lagique), Balushkin renomme *Pseudotrematomus* presque toutes les esp ces alors int gr es sous le nom *Trematomus*   l'exception de *T. newnesi* (Figure III.1). En 1986, ses crit res morphologiques sont repris par Voskoboinikova, et sont combin s   des analyses des structures axiales et des d tails morphologiques de l' cil (Balushkin 1982; Balushkin 2000), le genre

*Pseudotrematomus* avec cette nouvelle composition ne repose sur aucune synapomorphie. Suite à ces études, le nom de genre n'est pas accepté par de nombreux auteurs (DeWitt et al. 1990). Le terme *Trematomus* est donc resté couramment employé dans la littérature, même si le genre *Pseudotrematomus* n'a jamais été officiellement invalidé.

Au sein des *Trematomus*, deux espèces ont été individualisées en raison de leur mode de vie cryopélagique. Il s'agit de *Pagothenia borchgrevinki* et de *Pagothenia brachysoma*. *P. borchgrevinki* avait dans un premier temps été placé dans les *Trematomus* pour être ensuite placé dans un genre créé spécialement pour lui par Nichols et La Monte (1936) : le genre *Pagothenia* (Tableau I.1, Figure III.1). Son positionnement tantôt au sein des *Trematomus*, des *Pseudotrematomus* ou des *Pagothenia* a alterné durant de nombreuses années (Andriashev and Jakubowski 1971; Balushkin 1982). Actuellement, son appartenance au genre *Trematomus* a été reconfirmée, et son remplacement dans ce dernier a été proposée (Sanchez et al. 2007; Kuhn and Near 2009; Lautrédou et al. 2012). Toutefois, dans la littérature actuelle, le terme *Pagothenia* (Bargelloni et al. 1994; Ritchie et al. 1996; Ritchie et al. 1997; Balushkin 2000; Bargelloni et al. 2000; Near 2004; Cziko and Cheng 2006; Lautrédou et al. 2010; Janko et al. 2011; Lautrédou et al. 2012; Murphy 2015), ou celui de *Trematomus* (Sanchez et al. 2007; Near and Cheng 2008; Kuhn and Near 2009; Dettai et al. 2012; Colombo et al. 2015) sont employés. Pour *P. brachysoma* (Pappenheim, 1912), aucun adulte ou juvénile n'a été retrouvé depuis la description de l'espèce. Une larve capturée récemment en mer de Ross (Murphy 2015) a été identifiée comme étant probablement *P. brachysoma*, grâce à sa position phylogénétique en tant qu'espèce sœur de *P. borchgrevinki* sur la base de l'analyse du marqueur moléculaire ND2 (divergence nucléotidique de 5,4 % en moyenne entre la larve et les autres *P. borchgrevinki*, inférieure à la divergence moyenne de 11% entre la larve et les autres spécimens de *Trematomus* examinés). Cependant, l'analyse d'un seul marqueur est insuffisante pour tirer des conclusions quant à l'identification et la séparation de ces deux espèces.

Le genre *Cryothernia* décrit par Daniels en 1981 compte deux espèces (Tableau I.1, Figure III.1). La première, *Cryothernia peninsulae* Daniels, 1981 (collectée en péninsule antarctique sur la base de Palmer) a donné son nom au genre. Elle a premièrement été incluse dans la sous famille des Pleuragramminae aux côté des genres *Pleuragramma* et *Aethotaxis* (Anderson 1984; DeWitt et al. 1990) sur la base de critères morphologiques. Une vingtaine de spécimens ont été récoltés en 1981, mais ils ont été conservés dans le formol, rendant impossible des études sur des caractères moléculaires. La deuxième espèce est *Cryothernia amphitrete* Cziko et Cheng (2006) (collectée en mer de Ross sur la station de McMurdo). Elle a été placée dans le genre *Cryothernia* en raison de ses grandes ressemblances morphologiques avec *C. peninsulae*. Des études phylogénétiques employant des marqueurs moléculaires (mitochondriaux : ND2, COI, 16S, ou

nucléaires : RPS7, PkD1, HECW, Rhodopsine, SSRP) ont intégré ce spécimen dans un clade monophylétique avec les *Trematomus* et les *Pagothenia* (Cziko and Cheng 2006; Kuhn and Near 2009; Lautrédou et al. 2010). Comme pour les *Pagothenia*, un changement de genre a également été proposé par Lautrédou *et al.* (2010) pour remplacer les *Cryothernia amphitreta* au sein du genre *Trematomus*.

Enfin, le genre *Indonotothenia* a été décrit par Balushkin en 1984 (Tableau I.1, Figure III.1). Il est constitué d'une seule espèce : *I. cyanobrancha* (Richardson, 1844). Cette espèce est exclusivement subantarctique (endémique des plateaux continentaux des îles Kerguelen et Heard) (Figure I.2, Duhammel *et al.* (2014)). D'abord placé dans le genre *Notothenia* (en tant que *N. cyanobrancha*) (Hureau 1970; Bargelloni et al. 2000), cette espèce a ensuite été assignée à ce nouveau genre sur la base de critères morphologiques (Balushkin 1982). Elle est assez peu représentée dans les études de phylogénie moléculaire (Bargelloni and Lecointre 1998; Bargelloni et al. 2000; Dettai et al. 2012) (marqueurs employés 12S, 16S, COI, Rhodopsine, SSRP1, PPM1d, PKD1, HECW2), mais celles-ci la placent dans les Trematominae avec d'autres *Trematomus/Pagothenia*, et non dans les *Nototheniinae*.

Le groupe "*Trematomus*", tel que nous l'employons dans cette étude, est constitué du genre *Trematomus sensus stricto* Boulenger, 1902 (onze espèces), et des genres *Pagothenia* Nichols et La Monte, 1936 (deux espèces), *Cryothernia* Daniels, 1981 (deux espèces), et *Indonotothenia* Balushkin, 1982 (une espèce). Ces quatre genres forment un ensemble monophylétique bien soutenu (Balushkin 2000; Near 2004; Cziko and Cheng 2006; Sanchez et al. 2007; Near and Cheng 2008; Janko et al. 2011; Dettai et al. 2012; Near et al. 2012; Colombo et al. 2015). La divergence de cette radiation adaptative a été datée de manières diverses :  $9,8 \pm 0,4$  Ma (Near *et al.* (2004)),  $9,1 \pm 2,4$  Ma, Near *et al.* (2012), entre 8 et 2 Ma, Bargelloni *et al.* (2000) et  $4,26 \pm 1,5$  Ma, Colombo *et al.* (2015). Ce groupe des « *Trematomus* » forme, avec ses groupes frères (les genres *Lepidonotothen* et *Patagonotothen*), la sous-famille des Trematominae (Near 2004; Sanchez et al. 2007; Near and Cheng 2008; Rutschmann et al. 2011; Dettai et al. 2012; Colombo et al. 2015) (Tableau I.1, Figure III.1).

### **III.A.2 : Phylogénie du groupe « *Trematomus* »**

Plusieurs études ont tenté de résoudre les relations phylogénétiques entre les Trematominae, et plus particulièrement au sein du groupe « *Trematomus* », soit au moyen de caractères morphologiques, soit sur la base de marqueurs moléculaires mitochondriaux et/ou nucléaires.

Bien que ces différentes études présentent une grande variabilité dans les topologies reconstruites, un consensus existe sur la monophylie du groupe (Bargelloni et al. 1994; Ritchie

et al. 1996; Balushkin 2000; Near 2004; Cziko and Cheng 2006; Sanchez et al. 2007; Near and Cheng 2008; Kuhn and Near 2009; Lautrédou et al. 2010; Janko et al. 2011; Rutschmann et al. 2011; Dettai et al. 2012; Lautrédou et al. 2012; Colombo et al. 2015; Murphy 2015). D'autre part, un certain nombre de noeuds apparaissent communs. Ainsi, *T. scotti* est le premier « *Trematomus* » à diverger nettement avant les autres espèces du genre (Bargelloni et al. 1994; Ritchie et al. 1996; Ritchie et al. 1997; Bargelloni et al. 2000; Stankovic et al. 2002; Near 2004; Sanchez et al. 2007). Quand *I. cyanobrancha* est inclus dans l'analyse, il est retrouvé également en position basale, dans une polytomie ou en groupe frère de *T. scotti* (Bargelloni et al. 2000; Near 2004; Dettai et al. 2012). *T. lepidorhinus* et *T. loennbergii* sont retrouvées groupes frères dans la plupart des topologies reconstruites (Ritchie et al. 1997; Sanchez et al. 2007; Near and Cheng 2008; Lautrédou et al. 2010; Janko et al. 2011; Dettai et al. 2012; Lautrédou et al. 2012; Murphy 2015). Ces deux espèces sont en effet très proches génétiquement pour les marqueurs utilisés jusqu'à présent, au point que certains auteurs ont proposé qu'il s'agissait d'une seule et même espèce dont les spécimens pourraient présenter différents morphes et distributions bathymétriques. Les espèces *T. bernacchii* et *T. hansonii* sont également positionnées en groupes frères de manière assez récurrente (Ritchie et al., 1996; Near et al., 2004; Sanchez et al., 2007; Kuhn et Near, 2009: 16S/ND2; Janko et al., 2011; Lautrédou et al., 2012: COI/HECW/Pkd1/Rhodopsine; Colombo et al., 2015). Cependant, la fiabilité de ces relations phylogénétiques reste à évaluer, car elles ne sont pas identiques dans toutes les topologies publiées. Par exemple, les analyses de Lautrédou *et al.* (2010) et Dettai *et al.* (2012) avec une combinaison de marqueurs mitochondriaux et nucléaires ne placent pas *T. bernacchii* et *T. hansonii* ensemble. Il arrive également qu'au sein d'une même étude, selon le marqueur employé, des discordances apparaissent entre topologies (Ritchie et al. 1996; Near and Cheng 2008; Kuhn and Near 2009; Lautrédou et al. 2010; Janko et al. 2011; Lautrédou et al. 2012). C'est par exemple le cas de l'étude de Kuhn et Near (2009) qui regroupe ensemble les espèces *T. lepidorhinus* et *T. loennbergii* sur la base des marqueurs mitochondriaux: NADH déshydrogénase 2 (ND2) et 16S, mais pas en utilisant le marqueur nucléaire: protéine ribosomale S7 (RPS7).

Le manque de résolution et l'incongruence entre certaines topologies reconstruites peuvent s'expliquer par différents facteurs. Premièrement, la diversification du groupe « *Trematomus* » est une radiation rapide et récente (Janko et al. 2011; Lautrédou et al. 2012; Near et al. 2012; Lecointre et al. 2013; Colombo et al. 2015). Ces génomes présentent donc une divergence faible sur les marqueurs employés et en particulier sur les noeuds à la base du groupe, avec un signal phylogénétique insuffisant pour résoudre de manière fiable les liens de parentés (Janko et al. 2011; Lautrédou et al. 2012). Deuxièmement, l'échantillonnage de certaines études est insuffisant, ne prenant en compte que très peu de spécimens (souvent un seul) par espèce (Near

2004; Sanchez et al. 2007; Janko et al. 2011; Dettai et al. 2012), voire un individu différent selon le marqueur employé (Colombo et al. 2015). Lors d'une radiation en particulier, le temps de coalescence d'un gène peut être supérieur au temps de divergence des espèces. Dans ce cas de figure, l'histoire de ce gène ne reflétera donc pas l'histoire des espèces. Faute d'un échantillonnage suffisant chez les «*Trematomus*», des biais d'analyse liés à des problèmes de coalescence pourraient expliquer en grande partie cette irrésolution et incongruence (Lautrédou et al. 2010; Lautrédou et al. 2012). Troisièmement, la nature (mitochondriale vs nucléaire) ainsi que le nombre de marqueurs (un seul ou plusieurs) pris en compte dans les reconstructions varient beaucoup d'une étude à l'autre. Un faible nombre de marqueurs (Near 2004; Kuhn and Near 2009; Lautrédou et al. 2010), qui plus est peu polymorphes et partagés par un nombre insuffisant d'individus (Sanchez et al. 2007; Near and Cheng 2008; Janko et al. 2011; Colombo et al. 2015), ne fournit pas de signal suffisamment résolutif (valeurs de robustesse comme de bootstrap assez faibles, incongruences entre topologies selon le/les marqueur(s) employé(s)).

### **III.A.3 : Analyse phylogénomique**

Afin d'obtenir un signal phylogénétique suffisant, il est nécessaire d'augmenter le nombre de sites polymorphes étudiés entre les différentes espèces (si possible partagés entre les individus). Pour cela, il est possible de jouer sur deux paramètres : le nombre et la nature (conservée ou variable) des marqueurs employés. C'est dans ce sens que des études récentes ont non seulement augmenté le nombre mais aussi combiné les marqueurs moléculaires d'origine mitochondriale et nucléaire (Janko et al. 2011; Rutschmann et al. 2011; Dettai et al. 2012; Lautrédou et al. 2012; Near et al. 2012; Colombo et al. 2015). Ainsi, la prise en compte du cytochrome b (cytb), ND2, ND4, et parfois d'une partie de la région de contrôle pour les marqueurs mitochondriaux, ou encore de marqueurs nucléaires tels que *enc1*, *myh6*, *PTCHD4*, *rps7*, *snx33*, *trb1*, *zic1*, *HECW2*, *SSRP1*, *PPm1d*, *Pkd1* (Near 2004; Sanchez et al. 2007; Dettai et al. 2012; Lautrédou et al. 2012; Near et al. 2012; Colombo et al. 2015) a permis l'identification de nœuds communs. Cependant, l'augmentation et la diversification des jeux de données employés, ainsi que la diversité des méthodes de reconstruction (parcimonie, maximum de vraisemblance, distance, inférence bayésienne) n'ont pas conduit à une congruence des topologies de «*Trematomus*» publiées. Lautrédou *et al.* (2012) concluent alors à une polytomie des «*Trematomus*». Elle est qualifiée de «*soft*», c'est-à-dire que la topologie n'est pas résolue, faute de signal phylogénétique suffisant. Ce manque de signal peut être intrinsèque à la radiation, ou bien le fait d'études incluant trop peu de données.



Afin de résoudre les relations de parentés au sein des « *Trematomus* », nous avons cherché à augmenter l'information phylogénétique en utilisant un grand nombre de marqueurs polymorphes informatifs. Nous avons donc choisi une approche de ddRAD-seq (double digest Restriction site Associated DNA marker-sequencing) en séquençage haut débit (Baird et al. 2008). Il s'agit d'un balayage génomique permettant d'échantillonner de manière pseudo-aléatoire une partie des génomes d'intérêt. Les fragments obtenus sont en effet bornés par des sites de restriction que l'on espère conservés entre génomes apparentés, puis sélectionnés une seconde fois selon leur taille pour obtenir un nombre de locus en théorie proportionnel à la couverture de chaque banque (Peterson et al. 2012; Cruaud et al. 2014). Ces méthodes sont donc adaptées aux taxons dont la divergence est récente car partageant des mêmes sites de restriction (Emerson et al. 2010; Jones et al. 2013; Nadeau et al. 2013; McCluskey and Postlethwait 2015), comme c'est le cas chez les « *Trematomus* ». Cette technique permet d'étudier des régions géniques, mais aussi des régions non codantes, moins contraintes, et donc probablement plus polymorphes. De plus, la multiplicité des marqueurs employés permettra notamment d'explorer les problèmes de coalescence dans les génomes étudiés (Baird et al. 2008; Cutter 2013), surtout avec l'utilisation de méthodes de congruence qui tiendra compte de l'histoire de tous les locus reconstitués.

### **III.B : Matériels et méthodes**

#### **Matériel :**

Nous avons disposé pour cette étude d'une large diversité d'espèces de « *Trematomus* », incluant *I. cyanobrancha*, et *P. borchgrevinki*. Cependant, il n'a pas été possible d'inclure *P. brachysoma* ni un représentant du genre *Cryothermia* (*C. peninsulae* ou *C. amphitreta*). Dans la mesure du possible, nous avons intégré deux spécimens pour chaque espèce du groupe. Les groupes externes ont été choisis de manière à avoir deux extragroupes non monophylétiques (Figure I.1) : le Trematominae *L. squamifrons* et le Pleuragramminae *P. antarctica*.

Vingt-trois spécimens de quatorze espèces ont ainsi été prélevés (*I. cyanobrancha*, *T. scotti*, *T. nicolai*, *T. bernacchii*, *T. lepidorhinus*, *T. borchgrevinki*, *T. hansonii*, *T. newnesi*, *T. loennbergii*, *T. pennellii*, *T. eulepidotus*, *L. squamifrons*, *P. antarctica*). Ces échantillons ont été récoltés lors des différentes campagnes POKER (POissons de KERguelen, 2006, 2010 and 2013), CEAMARC (Collaborative East Antarctic Marine Census for the Census of Antarctic Marine Life, 2007-2008), et REVOLTA (Radiation EVOLutives en Terre Adélie, 2010-2014). Les spécimens proviennent tous de la même localité géographique (Terre Adélie), excepté *I. cyanobrancha* qui est issu du

plateau continental subantarctique Kerguelen-Heard. Pour chacun des spécimens, les échantillons de tissus sont référencés en annexe (annexe 1).

## **Méthodes :**

### Collection du matériel :

Les individus ont été anesthésiés avec du MS 222, puis sacrifiés. Des échantillons de muscles et de nageoires ont été prélevés et conservés dans l'éthanol 85 % à 4°C. Les spécimens correspondants sont conservés pour référence dans les collections du MNHN. Les spécimens ont été choisis pour refléter un maximum la divergence inter-individuelle au sein de chaque espèce sur la base des divergences observées avec le marqueur moléculaire COI (Lautrédou et al. 2012) et, lorsque cela était possible, nous avons sélectionné deux spécimens de la même espèce provenant de deux campagnes antarctiques différentes (CEAMARC et REVOLTA).

### Extractions :

Les ADN génomiques ont été extraits par un protocole au CTAB et au chloroforme-isoamyle alcool (CIA) (Winnepenincks et al. 1993). La présence de fragments de haut poids moléculaire dans les extraits a été vérifiée par dépôt sur gel d'agarose 1 %. Pour chacun des extraits, la concentration d'ADN a été mesurée par dosages Qubit HS (kit DNA Broad Range, Thermofisher Scientific).

### Préparations des librairies :

Les banques de séquençages Illumina ont été préparées en s'inspirant des protocoles proposés dans la littérature (Baird et al. 2008; Hohenlohe et al. 2010; Peterson et al. 2012). Ces protocoles ont été affinés par l'équipe du Service de Systématique Moléculaire (SSM) du Muséum National d'Histoire Naturelle (MNHN). Les détails de ces protocoles complets sont présentés en annexe (annexe 2).

### *Digestion*

Les enzymes de restriction pour la double digestion des extraits ont été choisies par rapport à la littérature existante sur des espèces phylogénétiquement proches des Nototheniidae tels que *Gasterosteus aculeatus* (Hohenlohe et al. 2010), *Perca fluviatilis* (Henning et al. 2014; Pukk et al.

2015) et d'autres téléostéens Anguiliformes (Kai et al. 2014). Les fréquences de coupes ont été estimées également par des simulations *in silico* sur les contigs de *G. aculeatus* (Jones et al. 2013) et *N. coriiceps* (Shin et al. 2014). Cela nous a permis d'estimer le nombre de fragments de taille moyenne entre 400 et 1500 pb après coupe et d'en déduire par calcul le facteur de dilution fonctionnelle nécessaire pour chacun des deux adaptateurs et nos concentrations en ADN génomique (fichier de calcul issu de Peterson *et al.* (2012)).

420 ng d'ADN a fait l'objet d'une double digestion par EcoR1 (« rare cut site ») et MspI (« common cut site ») (Peterson et al. 2012) (Promega). Les produits de digestion ont ensuite été purifiés par l'utilisation de billes magnétiques (Nucleomag, Macherey-Nagel).

### *Ligation*

Les fragments digérés ont été ligués à l'aide d'une T4 ligase (1 MU/ml, marque) aux adaptateurs EcoR1 et MspI (Proméga) dilués selon la concentration des fragments d'ADN à liguer. Les 24 banques ont été liguées à 24 adaptateurs portant des barcodes différents. Les banques ont ensuite été de nouveau dosées une à une par qPCR (CFX96 Touch Real-time PCR Detection System, Bio-Rad, « antibody-mediated hot-start Sso7d fusion polymerase »). Deux pools équimolaires contenant chacun 12 banques ont été réalisés d'après ces dosages.

### *Sélection de taille des fragments digérés et ligués*

Les fragments présents dans les pools ont été sélectionnés entre 350 et 600 pb au moyen d'une double purification par billes magnétiques (Nucleomag, Macherey-Nagel).

### *PCR d'indexage*

Les pools des banques Illumina ont été indexés par PCR (index 5 pour le pool 1 et index 19 pour le pool 2), puis purifiés à nouveau par billes magnétiques magnétiques (Nucleomag, Macherey-Nagel). Les deux pools ont été dosés au Qubit HS (kit DNA Broad Range, Thermofisher Scientific), et réamplifiés pour obtenir un échantillon final de 30 µl à 10 ng/µl.

Le séquençage Illumina a été externalisé à la Station Biologique de Roscoff (Sorbonne Université). Il a été réalisé après vérification de la taille des fragments à séquencer sur une puce Bioanalyzer (marque kit) au moyen d'une cartouche MiSeq Reagent Kit v2 300cycles (MS-102-2002, Illumina) en paired-end avec le kit correspondant.

### Démultiplexage et filtres tailles/qualité :

Les bibliothèques générées par le séquençage Illumina nous ont été communiquées séparées par index. Elles ont ensuite été démultiplexées au moyen de l'étape 1 du « pipeline » PyRAD (Eaton and Ree 2013; Eaton 2014). Pour s'assurer de la fiabilité de la séparation par banque, et donc des individus, aucune différence n'a été permise par rapport aux séquences d'adaptateurs et de barcodes P1. Lors de cette étape de démultiplexage, une erreur de barcode correspondant à un *T. pennellii* (CE3317) a été détectée chez un des deux *T. borchgrevinki* (CE5466). L'individu porteur du mauvais barcode a été identifié grâce à son index (index 5), différent des deux autres *T. pennellii* (index 19). Une erreur de manipulation lors de la constitution des lots d'indexage a donc été mise en évidence et rectifiée. Les lectures présentent dans les deux banques portant le barcode commun ont ainsi été regroupées en un seul fichier représentant l'individu *T. pennellii* (CE3317).

Les séquences de chaque banque ont ensuite été filtrées. Premièrement par un filtre « enzyme de restriction » ne retenant que les lectures présentant à leur deux extrémités respectivement les deux sites de restriction EcoR1 et MspI. Deuxièmement par un filtre de longueur ne retenant que les lectures de 141 pb (Geneious, v9.0.2, (Kearse et al. 2012)). Troisièmement par un filtre de qualité ne retenant que les lectures présentant un maximum de 4 bases de faible qualité (Qscore inférieur à 20) dans leur séquence (outil «NGS: QC and manipulation, filter fastQ » dans le portail Galaxy : <https://usegalaxy.org>). Et enfin quatrièmement par un filtre ciblant les éléments transposables (présents en multicopies dans les génomes) des « locus ETs » non écartés par PyRAD ou par Stacks, par blastx (2.2.28 + blastpackage, E-value de  $10^{-10}$ ) avec une base de données Repbase enrichie avec les éléments transposables que nous avons identifiés chez les Nototheniidae. Ces deux « pipelines » utilisés sont également composés d'un filtre éliminant les éventuels paralogues détectés, sans toutefois donner d'information précise concernant le moment ni la nature du tri.

Les données paired-end issues du séquençage Illumina ont été assemblées par FLASH (v1.2.11). Etant donné le faible recouvrement des données paired-end issues du séquençage (inférieur à 15 %) pouvant être dû à une deuxième sélection de taille (300-800pb) avant le séquençage, l'ensemble du jeu de données a été traité comme un séquençage single-end. Les lectures dans les deux sens ont ainsi été rassemblées en un seul fichier pour chacun des 23 individus examinés. Des paires de locus appartenant à la même région génomique (grande proximité géographique avec possibilité de recouvrement) ont donc pu être constituées au cours des analyses. Cependant, au vu du faible chevauchement obtenu et de la petite taille des lectures (Peterson et

al. 2012) par rapport à ces fragments, nous avons considéré qu'il s'agit de deux locus distincts d'une région génomique donnée.

### Reconstruction de locus :

L'identification de SNPs informatifs dans un cadre phylogénomique est assurée par la reconstruction de locus polymorphes chez chaque individu (locus intraindividuels), puis partagés entre individus (locus interindividuels) au moyen de « pipelines » adaptés et ajustés en fonction des jeux de données. Nous avons utilisé les deux « pipelines » les plus couramment employés dans la littérature : Stacks (Catchen et al. 2011) et PyRAD (Eaton and Ree 2013; Eaton 2014).

#### *PyRAD*

Les données générées par le séquençage Illumina ont été analysées avec le « pipeline » d'analyse PyRAD v3.0.66 (Eaton and Ree 2013; Eaton 2014). Nous avons testé les différents paramètres proposés par le « pipeline » pour estimer leurs impacts respectifs sur le jeu de données et retenu les suivants : minimum de quatre lectures pour former un locus (Mindepth), minimum de 85 % d'identité entre les lectures pour former un locus (WClust), minimum de 16 ou 20 individus pour former un locus partagé (MinCov) afin de minimiser les données manquantes. Nous estimons en effet que sous l'hypothèse que les deux spécimens de la même espèce soient représentés par les mêmes locus, un même locus comprendra les représentants de 8 ou 10 espèces différentes, soit un minimum de 70 ou 87 % de la diversité totale échantillonnée. Le reste des paramètres a été laissé avec les valeurs par défaut.

#### *Stacks*

Les données générées par le séquençage Illumina ont également été analysées avec le « pipeline » d'analyse Stacks v1.47 (Catchen et al. 2011). Nous avons testé les différents paramètres et retenu les suivants pour correspondre aux paramètres aussi retenus avec PyRAD : m (minimum de lectures pour former un locus, équivalent de Mindepth), M (distance autorisée entre piles, équivalent en partie de WClust), et n (distance autorisée par au catalogue de locus, équivalent en partie de WClust). Pour la constitution des locus intraindividuels (denovomap.pl) : m = 4, M = 20 (correspondant à 85 % d'identité de PyRAD) et n = 20. Pour la constitution des locus inter individuels (populations.pl) : (chaque spécimen correspondant respectivement à une population dans popmap.txt), m = 4 (équivalent de Mindepth), r = 1 et p = 16, 20 (équivalent de MinCov). Le reste des paramètres a été laissé avec les valeurs par défaut.

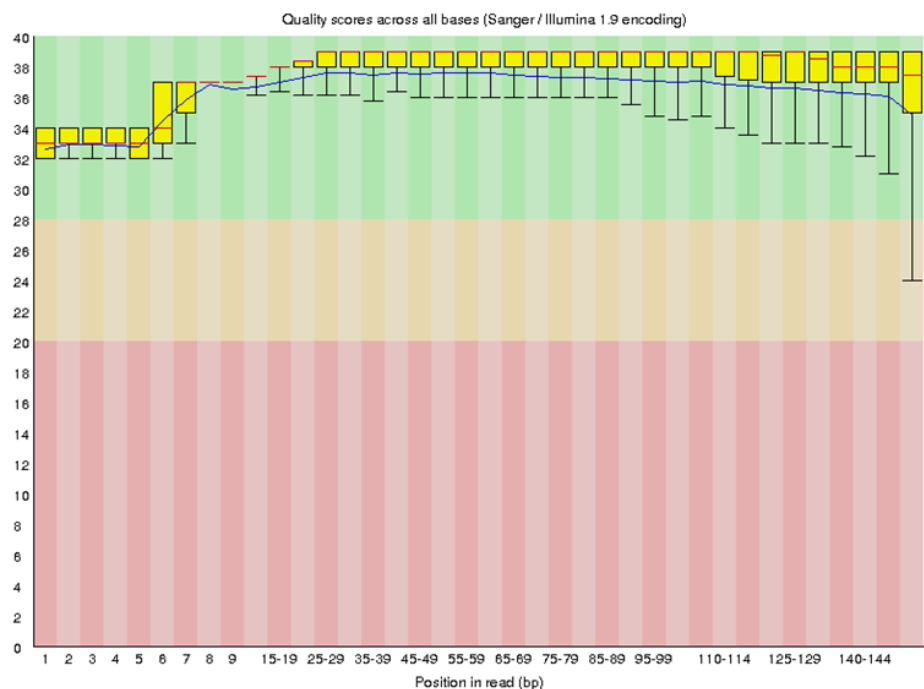
### Reconstructions phylogénétiques :

Comme conseillé pour les « pipelines » employés, les arbres phylogénétiques ont été générés par la méthode de maximum de vraisemblance avec le jeu de données assemblées à partir des multiples locus polymorphes partagés et triés au moyen du logiciel RaxML v8.2.12 (modèle de substitution pour séquences nucléotidiques GTRGAMMA, 1000 réplicats de bootstrap) (Stamatakis 2014). Deux méthodes différentes ont été employées : la concaténation (jeu de données de locus concaténés au format phylip généré à l'aide de GENEIOUS v9.0.2 (Kearse et al. 2012)) ou la congruence (arbres de gènes reconstruits pour chaque locus par RaxML v8.2.12, puis un super-arbre généré par l'outil ASTRALII (Mirarab and Warnow 2015). Des scripts élaborés grâce à l'aimable participation de Sophie Brouillet et Jawad Abdelkrim ont été mis au point pour traiter les étapes intermédiaires depuis la production des locus jusqu'aux reconstructions des topologies (données non publiées). Il s'agit des étapes de séparation des locus polymorphes parmi l'ensemble des locus sortis par PyRAD à partir du fichier de sortie .locus, de séparation des locus à partir du fichier de sortie fullseq.phylip de Stacks, et d'automatisation de la création d'arbres par RAXML pour chacun des locus séparés en vue d'analyses de congruence.

## III.C : Résultats

### III.C.1 : Génération du jeu de données pour former les locus

Un total de 18 430 943 lectures brutes distribuées parmi les vingt-quatre banques ont été obtenues par ce séquençage Illumina. La qualité des lectures a directement été évaluée au moyen d'analyses FastQC (Figure III.2). Chacune des 151 positions de ces lectures brutes présentent un score « confiance qualité » (ou « phred score ») supérieur à 32, situé dans la gamme la plus haute de cet indice (entre 28 et 40). Les cinq premières bases de la séquence correspondent aux adaptateurs P1 ligués au site de restriction EcoR1 permettant de reconnaître chaque individu au sein des pools. Elles semblent présenter un indice plus faible (entre 32 et 34) que le reste de la séquence, dont la majorité des bases ont un indice maximal (entre 38 et 40). On note également, comme attendu (Andrews 2016), une légère décroissance pour les dernières bases des lectures, dont l'indice qualité est plus variable.



**Figure III.2:** exemple d'un résultat d'une analyse FastQC sur lectures brutes à la sortie du run. Sortie des lectures « sens » des douze banques de ddRAD-seq portant l'index cinq analysées au moyen de l'outil FastQC sur le portail Galaxy. Les mêmes résultats ont été obtenus avec les lectures « antisens » et avec l'index 19 dans les deux sens.

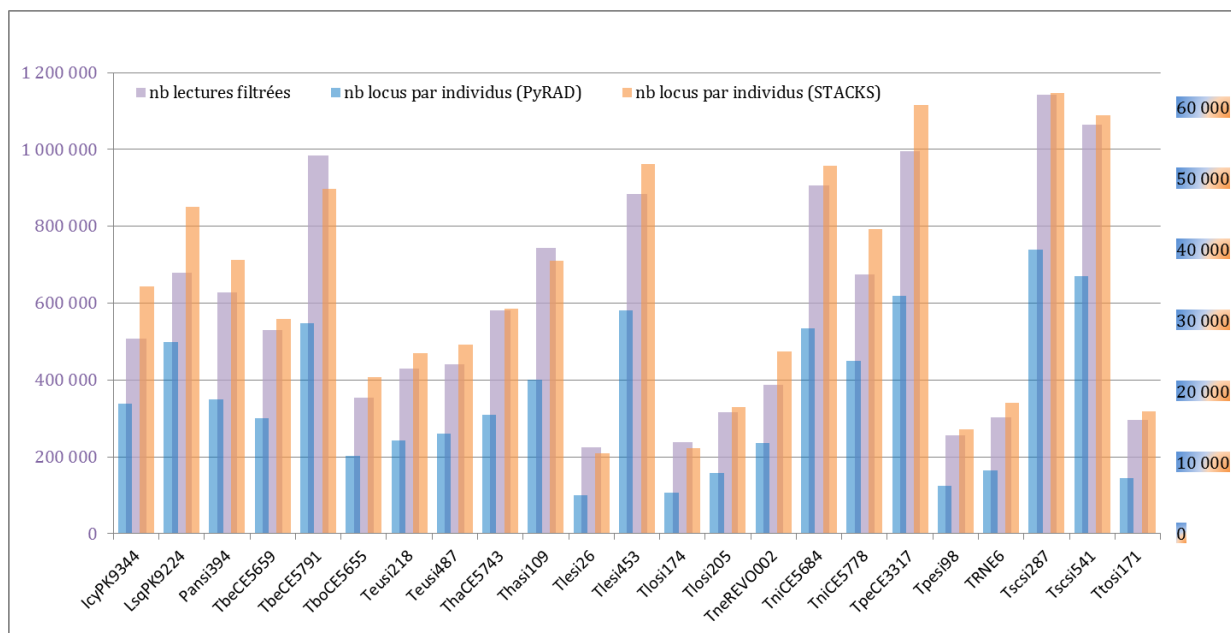
Les distributions de lectures brutes puis ayant passé les différentes étapes de filtres pour chaque individu sont présentées Tableau III.1.

| Espèce                  | Codes terrain | Nb reads bruts | filtre enzyme de restriction | filtre taille | filtre qualité |
|-------------------------|---------------|----------------|------------------------------|---------------|----------------|
| <i>T. bernachii</i>     | CE5791        | 2 658 924      | 2 634 396                    | 1 691 255     | 984 469        |
| <i>T. bernachii</i>     | CE5659        | 1 311 092      | 1 305 926                    | 909 483       | 529 655        |
| <i>T. pennellii</i>     | si98          | 630 110        | 628 092                      | 433 842       | 255 884        |
| <i>T. pennellii</i>     | CE3317        | 2 516 430      | 2 502 378                    | 1 676 515     | 996 239        |
| <i>T. borchgrevinki</i> | CE5655        | 953 156        | 949 184                      | 605 664       | 352 869        |
| <i>T. eulepidotus</i>   | si218         | 1 131 634      | 1 125 240                    | 717 233       | 430 313        |
| <i>T. eulepidotus</i>   | si487         | 1 100 228      | 1 095 938                    | 724 377       | 441 258        |
| <i>T. hansonii</i>      | si109         | 1 974 006      | 1 962 586                    | 1 286 715     | 744 568        |
| <i>T. hansonii</i>      | CE5743        | 1 535 896      | 1 529 854                    | 1 016 142     | 581 779        |
| <i>T. lepidorhinus</i>  | si26          | 624 712        | 619 592                      | 384 955       | 225 325        |
| <i>T. lepidorhinus</i>  | si453         | 2 205 020      | 2 196 154                    | 1 425 260     | 882 855        |
| <i>T. loennbergii</i>   | si205         | 862 694        | 856 274                      | 522 243       | 316 657        |
| <i>T. loennbergii</i>   | si174         | 651 430        | 647 378                      | 408 069       | 237 276        |
| <i>T. newnesi</i>       | TRNE6         | 856 568        | 849 122                      | 524 078       | 302 726        |
| <i>T. newnesi</i>       | REVO002       | 1 020 034      | 1 012 472                    | 627 377       | 387 190        |
| <i>T. nicolai</i>       | CE5684        | 2 465 632      | 2 451 666                    | 1 504 273     | 906 167        |
| <i>T. nicolai</i>       | CE5778        | 1 857 386      | 1 852 258                    | 1 121 141     | 674 246        |
| <i>T. scotti</i>        | si287         | 3 154 258      | 3 144 774                    | 1 898 820     | 1 142 945      |
| <i>T. scotti</i>        | si541         | 2 772 686      | 2 740 166                    | 1 803 205     | 1 064 658      |
| <i>T. tokarevi</i>      | si171         | 749 358        | 745 042                      | 485 142       | 296 467        |
| <i>I. cyanobranca</i>   | PK9344        | 1 368 946      | 1 364 778                    | 841 500       | 506 897        |
| <i>L. squamifrons</i>   | PK9224        | 1 802 444      | 1 790 348                    | 1 130 898     | 679 546        |
| <i>P. antarctica</i>    | si394         | 1 879 220      | 1 874 250                    | 1 051 368     | 627 297        |

**Tableau III.1 : distribution des lectures pour chacun des 23 individus échantillonnés.** La troisième colonne correspond aux lectures brutes directement issues de l'étape de démultiplexage. La quatrième colonne indique les lectures contenant le site de restriction EcoR1 (AATTC) en 5' des lectures « sens » (R1) et le site de restriction MspI (CGG) en 3' des lectures « inverses » (R2). Les cinquièmes et sixièmes colonnes indiquent le nombre de lectures constituant chaque banque à l'issue des étapes de filtres longueur et qualité. La sixième et dernière colonne (filtre qualité) correspond aux lectures filtrées constituant les banques utilisées pour les reconstructions de locus puis les analyses phylogénétiques.

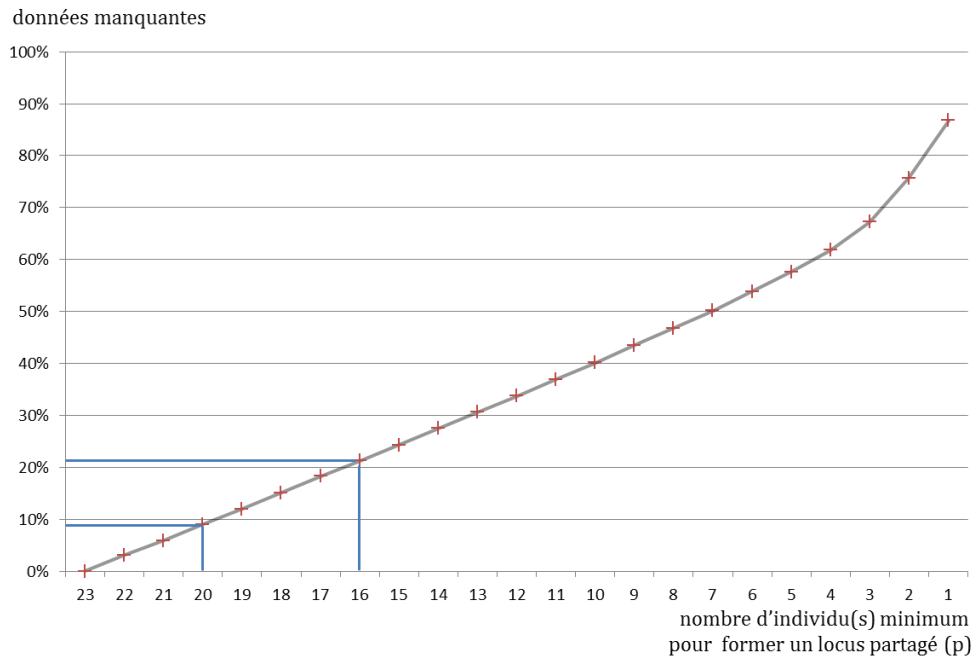
A l'issue des différents filtres (sites de restriction présents aux deux extrémités, taille, qualité), un minimum de 35,3 % (*T. newnesi*, TRNE6) à un maximum de 40,6 % (*T. pennellii*, si98) du nombre de lectures brutes ont été gardés pour la constitution des locus. Les couvertures des différentes banques filtrées sont relativement hautes (15 banques présentent une couverture supérieure à 400 000 lectures) et homogènes (moyenne de 565 304 lectures par banques), avec des minima situés autour de 240 000 (*T. lepidorhinus* si26, *T. pennellii* si98) et des maxima autour de 1 100 000 lectures. Le nombre de locus formés pour chaque individu est directement corrélé au nombre de lectures filtrées ( $r_{\text{Pearson}} = 0,9856$ ,  $p = 5.365 \text{ e-}18$  pour une corrélation positive avec les données pyRAD et  $r_{\text{Pearson}} = 0.9801$ ,  $p = 1.582 \text{ e-}16$  pour les données Stacks) (Figure III.3).





**Figure III.3 :** histogramme présentant la distribution des couvertures des banques filtrées (en nombre de lectures) et le nombre de locus correspondants par individus constitués par Pyrad ou Stacks. Le nombre de lectures filtrées est indiqué en violet (échelle de gauche) et le nombre de locus résultants obtenus pour chaque banque avec PyRAD ou Stacks sont indiqués respectivement en bleu et orange (échelle de droite).

Afin de pouvoir tester l'impact des données manquantes dans le jeu de données et en fixant les autres paramètres, nous avons fait varier le paramètre « Mindepth » (Pyrad) ou « p » (Stacks). La proportion en données manquantes a été calculée sur la base de l'histogramme de distribution des locus partagés à p donné (Figure III.4). Quelle que soit la méthode employée de reconstitution des locus, le nombre retenu de locus partagés est d'autant plus important que le paramètre p est faible. Dans ce cas, les données manquantes augmentent en proportion du jeu de données. Au contraire, le nombre de locus partagés retenus diminue lorsque p augmente (en moyenne 75 % de perte entre p16 et p20 pour PyRAD comme pour Stacks), réduisant cependant le nombre de données manquantes dans la partition (70 % des individus représentés au minimum par locus pour p16 et 87 % pour p20) (Figure III.4). En-dessous du seuil de 10 % de données manquantes, nous avons considéré que les données utilisées pour les reconstructions phylogénétiques risquent de ne pas représenter l'ensemble des sites polymorphes partagés pour les individus pris en compte dans l'analyse. Pour PyRAD comme pour Stacks, nous avons ainsi retenu les locus partagés par un minimum de vingt individus pour les analyses phylogénétiques.

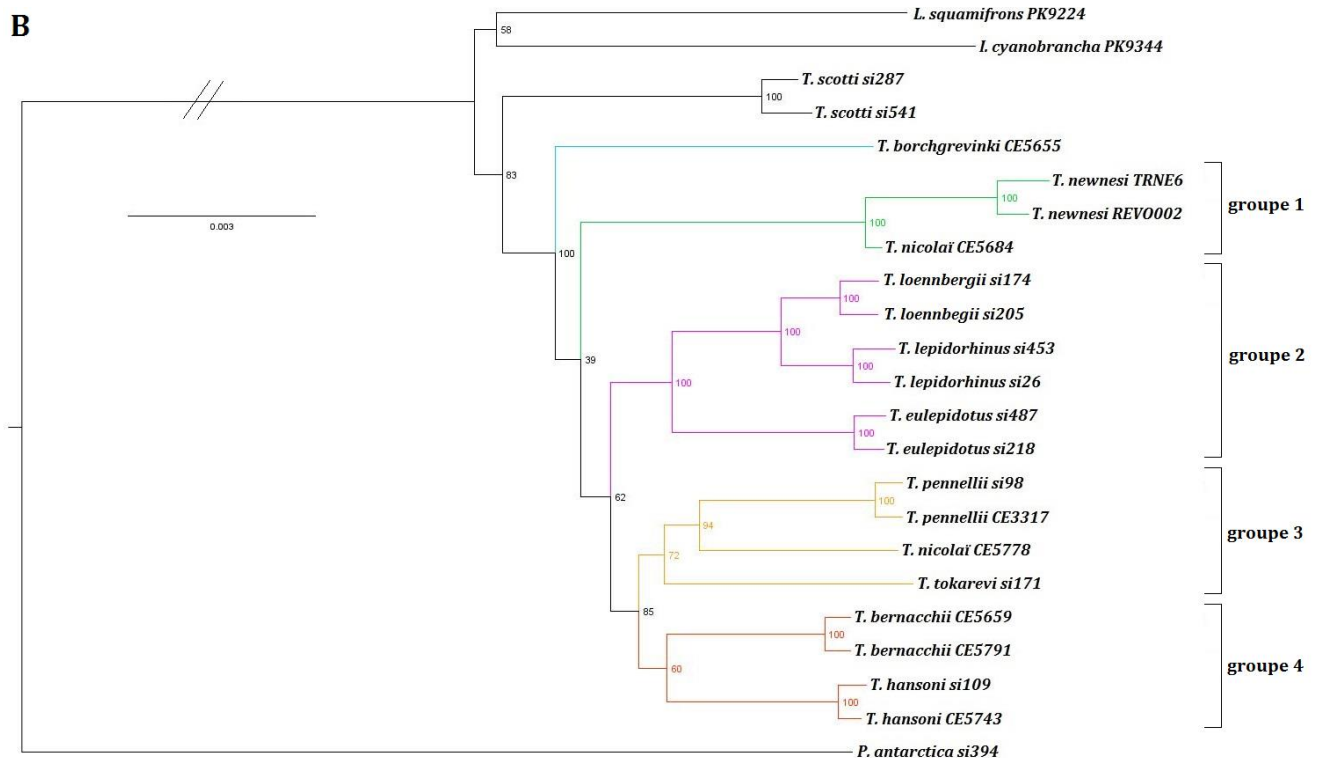
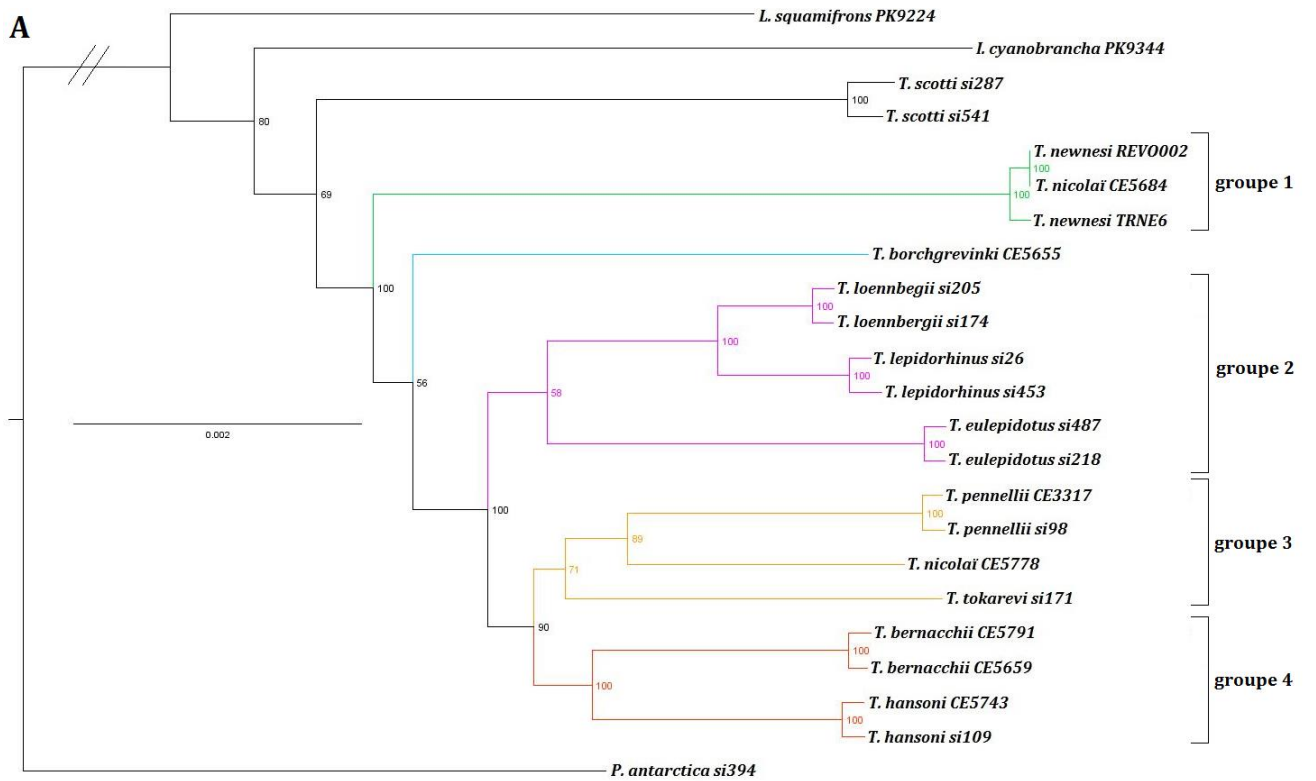


**Figure III.4 :** proportion des données manquantes (en %) en fonction de la valeur  $p$  choisie avec Stacks. L'évolution semble linéaire jusqu'à  $p_4$ . Sont représentés en bleu le paramètre choisi dans notre étude :  $p_{20}$  (9 % de données manquantes), ainsi que celui choisi à titre de comparaison :  $p_{16}$  (21 % de données manquantes).

Un fois les locus partagés obtenus, nous avons retenu uniquement les locus polymorphes qui ont fait l'objet d'un dernier filtre visant à éliminer les éléments transposables restants. Il semble en effet que PyRAD comme Stacks ne suppriment pas complètement les ETs. Cependant, le nombre d'éléments transposables retrouvés (en moyenne, de l'ordre de 1 % pour Pyrad et de 0,5 % pour Stacks) lors de cette étape est relativement faible comparé à leur représentation attendue dans les génomes de Nototheniidae (13 à 14 %, Detrich *et al.* (2010)) (Tableau III.2).

### III.C.2 : Reconstructions phylogénétiques

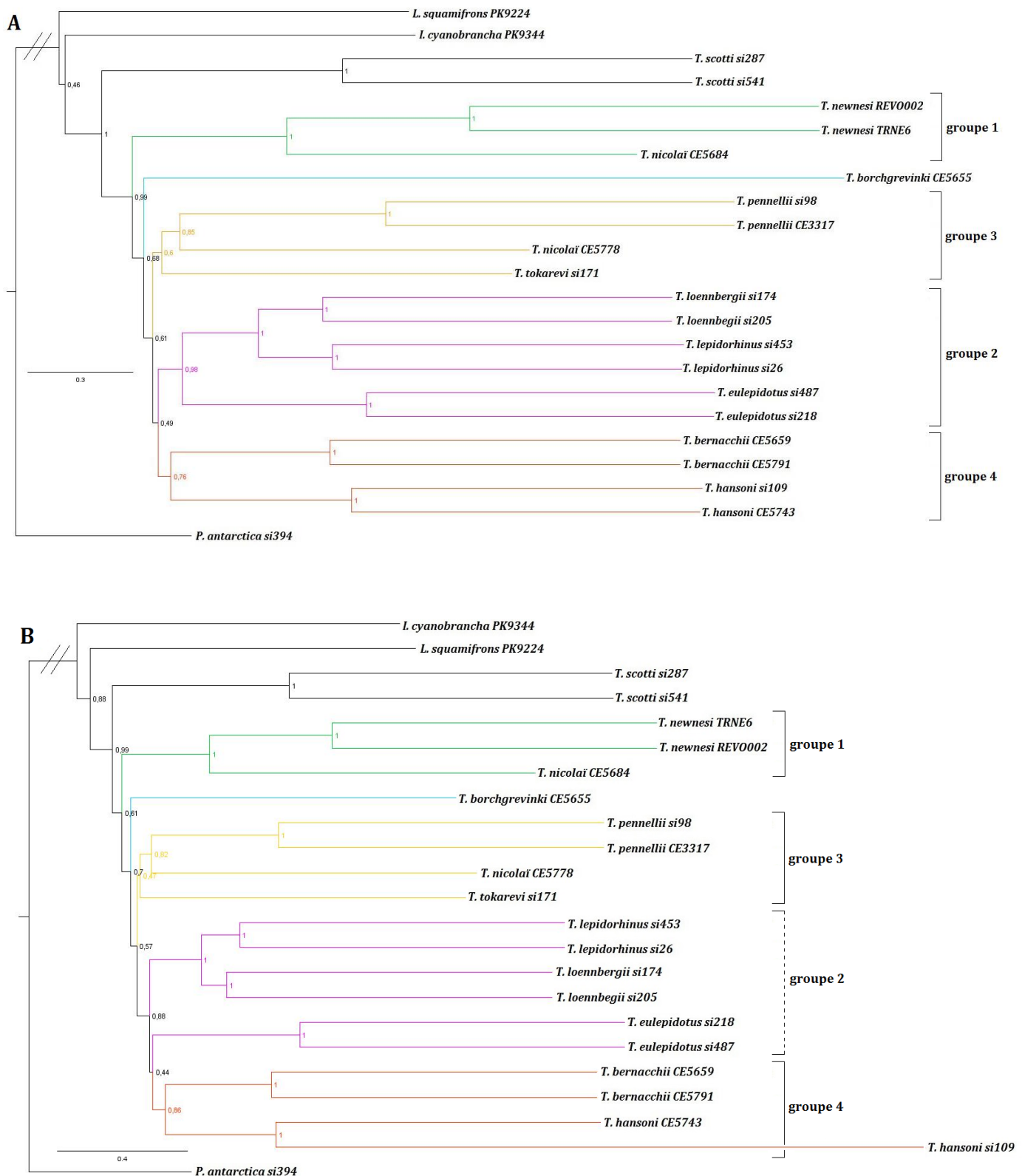
Les quatre topologies ont été obtenues à l'aide de RaxML (maximum de vraisemblance), sur les jeux de données générés par PyRAD ou Stacks, et par reconstruction avec des méthodes de concaténation (Figure III.5) ou de congruence (Figure III.6). Nos deux groupes externes sont représentés par le Trematominae *L. squamifrons* (PK9224) et le Pleuragraminae *P. antarctica* (si394). Nous avons raciné ces arbres phylogénétiques sur *P. antarctica* (si394), car c'est l'espèce la plus éloignée phylogénétiquement des « *Trematomus* » parmi ces deux groupes externes (Rutschmann et al. 2011; Dettai et al. 2012; Near et al. 2012; Colombo et al. 2015).



**Figure III.5 :** topologies obtenues par reconstructions avec les jeux de données issus de (A) Stacks ou (B) PyRAD et en employant la méthode de concaténation. Les bootstraps sont indiqués au niveau des nœuds. Pour faciliter la lecture, *T. borchgrevinki* est indiqué en bleu, le groupe un (*T. newnesi*/*T. nicolai* CE5684) en vert, le groupe deux (*T. eulepidotus*, *T. loennbergii*/*T. lepidorhinus*) en rose, le groupe trois (*T. pennellii*/*T. nicolai* CE 5778, *T. tokarevi*) en orange et le groupe quatre (*T.bernacchii* et *T. hansonii*) en rouge.

La reconstruction phylogénétique avec le jeu de données issu de Stacks suivie de la méthode de concaténation (Figure III.5.A) positionne à la base *L. squamifrons* (second groupe externe), représentant du groupe frère des « *Trematomus* » (*Lepidonotothen*/*Patagonotothen*). La monophylie du groupe « *Trematomus* » est soutenue (bootstrap de 80 %). La première espèce à émerger au sein du groupe est *L. cyanobrancha*, suivie par *T. scotti*, puis par un grand clade (soutenu à 100 %) formé par plusieurs groupes. Au sein de ce grand clade, on retrouve un ensemble monophylétique nommé « groupe un » formé par *T. nicolai* CE5684 et *T. newnesi*. Dans cette topologie, les deux individus *T. newnesi* ne forment pas un groupe monophylétique car *T. nicolai* CE5684 est positionné en groupe frère de l'individu *T. newnesi* REVO002. Emergent ensuite l'individu *T. borchgrevinki* CE5655, puis un autre ensemble monophylétique (soutenu à 100 %). Ce dernier est constitué d'un côté par un ensemble monophylétique nommé « groupe deux » et formé par les deux *T. loennbergii*, *T. lepidorhinus* et *T. eulepidotus*, et de l'autre côté par deux clades : un ensemble nommé « groupe trois » et formé par les deux *T. pennellii*, l'individu *T. nicolai* CE3317 et *T. tokarevi* si171, et enfin un dernier ensemble nommé « groupe quatre » et formé par les deux *T. bernacchii* et *T. hansonii*. Au sein du groupe deux, *T. lepidorhinus* et *T. loennbergii* sont groupes frères. De même pour le groupe trois avec *T. pennellii* et *T. nicolai* CE5778 formant un clade dont le groupe frère est *T. tokarevi*. Les individus appartenant à la même espèce sont toujours ensemble dans les arbres, excepté pour le cas de *T. newnesi* explicité ci-dessus et de *T. nicolai* dont un individu (CE5778) est retrouvé dans le groupe un, mais l'autre individu (CE5684) est retrouvé dans le groupe trois.

La reconstruction phylogénétique avec le jeu de données issu de PyRAD suivie de la méthode de concaténation (Figure III.5.B) positionne à la base *L. cyanobrancha* en groupe frère de *L. squamifrons* (deuxième groupe externe). Cependant, le clade formé par ces deux espèces est faiblement soutenu avec un nœud à 58 %. On ne retrouve donc pas de monophylie du groupe « *Trematomus* » ici. Les deux *T. scotti* se séparent ensuite, puis l'individu *T. borchgrevinki* CE5655. Cependant, le grand clade formé par les groupes un à quatre est beaucoup plus faiblement soutenu (39 %). Au sein du groupe un, les deux individus *T. newnesi* sont cette fois-ci monophylétiques, et groupes frères de l'individu *T. nicolai* CE5684. Les relations phylogénétiques entre et au sein des groupes deux, trois et quatre restent inchangées. Les valeurs de soutien des nœuds regroupant ces trois groupes, ainsi que les groupes trois et quatre sont plus faibles (62 et 85 % vs 100 et 90 %) que celles obtenues avec le jeu de données issu de Stacks (Figure III.5.A).



**Figure III.6 :** topologies obtenues par reconstructions avec les jeux de données issus de (A) PyRAD ou (B) Stacks et en employant la méthode de congruence. Les valeurs de soutien sont indiquées au niveau des nœuds. Pour faciliter la lecture, *T. borchgrevinki* est indiqué en bleu, le groupe un (*T. newnesi*/*T. nicolai* CE5684) en vert, le groupe deux (*T. eulepidotus*,*T. loenbergii*/*T. lepidorhinus*) en rose, le groupe trois (*T. pennellii*/*T. nicolai* CE 5778, *T. tokarevi*) en orange et le groupe quatre (*T. bernacchii* et *T. hansonii*) en rouge.

La reconstruction phylogénétique avec le jeu de données issu de PyRAD suivie de la méthode de congruence (Figure III.6.A) positionne à la base *L. squamifrons* (second groupe externe), représentant du groupe frère des « *Trematomus* » (*Lepidonotothen*/*Patagonotothen*). La monophylie du groupe « *Trematomus* » est cependant faiblement soutenue (bootstrap de 46 %) en comparaison du groupe excluant *I. cyanobrancha* (soutien du nœud regroupant toutes les autres espèces du groupe à 100 %). La première espèce à émerger au sein du groupe est *I. cyanobrancha*, suivie par *T. scotti*, puis par un grand clade (soutenu à 99 %) formé par plusieurs groupes. Au sein de ce grand clade, on retrouve le groupe un formé par *T. nicolai* CE5684 et les deux *T. newnesi*. Dans cette topologie, les deux individus *T. newnesi* sont monophylétiques. Emergent ensuite l'individu *T. borchgrevinki* CE5655, puis un autre ensemble monophylétique (soutenu à 61 %). Ce dernier est constitué d'un côté par le groupe trois (rassemblant les deux *T. pennellii*, l'individu *T. nicolai* CE3317 et *T. tokarevi* si171) et de l'autre côté par deux clades (soutien du nœud à 49 %) : le groupe deux (rassemblant les deux *T. loennbergii*, *T. lepidorhinus* et *T. eulepidotus*), et enfin le groupe quatre (rassemblant les deux *T. bernacchii* et *T. hansonii*). Au sein du groupe deux, *T. lepidorhinus* et *T. loennbergii* sont groupes frères. De même pour le groupe trois avec *T. pennellii* et *T. nicolai* CE5778 formant un clade dont le groupe frère est *T. tokarevi*. Les individus appartenant à la même espèce sont toujours ensemble dans les arbres, excepté *T. nicolai* dont un individu (CE5778) est retrouvé dans le groupe un, mais l'autre individu (CE5684) est retrouvé dans le groupe trois.

La reconstruction phylogénétique avec le jeu de données issu de PyRAD suivie de la méthode de congruence (Figure III.6.B) fait diverger *I. cyanobrancha* avant *L. squamifrons*. La monophylie du groupe « *Trematomus* » excluant *I. cyanobrancha* est toujours fortement soutenue (99 %). Emergent ensuite toujours le groupe un au sein duquel les relations phylogénétiques restent inchangées, suivi de l'individu *T. borchgrevinki* CE5655. Les relations phylogénétiques entre les groupes deux, trois et quatre restent également inchangées. Cependant, il est à noter que le groupe deux est éclaté car *T. eulepidotus* est positionné en groupe frère du groupe quatre (nœud faiblement soutenu à 44 %) et non du clade formé par les deux *T. lepidorhinus* et *T. loennbergii* dans cette topologie.

## III.D : Discussion

### III.D.1 : Jeux de données et paramètres d'analyse

L'approche de phylogénomique ddRAD-seq couplée au séquençage Illumina nous a permis de générer un jeu de données important et de qualité. La qualité des séquences obtenues, ainsi que l'homogénéité des longueurs des lectures et des couvertures de chaque banque sont autant de critères optimaux pour l'utilisation des outils d'analyse et des reconstructions phylogénétiques associées.

Nous avons utilisé les « pipelines » de reconstruction de locus les plus couramment employés dans la littérature : PyRAD (Eaton and Ree 2013; Eaton 2014) et Stacks (Catchen et al. 2011). Ces deux outils ne fonctionnent pas exactement de la même manière. Par exemple, PyRAD aligne les lectures alors que Stacks constitue des locus par simple empilement des lectures. De cette manière, PyRAD peut gérer les insertions/délétions, contrairement à Stacks. De plus, les étapes de constitutions des locus par individus puis des locus partagés s'enchaînent directement avec PyRAD, alors qu'elles sont entrecoupées par des étapes de constitution d'un catalogue général des locus identifiés par individus suivi d'un enrichissement par comparaison des lectures de chaque individu à ce catalogue avec Stacks. Cependant, dans la mesure du possible, nous avons choisi des paramètres communs comme critères pour former des locus, tels que le seuil minimum d'identité entre lectures (85 %) ou bien le nombre minimum de lectures requises (4) pour former un locus, ou encore le nombre d'individus minimum (20) requis pour former un locus partagé. En partant d'un jeu de données similaires en lectures filtrées et en réglant ces paramètres communs, les topologies obtenues à partir des jeux de données issus de ces deux approches sont les plus comparables possibles.

Nous avons choisi de ne pas dépasser les 10 % de données manquantes au sein du jeu de locus reconstitués avec Stacks en retenant un nombre minimum de vingt individus (p20) pour former un locus partagé. Etant donné que PyRAD ne fournit pas directement la distribution des locus en fonction du paramètre  $p$  et qu'il a généré environ la moitié de locus construit par Stacks (Tableau III.2), nous avons voulu estimer le chevauchement entre les jeux de locus constitués par ces deux logiciels. 71% des locus construits par PyRAD trouvent un équivalent dans le jeu de locus reconstitué par Stacks. Le jeu de locus PyRAD constitue donc, en majorité, un sous-ensemble du jeu de locus Stacks. La distribution en données manquantes estimées précédemment (Figure III.4) peut donc selon cette hypothèse aussi s'appliquer au jeu de locus PyRAD, pour lequel nous avons également utilisé le paramètre p20. Afin d'évaluer l'impact de la proportion des données manquantes sur le nombre de locus reconstitués et en conséquence sur

la reconstruction des relations phylogénétiques au sein du groupe « *Trematomus* », nous avons également fait des analyses avec le paramètre p16 (Tableau III.2).

|                                | p16    |        | p20   |        |
|--------------------------------|--------|--------|-------|--------|
|                                | PyRAD  | STACKS | PyRAD | STACKS |
| <b>nb locus partagés total</b> | 2 983  | 5 781  | 719   | 1 355  |
| <b>nb locus polymorphes</b>    | 2 807  | 4 256  | 666   | 1 001  |
| <b>nb locus ETs</b>            | 34     | 29     | 7     | 3      |
| <b>nb locus finaux</b>         | 2 773  | 4 227  | 659   | 998    |
| <b>nb SNPs total</b>           | 11 881 | 57 218 | 2828  | 13 166 |

**Tableau III.2 : récapitulatif des locus partagés constitués par PyRAD ou Stacks et du nombre de SNPs total du jeu de donné résultant en fonction des paramètres utilisés (minimum de p16 ou p20 individus sur 23 pour former un locus partagé).** Seuls les locus polymorphes partagés (locus finaux) ont été retenus après retrait des potentiels éléments transposables identifiés lors de la dernière étape de filtre. Le nombre total de SNPs correspondant à ces locus est indiqué dans la dernière ligne.

Notre filtre éléments transposables a permis d'éliminer des ETs putatifs non filtrés par les logiciels. Ils sont certes peu nombreux par rapport au nombre total de locus obtenus (0,5 à 1 %) pour p16 comme pour p20 (Tableau III.2), mais auraient pu constituer des biais dans les analyses de par leur capacité à bouger par transferts horizontaux (Sánchez-Gracia et al. 2005; Bartolomé et al. 2009; Schaack et al. 2010) ou leur présence en multiples copies dans les génomes étudiés (Detrich et al. 2010; Detrich and Amemiya 2010; Auvinet et al. 2018).

D'une manière générale, on note une proportion plus importante des locus polymorphes par rapport au nombre total de locus reconstitués par PyRAD (93 %) que par Stacks (73 %). Comme explicité précédemment, le nombre de locus reconstruits à l'aide de Stacks est environ deux fois plus élevé que celui reconstruits par PyRAD pour p16 comme pour p20. Cela peut en partie expliquer le grand nombre total de SNPs retrouvés dans les locus reconstruits par Stacks en comparaison à PyRAD (de l'ordre de cinq fois plus, Tableau III.2). De plus, il est à noter que si seuls les SNPs informatifs pour la phylogénie (partagés par au moins deux individus) sont indiqués par PyRAD, aucune précision n'est donnée dans Stacks. Le nombre de SNPs indiqué par Stacks dans le tableau III.2 peut donc être sur-évalué par rapport à celui indiqué par PyRAD, reflétant alors le nombre total de SNPs du jeu de données (informatifs et non informatifs). Quoiqu'il en soit, le nombre de sites variables informatifs retrouvé au moyen de ces méthodes haut débit est de l'ordre de cent à plusieurs dizaine de milliers de fois plus élevé que celui trouvé par des études uni ou multigènes publiées jusqu'à présent (Baird et al. 2008; Mardis 2008; Emerson et al. 2010; Hohenlohe et al. 2010; Davey et al. 2011), y compris chez les Nototheniidae (Near and Cheng 2008; Rutschmann et al. 2011; Dettai et al. 2012; Lautrédou et al. 2012). Ce haut degré de polymorphisme est crucial pour obtenir un signal phylogénétique et résoudre des nœuds au sein de cette radiation.



### III.D.2 : Influence des données manquantes

On retrouve une cohérence dans les résultats des analyses en fonction du paramètre  $p$ . Plus le nombre d'individus requis pour former un locus partagé est élevé ( $p20$  vs  $p16$ ), plus le nombre de locus reconstruits est faible (perte de 77 % des locus en moyenne dans les deux « pipelines », Tableau III.2), mais les données manquantes moindres (réduction de plus de la moitié en proportions, Figure III.4).

L'importance de l'influence des données manquantes lors de l'emploi d'approches NGS de type RAD-seq a été beaucoup discutée (Rubin et al. 2012; Eaton and Ree 2013; Anderson et al. 2017). Lorsqu'elle sont importantes en proportion dans le jeu de données, ces données manquantes peuvent influencer considérablement la résolution et le soutien des nœuds au sein d'une analyse phylogénétique, *a fortiori* dans le cadre d'une radiation (Hejnal et al. 2009; Wagner et al. 2013). Cependant, leur influence varie entre études suivant leurs proportions par rapport à l'échantillonnage, leurs distributions, la taille du jeu de données ou encore la rétention de polymorphisme ancestral (Xi et al. 2016). Elles peuvent être dues à plusieurs facteurs inhérents à ce type d'approche, comme par exemple les mutations dans les sites de restriction, la qualité des séquences, la variabilité permise pour un site donné au sein d'un locus, ou encore la couverture minimale paramétrée pour former un locus (MinCov/m) (Huang and Knowles 2016). Cette influence a été testée ici par comparaison entre les résultats obtenus avec les paramètres MinDepth/ $p$  de 16 ou 20 (21 ou 9 % de données manquantes, Figure III.4) (Tableau III.3). Les topologies reconstruites avec le paramètre  $p16$  sont présentées en annexe 3.

| méthode de reconstruction   | concaténation / congruence |           |            |            |
|---|----------------------------|-----------|------------|------------|
|   | PyRAD p16                  | PyRAD p20 | Stacks p16 | Stacks p20 |
| 1 : <i>T. newnesi</i> , <i>T. nicolai</i> CE5684                              | 100 / 1                    | 100 / 1   | 100 / 1    | 100 / 1    |
| 2 : <i>T. eulepidotus</i> ,<br><i>T. lepidorhinus</i> / <i>T. loennbergii</i> | 100 / 1                    | 100 / 0,6 | 100 / 0,96 | 58 / non   |
| 3 : <i>T. nicolai</i> CE5778 / <i>T. pennellii</i> ,<br><i>T. tokarevi</i>    | 100 / 0,78                 | 72 / 0,98 | 100 / non  | 71 / 0,47  |
| 4 : <i>T. bernacchii</i> / <i>T. hansonii</i>                                 | 100 / 0,93                 | 60 / 0,76 | 100 / 1    | 100 / 0,86 |

**Tableau III.3 :** présence des groupes monophylétiques un à quatre dans les huit analyses phylogénétiques. Les sous-clades au sein d'un même groupe sont séparés par des slash. Lorsque ces groupes sont présents, le soutien des nœuds est indiqué dans la case correspondante en rouge pour les reconstructions par méthode de concaténation et en bleu par la méthode de congruence. Un non signifie que ce groupe est paraphylétique pour l'analyse donnée.

Ces valeurs de données manquantes sont très faibles comparées aux autres études en RAD et ddRAD-sequencing, où les données manquantes représentaient 28 à 81 % sont observées des jeux de données (Dunn et al. 2008; Hejnol et al. 2009; Wickett et al. 2014; Xi et al. 2014). Cette faible proportion s'explique sans doute par la qualité des lectures obtenues par le séquençage Miseq Illumina, la conservation des sites de restriction entre espèces et par nos paramètres choisis pour constituer les locus.

Les topologies reconstruites à l'aide du même logiciel (pyRAD ou Stacks) et de la même méthode (concaténation ou congruence) ne sont pas toujours identiques selon qu'on a choisi le paramètre p16 ou p20. Par exemple, la position de *T. borchgrevinki* varie selon le nombre d'espèces minimum devant figurer dans les locus partagés. *T. borchgrevinki* est ainsi tantôt retrouvée émergente juste après *T. scotti* et *I. cyanobranca* (Stacks concaténation p16, annexe 3), tantôt incluse dans le clade défini par les autres espèces du groupe « *Trematomus* » comprenant *T. newnesi*, *T. nicolai*, *T. pennellii*, *T. tokarevi*, *T. eulepidotus*, *T. loennbergii*, *T. lepidorhinus*, *T. bernacchii* et *T. hansonii* (Stacks concaténation p20, Figures III.5.A). C'est également le cas pour les positions relatives des groupes deux (*T. pennellii*/*T. nicolai* CE5778, *T. tokarevi*), trois (*T. lepidorhinus*/*T. loennbergii*, *T. eulepidotus*) et quatre (*T. bernacchii*, *T. hansonii*) (PyRAD congruence p16/p20, Figure III.6.A et annexe 3).

On note des valeurs de bootstrap inférieures à 86 % pour les nœuds mettant en jeu les groupes deux, trois et quatre dans les arbres reconstruits avec les jeux de données p16 comme p20 (Tableau III.3). Ces valeurs de soutien paraissent d'autant plus faibles que le paramètre p augmente, ce qui n'empêche pas d'observer des nœuds non résolus avec p16 (groupe 3, Stacks, congruence). Les locus partagés ne contiennent en effet pas forcément le même signal phylogénétique et peuvent même parfois se contredire (Tableau III.3), donnant lieu à la

diminution des valeurs de bootstrap observées pour ces nœuds et à l'augmentation du nombre de polytomies (Tableau III.3, Figures III.5 et III.6). Ces problèmes de congruence entre ces locus polymorphes partagés par un grand nombre d'individus pourraient refléter un problème de coalescence (rétention de polymorphisme ancestral) propre aux génomes de la radiation du groupe « *Trematomus* », déjà explicité dans la littérature (Janko et al. 2011; Lautrédou et al. 2012).

C'est pour cette raison que nous avons également voulu faire appel à des méthodes de reconstruction phylogénétiques par des méthodes de congruence et non seulement par concaténation. Par cette méthode, les locus ne sont pas fusionnés les uns à la suite des autres. Au contraire, ils sont pris en compte un à un au moyen de la construction « d'arbres de gènes ». La constitution puis la confrontation de combinaisons de sous-arbres (probabilité pour un quartet d'arbres de gènes d'être associé à une tripartition) permettent théoriquement de dégager un signal pondéré sur la totalité du jeu de données (Mirarab et al. 2014; Mirarab and Warnow 2015). A travers les quatre analyses par congruence, nous observons que les groupes un à quatre sont retrouvés dans la plupart des cas (exceptés pour le groupe trois dans l'analyse Stacks, congruence, p16, et le groupe deux dans l'analyse Stacks, congruence, p20 ; Tableau III.3, Figures III.6 et annexe 3). Le support des nœuds est globalement plus faible que celui observé par la méthode de concaténation, mais reste supérieur à 78 % (exceptés pour le groupe deux soutenu à 60 % avec l'analyse PyRAD, congruence, p20, et le groupe trois soutenu à 47 % avec l'analyse Stacks, congruence, p20, Tableau III.3). N'étant pas biaisé par un signal majoritaire risquant de masquer des signaux discordants, ces reconstructions pourraient refléter de manière plus juste le support « réel » de chaque nœud (Mirarab et al. 2014; Mirarab and Warnow 2015).

L'importance des données manquantes est à prendre en compte dans les analyses car il influe au même titre que le nombre total de sites polymorphes partagés sur la reconstruction des locus et des liens phylogénétiques. Il s'agit donc dans ce type d'approche de trouver un compromis entre l'importance du signal (nombre de locus polymorphes reconstitués) et le partage de ce signal (nombre de locus partagés par un maximum d'individus). La comparaison des topologies obtenues en fonctions des différents paramètres testés au cours de cette étude permet de dégager un signal phylogénétique robuste au changement de jeu de données.

### III.D.3 : Résolution des relations phylogénétiques chez les « *Trematomus* »

Le pourcentage de nœuds retrouvés d'une analyse à l'autre (supérieur à 75 %) est bien supérieur à celui des nœuds où il subsiste une ambiguïté en prenant en compte tous les types d'analyses confondues (Figures III.5, III.6 et annexe 3). Les nœuds terminaux qui regroupent deux individus de la même espèce sont quasiment tous résolus avec 100 % de valeur de bootstrap (excepté pour le regroupement des deux *T. newnesi* pour l'analyse Stacks, concaténation, p16, annexe 3). On note toutefois une exception récurrente avec la non-monophylie des *T. nicolai* examinés. La position de cette espèce dans l'arbre est impossible à établir à partir de la littérature étant donné les multiples topologies contradictoires publiées. *T. nicolai* est tantôt positionné groupe frère de *T. eulepidotus* (Bargelloni et al. 1994), de *T. pennellii* (Kuhn and Near 2009), de *T. borchgrevinki* (Sanchez et al. 2007), ou encore de *T. bernacchii* (COI, Lautrédou et al. (2010)), et parfois même retrouvé au sein d'une polytomie (Ritchie et al. 1996; Near 2004; Near and Cheng 2008). En prenant en compte un nombre plus faible d'individus pour former un locus comme p6 ou p12, *T. nicolai* CE5778 se retrouve parfois groupe frère du groupe un (*T. nicolai* CE5684, *T. newnesi* REVO002/*T. newnesi* TRNE6), alors que la position de l'individu *T. nicolai* CE5684 ne bouge pas (topologies non montrées). Ces deux spécimens font partie des banques ayant des couvertures hautes (supérieures à 650 000 lectures filtrées, Tableau III.1), mais la banque représentant l'individu *T. newnesi* CE5684 est environ un tiers plus couverte, contrairement au deuxième individu *T. nicolai* CE5778. La possibilité d'un hybride entre *T. pennellii* et *T. nicolai* n'est pas à exclure pour ce dernier. La vérification de l'identification des individus a été refaite par une amplification du cytochrome oxydase I (COI) des individus *T. nicolai* CE5684 et CE5778, *T. newnesi* TRNE6 et REVO002, et *T. pennellii* CE3317 et si98 (mêmes extraits que ceux ayant servi pour la constitution des banques de ddRAD-seq) et puis la comparaison des séquences par BLAST avec la base de données existante chez les « *Trematomus* » dans la base de données « Barcode of Life » ou « National Center for Biotechnology Information » (ncbi) (Lautrédou et al. 2010). Afin de pouvoir faire des hypothèses argumentées sur les positions phylogénétiques des deux individus *T. nicolai*, il faudrait analyser plus précisément l'histoire racontée par chaque locus. Cela permettrait d'estimer le pourcentage de locus positionnant *T. nicolai* en groupe frère de *T. newnesi* et celui positionnant *T. nicolai* en groupe frère de *T. pennellii*.

Grâce à cette approche de ddRAD-seq, nous avons identifié, pour la première fois dans les reconstructions des relations phylogénétiques de cette radiation des liens de parentés robustes et fiables (peu d'influence des paramètres variables sur la présence et la résolution des nœuds définissant les liens phylogénétiques au sein et entre les groupes définis, valeurs de bootstrap

des nœuds supérieures à 75 % dans la majorité des cas, Tableau III.3, Figure III.5, III.6 et annexe 3). *I. cyanobrancha* et *T. scotti* semblent bien à la base du groupe « *Trematomus* », ce qui va tout à fait dans le sens des études précédemment publiées. Cependant, *I. cyanobrancha* est parfois retrouvé avant ou avec le deuxième groupe externe *L. squamifrons*. Dans ces cas, le support des nœuds est plus faible. Un échantillonnage des groupes externes plus dense pourrait aider à lever cette ambiguïté concernant sa position ou non au sein du groupe « *Trematomus* ». La distribution subantarctique de cette espèce, contrairement à la répartition des autres « *Trematomus* » en grande majorité inféodés au plateau continental antarctique pourrait expliquer le faible nombre d'études phylogénétiques l'incluant dans l'échantillonnage. Emergent ensuite *T. borchgrevinki* ainsi que le clade des *T. newnesi* et du *T. nicolai* CE5684 (groupe un). Un grand clade regroupe les espèces restantes, distribuées dans les groupes deux, trois et quatre. Au sein du groupe deux, les *T. eulepidotus* sont systématiquement groupes frères du clade formé par les *T. loennbergii* et les *T. lepidorhinus*. Dans le groupe trois, *T. tokarevi* est groupe frère du clade formé par *T. nicolai* CE5778 et les *T. pennellii*. Enfin, dans le groupe quatre, les *T. bernacchii* sont associés aux *T. hansonii*. Nous retrouvons donc bien les rares regroupements consensus publiés : *T. scotti* et *I. cyanobrancha* en position basale dans l'arbre des « *Trematomus* » (Bargelloni et al. 1994; Ritchie et al. 1996; Ritchie et al. 1997; Bargelloni et al. 2000; Stankovic et al. 2002; Near 2004; Sanchez et al. 2007; Dettai et al. 2012), positionnement en groupes frères des espèces *T. loennbergii* et *T. lepidorhinus* (Ritchie et al. 1997; Sanchez et al. 2007; Near and Cheng 2008; Lautrédou et al. 2010; Janko et al. 2011; Dettai et al. 2012; Lautrédou et al. 2012; Murphy 2015), ainsi que *T. bernacchii* et *T. hansonii* (Ritchie et al., 1996; Near et al., 2004; Sanchez et al., 2007; Kuhn et Near, 2009 : 16S/ND2, Janko et al., 2011; Lautrédou et al., 2012 : COI/HECW/Pkd1/Rhodopsine; Colombo et al., 2015). Les nœuds des groupes un à quatre sont soutenus (Figure III.3) et retrouvés dans la grande majorité des topologies reconstruites.

Les « *Trematomus* » présentent la plus grande diversité écologique observée parmi les Nototheniidae (di Prisco et al. 1991; Eastman and Devries 1997; Eastman and McCune 2000; Eastman 2005). Certains regroupements de notre étude correspondent à des espèces partageant des caractéristiques écologiques précises associées à leur mode de vie. C'est le cas de *T. bernacchii* et *T. hansonii* (constituant le groupe quatre), toutes deux épibenthiques. Ce sont des prédateurs pourvus de grands yeux, et nageant lentement et souvent sur le substrat (di Prisco et al. 1991; Eastman and Devries 1997; La Mesa et al. 2008). C'est également le cas pour *T. pennellii*, *T. nicolai* et *T. tokarevi* (constituant le groupe trois) qui sont des espèces plutôt côtières, peu profondes et démersales (di Prisco et al. 1991; Eastman and Devries 1997; Eastman 2017). Enfin, les espèces *T. eulepidotus*, *T. loennbergii* et *T. lepidorhinus* (constituant le groupe deux) sont toutes les trois benthopélagiques (di Prisco et al. 1991), même si *T. loennbergii* a une distribution bathymétrique différente (nettement plus profonde) que les

deux autres (Eastman 2017). Certaines espèces présentent un mode de vie très particulier et sont retrouvées dans des clades à part dans nos analyses. C'est le cas avec *T. newnesi* qui fait preuve d'une grande plasticité phénotypique et trophique, présentant des particularités morpho-anatomiques, physiologiques et comportementales (Eastman and Devries 1997; Eastman and Hubold 1999) qui sont en accord avec son branchement dans un groupe à part (groupe un). Epibenthique, il se nourrit tantôt sur le fond, tantôt en pleine eau (Eastman and Devries 1997). Enfin, *T. borchgrevinki* présente la particularité d'être cryopélagique, se déplaçant sous la glace et avec la glace dérivante en s'y nourrissant (Eastman and DeVries 1985).

Au sein du grand clade formé par *T. newnesi*, *T. nicolai*, *T. borchgrevinki*, *T. pennellii*, *T. tokarevi*, *T. eulepidotus*, *T. lepidorhinus*, *T. loennbergii*, *T. bernacchii* et *T. hansonii*, les résultats sont congruents et positionnent ensemble les groupes trois et quatre en groupes frères du groupe deux. Cependant, avec les analyses PyRAD comme avec Stacks, les analyses de congruence à p20 montrent plus d'incertitude des nœuds pour ce grand clade. Les groupes deux (*T. eulepidotus*, *T. lepidorhinus*/*T. loennbergii*) et quatre (*T. bernacchii*, *T. pennellii*) sont associés avec des valeurs de bootstrap faible (49 % pour l'analyse PyRAD, Figure III.6.A). Le groupe deux est même éclaté avec l'analyse Stacks, mais *T. eulepidotus* est rattaché au groupe quatre avec une valeur de soutien très faible de 44 % (Figure III.6.B). Cela pourrait être dû aux problèmes de coalescence mis en évidence dans ces génomes auparavant (Lautrédou et al., 2012). De plus, le filtre effectué par PyRAD comme par Stacks a peut être laissé passer des locus paralogues, biaisant ainsi les reconstructions phylogénétiques. Sans génomes de référence, nous n'avons pas pu vérifier cette étape de filtre.

Il semble que le polymorphisme obtenu avec cette approche haut débit soit suffisant pour dégager un signal phylogénétique clair et relativement robuste pour la plupart des nœuds (Tableau III.3, Figures III.5, III.6 et annexe 3). Cependant, plusieurs approches pourraient permettre d'affiner ces relations de parenté et d'identifier précisément les conflits entre locus imputables aux temps de coalescence différents des temps de divergence des espèces. Il conviendrait par exemple d'enrichir cet échantillonnage en augmentant le nombre d'individus par espèce, tout en gardant une couverture suffisante des banques par individus pour obtenir un maximum de locus polymorphes partagés. De plus, il serait aussi intéressant d'estimer la proportion de locus plus proches de chaque groupe précédemment défini pour chaque individu (comme proposé plus tôt pour *T. nicolai*). Enfin, la disponibilité d'un génome de référence bien assemblé et annoté permettrait d'effectuer un tri plus fiable des locus contenant d'éventuels paralogues non filtrés par PyRAD et Stacks (Emerson et al. 2010; Hohenlohe et al. 2010; Pfender et al. 2011; McCluskey and Postlethwait 2015).



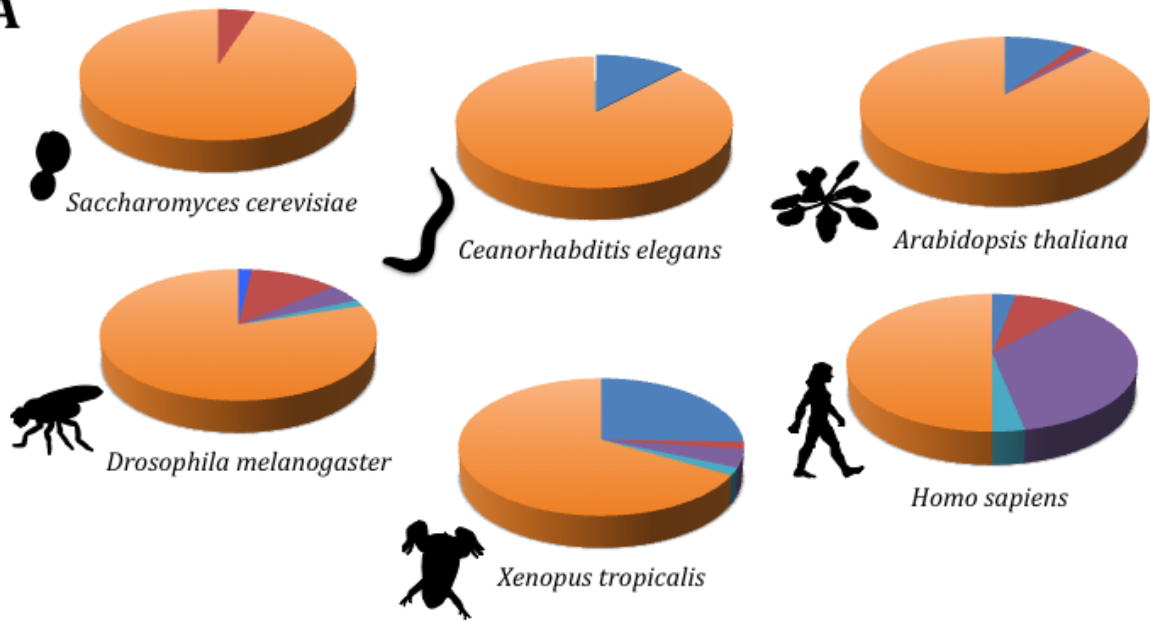
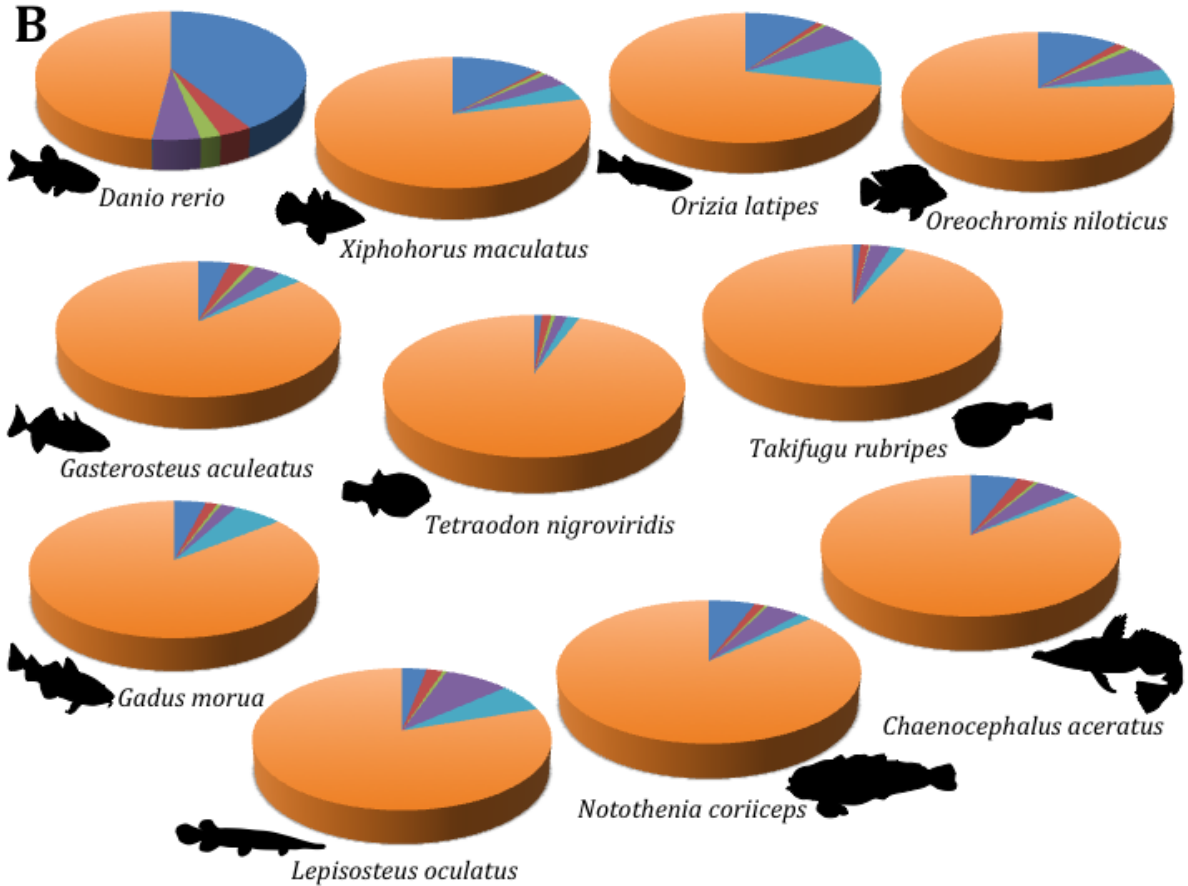
## IV) Chapitre 3 : les éléments transposables, facilitateurs de ces remaniements au cours de l'évolution?

### IV.1 : Introduction

Les éléments transposables (ETs) sont une composante majeure des génomes Eucaryotes (Brosius 2003; Deininger et al. 2003; Kazazian 2004; Oliver and Greene 2009; Rebollo et al. 2010). Ils sont ubiquitaires et représentent une part variable de ces génomes (Kidwell 2002; Biémont and Vieira 2005; Sotero-Caio et al. 2017). Cette abondance en ETs est inférieure à 2 % dans les génomes des poissons-globes *T. rubripes* et *T. nigroviridis* (Elgar 1999; Crollius 2000; Volff et al. 2003), qualifiés de génomes « compacts ». Elle varie entre 2 et 16 % chez la levure *S. cerevisiae* (Kim et al. 1998), le nématode *C. elegans* (The *C. elegans* Sequencing Consortium 1998), les angiospermes *A. thaliana* (The Arabidopsis Genome Initiative 2000) et *O. sativa* (Yu 2002; Turcotte et al. 2008), et la drosophile *D. melanogaster* (Biémont and Vieira 2005; Biémont and Vieira 2006). Les ETs peuvent aussi être beaucoup plus abondants : ils constituent près de 55 % des génomes du poisson zèbre *D. rerio* (Howe et al. 2013) ou de l'homme *Homo sapiens* (de Koning et al. 2011), 77 % du génome de la grenouille verte *Rana esculenta* (Pontecorvo et al. 2000) et jusqu'à 85 % des génomes, comme c'est le cas chez le maïs *Z. mays* (Schnable et al. 2009) (Figures IV.1.A et B).

**Figure IV.1 (page suivante): diversité et représentation des grandes super-familles d'éléments transposables retrouvées dans les génomes (A) de divers organismes Eucaryotes (B) en particulier chez les téléostéens.** Les éléments de classe II sont regroupés en une seule catégorie. Les éléments de classe I sont partagés en différents grands ordres. Les données des proportions en différentes catégories sont extraites de la littérature pour la levure (Kim et al. 1998), le nématode (Stein et al. 1998; The *C. elegans* Sequencing Consortium 1998), l'arabette (The Arabidopsis Genome Initiative 2000), la drosophile (Biémont and Vieira 2005), le xénope (Hellsten et al. 2010), l'homme (The Chimpanzee Sequencing and Analysis Consortium 2005; The 1000 Genomes Project Consortium 2012), le poisson zèbre (Howe et al. 2013), ainsi que le platy, le medaka, le tilapia, l'épinoche, les poissons globes, la morue et la lépisostée (Chalopin et al. 2015), et enfin les deux téléostéens antarctiques « black rockcod » et « blackfin icefish » (Detrich et al. 2010; Detrich and Amemiya 2010).



**A****B**

Lorsqu'ils sont mobilisés, les ETs se déplacent et se multiplient au sein des génomes (McClintock 1950; McClintock 1984), on les retrouve donc naturellement dans différents chromosomes d'une même espèce. Longtemps considérés comme des éléments égoïstes et parasites des génomes (« junk » ou « selfish DNA ») (Biémont and Vieira 2006; Muotri et al. 2007; Sotero-Caio et al. 2017), il est avéré aujourd'hui qu'ils ont un impact fort sur leur structuration, leur fonction, et leur plasticité (Wilhelm and Wilhelm 2001; Volff 2005; Böhne et al. 2008; Raskina et al. 2008; Kraaijeveld 2010; Howe et al. 2013). Par leur propriété auto-répliatrice, ils génèrent ainsi un large panel de modifications à l'échelle génomique allant de la simple insertion à des remaniements importants (Bourque 2009). Leur mobilité génère ainsi de la variabilité génétique sur laquelle différentes forces évolutives peuvent s'exercer (van de Lagemaat et al. 2003; Sela et al. 2010; Warren et al. 2015). Si leurs insertions dans ou à proximité des gènes sont souvent défavorable et donc susceptible d'être éliminée par la sélection naturelle (Kidwell 2002; Brookfield 2005; Chénais et al. 2012), les ETs peuvent aussi créer de la nouveauté génétique qui peut être positivement sélectionnée (Böhne et al. 2008; Kraaijeveld 2010; Rebollo et al. 2010). C'est le cas par exemple chez les mammifères avec l'« exonisation » de certains ETs (Sorek 2007; Sela et al. 2010) (Sorek, 2007, Sela et al., 2010) enrichissant le transcriptome, ou encore la « domestication » : recombinaison V(D)J chez l'homme, (Sadofsky 2001) ou la fonction télomérase chez la drosophile (Levis et al. 1993; Zhang and Rong 2012). Les ETs peuvent aussi modifier l'expression génique par « cis-régulation » (Feschotte 2008; Bourque 2009; Chénais et al. 2012) en s'insérant dans des régions 5' UTR (van de Lagemaat et al. 2003; Teng et al. 2011) ou en jouant le rôle de nouveaux promoteurs (Cohen et al. 2009). Enfin, leur mobilité peut induire de la recombinaison ectopique due à des cassures localisées d'ADN double brins associées à des mécanismes de réparations (Volff 2005; Bonnivard and Higué 2009). Dans ce dernier cas, ils sont alors susceptibles de jouer un rôle dans les réarrangements chromosomiques (Lönnig and Saedler 2002; Raskina et al. 2008; Rebollo et al. 2010; Valente et al. 2011; Chénais et al. 2012) qui peuvent être positivement ou négativement sélectionnés.

Le contrôle de la mobilisation des ETs est principalement d'ordre épigénétique (Biémont and Vieira 2006; Mansour 2007; Slotkin and Martienssen 2007; Rebollo et al. 2010; Rey et al. 2016). Ces séquences ont tendance à former des îlots d'accumulations (ou « hot spot d'insertion ») dans des régions hétérochromatiques des chromosomes (Pimpinelli et al. 1995; Biémont and Vieira 2005; Biémont and Vieira 2006; Mazzuchelli and Martins 2009; Schneider et al. 2013). Leur régulation inhibitrice s'effectue au niveau transcriptionnel par la méthylation de l'ADN et la compaction de la chromatine (Rebollo et al. 2010). Les ETs sont également inactivés au niveau post-transcriptionnel par l'intervention de petits ARN non codants interférants de type piRNA (Lisch 2009; Obbard et al. 2009). Cependant, ces processus sont coûteux en énergie. Lors d'un stress, notamment de type thermique, les rétrotransposons sont fortement mobilisés (Giraud

and Capy 1996; Mansour 2007; Piednoël and Bonnivard 2009). Ces mécanismes de verrou sont alors partiellement levés, ce qui peut entraîner de nouvelles vagues de transpositions dans les génomes (appelés « burst ») (Böhne et al. 2008; Oliver and Greene 2009; Rebollo et al. 2010; Belyayev 2014). C'est par exemple le cas chez le Medaka *O. latipes* avec la forte accumulation des éléments *Tol1* et *Tol2* (Koga et al. 2007; Watanabe et al. 2014). Ces accumulations peuvent être détectées sur les chromosomes avec des « points chauds » d'insertions (Bennetzen 2000; Capriglione et al. 2002; Slotkin and Martienssen 2007; Mazzuchelli and Martins 2009; Valente et al. 2011; Schneider et al. 2013) constituant en fait des points chauds de cassures suivies de réparations (Bourque 2009; Oliver and Greene 2009). C'est dans ce contexte que les ETs peuvent potentiellement être associées à des phénomènes de remaniements chromosomiques et renforcer une barrière reproductive menant à la spéciation (Lim and Simmons 1994; Lönnig and Saedler 2002; Raskina et al. 2008). Il est cependant difficile de corrélérer *a posteriori* une mobilisation massive d'ETs et des phénomènes de diversifications, leur rôle en tant que médiateur de la spéciation est encore assez débattu dans la littérature.

Les éléments transposables présentent une grande diversité (Finnegan 1989; Kapitonov and Jurka 2008). On peut les regrouper en deux grandes catégories selon leur structure et leur mode de transposition (Curcio and Derbyshire 2003). Les ETs de classe II, appelés transposons à ADN, transposent par un mécanisme de type couper/coller. Dans cette première catégorie, on distingue les éléments autonomes (TIR : *Tc1-Mariner*, *hAT*, *Mutator*, *Merlin*, *Transib*, *P*, *PiggyBac*, *PIF-Harbinger*, *CACTA*), Crypton, Helitron et Marevick/Polintron) des éléments non autonomes (les ETs à répétitions inversées miniatures : MITEs pour « Miniatures Inverted repeats Transposable Elements » (Wessler 1995), mobilisés par l'intermédiaire des transposases des éléments autonomes de classe II). Les ETs de classe I, appelés rétrotransposons, transposent par un mécanisme de type copier/coller par un intermédiaire ARN reverse-transcrit. Dans cette deuxième catégorie, il existe également des éléments autonomes (à longues répétitions terminales (« Long Terminal Repeats » ou LTR) : *Copia*, *Gypsy*, *BEL/Pao*), à tyrosine recombinase (YR) : *DIRS*, *Ngaro*, *VIPER*), les *Penelope* (PLE), et les LINEs : *R1*, *RTE*, *Jokey*, *L1*, *I*) et des éléments non autonomes (les SINEs (tRNA, 7SL, 5S), mobilisés grâce aux LINEs) (Eickbush and Jamburuthugoda 2008). Parmi toutes les classifications proposées, nous adopterons celle de Wicker *et al.* (2007), largement utilisée dans la littérature. Elle prend en compte des critères structuraux et fonctionnels (mécaniques et enzymatiques) pour établir les groupes, et subdivisent en hiérarchisant les ETs par classe, sous-classes, ordres, et superfamilles (indiquées en italique). Au sein des différentes super-familles d'éléments, on retrouve une grande diversité avec des regroupement par familles et sous-familles (Rho et al. 2010; Piednoël et al. 2011; Auvinet et al. 2018).

Les difficultés d'assemblages des ETs et l'intérêt prépondérant pour les séquences géniques rendent difficiles leur étude (Hoen et al. 2015; Platt et al. 2016). Leur représentation dans les génomes est, de ce fait, souvent sous-estimée (Sotero-Caio et al. 2017). Cependant, le répertoire en ETs dans différents génomes Eucaryotes est de mieux en mieux caractérisé. La part des ETs dans les génomes des téléostéens peut varier entre 6 % (poissons globes), 14-28 % (morue, lépisostée, platy, tilapia, medaka) et atteindre jusqu'à 55 % (poisson zèbre) (Chalopin et al. 2015) (Figure IV.3.B).

La diversité et la distribution en ETs peut varier considérablement entre espèces présentant un fort degré d'apparentement (Sotero-Caio et al. 2017). Les téléostéens présentent une diversité en ETs qui est globalement plus importante que chez les autres vertébrés (Duvernell et al. 2004; Volff 2005; Chalopin et al. 2015). Ils sont largement distribués dans les génomes des téléostéens (Aparicio et al. 2002). Vingt-sept super-familles d'éléments ont par exemple été recensées chez le poisson zèbre (Howe et al. 2013). On retrouve des ETs de classe II, majoritaires chez les cichlidés et le poisson zèbre (Howe et al. 2013; Brawand et al. 2014) ; ainsi que des éléments de classe I, majoritaires chez le fugu et la lépisostée (Volff et al. 2003; Braasch et al. 2016) (Figure IV.1.B). De nombreux cas de pertes d'ETs ont été recensés dans des clades entiers. Par exemple, les LINEs *Rex3* sont largement distribués chez la plupart des téléostéens, mais totalement absents chez les salmonidés (Volff et al. 2001). Il en est de même avec certains éléments *Copia*, rares ou non retrouvés dans les génomes de crustacés (Piednoël et al. 2013). Aucun téléostéen ne semble présenter une prédominance forte en rétrotransposons à LTR de type *Gypsy*, *Copia* et *BEL*, mais une grande partie de ces ETs ont été recensés en nombre chez plusieurs espèces (Figure IV.3.B) (Herniou et al. 1998; Shen and Steiner 2004; Volff 2005). Les éléments à tyrosine recombinase de type DIRS ont également été mis en évidence chez les téléostéens (Goodwin and Poulter 2001; Goodwin and Poulter 2004; Piednoël et al. 2011). Ils sont retrouvés relativement nombreux et divers chez le poisson zèbre *D. rerio*, mais aussi chez l'épinoche *G. aculeatus*, le medaka *O. latipes* et les poissons globes *T. rubripes* et *T. nigroviridis* (Piednoël et al. 2011) (Figure IV.3.B).

Il existe de manière générale peu de données génomiques et d'études moléculaires sur le contenu en ETs chez les Nototheniidae. Detrich *et al.* (2010) dressent pour la première fois un paysage en ETs dans les génomes du Nototheniinae *N. coriiceps* et du Channichthyinae *C. aceratus* par extrapolation du nombre de copies en ETs après séquençages aléatoires de BACs contenant des fragments d'ADN génomiques de ces deux espèces. Les ETs et autres séquences répétées (Single-Sequence Repeats ou « SSRs » et ADN satellite) représenteraient respectivement 13,4 et 14,5 % de leur génome (Figure IV.3.B). Trois autres études se sont également intéressées aux ETs chez les Nototheniidae, en se focalisant plus spécifiquement cette

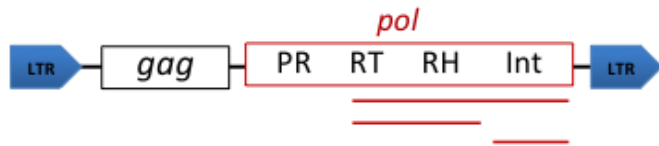
fois sur leur rôle dans les remaniements chromosomiques observés dans cette famille. Les deux premières concernent le transposon à ADN de type *Tc1-Mariner* et *Helitron* (Helinoto), étudié principalement chez le Channichthyinae *C. hamatus* (Capriglione et al. 2002; Capriglione et al. 2014). La troisième concerne les rétrotransposons LINEs de type *Rex1* et *Rex3* sur une diversité d'espèces plus importante : une espèce de « *Trematomus* » (*T. newnesi*), un Dissostichinae (*D. mawsoni*), un Gymnodraconinae (*G. acuticeps*), un Nototheniidae (*N. coriiceps*), et un Channichthyinae (*N. ionah* et *C. hamatus*) (Ozouf-Costaz et al. 2004). Elles impliquent une étude approfondie de ces éléments (caractérisation, diversité, distribution dans les génomes examinés), ainsi que leur localisation chromosomique par *FISH* sur les chromosomes de ces espèces (éléments *Tc1-Mariner* et *Rex*). Si ces éléments ont été trouvés dispersés sur la plupart des chromosomes des différentes espèces, les auteurs notent leur tendance à s'accumuler dans certaines régions chromosomiques. Des points chauds d'insertions d'éléments *Tc1-Mariner* et *Rex3* ont par exemple été retrouvés sur le chromosome Y des mâles *C. hamatus* en position centromérique/péricentromérique et intercalaire dans le bras long. Leur rôle est ainsi mentionné dans la fusion en tandem entre un autosome et un gonosome (Pisano and Ozouf-Costaz 2000) ayant donné lieu à ce chromosome sexuel différencié.

#### **IV.2 : Diversité, quantification, localisation chromosomiques des rétrotransposons à LTR Gypsy et Copia et rétrotransposons à YR DIRS1 chez les « *Trematomus* »**

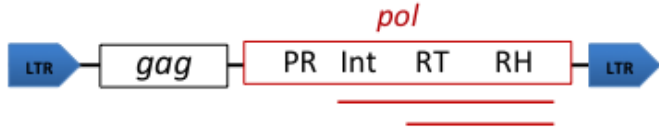
Dans le but d'identifier le rôle éventuel des ETs dans les réarrangements chromosomiques, nous les avons recherchés dans les génomes de divers Nototheniidae à l'aide des domaines conservés (Reverse Transcriptase/RNaseH et Intégrase) (Figure IV.2). Notre investigation s'est portée plus particulièrement chez le groupe « *Trematomus* » et sur trois types de rétrotransposons non caractérisés jusqu'à présent dans ces génomes: les rétrotransposons à LTR de type *Gypsy* et *Copia* et les ETs à tyrosine recombinase de type *DIRS1* (Figure IV.2). Nous avons ainsi estimé leur diversité, distribution, abondance dans ces génomes, et tenté de caractériser leur localisation chromosomique.

## rétrotransposons à LTR

### A *Gypsy*

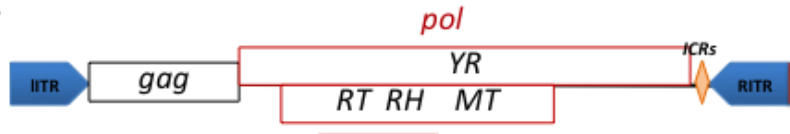


### B *Copia*



## rétrotransposons à YR

### C *DIRS1*



**Figure IV.2 :** structure des ETs recherchés dans les génomes Nototheniidae (A) rétrotransposon à LTR de type *Gypsy* (B) rétrotransposon à LTR de type *Copia* (C) rétrotransposon à YR de type *DIRS1*. Adapté de Eickbush et Jamburuthugoda (2008) et Capello *et al.* (1985). Les rétrotransposons à LTR sont bordés par des « Long terminal Repeats », alors que les rétrotransposons à YR sont bordés par des régions répétées inversées ICR et ITR pour « internal complimentary repeats » et « internal terminal repeats ». Les traits rouges signalent des domaines conservés présents dans la partie *pol* et amplifiés dans la présente étude. Ils recouvrent les domaines Reverse Transcriptase (RT), RNaseH (RH) et/ou Intégrase (Int) pour les *Gypsy* et les *Copia* (fragments de 0,6 à 1,6 kb), et les domaines RT/RH et tyrosine recombinase (YR), présents dans un autre cadre de lecture pour les *DIRS1* (fragments de 1,3 kb).

Cette étude a fait l'objet d'une publication parue dans BMC Genomics le 9 mai 2018. Afin de faciliter la lecture, nous avons choisi d'intégrer directement dans le corps du texte les figures additionnelles.

### IV.2.A : Résumé des principaux résultats

Parmi les trois super-familles de rétrotransposons recherchées, nous avons identifié dans les génomes de huit « *Trematomus* », ainsi que *N. coriiceps* et *D. mawsoni* des éléments appartenant à quatre familles de *DIRS1* (*YNotoJ*, *V*, *B* et *R*), neuf familles de *Gypsy* (*GyNotoA*, *B*, *D*, *E*, *F*, *H*, *I*, *J*, *RT*) et deux familles de *Copia* (*CoNotoA* et *B*) (Tableau IV.2). Seules deux familles de *Gypsy* (*GyNotoD* et *J*) et une famille de *Copia* (*CoNotoA*) n'ont pas été retrouvées dans tous les génomes étudiés (Figure additionnelle IV.2). Les rétrotransposons *BEL-Pao* n'ont pas été détectés par ces méthodes.

D'une manière générale, ces séquences ont été repositionnées au sein d'une phylogénie plus large des éléments décrits chez d'autres Eucaryotes et sont retrouvées parmi les éléments décrits chez les téléostéens, ou plus largement chez les actinoptérygiens (Figures IV.3, 4 et 5).

Les ETs identifiés se caractérisent tous par une conservation intrafamille, indépendamment des espèces où ils sont présents. Les séquences présentent ainsi des pourcentages d'identité moyens de 89,5 % pour les *DIRS1*, 94,1 % pour les *Gypsy* et 97,2 % pour les *Copia* d'une même famille toutes espèces confondues (Tableau IV. 3).

Le nombre de copies de certaines familles d'ETs a été estimé. Nous avons détecté entre sept et plusieurs centaines d'éléments par famille dans les génomes de trois « *Trematomus* » (*T. pennellii*, *T. hansonii* et *T. eulepidotus*), *N. coriiceps* et *D. mawsoni* (Tableau IV.4). On note une nette prédominance des copies de *DIRS1* dans les génomes des espèces examinées, et *a fortiori* dans les génomes des trois « *Trematomus* » étudiés : *T. eulepidotus*, *T. hansonii* et *T. pennellii* (moyenne de 166 copies par familles de *DIRS1* vs 37 par famille de *Gypsy* ou *Copia*).

Deux familles de *Gypsy* et *DIRS1*, ainsi qu'une famille de *Copia* ont également été localisés sur les chromosomes de ces trois « *Trematomus* » ainsi que les deux autres espèces de Nototheniidae étudiées par Hybridation *In Situ* en Fluorescence (*FISH*) (Figure IV.6). Les *Gypsy* et les *Copia* examinés sont retrouvés dispersés dans les chromosomes chez l'ensemble des espèces. Ils ne semblent pas s'accumuler dans certaines localités particulières. Au contraire, nous avons noté une forte accumulation des *DIRS1* dans les régions hétérochromatiques et plus précisément, dans les régions centromériques et péri-centromériques, et parfois télomériques des chromosomes acrocentriques et métacentriques. Ces régions constituent des points chauds d'insertions ayant pu faciliter les remaniements particulièrement importants observés dans ce groupe au cours de la radiation des « *Trematomus* ». Ces accumulations caractéristiques de *DIRS1* ont été retrouvées chez d'autres Nototheniidae, mais pas sur les chromosomes d'espèces subantarctiques ou tempérées n'ayant pas subi des épisodes de réchauffements ou glaciations (Figure IV.7).

**IV.2.B : Mobilization of retrotransposons as a cause of chromosomal diversification and rapid speciation: the case for the Antarctic teleost genus *Trematomus* (article)**

**Research article**

**Mobilization of retrotransposons as a cause of chromosomal diversification and rapid speciation: the case for the Antarctic teleost genus *Trematomus***

J. Auvinet<sup>1,3,\*</sup>, P. Graça<sup>1</sup>, L. Belkadi<sup>4</sup>, L. Petit<sup>2</sup>, E. Bonnivard<sup>1</sup>, A. Dettai<sup>3</sup>, H. William Detrich, III<sup>5</sup>, C. Ozouf-Costaz<sup>1</sup>, D. Higuët<sup>1</sup>

<sup>1</sup> Sorbonne Université, Univ Antilles, CNRS, Institut de Biologie Paris Seine (IBPS), Laboratoire Evolution Paris Seine, F-75005 Paris, France.

<sup>2</sup> Sorbonne Université, CNRS, - Institut de Biologie Paris-Seine (BDPS - IBPS), Plateforme d'Imagerie et Cytométrie en flux, F-75005 Paris, France.

<sup>3</sup> Institut de Systématique, Evolution, Biodiversité (ISYEB), Museum National d'Histoire Naturelle, CNRS, Sorbonne Université, EPHE, 57, rue Cuvier, 75005 Paris, France.

<sup>4</sup> Institut Pasteur, Laboratoire Signalisation et Pathogénèse, UMR CNRS 3691, Bâtiment DARRE, 25-28 rue du Dr Roux, Paris, 75015, France.

<sup>5</sup> Department of Marine and Environmental Sciences, Marine Science Center, Northeastern University, Nahant, MA 01908, USA.

\* Corresponding author: Juliette Auvinet, e-mail address:

Author email addresses: [paula.graca@sorbonne-universite.fr](mailto:paula.graca@sorbonne-universite.fr), [lbelkadi95@gmail.com](mailto:lbelkadi95@gmail.com), [laurence.petit@sorbonne-universite.fr](mailto:laurence.petit@sorbonne-universite.fr), [eric.bonnivard@sorbonne-universite.fr](mailto:eric.bonnivard@sorbonne-universite.fr), [adettai@mnhn.fr](mailto:adettai@mnhn.fr), [iceman@neu.edu](mailto:iceman@neu.edu), [catherine.ozouf@orange.fr](mailto:catherine.ozouf@orange.fr), [dominique.higuet@sorbonne-universite.fr](mailto:dominique.higuet@sorbonne-universite.fr)



## Abstract

### Background:

The importance of transposable elements (TEs) in the genomic remodeling and chromosomal rearrangements that accompany lineage diversification in vertebrates remains the subject of debate. The major impediment to understanding the roles of TEs in genome evolution is the lack of comparative and integrative analyses on complete taxonomic groups. To help overcome this problem, we have focused on the Antarctic teleost genus *Trematomus* (Notothenioidei:Nototheniidae), as they experienced rapid speciation accompanied by dramatic chromosomal diversity. Here we apply a multi-strategy approach to determine the role of large-scale TE mobilization in chromosomal diversification within *Trematomus* species.

### Results:

Despite the extensive chromosomal rearrangements observed in *Trematomus* species, our measurements revealed strong interspecific genome size conservation. After identifying the *DIRS1*, *Gypsy* and *Copia* retrotransposon superfamilies in genomes of 13 nototheniid species, we evaluated their diversity, abundance (copy numbers) and chromosomal distribution. Four families of *DIRS1*, nine of *Gypsy*, and two of *Copia* were highly conserved in these genomes; *DIRS1* being the most represented within *Trematomus* genomes. Fluorescence *in situ* hybridization mapping showed preferential accumulation of *DIRS1* in centromeric and pericentromeric regions, both in *Trematomus* and other nototheniid species, but not in outgroups: species of the Sub-Antarctic notothenioid families Bovichtidae and Eleginopsidae, and the non-notothenioid family Percidae.

### Conclusions:

In contrast to the outgroups, High-Antarctic notothenioid species, including the genus *Trematomus*, were subjected to strong environmental stresses involving repeated bouts of warming above the freezing point of seawater and cooling to sub-zero temperatures on the Antarctic continental shelf during the past 40 millions of years (*My*). As a consequence of these repetitive environmental changes, including thermal shocks; a breakdown of epigenetic regulation that normally represses TE activity may have led to sequential waves of TE activation within their genomes. The predominance of *DIRS1* in *Trematomus* species, their transposition mechanism, and their strategic location in “hot spots” of insertion on chromosomes are likely to have facilitated nonhomologous recombination, thereby increasing genomic rearrangements.

The resulting centric and tandem fusions and fissions would favor the rapid lineage diversification, characteristic of the nototheniid adaptive radiation.

## Keywords

*Trematomus*, chromosomal rearrangements, speciation, Nototheniidae, retrotransposons, *FISH*, *DIRS1* insertion hot spots

## Background

Chromosomal changes are considered by many as a major driving force behind speciation (Böhne et al. 2008; Kraaijeveld 2010; Rebollo et al. 2010). They are major sources of accumulation of genetic incompatibilities, and their fixation is the first stage toward complete reproductive isolation (Ozouf-Costaz et al. 1997; Garagna et al. 2001). Through recombination events (Devos et al. 2002; Bonnivard and Higuete 2009), transposable element (TE) activity can lead to chromosomal rearrangements (Kehrer-Sawatzki and Cooper 2007; Raskina et al. 2008; Kraaijeveld 2010) that may drive lineage-specific diversification (Böhne et al. 2008; Warren et al. 2015), although the relative importance of TE-mediated rearrangements to speciation continues to be debated (Mahé 2009; Ferreira et al. 2010). Therefore, the characterization of TE content and chromosomal organization among diverging lineages in natural species ensembles is an essential step toward understanding the role of mobile elements in species diversification.

TEs are major components of eukaryote genomes: they impact genome structure and plasticity (Volf 2005; Bonnivard and Higuete 2009; Chénais 2015), generating genetic variability on which different evolutionary forces can act (Sela et al. 2010). Depending on their mode of transposition, TEs are divided in two categories: the DNA transposons (class II) move through a “cut-and-paste” mechanism, whereas the retrotransposons (class I) replicate via a “copy-and-paste” process involving an RNA intermediate. Five major orders of class I TEs have been identified (Wicker et al. 2007): Long Interspersed Elements (LINEs), Short Interspersed Elements (SINEs), Penelope (PLEs), Long Terminal Repeat (LTR) retrotransposons and tYrosine Recombinase (YR) encoding elements (also known as DIRS, Dictyostelium Intermediate Repeat Sequence). Within each order, TEs have been categorized in different superfamilies. In animal genomes, the *Gypsy*, *Copia* and *BEL/Pao* superfamilies of the LTR order are the most widely distributed and diversified (Eickbush and Malik 2002; Kidwell 2002; Volf et al. 2003). These elements are flanked by two direct LTRs that encompass the promoter and regulatory regions. Between the LTRs two open reading frames (ORFs) are usually found (Hua-Van et al. 2005): 1)

*gag*, encoding virus-like particles; and 2) *pol*, encoding a reverse transcriptase (RT), an RNase H (RH), and an integrase (INT), necessary for insertion activity. The *pol* region of DIRS elements are generally similar to LTR retrotransposons, except for the substitution of a tyrosine recombinase ORF for the integrase ORF of LTRs. The DIRS order contains Ngaro, VIPER, and DIRS superfamilies; the last is further subdivided into *DIRS1* and *PAT* elements. In contrast to LTR retrotransposons found in a wide continuum of species, DIRS elements have a more patchy taxonomic distribution (Piednoël and Bonnivard 2009). While DIRS are absent in model organisms, they have been detected in teleosts (*Danio rerio* and *Takifugu rubripes*) (Goodwin and Poulter 2001; Goodwin 2004; Howe et al. 2013; Poulter and Butler 2015). Moreover, the small number of DIRS families in the species genomes (Poulter and Goodwin 2005; Piednoël et al. 2011) stands in strong contrast to the abundant representation of LTR families (Wicker et al. 2007; Wicker and Keller 2007; Wang and Liu 2008; Rho et al. 2010).

Bursts of TE activity and amplification in the genomes of plants (Raskina et al. 2008), *D. melanogaster* (Biémont and Vieira 2005), and vertebrates (Volf 2005; Symonová et al. 2013; Warren et al. 2015) are known to have participated in speciation events (McClintock 1984), for example by promoting genomic destabilization and incompatibilities (postzygotic barriers) in hybrid generations (Dion-Cote et al. 2014). Furthermore, environmental changes, including thermal stress, can cause the breakdown of the epigenetic control that normally represses TE activity (Oliver and Greene 2009). This may cause massive bursts of transposition during species radiations (Mansour 2007; Rebollo et al. 2010; Symonová et al. 2013). The activation and mobility of TEs increase chromosomal diversification and potentially speciation rate in some groups, as shown in *D. melanogaster* (Lim and Simmons 1994; Aulard et al. 2004; Ladevèze et al. 2012), and in primates (Pace and Feschotte 2007). However, integrative analyses of chromosomal rearrangements and their linkage to retrotransposition in complete taxonomic groups have in general been hindered by lack of the requisite molecular, chromosomal and environmental data.

During the last 40 millions of years (*My*), the Antarctic fauna have experienced multiple glacial-interglacial cycles, leading to habitat disturbance by iceberg scouring and habitat fragmentation during glacial maxima (Kennett 1986; Kennett and Barker 1990; Stott et al. 1990). The Antarctic teleosts fish family Nototheniidae (antifreeze glycoprotein-(AFGP-) producing “cod icefishes” (Near et al. 2012)) rapidly diversified into several species flocks and now constitute the dominant group of Antarctic teleosts (Eastman and McCune 2000; Near et al. 2012; Lecointre et al. 2013) with well-documented phylogenetic relationships (Rutschmann et al. 2011; Dettai et al. 2012; Lautrédou et al. 2012; Near et al. 2012). The nototheniid genus *Trematomus* (including

*Indonotothenia cyanobrancha* (Dettai et al. 2012)) is an example of a relatively recent (–9.1 My (Near et al. 2012); –4.3 My (Colombo et al. 2015)) and rapid marine adaptive radiation. All are endemic to coastal waters of the Antarctic continental shelf. Although their ancestral diploid karyotype was inferred to have 48 acrocentric chromosomes (the typical karyotype of most modern teleosts), trematomines exhibit the highest chromosomal diversity among the Nototheniidae (Ozouf-Costaz et al. 1991; Ghigliotti et al. 2015). Diploid chromosome numbers range between 24 and 58, and chromosomal rearrangements of all types are plentiful (Morescalchi et al. 1992; Pisano and Ozouf-Costaz 2000; Pisano et al. 2003), which is rare among other marine teleosts (Molina and de Freitas Bacurau 2006). Sex chromosome differentiation also occurred via genomic restructuring (Ozouf-Costaz et al. 1991; Ozouf-Costaz et al. 1997; Pisano and Ozouf-Costaz 2000; Ghigliotti et al. 2015), as at least five *Trematomus* species possess a multiple sex chromosome system of the female  $X_1X_1X_2X_2$  / male  $X_1X_2Y$  type. Males have an odd chromosome number due to the tandem fusion between one  $X_1$  and one  $X_2$  autosome to form the Y element.

The tractable species number of the monophyletic genus *Trematomus*, the history of thermal perturbations of the taxon, its remarkable inter- and intraspecific chromosomal diversity, and its plethora of TEs (Capriglione et al. 2002; Ozouf-Costaz et al. 2004; Detrich et al. 2010; Cocca et al. 2011; Chen et al. 2017) make the group ideal to investigate whether mobilization of TEs in response to environmental change may have played a role in the chromosomal diversification that often accompanies rapid speciation events. In the present multi-strategy study, we focus on DIRS- and LTR-order retrotransposons, specifically the diversity, distribution, chromosomal locations, and quantification of the pol region of the *DIRS1*, *Gypsy*, and *Copia* elements, which we find to be widely distributed in the genomes of *Trematomus* and other nototheniid species. The results obtained were compared to related temperate and Sub-Antarctic non AFGP-bearing species that have not experienced such environmental changes.

## Results

### Genome sizes estimations

We determined a mean genome size of  $1.19 \pm 0.12$  pg (s.e.m.) for *Trematomus* species, with C-values ranging from 1.09 to 1.26 pg. Our values from the nototheniids *D. mawsoni* (1.02 pg) and *N. coriiceps* (1.36 pg) flanked the trematomine range. Variance between technical triplicates (0.015 pg, s.e.m.) and biological triplicates (0.017 pg, s.e.m.) were correspondingly low (one to seventeen specimens per species, see Additional file IV.1). In general, our results are in close agreement with those C-values published previously (Table IV.1).

| Sub-families   | Genus. species<br>(abbreviation) | Chromo-<br>somal<br>numbers<br>(2n) | No.<br>exami-<br>ned | C-values<br>(pg) ± SE | Prior<br>published<br>results (ref) |
|----------------|----------------------------------|-------------------------------------|----------------------|-----------------------|-------------------------------------|
|                | <i>T. eulepidotus</i> (Teu)      | 24                                  | 1                    | 1.26                  |                                     |
|                | <i>T. pennellii</i> (Tpe)        | 32                                  | 1                    | 1.15                  |                                     |
|                | <i>T. borchgrevinki</i> (Tbo)    | <u>45 - 46*</u>                     | 8                    | 1.09 ± 0.039          | 1.28 (1)                            |
|                | <i>T. hansonii</i> (Tha)         | <u>45 - 46</u><br>/46/48*           | 12                   | 1.26 ± 0.025          |                                     |
| Trematominae   | <i>T. bernacchii</i> (Tbe)       | 48                                  | 17                   | 1.12 ± 0.019          | 1.20 (2, 3)                         |
|                | <i>T. loennbergii</i> (Tlo)      | <u>47 - 48*</u>                     | 1                    | 1.34                  |                                     |
|                | <i>T. lepidorhinus</i> (Tle)     | <u>47 - 48*</u>                     | 0                    | ND                    |                                     |
|                | <i>T. newnesi</i> (Tne)          | <u>45 - 46*</u>                     | 2                    | 1.15 ± 0.141          | 1.01 (4)                            |
|                | <i>T. scotti</i> (Tsc)           | 48                                  | 0                    | ND                    |                                     |
|                | <i>T. nicolai</i> (Tni)          | <u>57 - 58*</u>                     | 2                    | 1.17 ± 0.04           |                                     |
|                | <i>I. cyanobranca</i> (Icy)      | 48                                  | 0                    | ND                    |                                     |
| Nototheniinae  | <i>N. coriiceps</i> (Nco)        | 22                                  | 3                    | 1.36 ± 0.118          | 1.13 (4)                            |
| Dissostichinae | <i>D. mawsoni</i> (Dma)          | 48                                  | 3                    | 1.02 ± 0.016          | 0.97 (1); 1.03<br>(2, 3); 1.20 (4)  |

**Table IV.1: C-values (pg) of nototheniid species.** Several different karyotypes per species due to sex chromosome differentiation (either an X<sub>1</sub>X<sub>1</sub>X<sub>2</sub>X<sub>2</sub> female or an X<sub>1</sub>X<sub>2</sub>Y male). The possible combinations from our sampling are underlined. ND: not determined, no erythrocytes samples available. 1: (Chen et al. 2008), 2: (Hardie and Hebert 2003), 3: (Hardie and Hebert 2004), 4: (Detrich et al. 2010).

| Sub-family     | Genus. species | Field reference | C value (pg) mean / triplicate | Measure STDEVA | Mean STDEVA |
|----------------|----------------|-----------------|--------------------------------|----------------|-------------|
| Trematominae   | <i>Teu</i>     | CE 4011         | 1.261                          | 0.013          | 0.017       |
|                | <i>Tpe</i>     | CE 4313         | 1.145                          | 0.016          |             |
|                | <i>Tbo</i>     | CE 5560         | 1.052                          | 0.024          |             |
|                |                | CE 5561         | 1.078                          | 0.023          |             |
|                |                | CE 5562         | 0.994                          | 0.018          |             |
|                |                | CE 5656         | 1.061                          | 0.018          |             |
|                |                | CE 5655         | 0.970                          | 0.011          |             |
|                |                | CE 5662         | 1.174                          | 0.004          |             |
|                |                | CE 5664         | 1.313                          | 0.031          |             |
|                |                | CE 5662         | 1.124                          | 0.003          |             |
|                | <i>Tha</i>     | CE 4002         | 1.257                          | 0.015          |             |
|                |                | CE 4003         | 1.271                          | 0.016          |             |
|                |                | CE 4004         | 1.331                          | 0.012          |             |
|                |                | CE 4316         | 1.193                          | 0.008          |             |
|                |                | CE 5708         | 1.154                          | 0.009          |             |
|                |                | CE 5711         | 1.052                          | 0.016          |             |
|                |                | CE 5738         | 1.248                          | 0.007          |             |
|                |                | CE 5856         | 1.320                          | 0.005          |             |
|                |                | CE 5880         | 1.307                          | 0.005          |             |
|                |                | CE 5660 (M)     | 1.288                          | 0.008          |             |
|                | CE 5661 (F)    | 1.304           | 0.008                          |                |             |
|                | CE 5664        | 1.364           | 0.010                          |                |             |
|                | <i>Tbe</i>     | Ich 1098        | 1.059                          | 0.005          |             |
|                |                | Ich 1099        | 1.231                          | 0.016          |             |
|                |                | Ich 1100        | 1.081                          | 0.020          |             |
|                |                | Ich 1101        | 1.257                          | 0.010          |             |
|                |                | Ich 1102        | 1.141                          | 0.021          |             |
|                |                | Ich 1114        | 1.213                          | 0.000          |             |
|                |                | CE 4315         | 1.087                          | 0.031          |             |
|                |                | CE 4314         | 1.178                          | 0.007          |             |
|                |                | CE 5709         | 1.096                          | 0.016          |             |
|                |                | CE 5737         | 1.051                          | 0.026          |             |
|                |                | CE 6042         | 1.093                          | 0.045          |             |
| CE 6043        |                | 1.076           | 0.003                          |                |             |
| CE 6091        |                | 1.118           | 0.002                          |                |             |
| CE 6092        |                | 1.033           | 0.005                          |                |             |
| CE 5658        |                | 1.100           | 0.014                          |                |             |
| CE 5669        |                | 1.238           | 0.010                          |                |             |
| CE 5685        |                | 0.991           | 0.004                          |                |             |
| <i>Tlo</i>     |                | CE 5916         | 1.340                          | 0.004          |             |
| <i>Tne</i>     |                | CE 4319         | 1.007                          | 0.021          |             |
|                |                | TA 456          | 1.289                          | 0.014          |             |
| <i>Tni</i>     | Ich 1115 (F)   | 1.206           | 0.007                          |                |             |
|                | CE 5684        | 1.127           | 0.009                          |                |             |
| Nototheniinae  | <i>Nco</i>     | Ich 1095 (F)    | 1.432                          | 0.099          |             |
|                |                | Ich 1096 (F)    | 1.404                          | 0.034          |             |
|                |                | Ich 1097 (F)    | 1.236                          | 0.013          |             |
| Dissostichinae | <i>Dma</i>     | CE 4054 (M)     | 1.001                          | 0.036          |             |
|                |                | CE 4055 (F)     | 1.004                          | 0.009          |             |
|                |                | CE 4056 (F)     | 1.050                          | 0.011          |             |
| <b>Mean</b>    |                |                 | 1.168                          | 0.015          | 0.017       |

**Additional file IV.1: Repeatability of measurements for flow cytometry values determining genome sizes.**

When identified, the sex of the specimen is indicated next to the field number M: male, F: female. Nototheniid retrotransposons: identification, genomic distribution and characterization

Four families of *DIRS1* (named *YNotoJ*, *V*, *B* and *R*) have been identified based on the 1.25 kb fragments overlapping the RT/RH domains. For the *Gypsy* elements, nine families (named *GyNotoA*, *B*, *D*, *E*, *F*, *H*, *I*, *J*, *RT*) have been identified; five with 1.5 kb fragments than span the RT/RH/INT domains (region 2), the other four with 0.6 kb fragments overlapping the RT/RH or the INT domain (region 1 or 3), as it was not possible to extend the seed by TE-walking. Two families of *Copia* (named *CoNotoA* and *CoNotoB*) were detected with sequences of 1 or 1.4 kb overlapping either the RT/RH or the RT/RH/INT domains (Table IV.2). All identified TE families appeared ubiquitous, in all investigated genomes, with three exceptions: *GyNotoD* and *GyNotoJ*, only detected in one or two genomes; and *CoNotoA* not detected in *Trematomus* genomes (Additional file IV.2).

| TE superfamily               | Family          | Amplification size (kb) <sup>1</sup> | Region <sup>2</sup> |
|------------------------------|-----------------|--------------------------------------|---------------------|
| <b><i>DIRS1 (YNoto)</i></b>  | <i>YNotoJ</i>   | 1.25                                 | 1                   |
|                              | <i>YNotoV</i>   | 1.26                                 | 1                   |
|                              | <i>YNotoB</i>   | 1.25                                 | 1                   |
|                              | <i>YNotoR</i>   | 1.25                                 | 1                   |
| <b><i>Gypsy (GyNoto)</i></b> | <i>GyNotoA</i>  | 1.54                                 | 2                   |
|                              | <i>GyNotoB</i>  | 1.54                                 | 2                   |
|                              | <i>GyNotoD</i>  | 1.55                                 | 2                   |
|                              | <i>GyNotoE</i>  | 1.62                                 | 2                   |
|                              | <i>GyNotoF</i>  | 0.61                                 | 3                   |
|                              | <i>GyNotoH</i>  | 0.61                                 | 3                   |
|                              | <i>GyNotoI</i>  | 0.61                                 | 3                   |
|                              | <i>GyNotoJ</i>  | 1.22                                 | 2                   |
|                              | <i>GyNotoRT</i> | 0.65                                 | 1                   |
| <b><i>Copia (CoNoto)</i></b> | <i>CoNotoA</i>  | 0.94                                 | 1                   |
|                              | <i>CoNotoB</i>  | 1.38                                 | 2                   |

**Table IV.2: Characteristics of the *DIRS1*, *Gypsy* and *Copia* sequences (TE dataset) detected in nototheniid genomes.** <sup>1</sup>Size of TE consensus from each family found in nototheniid genomes. <sup>2</sup>Either for *DIRS1*, *Gypsy* or for *Copia* elements, we focused on the *pol* region. The TE sequences are then overlapping on 1: RT/RH, 2: RT/RH/INT or 3: INT domains.

| Species                 | <i>DIRS1</i><br>( <i>YNoto</i> ) |           |           |           | <i>Gypsy</i><br>( <i>GyNoto</i> ) |            |            |            |            |            |            |            |             | <i>Copia</i><br>( <i>CoNoto</i> ) |            |
|-------------------------|----------------------------------|-----------|-----------|-----------|-----------------------------------|------------|------------|------------|------------|------------|------------|------------|-------------|-----------------------------------|------------|
|                         | <i>YB</i>                        | <i>YJ</i> | <i>YR</i> | <i>YV</i> | <i>GyA</i>                        | <i>GyB</i> | <i>GyD</i> | <i>GyE</i> | <i>GyF</i> | <i>GyH</i> | <i>GyI</i> | <i>GyJ</i> | <i>GyRT</i> | <i>CoA</i>                        | <i>CoB</i> |
| <i>T. eulepidotus</i>   | +                                | +         | +         | +         | +                                 | +          | 0          | +          | +          | +          | +          | 0          | +           | 0                                 | +          |
| <i>T. pennellii</i>     | +                                | +         | +         | +         | +                                 | +          | 0          | +          | +          | +          | +          | 0          | +           | 0                                 | +          |
| <i>T. lepidothinus</i>  | +                                | +         | +         | +         | +                                 | +          | 0          | +          | +          | +          | +          | +          | +           | 0                                 | +          |
| <i>T. borchgrevinki</i> | +                                | +         | +         | +         | +                                 | +          | 0          | +          | +          | +          | +          | +          | +           | 0                                 | +          |
| <i>T. hansonii</i>      | +                                | +         | +         | +         | +                                 | +          | 0          | +          | +          | +          | +          | 0          | +           | 0                                 | +          |
| <i>T. bernacchii</i>    | +                                | +         | +         | +         | +                                 | +          | 0          | +          | +          | +          | +          | 0          | +           | 0                                 | +          |
| <i>T. loennbergi</i>    | +                                | +         | +         | +         | +                                 | +          | 0          | +          | +          | +          | +          | 0          | +           | 0                                 | +          |
| <i>T. newnesi</i>       | +                                | +         | +         | +         | +                                 | +          | 0          | +          | +          | +          | +          | 0          | +           | 0                                 | +          |
| <i>T. scottii</i>       | +                                | +         | +         | +         | +                                 | +          | 0          | +          | +          | +          | +          | 0          | +           | 0                                 | +          |
| <i>T. nicolai</i>       | +                                | +         | +         | +         | +                                 | +          | 0          | +          | +          | +          | +          | 0          | +           | 0                                 | +          |
| <i>I. cyanobrancha</i>  | +                                | +         | +         | +         | +                                 | +          | 0          | +          | +          | +          | +          | +          | +           | 0                                 | +          |
| <i>N. coriiceps</i>     | +                                | +         | +         | +         | +                                 | +          | +          | +          | +          | +          | +          | 0          | +           | +                                 | +          |
| <i>D. mawsoni</i>       | +                                | +         | +         | +         | +                                 | +          | 0          | +          | +          | +          | +          | 0          | +           | +                                 | +          |

**Additional file IV.2: Detection of identified TEs among *Trematomus*, *N. coriiceps* and *D. mawsoni* genomes.**

+ presence, 0 undetected

All TE families sequenced showed a strong intrafamily conservation across nototheniid species. We found a high nucleotide sequence identity within each *DIRS1* family (88.2 to 91.4 %) (Table IV.3.a), as well as within each *Gypsy* family (88.3 to 96.6 % for the fragments that span the RT/RH portion and 89.7 to 97.8 % for the fragments overlapping on INT portion) (Table IV.3.b). *Copia* sequences within a family were similarly conserved (95.9 to 98.5 %) (Table IV.3.c). The maximum divergence between TE families across nototheniid species was 50.3% for *DIRS1*, followed by 37.8 % interfamily divergence for *Gypsy*. *CoNotoA* (*GalEa*) and *CoNotoB* (*Hydra*) could not be reliably aligned for comparison.



**a**

| RT/RH         | <i>YNotoB</i> | <i>YNotoR</i> | <i>YNotoJ</i> | <i>YNotoV</i> |
|---------------|---------------|---------------|---------------|---------------|
| <i>YNotoB</i> | <b>88.2</b>   | 48.3          | 49.4          | 49.6          |
| <i>YNotoR</i> |               | <b>91.4</b>   | 45.7          | 50.1          |
| <i>YNotoJ</i> |               |               | <b>88.4</b>   | 55.3          |
| <i>YNotoV</i> |               |               |               | <b>89.8</b>   |

**b**

| RT/RH<br>INT   | <i>GyNotoA</i> | <i>GyNotoB</i> | <i>GyNotoD</i> | <i>GyNotoE</i> | <i>GyNotoF</i> | <i>GyNotoH</i> | <i>GyNotoJ</i> |
|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|
| <i>GyNotoA</i> | <b>96.6</b>    | 59.4           | 60.9           | 62.8           |                |                | 62.6           |
| <i>GyNotoB</i> | <b>95.7</b>    | <b>92.2</b>    | 68.9           | 64.5           |                |                | 57.9           |
| <i>GyNotoD</i> | 62.6           | <b>89.7</b>    | NA             | 64.8           |                |                | 59.4           |
| <i>GyNotoE</i> | 64.4           | 70.6           | NA             | <b>96.6</b>    |                |                | 60.4           |
| <i>GyNotoF</i> | 65.7           | 66.3           | 69.2           | <b>96.0</b>    | <b>97.8</b>    |                |                |
| <i>GyNotoH</i> | 65.2           | 66.2           | 68.9           | 74.4           |                | <b>93.8</b>    |                |
| <i>GyNotoJ</i> | 65.5           | 61.7           | 63.3           | 62.6           | 63.8           |                | <b>88.3</b>    |

**c**

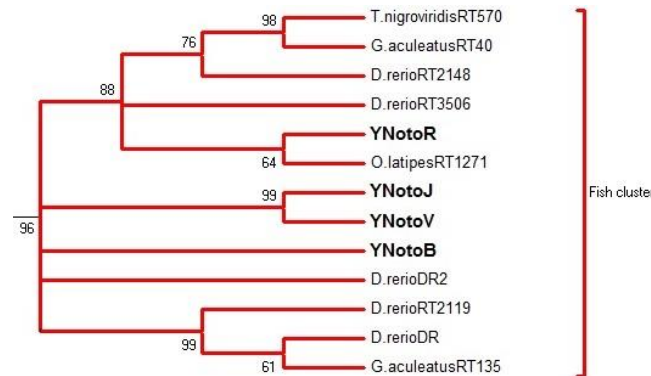
| RT/RH±INT      | <i>CoNotoA</i> | <i>CoNotoB</i> |
|----------------|----------------|----------------|
| <i>CoNotoA</i> | <b>98.5</b>    |                |
| <i>CoNotoB</i> |                | <b>95.9</b>    |

**Table IV.3: Nucleic acid identity matrices of *DIRS1* (a), *Gypsy* (b) and *Copia* (c) families across nototheniid species.** Identity percentages are based on the nucleotide sequence alignments of identified nototheniid retrotransposons (see Additional file IV.9 for taxonomic sampling). TE intrafamily percentage identities across nototheniid species are indicated in **bold**. For *Gypsy* elements, identities calculated from RT/RH are presented above the table diagonal line and those based on INT are shown below. NA: not applicable because only one sequence identified for the family.

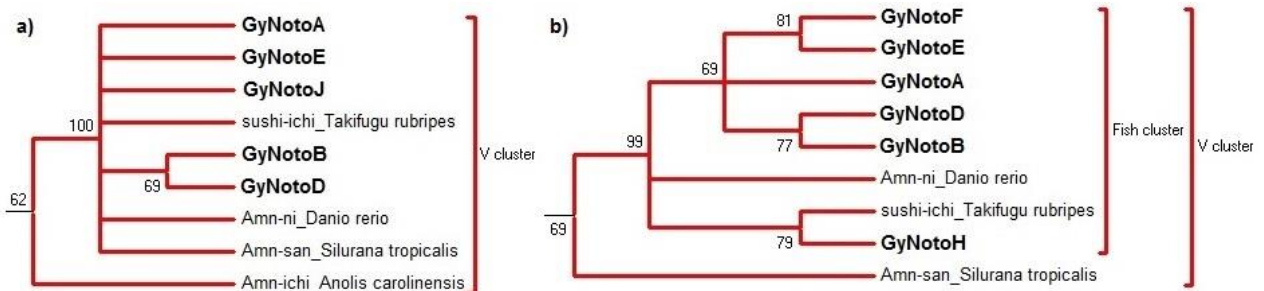
#### Nototheniid retrotransposon positioning among eukaryote TEs

We ran phylogenetic analyses for *DIRS1*, *Gypsy* and *Copia* TEs in order to position our TE family among TE families described in eukaryote genomes. The same topologies were observed whether Distance or Maximum Likelihood reconstruction methods were used. The large majority of our sequences cluster together with sequences previously described in bony fish or vertebrate groups (Figure IV.3 to IV.5, Additional files IV.3 to IV.5 for full trees). *DIRS1* nototheniid consensus sequences cluster with other bony fish *DIRS1* sequences from *Tetraodon nigroviridis*, *Gasterosteus aculeatus*, *Danio rerio*, and *Oryzias latipes* (0.96 support). Except for

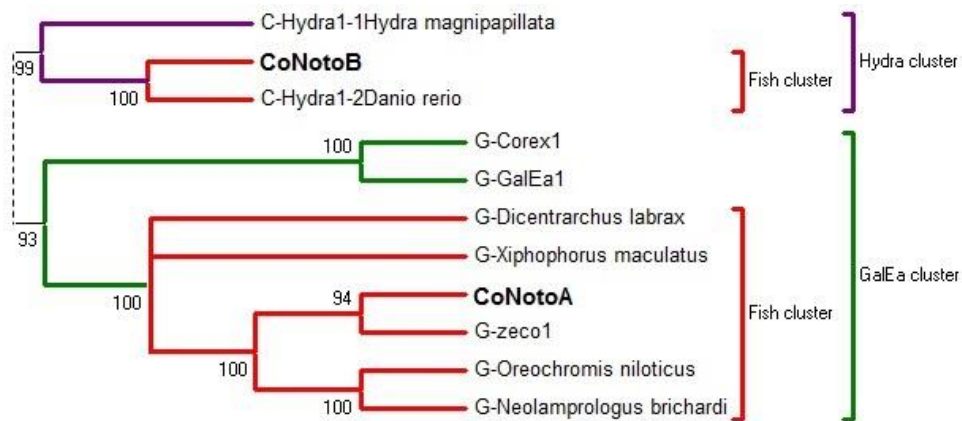
*GyNotoI* and *GyNotoRT*, *Gypsy* nototheniid consensus sequences cluster with other vertebrate and bony fish *Gypsy* sequences from *Tetraodon rubripes* and *Danio rerio* (0.62 and 0.69 support, respectively). *Copia* nototheniid consensus sequences cluster with other bony fish - *GalEa* sequences for *CoNotoA* or *Hydra* sequences for *CoNotoB* - from *Danio rerio*, *Dicentrarchus labrax* and *Oreochromis niloticus* (support = 1) (Figure IV.3, IV.4.a, b and IV.5).



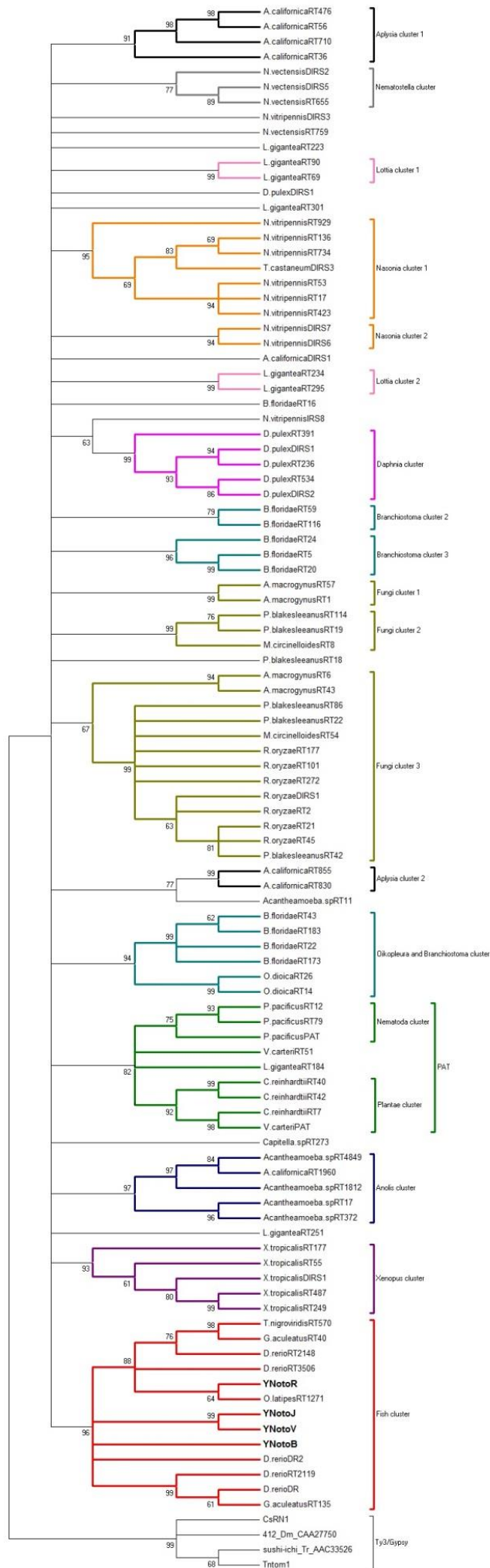
**Figure IV.3:** NJ bootstrap consensus tree for *DIRS1* based on the RT/RH amino acid sequences. Only the branch containing the nototheniid *DIRS1* families and the closest related sequences are shown. The four nototheniid *DIRS1* families (**bold font**) group with the other bony fish sequences: *Tetraodon nigroviridis*, *Gasterosteus aculeatus*, *Danio rerio*, *Oryzias latipes*. Distances were calculated with the JTT model and a gamma distribution correction for amino acid. Support for individual clusters was evaluated using non-parametric bootstrapping with 1 000 replicates. Only bootstraps over 60 are presented. Nodes with bootstraps <60 % were collapsed. See full tree in Additional file IV.3.



**Figure IV.4:** NJ bootstrap consensus tree for *Gypsy* based on the RT/RH (a), INT (b) regions. Only the branch containing the nototheniid *Gypsy* and the closest related amino acid sequences are presented. Except for *GyNotoI* and *GyNotoRT*, the seven nototheniid *Gypsy* families shown (**bold font**) group with bony fish sequences: *Takifugu rubripes* -sushi-ichi and *Danio rerio* -Amn-ni in addition to other vertebrate *Gypsy* sequences: *Xenopus/Silurana tropicalis*. Distances were calculated with the JTT model and a gamma distribution correction for amino acid. Support for individual clusters was evaluated using non-parametric bootstrapping with 1 000 replicates. Only bootstraps over 60 are presented. Nodes with bootstraps <60 % were collapsed. See full tree in Additional file IV.4.

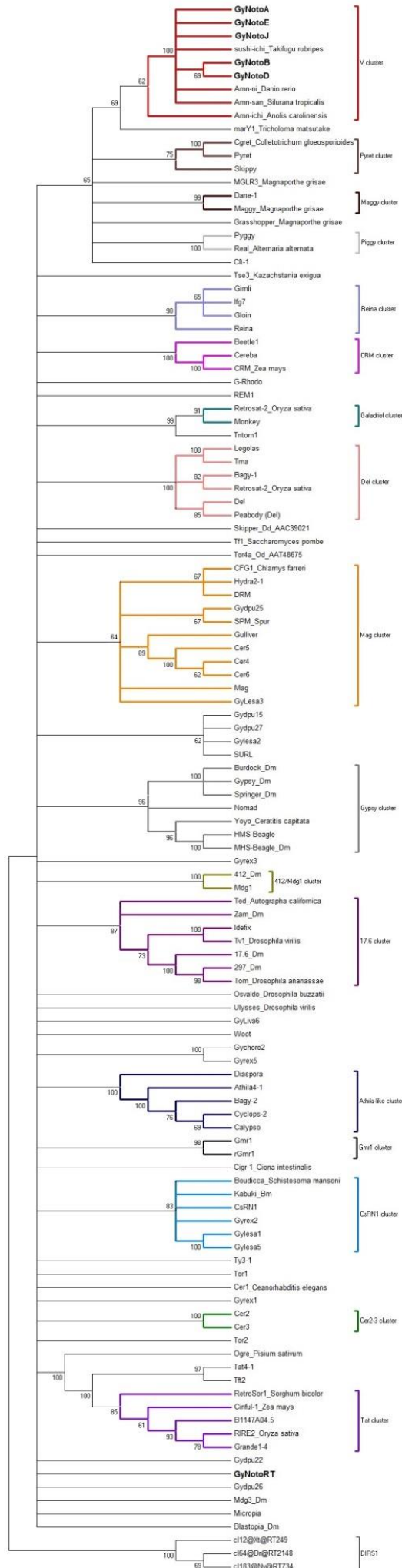


**Figure IV.5:** NJ bootstrap consensus tree for *Copia* based on the RT/RH amino acid sequences. Only the branch containing the nototheniid *Copia* families and the closest related sequences are shown. The two nototheniid *Copia* families (**bold font**) identified in nototheniid genomes group with bony fish sequences: *Dicentrarchus labrax*, *Xiphophorus maculatus*, *Danio rerio*, *Oreochromis niloticus* and *Neolamprologus brichardi*. Distances were calculated with the JTT model and the gamma distribution correction for amino acid. Support for individual clusters was evaluated using non-parametric bootstrapping with 1 000 replicates. Only bootstraps over 60 are presented. Nodes with bootstraps <60 % were collapsed. See full tree in Additional file IV.5.



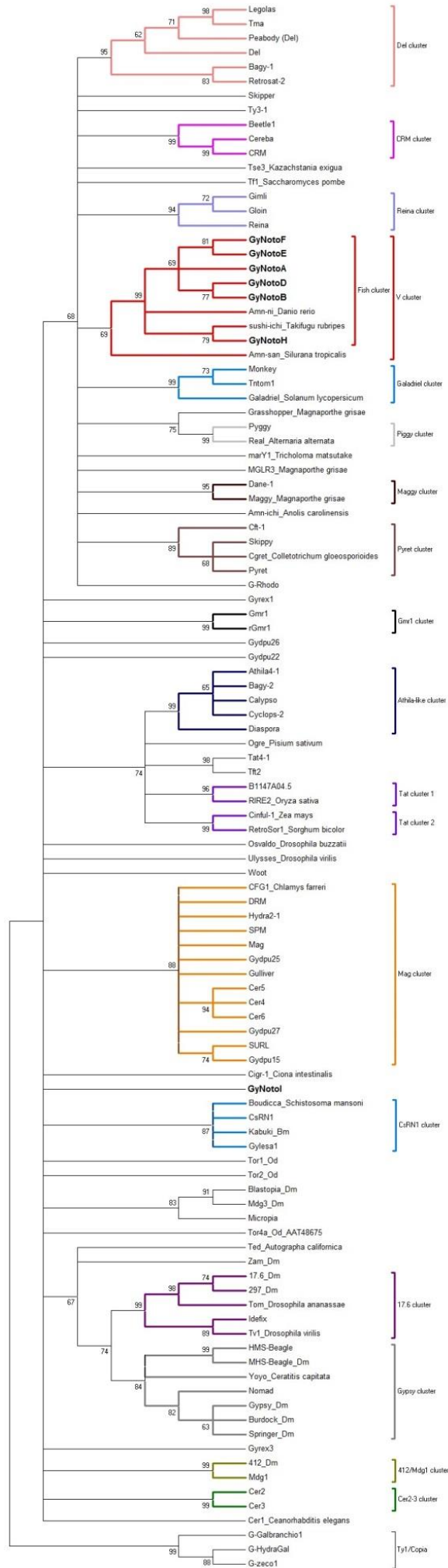
**Additional file 3: NJ bootstrap consensus tree for *DIRS1* based on the RT/RH amino acid sequences.** Complement of Figure IV.3. The four families identified in nototheniid genomes (**bold font**) group with the other bony fish *DIRS1* sequences: *Tetraodon nigroviridis*, *Gasterosteus aculeatus*, *Danio rerio*, *Oryzias latipes*. Distances were calculated with the JTT model and a gamma distribution correction for amino acid. Support for individual clusters was evaluated using non-parametric bootstrapping with 1 000 replicates. Only bootstraps over 60 are presented. Nodes with bootstraps <60 % were collapsed.

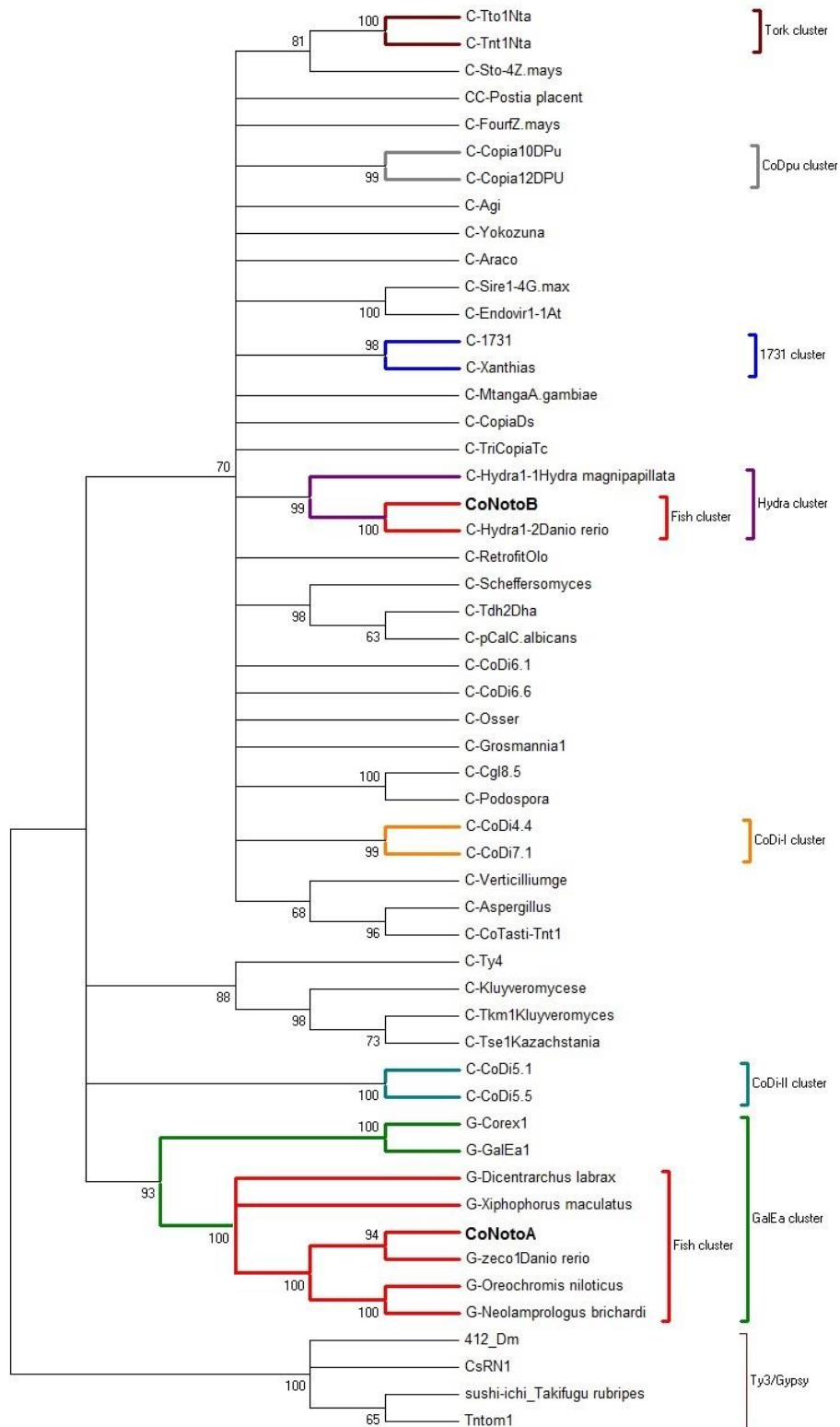
a)



**Additional file IV.4: NJ bootstrap consensus tree for Gypsy based on the amino acid sequences of the RT/RH (a) and INT (b) regions.** Complement of Figure IV.4. Except for *GyNotoI* and *GyNotoRT*, the families identified in nototheniid genomes (**bold font**) group with the other vertebrate. *Xenopus/Silurana tropicalis*; and particularly with bony fish *Gypsy* sequences: *Tetraodon rubripes*-sushi-ichi, *Danio rerio*-Amn-ni. Distances were calculated with the JTT model and the gamma distribution correction for amino acid. Support for individual clusters was evaluated using non-parametric bootstrapping with 1 000 replicates. Only bootstraps over 60 are presented. Nodes with bootstraps <60 % were collapsed.

b)





**Additional file IV.5: NJ bootstrap consensus tree for *Copia* based on the RT/RH amino acid sequences.**

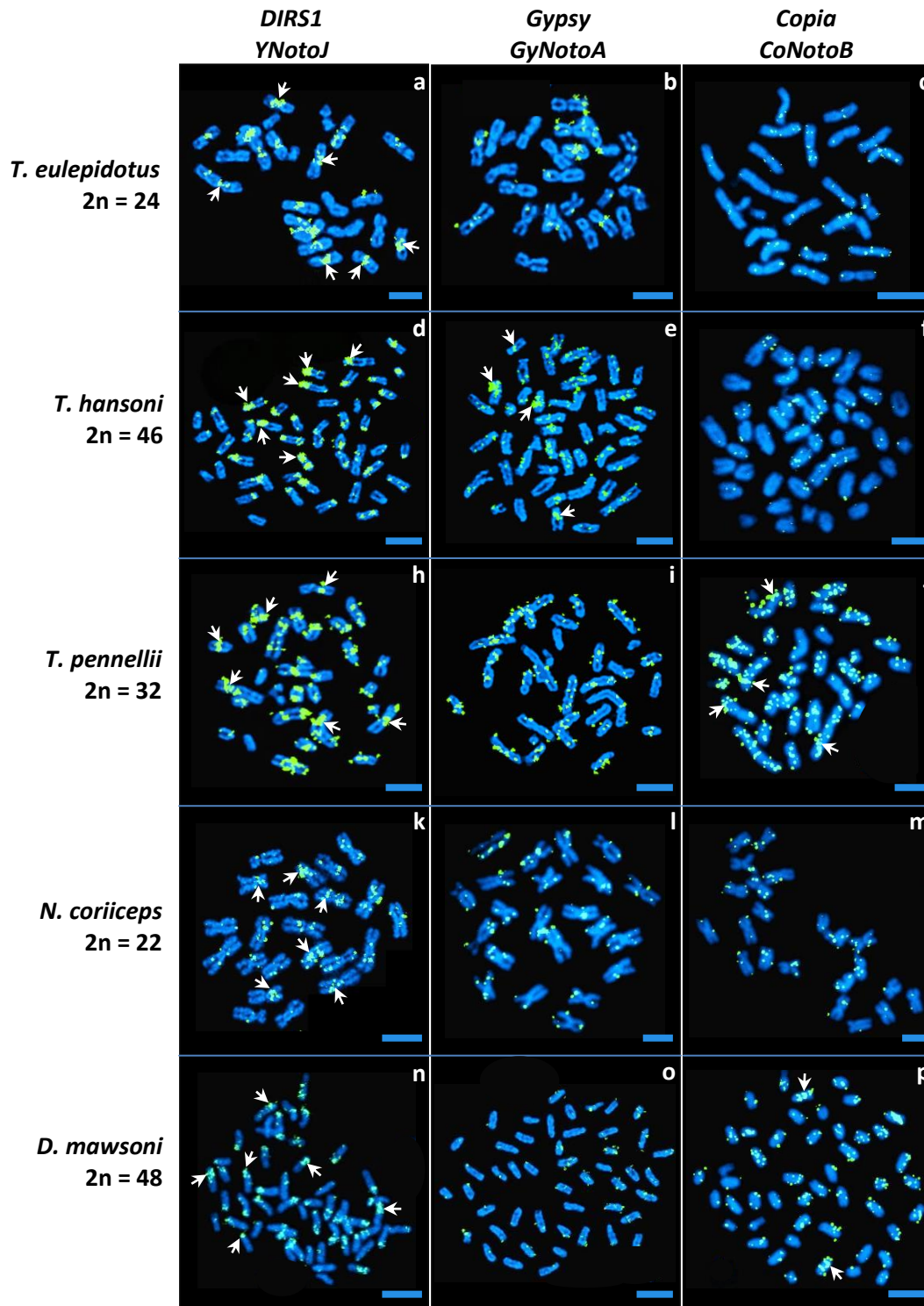
Complement of Figure IV.5. The *GalEa* and *Hydra* families identified in nototheniid genomes (**bold font**) group with the other bony fish *Copia* sequences: *Dicentrarchus labrax*, *Xiphophorus maculatus*, *Danio rerio*, *Oreochromis niloticus* and *Neolamprologus brichardi*. Distances were calculated with the JTT model and the gamma distribution correction for amino acid. Support for individual clusters was evaluated using non-parametric bootstrapping with 1 000 replicates. Only bootstraps over 60 are presented. Nodes with bootstraps <60 % were collapsed.

All *DIRS1* consensus were positioned among the described “FISH cluster” (“*DrDIRS1*” group) (Goodwin and Poulter 2001; Goodwin 2004; Piednoël et al. 2011). *YNotoJ* and *YNotoV* seemed to be more closely related between each other than with *YNotoR* and *YNotoB* (Figure IV.3). In the same way, *GyNotoB* and *GyNotoD* were closely related but no nototheniid *Gypsy* clade was detected: all *GyNoto* consensus sequences (except *GyNotoI* and *RT*) were grouped with other *Gypsy* sequences from *T. rubripes* and from *D. rerio* belonging to the described “V cluster” (Llorens et al. 2009) (Figure IV.4). *GyNotoI* and *GyNotoRT* even fell outside the bony fish TE clade (Additional file IV.4) and were much more difficult to align than the other family consensus sequences. Finally, the two *CoNoto* consensus sequences were placed within two distinct *Copia* bony fish sub-clades, both closely related to *Copia* sequences found in the zebrafish *D. rerio*: *Hydra1-2* for *CoNotoB* and *Galea-Zeco1* for *CoNotoA* (Figure IV.5).

#### Chromosomal location of nototheniid retrotransposon families

Fluorescent *In Situ* hybridization (*FISH*) is a powerful technology for imaging the distribution of repetitive gene families on the condensed chromosomes of metaphase spreads. To assess the potential role of nototheniid TEs in mediating chromosomal rearrangements in *Trematomus* and other High Antarctic species, we used *FISH* to map the locations of two families of *DIRS1*, two families of *Gypsy*, and one family of *Copia* (*Hydra*) on chromosome preparations from five nototheniid species that represent the diversity of nototheniid karyotypes ( $2n = 22$  metacentrics to  $2n = 48$  acrocentrics): *T. eulepidotus*, *T. hansonii*, *T. pennellii*, *N. coriiceps*, and *D. mawsoni*. For comparison, we chose chromosome preparations from three cool temperate/temperate outgroup species: the Sub-Antarctic notothenioids *B. diancanthus* and *E. maclovinus*, and the Eurasian perch *P. fluviatilis*, all of which feature 48 acrocentric chromosomes. The TE families were chosen for their cloned sequence sizes (>1 kb) and their ubiquity in the genomes of nototheniid species (Additional files IV.2 and IV.6). Figure IV.6 shows the two major types of TE distribution patterns: (1) dense accumulation (“hot spots”) mainly in centromeric and/or pericentromeric regions and sometimes in intercalary or telomeric positions; and (2) scattered, punctuate staining along chromosome arms. Combinations of these two patterns were also evident. The distributions (1) and/or (2) clearly depended upon the TE super-family, but within a super-family, different TE families produced comparable results (e.g., the nearly identical patterns observed between *YNotoJ* and *YNotoR* of the *DIRS1* superfamily as shown with the double *FISH*-mapping (Additional file IV.7 and IV.8), and between *GyNotoA* and *GyNotoE* of the *Gypsy* superfamily) (Figure IV.6 and Additional file IV.7). For a given TE family, there were small variations of signal intensity or distribution along chromosomes between species, but generally, the same location pattern was observed.

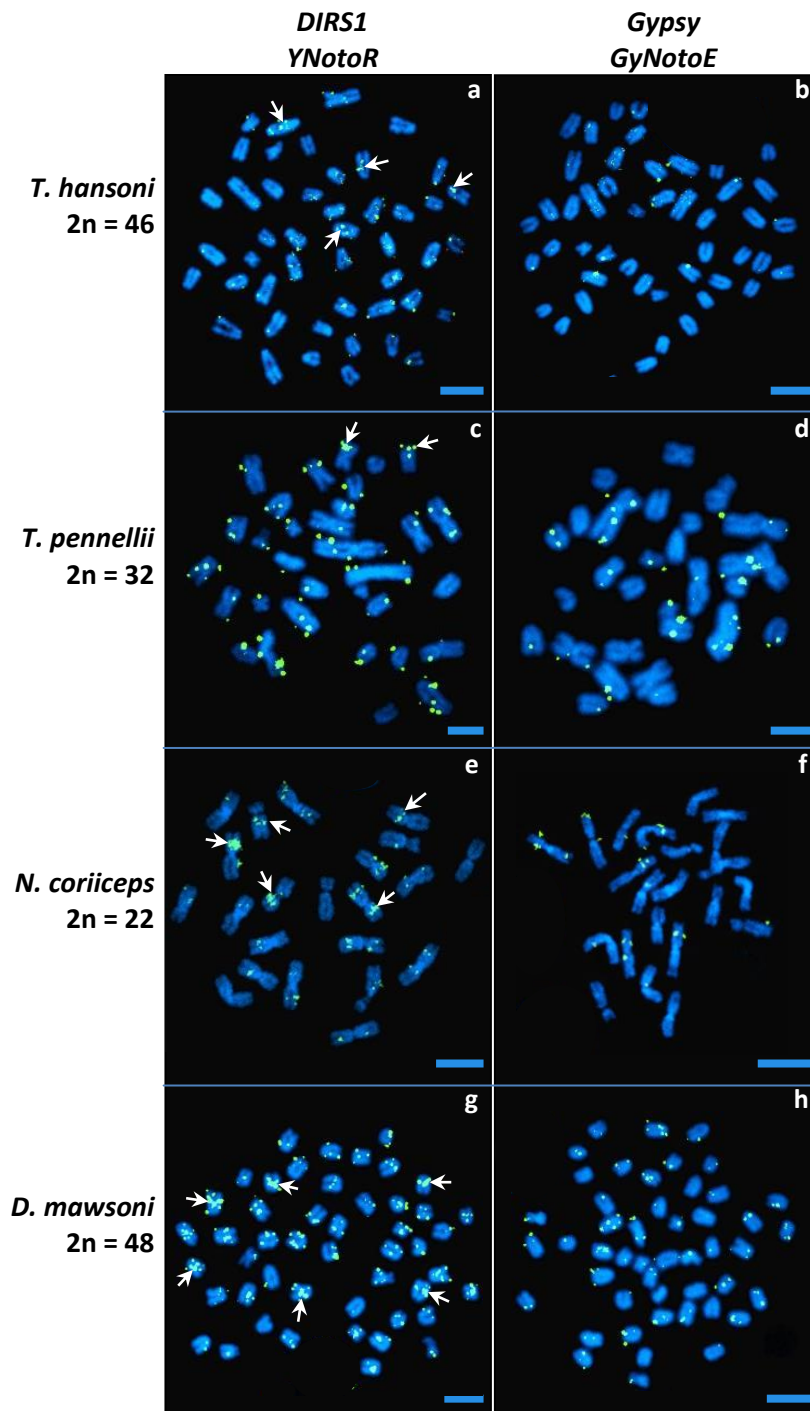




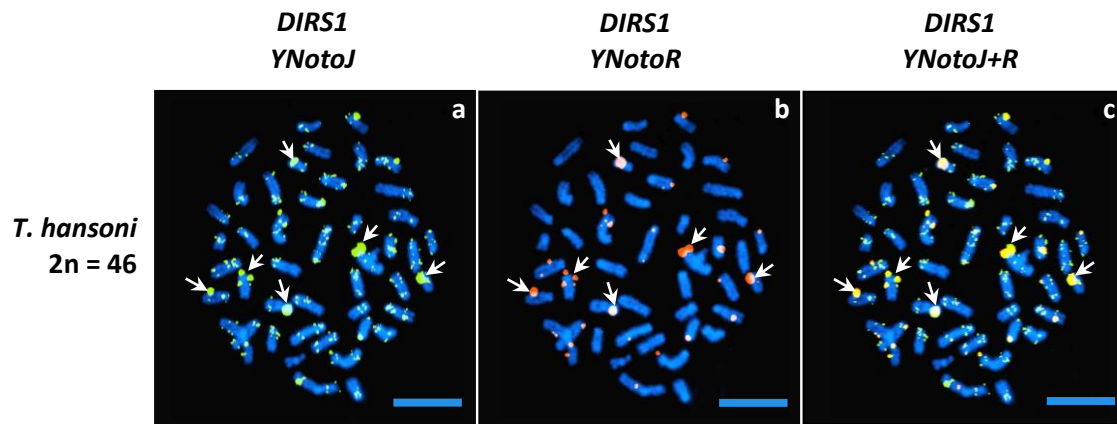
**Figure IV.6: Mapping of TEs on the chromosomes of five nototheniid species by FISH.** Each probe was labeled with biotin and bound probe was detected with incubation with Avidin-FITC (fluorescein, greenish spots). (Probe characteristics are indicated in Additional file IV.6). Chromosomal DNA was counterstained with DAPI. One family of each retrotransposon superfamily is represented in this figure: *YNotoJ* for *DIRS1*, *GyNotoA* for *Gypsy* and *CoNotoB* for *Copia* elements. (see Additional file IV.7 and IV.8 for FISH mapping with the second family of *DIRS1* (*YnotoR*) and *Gypsy* (*GyNotoE*)). Examples of TE distribution patterns for type 1: d, h; type 2: c, i; type 1+2: e, j, p. White arrows point examples of TE accumulations. TE Scale bars: 10  $\mu$ m.

| <b>TEs</b>   | <b>Family</b>          | <b>Sequence name</b> | <b>Species related</b> | <b>N° specimen*</b> | <b>Insert size (kb)</b> |
|--------------|------------------------|----------------------|------------------------|---------------------|-------------------------|
| <i>DIRS1</i> | <i>YNotoJ</i>          | <i>YTbeJ1</i>        | <i>T. bernacchii</i>   | TA364TRPE1          | 1.25                    |
|              | <i>YNotoR</i>          | <i>YThaR1</i>        | <i>T. hansonii</i>     | TA440TRHA2          | 1.10                    |
| <i>Gypsy</i> | <i>GyNotoA</i>         | <i>GyTpeA1</i>       | <i>T. pennellii</i>    | TA364TRPE1          | 1.54                    |
|              | <i>GyNotoE</i>         | <i>GyThaE1</i>       | <i>T. hansonii</i>     | TA440TRHA2          | 1.50                    |
| <i>Copia</i> | <i>CoNotoB (Hydra)</i> | <i>CoTpeB1</i>       | <i>T. pennellii</i>    | TA364TRPE1          | 1.10                    |

**Additional file IV.6: Summary of TEs probes for FISH.** \* for precisions (origin, survey), see Additional file IV.9 (taxonomic sampling). For precision about TE sequence characteristics, see Table IV.2.



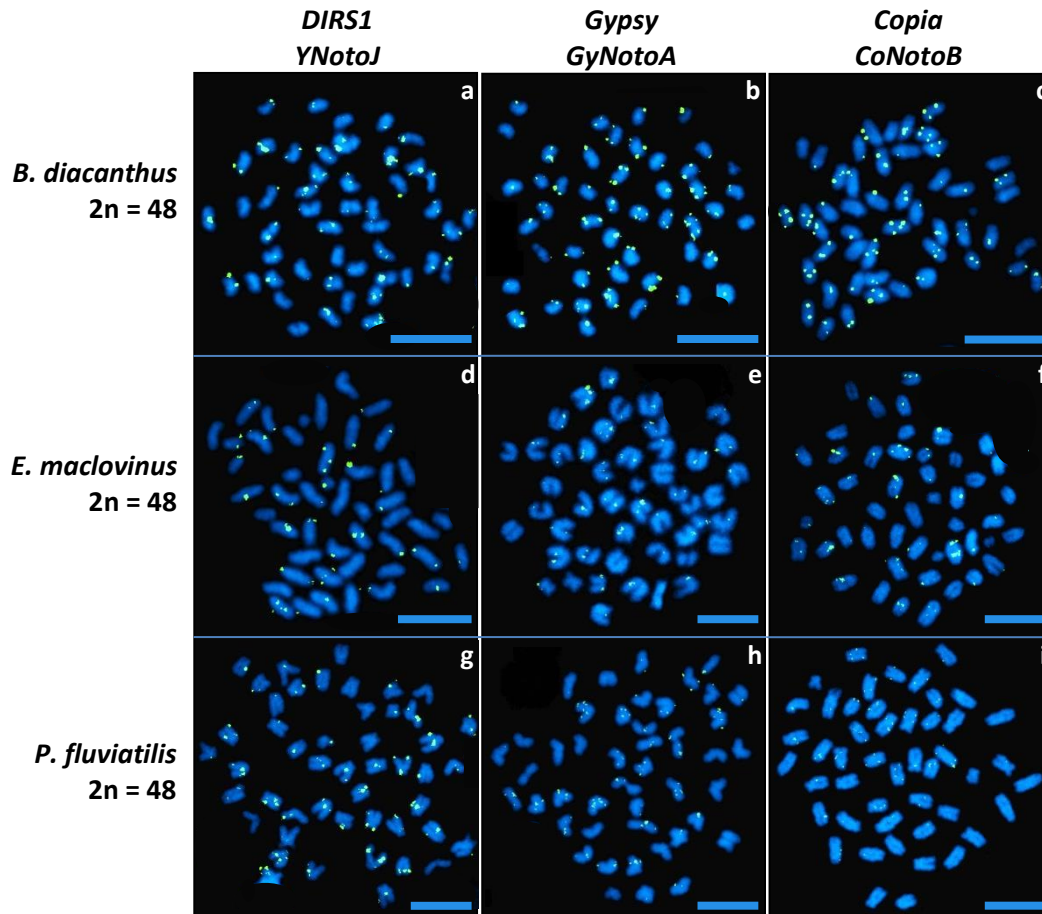
**Additional file IV.7: Mapping of TEs on the chromosomes of five nototheniid species by FISH.** Complement of Figure IV.6. Each probe was labeled with biotin and bound probe was detected with incubation with Avidin-FITC (fluorescein, greenish spots). (Probe characteristics are indicated in Additional file IV.6). Chromosomal DNA was counterstained with DAPI. One family of each retrotransposon superfamily is represented in this figure: *YNotoR* for *DIRS1*, *GyNotoE* for *Gypsy* elements. White arrows point examples of TE accumulations (distribution pattern corresponding to type 1). Scale bars: 10  $\mu$ m.



**Additional file IV.8: Double mapping of two *DIRS1* family representatives on chromosomes of *T. hansonii*.**

Complement of Figure IV.6 and Additional file IV.7. *YNotoJ* probe was directly labeled with fluorescein (greenish spots) and *YNotoR* probe, with Rhodamine (red spots). (Probe characteristics are indicated in Additional file IV.6). Chromosomal DNA was counterstained with DAPI. Two families of *DIRS1* elements are represented in this figure: *YNotoJ* (a), *YNotoR* (b), and the resulting signals of both of them superimposed (c). White arrows point examples of TE accumulations (distribution pattern corresponding to type 1), and the colocalization of the two *DIRS1* families. Scale bars: 10  $\mu\text{m}$ .

The two investigated families of *Gypsy* and *Copia Hydra* elements were mostly dispersed throughout nototheniid chromosome arms (type 2), forming multiple spots scattered everywhere on nototheniid chromosome arms (*T. pennellii*, *T. eulepidotus*, *N. coriiceps* and *D. mawsoni*) as well as in centromeres and telomeres of acrocentric (*T. hansonii*, *T. pennellii* and *D. mawsoni*) and metacentric (*T. eulepidotus*, *T. pennellii* and *N. coriiceps*) chromosomes. Signals occasionally accumulated on metacentric chromosomes (giving a type 1+2 pattern) (Figure IV.6.e, j, p). In the nototheniid species, the two *FISH*-investigated families of *DIRS1* elements appeared to accumulate (white arrows in Figure IV.6 and Additional file IV.7) in centromeric regions (type 1), mostly in short acrocentric pairs (Figure IV.6.d, n), in some telomeres (Figure IV.6.d, h) and were also aggregated in pericentromeric areas (type 1), especially in karyotypes with metacentric chromosomes (Figure IV.6.h, k). The *DIRS1* element pattern (type 1) was very distinct from the *Gypsy* and *Copia* patterns, with clear hot spots of insertion in all nototheniid species studied. In contrast, these *DIRS1* hot spots were not detected in the three outgroups *B. diancantus*, *E. maclovinus* and *P. fluviatilis*, where *FISH* signals were mostly scattered on chromosome arms (type 2), with some of them located on telomeres, but never aggregated. Even if the *Gypsy* and *Copia* *FISH* signals seemed less abundant, they were still scattered on chromosome arms (type 2) like for the five High Antarctic species (Figure IV.7).



**Figure IV.7: Mapping of TEs on the chromosomes of the three outgroups by FISH.** Each probe was labeled with biotin and bound probe was detected with incubation with Avidin-FITC (fluorescein, greenish spots). (Probe characteristics are indicated in Additional file IV.6). Chromosomal DNA was counterstained with DAPI. One family of each retrotransposon superfamily is represented in this figure: *YNotoJ* for *DIRS1*, *GyNotoA* for *Gypsy* and *CoNotoB* for *Copia* elements. Scale bars: 10  $\mu$ m.

#### Retrotransposon quantification

The same five TE families (*YNotoJ*, *YNotoR*, *GyNotoA*, *GyNotoE*, *CoNotoB*) were quantified in the same five nototheniid species (*T. pennellii*, *T. hansonii*, *T. eulepidotus*, *N. coriiceps*, *D. mawsoni*) using qPCR. Our quantification results indicated that *Gypsy* and *Copia* elements were present in 7-90 copies per family studied, while *DIRS1* elements were generally represented by 100-150 copies per family in the five genomes (Table IV.4). For the *DIRS1* retrotransposons, *YNotoJ* and *YNotoV* appeared the most represented families in *T. hansonii*, *T. eulepidotus*, *N. coriiceps* and *D. mawsoni* genomes. In *T. pennellii*, we estimated 370 *YNotoJ* but a roughly equal number for *YNotoV* (110) and *YNotoRB* (125). The *GyNotoA* (between 10 and 90 copies) were more represented than *GyNotoE* (< 10 copies, apart from for *T. hansonii*, with an estimate of 30 copies detected) except in *N. coriiceps* (8 *GyNotoA* vs 20 *GyNotoE*). *T. pennellii* carried the highest

number of *DIRS1* retrotransposons when counting all families together (total of 605 estimated copies) and *T. hansonii*, the highest number of *Gypsy* elements (total of 120 estimated copies). Both of those genomes also contained the highest copy number for *Copia Hydra* elements (45 estimated copies) whereas we estimated a low number of copies, 10 to 15 for the other three species *T. eulepidotus*, *N. coriiceps* and *D. mawsonii*.

| Species \ TEs         | <i>DIRS1</i>  |               |                | <i>Gypsy</i>   |                | <i>Copia</i>   |
|-----------------------|---------------|---------------|----------------|----------------|----------------|----------------|
|                       | <i>YNotoJ</i> | <i>YNotoV</i> | <i>YNotoRB</i> | <i>GyNotoA</i> | <i>GyNotoE</i> | <i>CoNotoB</i> |
| <i>T. pennellii</i>   | 370           | 110           | 125            | 50             | 8              | 45             |
| <i>T. hansonii</i>    | 275           | 200           | 30             | 90             | 30             | 45             |
| <i>T. eulepidotus</i> | 250           | 80            | 55             | 45             | 7              | 15             |
| <i>N. coriiceps</i>   | 80            | 135           | 15             | 8              | 20             | 10             |
| <i>D. mawsonii</i>    | 120           | 340           | 10             | 20             | 7              | 10             |

**Table IV.4:** TE copy numbers estimated in the nototheniid genomes.

## Discussion

### TE diversity

Our investigation revealed a high diversity of retrotransposons within nototheniid genomes, with four families of *DIRS1*, nine of *Gypsy* and two of *Copia* identified. This result corroborates previous studies on the large number and repartition of TEs in teleost species (Aparicio et al. 2002; Volf et al. 2003; Valente et al. 2011). Within nototheniid species, Detrich *et al.* (2010) have described TEs in *Notothenia coriiceps* and *Chaenocephalus aceratus* from partial genome sequencing data, finding respectively 13.4 and 14.5 % of their genomes composed of a wide TE diversity. Retrotransposons seem more represented than DNA transposons.

The present study shows retrotransposon sequence conservation between species, with levels of interspecific nucleotide identity similar to those observed between *Trematomus* species for *Rex1* and *Rex3* non-LTR retrotransposons (Ozouf-Costaz et al. 2004), or for nuclear coding genes like *Pkd1* (98.4 %), *RPS71* (98.2 %) and *Rhodopsin* (98.9 %) (Dettai et al. 2012; Near et al. 2012). This result suggests that these TEs are subject to strong selective pressure to maintain their activity (Radice et al. 1994; Valente et al. 2011). They also show that transposition events may have occurred recently in those genomes. These observations might be subject to some biases due to the method employed to detect the conserved domains on pol region and thus, preferentially potentially active, or recently active TEs.

The presence of twelve of the fifteen identified TE families in all the studied genomes points to a divergence of the TE families prior to the divergence of the nototheniid species (Capriglione et

al. 2002). Our sampling includes species representing the oldest divergence in the family, estimated between -22.4 and -13.4 My depending on the study (Near et al. 2012; Colombo et al. 2015). Their mobilization in the genomes however could have occurred more recently, perhaps during nototheniid radiation events.

Although our nototheniid TE consensus sequences belong systematically to the Vertebrate or the fish clade, they are not grouped together like in other studies (Ozouf-Costaz et al. 2004). All species studied were found to contain at least one copy of each identified TE family, except two families of *Gypsy* (*GyNotoD*, and *GyNotoJ*) and one of *Copia Galea* (*CoNotoA*). This last family was not recovered in the genus *Trematomus*, but it was successfully amplified in *N. coriiceps*, *D. mawsoni* and other Nototheniidae (Additional file IV.2). We extended the search for *Galea* to other trematomine species closely related to *Trematomus*: genera *Indonotothenia*, *Lepidonotothen* and *Patagonotothen* without being able to successfully amplify them by PCR. We are aware that a lack of PCR amplification of a given DNA fragment does not mean it is absent from the genome examined. However, given the strong interspecific conservation of *Galea* family in nototheniids and the tendency of *Galea* elements to be secondary lost in entire clades (for example within Fungal and Crustacean genomes (Piednoël et al. 2013; Donnart et al. 2017), this would indicate the presence of *Galea* retrotransposons in the last common ancestor of all Nototheniidae was followed by a secondary loss in Trematominae. Such result supports recent phylogenetic studies grouping together the genera *Cryothenia*, *Indonotothenia*, *Pagothenia*, *Trematomus*, *Lepidonotothen* and *Patagonotothen* in the Trematominae sub-family (Lautrédou et al. 2010; Rutschmann et al. 2011; Dettai et al. 2012; Lautrédou et al. 2012; Near et al. 2012; Duhamel et al. 2014; Colombo et al. 2015).

Extensive chromosomal rearrangements and conserved genome size within the *Trematomus*.

*Trematomus* C-values for the different studied species are very close to each other (1.0 to 1.2 pg). They fall also in the same range as the values obtained or published for other nototheniid sub-family representatives: *Notothenia* (*N. coriiceps*, 1.36 pg, Detrich et al. (2010)), *Dissostichus* (*D. mawsoni*, 1.02 pg, Detrich et al. (2010)), *Lepidonotothen* (*Lepidonotothen nudifrons*, 1.12 pg Detrich et al. (2010)), and *Gobionotothen* (*Gobionotothen gibberifrons*, 0.98 pg Detrich and Amemiya, (2010)). While there is no available genome size data for the sister groups of Nototheniidae (Bovichthidae, Eleginopsidae or Pseudaphritidae), or for the most recently proposed sister group of Notothenioidei, Percophidae (Near et al. 2015), the genome sizes of other families phylogenetically close to the suborder Notothenioidei like Percidae (*P. flavescens*, 0.92 pg, *P. fluviatilis*, 0.90 pg, *Percina caprodes*, 1.06 pg, *Sander lucioperca*, 1.14 pg, *Sander*

*vitreus*, 1.06 pg) (Hardie and Hebert 2004; Detrich et al. 2010), or Serranidae (1.09 pg mean genome size, 0.17 s.d) (Dettai and Lecointre 2004) also range between 0.90 and 1.23 pg. This suggests that a C value of 1-1.20 pg could be the ancestral state, conserved throughout the *Trematomus* radiation. Therefore, karyotype diversity within the genus *Trematomus* occurred without much variation in the amount of DNA per genome. Rearrangements involving changes in *Trematomus* chromosome numbers and formulae are structural modifications like tandem or Robertsonian centric fusions ( $2n = 24$  in *T. eulepidotus*,  $2n = 32$  in *T. pennellii* and  $2n = 45-46$  in *T. hansonii*, *T. borchgrevinkii* and *T. newnesii*) or fissions ( $2n = 57-58$  in *T. nicolai*) and maybe inversions. However, inversions could not be detected using *FISH* or classical chromosome banding tools (Morescalchi et al. 1992; Pisano and Ozouf-Costaz 2000).

Genome size stability is not observed in all nototheniid species. We also measured the C-values of some nototheniid species from the sub-families Channichthyinae, Gymnodraconinae and Bathydraconinae, phylogenetically closely related to each other, but further from the Trematominae. Although few species have been measured yet, they display higher C-values: 1.72 pg for *Chionodraco hamatus*, 1.54 pg for *Gymnodraco acuticeps*, and 1.34 pg for *Parachaenichtys charcoti*. This noticeable genome size augmentation could be explained by massive transpositions, segmental or genome duplication that could have happened during diversification of these groups (Kidwell 2002; Detrich and Amemiya 2010; Kraaijeveld 2010).

Did TE mobilization favor chromosomal diversification during the *Trematomus* radiation?

To evaluate the representation of some *DIRS1*, *Gypsy*, and *Copia* families in nototheniid genomes in the absence of a well-assembled genome available to run exhaustive computational analyses, we tried to estimate their copy number using qPCR approaches relative to two single copy gene standards. Despite the repeated nature of the TE sequences, this approach has already been used to estimate TE copy number on genomic DNA (Orsoly et al. 2011; Kraaijeveld et al. 2012; Chen et al. 2017). The same TE families quantified by qPCR were also mapped on nototheniid chromosomes using *FISH*. Even if *FISH* is not precise enough to estimate accurate TE copy numbers, we could estimate their relative number by the intensity of signals (Figure IV.6, IV.7, Additional file IV.7, IV.8). There is a consistency between the two approaches that is generalizable to the five nototheniids studied (numerous hot spots and high copy number, few interspersed spots and low number of copies) (Table IV.4, Figure IV.6, Additional file IV.7). In at least a few cases, there is probably a tendency for the qPCR to underestimate the real copy number, as based on their intensity most *FISH* signals correspond to multiple copies. This is particularly obvious for the *Hydra (CoNotoB)* probe as there are multiple spots (certainly more



than the estimated 10 copies) present in all chromosomes of *N. coriiceps* and *D. mawsoni* (Table IV.4, Figure IV.6). The regions amplified by qPCR are only a portion of the TE sequence used as probe for *FISH*; the signals are therefore not totally comparable. Whole or partial genome sequencing would be needed to get more precise estimations of relative weights of TE families in those genomes and their respective true copy numbers.

The copy number estimated for *DIRS1* elements is higher than for *Gypsy* and *Copia* retrotransposons in all five genomes quantified. It cannot, however, be excluded that the quantification of *Gypsy* elements is more biased by unequal amplification because of their higher family diversity (we quantified two of the four identified *DIRS1* families vs two of the nine identified *Gypsy* families). This tendency of *DIRS1* prevalence is even more pronounced for the three *Trematomus* species studied, *T. eulepidotus*, *T. hansonii* and *T. pennellii*. The predominance of *DIRS1* copies over *Gypsy* and *Copia* copies is contrary to the strong prevalence of *Gypsy* previously reported in the Takifugu genome (10 putative copies of *DIRS*, 35 copies of *BEL/Pao*, 50 copies of *Copia* and 2 500 copies of *Gypsy* identified) (Aparicio et al. 2002), in the zebrafish genome (46 806 copies of *DIRS*, 4 585 copies of *BEL/Pao*, 10 293 copies of *Copia* and 56 138 copies of *Gypsy* identified) (Howe et al. 2013), and in two nototheniid genomes based on partial sequencing of a BAC library (for *N. coriiceps*, 17 copies of *DIRS*, 1 copy of *Copia* and 49 copies of *Gypsy* identified, for *C. aceratus*, 46 copies of *DIRS*, 25 copies of *Copia* and 156 copies of *Gypsy* identified, no *BEL/Pao* identified in both cases) (Detrich et al. 2010; Detrich and Amemiya 2010). The *DIRS1* prevalence in *Trematomus* and other High Antarctic species investigated is consistent with the observed *FISH* signals (Figure IV.6). As it has been proposed in several studies, strong multiplication of retrotransposons could have happened during chromosomal diversification accompanying speciation events (Capriglione et al. 2002; Kidwell 2002; Lönnig and Saedler 2002; Curcio and Derbyshire 2003; Ozouf-Costaz et al. 2004; Volf 2005; Böhne et al. 2008; Schneider et al. 2013; Splendore de Borja et al. 2013). The bursts of retrotransposons could therefore have occurred during the radiation of *Trematomus* and other AFGP-bearing nototheniid relatives (Chaniichthyinae, Artedidraconinae).

The possible implication of massive TE mobilization in the observed *Trematomus* and nototheniid species chromosomal rearrangements is dependent on their location on the chromosomes. We noticed a major difference of *DIRS1* insertion patterns compared to those of *Gypsy* or *Copia* elements. *Gypsy* and *Copia* have numerous and scattered interstitial hybridization spots on chromosomes with no clear regions of accumulation, a pattern comparable to those previously reported for *Tc1*-like DNA transposons (Capriglione et al. 2002) and for *Rex*-like non-LTR retrotransposons (Ozouf-Costaz et al. 2004) with mappings on nototheniids chromosomes (including *T. bernacchii*, *T. hansonii*, *T. newnesi* and *T. pennellii*). In striking contrast, *DIRS1*

elements accumulate in hot spots of insertion mostly located in centromeric and pericentromeric regions (Figure IV.6) for the several nototheniid radiations studied, including in *Chionodraco hamatus* (sub-family Channichthyinae). In the biggest chromosome pair of *T. pennellii*, which certainly arose from two fusions (C. Ozouf-Costaz, unpublished data), there is an intercalary band of *DIRS1* in the long arm in addition to the pericentromeric hot spot. This pattern of insertion is reminiscent of the one highlighted by Ozouf-Costaz *et al.* (2004) with *Rex3* retrotransposon on the Y chromosome (originating from a tandem fusion between one X<sub>1</sub> and one X<sub>2</sub> autosome) of *C. hamatus* (Ozouf-Costaz *et al.* 2004). The intercalary band observed in the long arm of this Y chromosome suggest *Rex3* involvement in the tandem fusion (telomere-centromere) between the two X<sub>1</sub> and X<sub>2</sub> autosomes this chromosome originates from (Pisano *et al.* 2003; Ghigliotti *et al.* 2015).

*DIRS1* are not structured like LTR retrotransposons and this specificity may be linked with its particular origin. Discovered in the slim mold *Dictyostelium discoideum* (Goodwin and Poulter 2001), *DIRS1* might have arisen from the combination of a *crypton*-like transposon using tyrosine recombinase (YR) to cut and rejoin the recombining DNA molecules and the RT/RH part of a pre-existing *Gypsy* LTR retrotransposon (Poulter and Butler 2015). Various molecular mechanisms for *DIRS1* insertion in targeted sites (by integration, recombination) have been proposed but remain to be experimentally demonstrated (Goodwin and Poulter 2001; Goodwin 2004; Poulter and Goodwin 2005). A specific insertion mode was deduced (Rajeev *et al.* 2009) from this singular molecular structure: via a double-stranded circular pre-integrative DNA intermediate, catalyzed by the YR. This explains the absence of Target Site Duplication (TSD) following transposition (Duncan *et al.* 2002; Curcio and Derbyshire 2003). This is reminiscent of bacterial processes and could lead to ectopic recombination (Bonnivard and Higuët 2009; Kraaijeveld 2010) and thus, to rearrangements. Hot spots of insertion for DNA transposons (*Tc1*-like) and retrotransposons (*Rex*-like) have already been described a number of times in the literature for other teleost fish groups (Capriglione *et al.* 2002; Slotkin and Martienssen 2007; Mazzuchelli and Martins 2009; Valente *et al.* 2011; Schneider *et al.* 2013). *DIRS1* clustering in hot spots of integration is probably not random. Many *DIRS1* elements appear to have preferential target site selection and insert into their pre-existing copies in *Danio rerio* (Capello *et al.* 1985; Eickbush and Jamburuthugoda 2008). This process, also called “homing”, is usually reported for class II TEs, such as the *Drosophila P element* (Taillebourg and Dura 1999) and, to our knowledge, has been rarely reported for retrotransposons. Like the other TEs localized in teleost fish chromosomes (Capriglione *et al.* 2002; Slotkin and Martienssen 2007; Mazzuchelli and Martins 2009; Valente *et al.* 2011; Schneider *et al.* 2013), *DIRS1* inserts in « sheltered » heterochromatic regions (Pimpinelli *et al.* 1995; Biémont and Vieira 2005; Biémont and Vieira

2006; Mazzuchelli and Martins 2009; Schneider et al. 2013). Such locations in regions less affected by selection and with lower recombination rates might protect the TEs (Ferreira et al. 2011). It might also have been selected as a way to protect and structure genomes, ensuring a kind of self-control of their mobilization (Hua-Van et al. 2005; Slotkin and Martienssen 2007; Böhne et al. 2008; Valente et al. 2011).

A breakdown of epigenetic control mechanisms (Mansour 2007; Kraaijeveld 2010; Rebollo et al. 2010) could have favored *DIRS1*, *Gypsy*, and *Copia* activation, both in *Trematomus* species and other AFGP-bearing nototheniid relatives. According to recent time-calibrated molecular phylogeny analyses (Near et al. 2012; Colombo et al. 2015), the *Trematomus* species diversification occurred in the middle – late Miocene (-17 to -5 My), corresponding to a cooling of shallow Antarctic shelf waters, followed by rapid switches between glacial and interglacial conditions (Kennett 1986; Kennett and Barker 1990; Stott et al. 1990; Clark et al. 2004). This could have led to transpositional waves through genomes (Mahé 2009; Piednoël and Bonnivard 2009; Chen et al. 2017) possibly linked to species diversification (Kraaijeveld 2010; Rebollo et al. 2010). On the contrary, *B. diacanthus*, *E. maclovinus*, *P. fluviatilis* (non AFGP-bearing species chosen as outgroups) were not exposed to series of environmental changes with cooling at sub-zero temperatures (Kennett 1986; Kennett and Barker 1990), and do not reveal any regions of TE accumulation on their chromosomes for the studied TEs (Figure IV.7). TE hot spots were not detected in Bovichtidae (including one species from the subantarctic Southern Ocean), or in the two temperate species from the families Eleginopsidae and Percidae (including a few species that originated in South America, Southern Australia/Tasmania and New-Zealand) (Duhamel et al. 2014). Their karyotypes (2n = 48, all acrocentric chromosomes) (Nygren et al. 1968; Mazzei et al. 2006; Mazzei et al. 2008) are very stable across species (Pisano et al. 1995; Mazzei et al. 2006), suggesting few structural chromosomal rearrangements.

## Conclusion and perspectives

Transposable elements constitute key points to explore and understand genome evolution, especially in the recent divergence of Antarctic *Trematomus* species. While our TE exploration is wider than in previous studies, we cannot yet conclude the lack of detection of a given TE or family in a genome means they are absent. Targeted amplicon shotgun sequencing and untargeted genome sequencing are needed for more exhaustive exploration of TE populations and diversity.

No bursts linked with TE multiplication and accumulation on chromosomes were detected on the three temperate and subantarctic outgroups without AFGPs. On the contrary, exposure to strong environmental changes with cooling at sub-zero temperatures and series of glaciation and deglaciation cycles could have led to massive mobilization of retrotransposons (epigenetic deregulations), observed within *Trematomus* and other nototheniid genomes. A predominance of *DIRS1* (probably several hundred of copies) in hot spots of insertion suggests they could have facilitated repeated and localized DNA double strand breaks on centromeric and pericentromeric regions. Then, the cellular repair mechanisms, close to recombination processes, might have favored tandem or centric fusions and boosted other genomic changes (Bourque 2009; Oliver and Greene 2009). Chromosomal rearrangements reinforce reproductive isolation between distinct populations and drive vertebrate lineage diversification. They may have accompanied the *Trematomus* radiation.

Like the repeated independent emergences of the male Y-chromosome from centric fusions or tandem translocations, the chromosomal rearrangements observed in *Trematomus* species, and more generally in Nototheniidae, probably occurred independently as the result of convergent events. However, precise reconstruction of the evolutionary history of *Trematomus* chromosomal diversification requires identifying the inter-specific homologies between chromosomes. In this group, these homologies are generally impossible to establish due to the resemblance between chromosomes when using karyotype approaches based on chromosome morphology or size, and on DAPI counterstaining. Since it is impossible to recognize the chromosomes by classical techniques like banding, the development of tools like chromosome painting or BAC-*FISH* in these species would help to ascertain these inter-specific homologies (Ozouf-Costaz et al. 1997), permitting a more precise history of the genome reorganizations (evolutionary scenario) in this group, and their possible relation with the speciation events, to be established.

## Materials and Methods

### Fish specimens

Specimens of thirteen High Antarctic nototheniids (*Trematomus eulepidotus*, *T. pennelli*, *T. borchgrevinki*, *T. hansonii*, *T. bernacchii*, *T. loennbergi*, *T. lepidorhinus*, *T. newnesi*, *T. scotti*, *T. nicolai*, *I. cyanobranca*, *Notothenia coriiceps*, and *Dissostichus mawsoni*) and of two Sub-Antarctic notothenioid outgroups (*Bovichtus diacanthus*, *Eleginops maclovinus*) were collected during groundfish survey programs sponsored by the IPEV (Institut Polaire français Paul-Emile Victor) and the NSF (U.S. National Science Foundation): the ICEFISH 2004 Cruise (International Collaborative Expedition to collect and study Fish Indigenous to Sub-antarctic Habitats, Atlantic

sector of the Southern Ocean), ICOTA (Ichthyologie côtière en Terre Adélie, 1996-2008, Adélie Land), CEAMARC (Collaborative East Antarctic Marine Census for the Census of Antarctic Marine Life; 2007-2008, Eastern Antarctica, north of Adélie Land and George V Land), POKER (POissons de KERguelen, 2006, 2010 and 2013, shelf of Kerguelen-Heard islands) and REVOLTA (Radiation EVOLutives en Terre Adélie, 2010-2014, Adélie Land). Specimens of *Perca fluviatilis* were obtained in 2010 in Eaucourt (Somme, FRANCE) and are part of the MNHN collection. Tissue samples, chromosomal preparations and blood cells are referenced in Additional file IV.9. We used the nomenclature and classification of the SCAR atlas (Duhamel et al. 2014).

| Family (sub-family)            | Genus. Species          | Sampling type                            | Field reference  | Voucher reference   |
|--------------------------------|-------------------------|--|--|---|
| Nototheniidae (Trematominae)   | <i>T. eulepidotus</i>   | tissue                                   | si487n3152 <sup>4</sup>  | MNHN 2009-1358  |
|                                |                         | chromosomes                              | CE2381 <sup>5</sup> , m251 <sup>6</sup>  |   |
|                                |                         | erythrocytes                             | CE4011 <sup>5</sup>  |   |
|                                | <i>T. pennellii</i>     | tissue                                   | TA364TRPE1 <sup>3</sup>  | MNHN 1996-0319  |
|                                |                         | chromosomes                              | TA647 <sup>3</sup>   |   |
|                                |                         | erythrocytes                             | CE1434 <sup>5</sup>  |   |
|                                | <i>T. borchgrevinki</i> | tissue                                   | TA263PABO1 <sup>3</sup>  | MNHN 2002-1711  |
|                                |                         | erythrocytes                             | CE5560 / 5561 / 5562 / 5655 / 5656 / 5662 / 5664 <sup>5</sup>  |   |
|                                | <i>T. hansonii</i>      | tissue                                   | TA440TRHA2 <sup>3</sup>  | MNHN 1996-0307  |
|                                |                         | chromosomes                              | TA 646TRHA2 <sup>3</sup>   |   |
|                                |                         | erythrocytes                             | CE 4002 / 4003 / 4004 / 4316 / 5660 / 5664 / 5708 / 5711 / 5738 / 5856 / 5880 / 5661) <sup>5</sup>                                     |   |
|                                | <i>T. bernacchii</i>    | tissue                                   | Si352n2561 <sup>4</sup> / TA 337TRBE7-1/01/03 <sup>3</sup>   | MNHN 2009-1312  |
|                                |                         | erythrocytes                             | Ich1098 / 1099 / 1100 / 1101 / 1102 / 1114 / CE4314 / 4315 / 5658 / 5659 / 5685 / 5709 / 5737 / 6042 / 6043 / 6091 / 6092 <sup>5</sup> |   |
|                                | <i>T. loennbergii</i>   | tissue                                   | si259n923 <sup>4</sup> / si170n1265 <sup>4</sup>   | MNHN 2009-1246  |
|                                |                         | erythrocytes                             | CE5916 <sup>5</sup>  |   |
| <i>T. lepidorhinus</i>         | tissue                  | TNB238 <sup>6</sup>                      | MNHN 1999-0394   |   |
| <i>T. newnesi</i>              | tissue                  | TA401TRNE3 <sup>3</sup>                  | MNHN 2001-1148 <sup>3,5</sup>  |   |
|                                | erythrocytes            | CE4319 <sup>5</sup> / TA456 <sup>3</sup> |  |   |
| <i>T. scotti</i>               | tissue                  | si541n6317 <sup>4</sup>                  | MNHN 2009-1368   |   |
| <i>T. nicolaii</i>             | tissue                  | TNB214 <sup>6</sup>                      | MNHN 1999-0400 <sup>6</sup>  |   |
|                                | erythrocytes            | Ich1115 / CE5684 <sup>5</sup>            | MNHN 2002-1707 <sup>5</sup>  |   |
| <i>I. cyanobranca</i>          | tissue                  | Aus30 <sup>2</sup>                       | MNHN 2007-1844   |   |
| Nototheniidae (Nototheniinae)  | <i>N. coriiceps</i>     | tissue                                   | N°10RECOLTE2001 / TA421NOC08 <sup>3</sup>  | MNHN 2008-1884 <sup>3,4,5</sup><br>not vouchered <sup>1</sup> |
|                                |                         | chromosomes                              | TA449 <sup>3</sup> / NC4 <sup>1</sup>  |   |
|                                |                         | erythrocytes                             | Ich1095 / 1096 / 1097 <sup>5</sup>   |   |
| Nototheniidae (Dissostichinae) | <i>D. mawsoni</i>       | tissue                                   | REVO759 <sup>5</sup>   | MNHN 2001-1143 <sup>5</sup><br>not vouchered <sup>1</sup>     |
|                                |                         | chromosomes                              | DM1 <sup>1</sup>   |   |
|                                |                         | erythrocytes                             | CE4054 / 4055 / 4056 <sup>5</sup>  |   |
| Eleginopsidae                  | <i>E. maclovinus</i>    | chromosomes                              | RG <sup>1</sup>  | MNHN 2005-0093  |
| Bovichtidae                    | <i>B. diacanthus</i>    | chromosomes                              | BD1 / BD4 <sup>1</sup>   | MNHN 2005-0102  |
| Percidae                       | <i>P. fluviatilis</i>   | chromosomes                              |  | MNHN 2010-987   |

**Additional file IV.9: Taxonomic sampling for tissues, chromosomal suspensions and blood cells used in this**

**study.** Tissue samples, chromosomal preparations and blood cells were collected during the Antarctic campaigns of the IPEV programs: **ICEFISH2**<sup>1</sup> (2004, Southern Ocean and south atlantic), **POKER**<sup>2</sup> (2010 and 2013 Kerguelen-Heard, shelf and islands), **ICOTA**<sup>3</sup> (1996-2008, Adélie Land), **CEAMARC**<sup>4</sup> (2007/2008 north of Adélie Land and George V Land of Eastern Antarctica), **REVOLTA**<sup>5</sup> (2010-2014, Adélie Land). The fish identifier and collector is Catherine Ozouf-Costaz except for NC4, DM1, RG, BD1 and BD4: Guillaume Lecointre; TNB and m251: Eva Pisano, and sample of *P. fluviatilis*: Gaël Denys. TNB and m251 samples are originated from the **ICEFISH1**<sup>6</sup> campaign and conserved in the museum of Genoa (no voucher reference). A single voucher reference is provided per species/locality. Other specimens are deposited in the MNHN research collections.

## Sample collection and preparation

### Blood samples for flow cytometry

Fish were anesthetized with MS 222 before sampling blood (minimum 0.5 ml) by caudal venipuncture using heparinized syringes. Whole blood was diluted 1:10 with PBS, 50 µl aliquots of each blood sample were distributed dropwise to Eppendorf tubes containing 1 ml of 70 % ice cold ethanol, and the tubes were stored at -20°C.

### Tissues for DNA analyses

Muscle samples or fin clips for DNA analyses were stored in 85 % ethanol at -20°C. DNA was prepared following the protocol of Winnepenninckx *et al.* (1993).

### Chromosome preparations

Mitotic chromosome preparations were obtained from primary cell cultures of pronephric kidney or spleen according to Rey *et al.* (2015). Briefly, a suspension of cephalic kidney and/or spleen cells was prepared and cultured in L-15 Leibovitz culture medium without bicarbonate and supplemented with L-glutamine, fetal calf serum, lectins (concanavalin A and pokeweed mitogen) and antibiotics for periods up to one week at 0° to 2°C. Colchicine was added 6 h prior to harvesting cells. This was followed by a hypotonic treatment (1 h at 2°C) and conventional steps of fixation. Fixed cell suspensions were preserved as aliquots of 15 ml at -20°C. Prior to use, the suspensions were thawed and centrifuged at 1,500 rpm (Eppendorf 5430 microcentrifuge) for 10 min at 4°C. After decanting the supernatant, the cell pellet was resuspended in 0.8-1 ml of fresh fixative, and cells were spread onto Superfrost slides (pre-cleaned with absolute ethanol containing 1 % of 1N HCl). Slides were stored at -20°C until the FISH step.

Additional file IV.9 summarizes the blood samples, tissue samples, and chromosomal preparations used in this study.

## Retrotransposon amplification and fragment assembly

### Amplification of retrotransposons using degenerate primers

Retrotransposons were amplified from the genomes of all *Trematomus* species, with the exception of *T. tokarevi*, and from two nototheniid sister species, *N. coriiceps* and *D. mawsoni* (Additional file IV.9). Rather than pursue full-length elements (5 to 10 kb), we focused on their *pol* regions, as they include conserved domains and have a fundamental function in transposition. We amplified the retrotransposon *pol* fragments by PCR using degenerate oligonucleotide primer pairs designed for conserved protein motifs of the Reverse Transcriptase/RNaseH (RT/RH) domains (Additional file IV.10). For *Gypsy* retrotransposons, the GD1b/GD2b primer pair overlaps the PFLG/DASXXGW motifs. CD1 had been designed to amplify *Copia* elements when used with CD2, as previously employed for the galatheid squat lobsters (Terrat et al. 2008), but in nototheniid fishes, we found that it amplified the *Gypsy* Integrase gene, serving as both forward and reverse primer. For *Copia* elements, two primer pairs were used (CD3/CD4, CD5/CD6), which correspond to the DYCVR/DNQG and VDP/QLAD motifs, respectively. For *DIRS1* retrotransposons, we used the DD10/DD11 pair, which encode the DlkdAY/YafPPf motifs.

PCR was performed using 50 ng of genomic DNA, 2.5 U of Taq DNA polymerase (Promega) and 50 pmol of each degenerate primer in a final volume of 25  $\mu$ L for 35 cycles (94°C for 45 s, 50.2°C for 1 min and 72°C for 1 min). PCR products were visualized on 1 % agarose gels. Fragments of the expected molecular weights were excised, purified with the Nucleospin Extraction kit (Macherey\_Nagel), and cloned into the pGEM-T vector according to the supplier's recommendations (Promega, Madison, WI, USA). Cloned fragments were sequenced in both directions (<http://www.gatc-biotech.com>).

### TE Walking

Optimal DNA fragment sizes for classification of retrotransposons and for use as *FISH* probes, in our experience, range between 1.0 and 1.5 kb. The initial *Gypsy* amplicons (0.7 kb), in particular, did not meet this standard (Additional file IV.10). Therefore, we used "TE Walking" (Piednoël and Bonnivard 2009), anchored by specific primers designed from the original sequences and the GD4 degenerate primer (overlaps the YLDD motif), to extend the *Gypsy* fragments (Additional files IV.10 and IV.11). *Copia* and *DIRS1* elements were extended in similar fashion. Each new sequence was manually validated as an extension of the original fragment using a minimum overlap of 50 bp between the two sequences and a minimum DNA identity of 95 %. Final retrotransposon sequences were assembled from their fragments using the Cap contig assembly program included in BioEdit v7.2.5 (Hall 1999). We then designed specific primer pairs that we used to amplify the desired *pol* fragments (Additional file IV.11). We obtained DNA cloned fragments from 1.0 to 1.5 kb size, depending on the TE element. Consensus sequences



were deposited in BankIt (BankIt2016770: MF142597-MF142757). These larger DNA fragments were used for classification of retrotransposons and as *FISH* probes.

|                | TE                 | Primer              | Primer sequence 5' → 3' | Motif   | Fragment size (kb) |
|----------------|--------------------|---------------------|-------------------------|---------|--------------------|
| <b>deg</b>     | <i>Gypsy</i>       | GD1b                | KYYTBCCRTTYGGSYT        | PFGL    | 0.7                |
|                |                    | GD2b                | CCCCAGCCYKWSYKHCARGC    | DAS..GW |                    |
|                | <i>Copia Hydra</i> | CD1                 | ARRGCNMGNYTNGTNGC       | KARLVA  | 0.6                |
|                |                    | CD3                 | AYRTNGAYTAYTGYTAYMG     | DYCYR   | 1.4                |
|                |                    | CD4                 | GCDATNGYNCCYTGRTRRTC    | DNQG    |                    |
|                | <i>Copia GalEa</i> | CD5                 | GAYCCNGCNGTNTTYTAYTGG   | VDP     | 0.9                |
| CD6            |                    | TNARRCARTCNGCNARYTG | QLAD                    |         |                    |
|                | <i>DIRS1</i>       | DD10                | GAYYTNAARGAYGCNTAY      | DlkdAY  | 1.2                |
|                |                    | DD11                | RAANGGNGGRAANRYRTA      | YafPPf  |                    |
| <b>spe-deg</b> | <i>Gypsy</i>       | GD4                 | NYVDTAYMTVGAYGAY        | YLDD    | 1.3                |

**Additional file IV.10: Degenerated PCR primers used to amplify retroelements in nototheniid genomes.** deg: degenerate primer pairs used to amplify the different *DIRS1*, *Gypsy* and *Copia* TE superfamilies. spe-deg: degenerate primer coupled with specific primers (listed in Additional file IV.11) for the “TE walking”.

|             | TE                   | Primer | Primer sequence 5' → 3' | Fragment size (kb) |
|-------------|----------------------|--------|-------------------------|--------------------|
| <b>spe*</b> | <i>Gypsy</i>         | GS1    | ATGGGTAGAGGCTGAAGCAA    | 1.3                |
|             |                      | GS2    | GGAACCTTCCCTGGTGT       |                    |
|             |                      | GS3    | TCCACGGAAATGTATGACCA    |                    |
| <b>spe</b>  | <i>Gypsy</i>         | GS4    | GACGGAGAAGTGCGAGTTTC    | 1.6                |
|             |                      | GS5    | ATTGGCACAATGGAAAGGAG    |                    |
|             |                      | GS6    | GACGGAGAAGTGCGAGTTTC    | 1.5                |
|             |                      | GS7    | ATTGGCACAATGGAAAGGAG    |                    |
|             |                      | GSA1   | GGGTCAAGGAAGTGTGGAGA    | 1.5                |
|             |                      | GSA2   | CTCCACCCAMAGGAGTTGTT    |                    |
|             |                      | GSB1   | CCAGGAAGAAGGTYCARCAG    | 1.5                |
|             |                      | GSB2   | RCTCCAGGTRKTCTGGTTCYG   |                    |
|             |                      | GSE1   | CACGTTTCCACCGTCTTT      | 1.5                |
|             |                      | GSEF2  | CTGAAATGGGGAAAGACCAG    |                    |
|             |                      | GSF1   | AGGTGGGGGAGTATGTTGCT    | 0.6                |
|             |                      | GSH1   | TGTCAACGAGGACACTCAGG    | 0.6                |
|             |                      | GSH2   | CAAACGGGGAGAGACCAGTA    |                    |
|             |                      | GS11   | TGCTTAACCCACAACATCAGA   | 0.5                |
|             |                      | GS12   | AAATGGGGAGAAACCTGTCC    |                    |
|             |                      | GSJ1   | AGGCTGAAGCAAGCAGGTTA    | 1.5                |
|             |                      | GSJ2   | TAGAGGGGTTACGGTCAACG    |                    |
| GSRT1       | CTAAGGGATGAGTGCTGCTC | 0.6    |                         |                    |
| GSRT2       | GAAGGACGAAGGGCAA     |        |                         |                    |
|             | <i>Copia</i>         | HS1    | GCCCAGATATGGTTGCAAAAGG  | 1.1                |
|             | <i>Hydra</i>         | HS2    | TGACCCTCACTRTCCATGCTCT  |                    |
|             | <i>DIRS1</i>         | DSB1   | CTCTCCCTCTCTCCAAGGGT    | 1.0                |
|             |                      | DSB2   | CATYGGGCTGGTCACCTCTG    |                    |
|             |                      | DSR1   | CATCCARTATATTATCCAC     | 1.1                |
|             |                      | DSR2   | CGCATTGAAAAGAACAGCTG    |                    |
|             |                      | DSJ1   | TTCAGCAAATGTGTGGCCA     | 1.1                |
|             |                      | DSJ2   | CGTGTGCAAACGCGTCCAC     |                    |
|             |                      | DSV1   | CTGGCTCCTCGCACTTTTTC    | 1.0                |
|             |                      | DSV2   | GCGGTTCTCGCGTGCAGC      |                    |

**Additional file IV.11: Specific PCR primers used to amplify retroelements in nototheniid genomes (complement of Additional file IV.9).** spe\*: specific primers coupled with spe-deg primer (Additional file IV.10) for the “TE walking”. spe: specific primer pairs used to amplify the different families of *DIRS1*, *Gypsy* and *Copia* TEs.

## Retrotransposon classification, clustering, and phylogenetic analyses

### Classification and clustering of retrotransposon families

Nototheniid TEs were assigned to a specific retrotransposon superfamily using BLASTX analyses (Altschul et al. 1990) of a custom in-house TE database (2.2.28 + blastpackage), the NCBI protein database, and the *N. coriiceps* genome (Shin et al. 2014). *DIRS1*, *Gypsy*, or *Copia* sequences were clustered using BLASTClust toolkit v2.2.26 (Alva et al. 2016). A cluster of sequences was considered a separate family if its highest intra-group divergence was lower than the divergence between all putative families, without overlap of the two distributions (Piednoël and Bonnivard 2009).

The criteria for inclusion of a fragment in a cluster (i.e., family) were  $\geq 30\%$  sequence coverage and  $\geq 80\%$  sequence identity. We adopted the following nomenclature: *YNotoJ*, *V*, *R*, and *B* are the four identified families of *DIRS1*; *GyNotoA*, *B*, *D*, *E*, *F*, *H*, *I*, *J*, and *RT*, the nine identified families of *Gypsy*; *CoNotoA* and *CoNotoB*, the two identified families of *Copia*. Nucleic acid sequence identity matrices were calculated after alignment of sequences by MAFFT v7 (Katoh and Standley 2013) and gap removal using BioEdit v7.2.5.

### Phylogenetic analysis of nototheniid retrotransposons

The sequences of the nototheniid *DIRS1*, *Gypsy*, and *Copia* families were highly conserved across the genera examined. To establish their phylogenetic relationships with respect to TEs from other eukaryotes, we first generated a majority rule nucleotidic consensus for each family of *DIRS1*, *Gypsy* and *Copia* TEs using Geneious v9.0.2 (<http://www.geneious.com>, Kearse *et al.* (2012)), then translated them to generate corresponding amino acid sequences. For *DIRS1* elements, the four nototheniid consensus sequences were added to a dataset of *DIRS1* (Piednoël *et al.* 2011) that span 96 amino acid sequences of the RT/RH region and represent numerous eukaryote taxa (including Fungi, Crustaceans, Nematodes). Two phylogenetic analyses of the nine *Gypsy* consensus sequences were performed, the first covering the RT/RH region (110 amino acids) and the second the INT region (99 amino acids). Nototheniid consensus sequences were added to the *Gypsy* sequence compilation (Piednoël *et al.* 2013). Similarly, *CoNotoA* and *CoNotoB* sequences that span the RT/RH region were compared to a *Copia* dataset (Piednoël *et al.* 2013). The final datasets contained 282 informative sites for *DIRS1*, 167 for *Gypsy* RT/RH, 153 for *Gypsy* INT and 220 for *Copia*.

Multisequence alignments were performed with MAFFT v7 and ambiguously aligned sites were removed using Gblocks (Castresana 2000). Neighbor-joining (NJ) phylogenies were obtained

using MEGA 5.2.2 (Tamura et al. 2011). The best-fit model was selected with Topali v2, implemented using PhyML (Milne et al. 2004) and the JTT model (Jones et al. 1992) with gamma distribution. Maximum Likelihood (ML) phylogenetic reconstructions were obtained using RAxML 8.2.12 (Stamatakis 2014) and the evolution model PROTGAMMALG. Support for individual clusters was evaluated using non-parametric bootstrapping (Felsenstein 1985) and 1,000 bootstrap replicates. Nodes under 60 % were collapsed.

We selected outgroups for each reconstruction according to previously published retroelement relationships (Goodwin and Poulter 2001; Eickbush and Malik 2002; Eickbush and Jamburuthugoda 2008; Rho et al. 2010). Due to the proximity (sequence identity) of their RT/RH domains, we chose the DIRS1 outgroup for the Gypsy reconstruction and reciprocally, Gypsy as outgroup for the DIRS1 and Copia analyses. Because DIRS1 elements do not encode an Integrase (but a tYrosine Recombinase), Copia (GalEa) sequences were used as root for the Gypsy analysis based on Integrase consensus fragments.

## *FISH*

### TE probe preparation

*DIRS1*, *Gypsy* and *Copia* clones from *T. bernacchii*, *T. pennellii* and *N. coriiceps* greater than 1 kb in length were used as probes for *FISH* experiments (Additional file IV.6). TE probes were biotinylated by nick translation according to the manufacturer's instructions (Roche Diagnostics). For the double *FISH* mapping, we used probes in which fluorochromes were directly incorporated during the labeling step; Fluorescein for *YNotoJ*, and Rhodamine for *YNotoR*, using the ULS PlatinumBright Nucleic Acid Labeling Kit (Leica Biosystems). Each probe was dissolved at a final concentration of 20 ng/μl in high stringency hybridization buffer [65 % formamide, 2x SSC, 10 % dextran sulfate (pH 7)].

### *FISH with TE probes*

To ensure signal specificity for the three TE superfamilies, *FISH* was performed on chromosome preparations under high stringency conditions. Chromosome preparations were obtained from eight fish species, including five nototheniids (*T. pennelli*, *T. hansonii*, *T. eulepidotus*, *N. coriiceps*, *D. mawsoni*) and the three outgroup species (*B. diacanthus*, *E. maclovinus*, and *P. fluviatilis*). Biotinylated TE probes, denatured by heating at 85°C for 5 min, were applied to freshly thawed chromosome preparations, which were then incubated at 72°C for ten seconds to one minute.

Bound probes were detected using Avidin-FITC (fluorescein) according to the protocol published in the literature (Bonillo et al. 2015), which is optimized for repetitive sequences and multi-copy genes. Hybridization parameters were adjusted to each chromosome preparation: we pre-incubated the slides at 37°C to make chromosomes more resistant to denaturation, determined specific chromosome denaturation times for each specimen and tested several probe concentrations to select the one most adapted to our chromosomal material (Liehr 2009; Ozouf-Costaz et al. 2015).

#### Image acquisition and karyotyping

*FISH* signals were detected using a Zeiss Axioplan microscope equipped with a cooled CCD camera (Coolsnap Photometrics) and an XCite LED fluorescence light source. Karyotypes were processed using CytoVision 3.93.2/Genus *FISH*-imaging software for animal chromosomes (Leica Microsystems). Ten to forty metaphase spreads/species for each probe were examined.

#### Quantification

#### Genome size determination

The flow cytometry procedure was based on Detrich *et al.* (2010). 500 µl of each sample was washed twice by centrifugation (Eppendorf 5430 microcentrifuge, 1,500 rpm, 4°C, 5 min), first in 0.01 M PBS followed by 0.01 M PBS containing RNase (10 µg/ml, Miltenyi Biotec) and propidium iodide (PI; 50 µg/ml, Sigma Aldrich). Flow cytometry measurements were performed using a MACSQuant 10 flow cytometer. Unstained cells were used to determine autofluorescence thresholds and to calibrate the acquisition mode. Rainbow trout (*Oncorhynchus mykiss*) blood was collected at the Institut National de la Recherche Agronomique (INRA), Jouy-en-Josas ("synthetic INRA" strain, 2015). This rainbow trout provides our genome size reference, as it has been repeatedly estimated by flow cytometry with a mean calculated C-value of 2.695 pg (see also Volff *et al.* (2005): 2,600 Mbp with 978 Mbp = 1 pg). We systematically started each run with triplicate measurements of the reference rainbow trout, followed by *N. coriiceps* and *T. bernachii* prior to data acquisition for the other nototheniid samples (Additional file IV.1). At least 10 000 cells were measured per sample and each experiment was replicated 3X. Fluorescence intensities, which were analyzed using FlowJo v10 software, reflect PI intercalation into DNA and are directly proportional to genome size. The genome sizes were calculated by comparing the mean of fluorescence intensity (MFI) of each sample to the MFI of our reference.

### TE copy number determination by quantitative PCR (qPCR)

qPCR reactions were performed in a CFX96 Touch Real-time PCR Detection System (Bio-Rad). Specific primer pairs were designed with Oligo Analyzer v1.2, 1) to amplify 450 – 500 bp fragments overlapping the RT/RH conserved region of *DIRS1*, *Gypsy* and *Copia* TEs, and 2) to amplify fragments of similar size from the single copy *RAG1* and *Rhodopsin* genes (Additional file IV.12). The best practice recommendations (Bustin et al. 2009) were followed for sample preparation. PCR amplifications were performed in a final volume of 20 µl containing 4 µl of template DNA, 10 µmol of each primer, 10 µl of Sso Advanced Universal SYBR Green Supermix (Bio-Rad), and 4 µl of DNase free water. The PCR thermal cycling protocol began with polymerase activation and DNA denaturation at 98 °C for 4 min, followed by 40 cycles with denaturation for 15 s at 98 °C and annealing/extension for 1.5 min at 62 °C. After amplification, melting curve analyses were performed between 65 and 95 °C to determine amplification product specificity, with the temperature increasing of 0.1 °C/s. SYBR Green fluorescence was measured during the annealing/extension step. All samples were analyzed in triplicate with negative controls (different combinations: without DNA and primers, with DNA but without primers and with primers but without DNA).

For *DIRS1* elements, we created a primer pair to quantify simultaneously both the *YNotoR* and *YNotoB* families due to their high similarity in the RT/RH region (Additional file IV.12). Standard curves were generated for *RAG1* and *Rhodopsin* in each run. We performed qPCR for all fragments (*RAG1*, *Rhodopsin*, *YNotoJ*, *YNotoV*, *YNotoRB*, *GyNotoA*, *GyNotoE* and *CoNotoB*) for five species: *T. pennellii*, *T. hansonii*, *T. eulepidotus*, *N. coriiceps*, and *D. mawsoni*. PCR products were systematically checked by sequencing to verify primer specificity (Genewiz). Four to five tenfold dilution standards were prepared fresh each time, ranging from 1 to  $1 \times 10^{-4}$  ng/µl. For both single copy gene standards, at least four dilutions in triplicate were obtained and plotted against Ct values. We estimated copy numbers of *DIRS1*, *Gypsy* and *Copia* family relative to both single copy (*RAG1* on the one hand, and *Rhodopsin* on the other hand) calibration measurements using Bio-Rad CFX Manager v3.1, by comparison of their respective C(t) values, normalized to the DNA concentration initially present in the sample dilution. The slope of the standard curves was calculated and the amplification efficiency (E) was estimated as  $E = (10^{-1}/\text{slope})^{-1}$  for each amplicon and each dilution. We measured efficiencies for both standards (mean of  $98.2 \pm 2.6$  %), and for the whole TE families (mean of  $97.8 \pm 2.1$  %). *T<sub>m</sub>* of each studied amplicon was constant between runs (Additional file IV.12).

| Se-<br>quenc<br>e type  | Markers          | primer                         | Primer sequence 5' → 3'                           | Len<br>gth<br>(bp) | Region of<br>amplicon | Tm of<br>amplicon<br>(°C) |
|-------------------------|------------------|--------------------------------|---|--------------------|-----------------------|---------------------------|
| Single<br>copy<br>genes | <i>RAG1</i>      | RAG1F23<br>RAG1R458            | AGCCAGCTCTCAAGAATGTGTC<br>GTGAGTGTCTCATGGTCTGACT  | 457                | RAG1                  | 86.0                      |
|                         | <i>Rhodopsin</i> | RhodoF193<br>RhodoR563         | CCNTATGAATAYCCTCAGTACTACC<br>GTGATCCTCCCCGAAGCG   | 396                | Rhodo_N<br>7TM-GPCR   | 86.5                      |
| DIRS1                   | <i>YNotoj</i>    | DJF226<br>DJR676               | CGGGAACAGGCAATGTTTTG<br>TCCGTGAACACTGTGATGT       | 475                | RT/RH                 | 89.0                      |
|                         | <i>YNotoV</i>    | DVF229<br>DVR656               | GAAGTNGCTTTACAGACGGTACA<br>GTGACGTAACCTGCCAGG     | 447                | RT/RH                 | 88.5                      |
|                         | <i>YNotoR/B</i>  | DRBF185<br>DRBR696             | GCCGTATCTAGACRACACTGGC<br>CCCACCCASAGAGACATGC     | 530                | RT/RH                 | 89.3                      |
| <i>Gypsy</i>            | <i>GyNotoA</i>   | GF1all<br>GR480g1              | GACGGAGAAGTGGAGTTTC<br>CCAACGCCAACTTGACTGCT       | 503                | RT/RH                 | 88.0                      |
|                         | <i>GyNotoE</i>   | GyNotoEF330<br>GyNotoER78<br>5 | TTGACGCTTCWGATGTGGGCG<br>RACCCTCTCCTCAATGTCCCA    | 452                | RT/RH                 | 86.5                      |
| <i>Copia</i>            | <i>CoNotoB</i>   | HydraF951<br>HydraR1391        | ACAATATCGTGAAGTGGTAGGCA<br>TGACCCTCACTRTCCATGCTCT | 431                | RT/RH                 | 83.5                      |

**Additional file IV.12: Specific qPCR primer pairs**

## List of abbreviations

TEs: Transposable Elements; ORFs: open reading frames, RT: Reverse Transcriptase; RH: RNaseH; My: Million years; *FISH*: Fluorescent *In Situ* Hybridization; INRA: Institut National de la Recherche Agronomique; (q)PCR: (quantitative) Polymerase Chain Reaction; ICR: Internal Complementary Region; ITR: Inversed Terminal Repeat; MT: MethylTransferase; YR: tYrosine Recombinase; DAPI: 4',6-diamidino-2-phenylindole; PBS: phosphate buffered saline; NJ: Neighbor Joining; DIG: digoxigenin; FITC: fluorescein isothiocyanate; SSC: sodium chloride-sodium citrate buffer; MFI: mean on fluorescence intensity; BAC: bacterial artificial chromosome; AFGP: antifreeze glycoprotein; TSD: target site duplication.

## Declarations

Ethics approval and consent to participate

Ethical approval for all procedures was granted by the ethics committee of the Ministère de l'Environnement and the French Polar Research Institute (Institut Paul Emile Victor – IPEV), which approved all our fieldwork. The experiments complied with the Code of Ethics of Animal Experimentation in the Antarctic sector.

Consent for publication

“Not applicable”.

Availability of data and materials

The datasets (TE sequences and genome sizes) supporting the conclusions of this article are available in the GenBank NCBI repository (BankIt 2016770, refs MF142597 to MF142757) and on the genome size database repositories, [unique persistent identifier and hyperlink to dataset(s) in [http:// format](http://format)].

## **Competing interests**

“The authors declare that they have no competing interests”.

## **Funding**

This work, chromosome collection and preparation received financial support from the Institut Polaire français Paul-Emile Victor (IPEV) and the U.S. National Science Foundation (NSF) (programs Ichthyologie côtière en Terre Adélie (ICOTA 1996-2008) and Radiation EVOLutives en Terre Adélie (REVOLTA 2010-2014)); of Collaborative East Antarctic Marine Census for the Census of Antarctic Marine Life (CAML-CEAMARC 2007-2008, IPY project no. 53) cruise on board the RV Aurora Australis; of International Collaborative Expedition to collect and study Fish Indigenous to Sub-antarctic Habitats, Atlantic sector of the Southern Ocean (ICEFISH 2004), on board the R.V. Nathaniel Palmer (granted by National Science Foundation grant OPP 01-32032), of POissons de KERguelen, shelf of Kerguelen-Heard islands (POKER 2006, 2010, 2013) on board the “Austral” trawler (supported by the French National Research Agency ANR, and Ministry of Agriculture and Fisheries). The Université Pierre et Marie Curie (UPMC), the Centre National de la Recherche Scientifique (CNRS), the Museum National d’Histoire Naturelle (MNHN) and the Australian Antarctic Division (AAD) provided financial supports. HWD was also supported by the U.S. National Science Foundation (NSF) grant PLR-1444167. This is contribution 365 from the Marine Science Center at Northeastern University.

## **Authors' contributions**

Experiments were conceived, discussed and designed by DH, COC, AD, HWD and JA. The TE sequence dataset was generated with the kind help of PG and LB, and analyzed by DH and JA; *FISH* and qPCR experiments and analyses by JA, helped by LP, for flow cytometry experimental design, data acquisitions and analyses. The manuscript was prepared by JA and HWD. DH, COC,



HWD, EB and AD made a critical revision of the manuscript and participated in helpful discussions. All authors read and approved the final manuscript.

## **Acknowledgments**

We are grateful to Dimitri Rigaudeau (INRA of Jouy en Josas, FRANCE) for the Rainbow Trout blood sampling. We thank the crews and participants involved in the capture of the samples and chromosome preparations in different sectors of Antarctica and particularly M. Hautecoeur, G. Lecointre, H. D'Hont, S. Pavoine, C. Loots, O. Rey, J. Lanshere, T. Nebout, E. Pisano and C. Cheng. We thank Eva Pisano for additional samples, providing chromosomes of *T. eulepidotus* (stock of cytogenetic preparations for polar fishes organized by the CNR of Genoa, PNRA POLICY project). We thank the "Service de Systématique Moléculaire" (SSM, UMS 2700, MNHN/CNRS) for granting access to its technical platform and help. We are grateful to Pedro Bausero for his precious help and advices regarding our qPCR quantification of TE copy numbers. We also thank the editor and anonymous reviewers for their helpful comments.

## **Authors' information**

?: LP is in charge of the cytometers from the imaging platform of Sorbonne Université, 75 252 Paris, cedex 05

## **IV.3 : diversité des éléments transposables à l'échelle de la famille des Nototheniidae**

### **IV.3.A : recherche des familles des trois rétrotransposons chez les Nototheniidae**

#### ***Résultats et discussion***

Au cours de notre études des ETs au sein du groupe « *Trematomus* », nous avons utilisé une méthode ciblée, recherchant les rétrotransposons *DIRS1*, *Gypsy* et *Copia* à partir d'amorces dégénérées, puis spécifiques. Compte tenu du fort niveau d'identité au sein des familles d'ETs identifiées chez les « *Trematomus* », nous avons recherché ces même familles dans les génomes de dix-sept autres Nototheniidae, ainsi que deux autres représentants du sous-ordre des Notothenioidei : l'Eleginopsidae *E. maclovinus* et le Bovichtidae *B. diancanthus* (Tableau IV.5). Pour cela, nous avons utilisé les amorces spécifiques préalablement déterminées (voir Méthodes du IV.1 et le fichier additionnel IV.11). Les spécimens utilisés au cours de cette étude sont répertoriés en annexe (annexe 1).

| Super-familles d'ETs               | DIRS1  |        |        |        |         | Gypsy   |         |         |         |         |         |         |         |         |         | Copia    |         |         |         |    |
|------------------------------------|--------|--------|--------|--------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|----------|---------|---------|---------|----|
|                                    | YNotOB | YNotOR | YNotOI | YNotOV | GYNotoA | GYNotoB | GYNotoC | GYNotoD | GYNotoE | GYNotoF | GYNotoG | GYNotoH | GYNotoI | GYNotoJ | GYNotoK | GYNotoRT | CoNotoA | CoNotoB | CoNotoC |    |
| <i>Indonotothenia cyanobrancha</i> | +      | +      | +      | +      | +       | +       |         |         | +       | +       |         |         |         |         |         |          |         | +       | NT      | +  |
| <i>Trematomus scotti</i>           | +      | +      | +      | +      | +       | +       |         |         | +       | +       |         |         |         |         |         |          |         |         | NT      | +  |
| <i>Trematomus nicolai</i>          | +      | +      | +      | +      | +       | +       |         |         | +       | +       |         |         |         |         |         |          |         |         | NT      | +  |
| <i>Trematomus bernachii</i>        | +      | +      | +      | +      | +       | +       |         |         | +       | +       |         |         |         |         |         |          |         |         | NT      | +  |
| <i>Trematomus lepidorhinus</i>     | +      | +      | +      | +      | +       | +       |         |         | +       | +       |         |         |         |         |         |          |         |         | NT      | +  |
| <i>Trematomus borchgrevinkii</i>   | +      | +      | +      | +      | +       | +       |         |         | +       | +       |         |         |         |         |         |          |         |         | NT      | +  |
| <i>Trematomus hansonii</i>         | +      | +      | +      | +      | +       | +       |         |         | +       | +       |         |         |         |         |         |          |         |         | NT      | +  |
| <i>Trematomus newnesi</i>          | +      | +      | +      | +      | +       | +       |         |         | +       | +       |         |         |         |         |         |          |         |         | NT      | +  |
| <i>Trematomus boenberglii</i>      | +      | +      | +      | +      | +       | +       |         |         | +       | +       |         |         |         |         |         |          |         |         | NT      | +  |
| <i>Trematomus eulepidotus</i>      | +      | +      | +      | +      | +       | +       |         |         | +       | +       |         |         |         |         |         |          |         |         | NT      | +  |
| <i>Trematomus tokarevi</i>         | +      | +      | +      | +      | +       | +       |         |         | +       | +       |         |         |         |         |         |          |         |         | NT      | +  |
| <i>Lepidonotothen nudifrons</i>    | +      | +      | +      | +      | +       | +       |         |         | NT      | +       | +       |         |         |         |         |          |         |         | NT      | +  |
| <i>Lepidonotothen larseni</i>      | +      | +      | +      | +      | +       | +       |         |         | +       | +       |         |         |         |         |         |          |         |         | NT      | +  |
| <i>Lepidonotothen squamifrons</i>  | +      | +      | +      | +      | +       | +       |         |         | +       | +       |         |         |         |         |         |          |         |         | NT      | +  |
| <i>Patagonotothen guntheri</i>     | +      | +      | +      | +      | +       | +       |         |         | +       | +       |         |         |         |         |         |          |         |         | NT      | +  |
| <i>Pleurogramma antarctica</i>     | +      | +      | +      | +      | +       | NT      |         |         | NT      | +       | +       |         |         |         |         |          |         |         | NT      | +  |
| <i>Nototothen coriiceps</i>        | +      | +      | +      | +      | +       | +       |         |         | +       | +       |         |         |         |         |         |          |         |         | +       | +  |
| <i>Nototothen angustata</i>        | +      | +      | +      | +      | +       | +       |         |         | +       | +       |         |         |         |         |         |          |         |         | +       | +  |
| <i>Nototothen rossii</i>           | +      | +      | +      | +      | +       | +       |         |         | +       | +       |         |         |         |         |         |          |         |         | +       | +  |
| <i>Paranotothenia magellanica</i>  | +      | +      | +      | +      | +       | +       |         |         | +       | +       |         |         |         |         |         |          |         |         | NT      | +  |
| <i>Disostichus mawsonii</i>        | +      | +      | +      | +      | +       | +       |         |         | +       | +       |         |         |         |         |         |          |         |         | +       | +  |
| <i>Champsoscephalus gunnari</i>    | +      | +      | +      | +      | +       | +       |         |         | +       | +       |         |         |         |         |         |          |         |         | NT      | +  |
| <i>Chionodraco hamatus</i>         | +      | +      | +      | +      | +       | +       |         |         | +       | +       |         |         |         |         |         |          |         |         | NT      | NT |
| <i>Cyrodoraco antarcticus</i>      | +      | +      | +      | +      | +       | +       |         |         | +       | +       |         |         |         |         |         |          |         |         | NT      | +  |
| <i>Champsoscephalus esox</i>       | +      | +      | +      | +      | +       | +       |         |         | +       | +       |         |         |         |         |         |          |         |         | NT      | +  |
| <i>Cygnodraco mawsonii</i>         | +      | +      | +      | +      | +       | +       |         |         | +       | +       |         |         |         |         |         |          |         |         | NT      | +  |
| <i>Gymnodraco acuticeps</i>        | +      | +      | +      | +      | +       | NT      |         |         | +       | +       |         |         |         |         |         |          |         |         | +       | +  |
| <i>Pogonophryne scotti</i>         | +      | +      | +      | +      | +       | NT      |         |         | +       | +       |         |         |         |         |         |          |         |         | +       | +  |
| <i>Histiadraco velifer</i>         | +      | +      | +      | +      | +       | +       |         |         | +       | +       |         |         |         |         |         |          |         |         | NT      | +  |
| <i>Gobionotothen gibberifrons</i>  | +      | +      | +      | +      | +       | +       |         |         | NT      | +       | +       |         |         |         |         |          |         |         | +       | +  |
| <i>Eleginops maclovinus</i>        | +      | NT     | +      | NT     | NT      | NT      |         |         | NT      | NT      | +       | NT      | +       |         |         |          |         |         | NT      | NT |
| <i>Bovichtus diacanthus</i>        | NT     | NT     | NT     | NT     | NT      | NT      |         |         | +       | NT      | NT      | NT      | +       |         |         |          |         |         | NT      | NT |

**Tableau IV.5 : distributions de quatre familles de DIRS1, douze familles de Gypsy et trois familles de Copia dans les génomes de trente-et-une espèces de Nototheniidae et deux espèces de Notothenioidae (*E. maclovinus* et *B. diacanthus*). + : ETs recherchés et trouvés dans les génomes des différentes espèces (au moins une séquence par espèce), NT : ETs recherchés, mais non trouvés, case vide : ETs identifiés chez une espèce appartenant à une nouvelle famille, mais non recherchés chez les autres espèces.**

Les familles d'ETs ubiquitaires chez les « *Trematomus* » (*YNotoJ*, *V*, *R* et *B*, *GyNotoA*, *B*, *E*, *F*, *H*, *I* et *RT*) ont été retrouvées chez la quasi totalité des espèces de Nototheniidae examinées (Tableau IV.5). Dans certains cas, l'amplification nous a permis d'identifier de nouvelles familles de *Gypsy* (*GyNotoC*, *G*, et *K*) (Tableau IV.5). Tout comme *GyNotoD* et *GyNotoI* précédemment identifiées chez les « *Trematomus* », ces familles n'ont pas été recherchées dans les génomes des autres espèces. Une absence de séquence dans le tableau IV.5 (cases vides) ne signifie donc pas que ces familles sont absentes dans les génomes des différentes espèces. Comme *CoNotoB*, *CoNotoC* a été identifié comme un nouvel élément *Copia* appartenant à la famille des *Hydra*. Cette sous-famille a été retrouvée chez sept espèces différentes et présente un degré d'identité interspécifique similaire à celui trouvé pour les autres familles d'éléments *DIRS1* et *Gypsy* (96,7 %). Il est intéressant de souligner que lorsque la sous-famille *CoNotoB* a été identifiée dans le génome d'une espèce, la sous-famille *CoNotoC* n'a pas été trouvée, et inversement.

En ce qui concerne la recherche des ETs dans les génomes des deux groupes externes *E. maclovinus* et *B. diacanthus*, nous avons amplifié moins de familles que chez les Nototheniidae. On retrouve en effet deux familles de *DIRS1* et *Gypsy* (*YNotoJ* et *B*; *GyNotoG* et *I*), et une famille de *Copia* (*CoNotoB*) chez *E. maclovinus* et seulement une séquence de *Gypsy* (*GyNotoE*) chez *B. diacanthus* (Tableau IV.5). Ces génomes sont en effet phylogénétiquement plus éloignés des Nototheniidae et les séquences d'ETs présentent une divergence plus importante (70 % d'identité seulement pour les ETs d'une même famille identifiés entre les deux groupes externes et les autres espèces). Nous posons l'hypothèse que cette divergence peut également s'appliquer au niveau des sites d'accroches des amorces spécifiques. Toutefois, leur hybridation sur les chromosomes confirme leur présence dans les génomes d'*E. maclovinus* et *B. diacanthus*, mais tend à montrer un faible nombre de signaux (Figure VI.7) que l'on pourrait assimiler à un plus faible nombre de copies, rendant d'autant plus difficile leur amplification. Afin d'accéder à ces séquences chez les groupes externes et d'identifier potentiellement de nouvelles familles non trouvées dans les génomes des Nototheniidae, nous pourrions envisager d'utiliser d'autres approches ciblées, comme par exemple la capture ou la préparation de bibliothèques de RAD-seq.

La famille *CoNotoA*, correspondant aux éléments *Copia* de la famille des *GalEa*, a été identifiée chez un grand nombre de Nototheniidae. Cependant, ils ont été ici activement recherchés dans les génomes des Trematominae (genres *Trematomus*, *Patagonotothen*, *Lepidonotothen*) et du Pleuragramminae *P. antarctica*, mais sans succès. Les éléments *GalEa* sont connus pour avoir une distribution clairsemée et peuvent être perdus dans des clades entiers (Piednoël et al. 2013). Nos résultats semblent suggérer une perte de cette famille au cours de l'évolution des

Nototheniidae. Cela pourrait soutenir la position phylogénétique de *P. antarctica* en groupe frère de la sous-famille des Trematominae.

Les résultats obtenus par clonage ne sont pas exhaustifs. Ils correspondent aux familles d'éléments majoritairement représentées dans ces génomes. Afin d'identifier d'autres familles de *DIRS1*, *Gypsy* et *Copia*, nous avons amorcé des tests de recherche en ETs par séquençage NGS en shotgun de PCRs (Ion Torrent) à partir des amorces dégénérées de départ. Cependant, les ETs reconstitués à partir de l'assemblage des lectures de quelques centaines de paires de bases sont des éléments chimères et ne reflètent pas forcément la séquence dans son intégrité à cause de sa nature répétée. L'identification de séquences entières pourrait être établie en utilisant en complément un autre type de technologie comme le séquençage de type Nanopore (minION) permettant d'obtenir des lectures plus longues comprenant tout l'élément. Ce type de séquençage pose toutefois des problèmes persistants de taux d'erreur élevés et serait à compléter avec des séquençages plus fiables (Illumina).

### **IV.3.B : diversité des éléments transposables chez les « *Trematomus* »**

#### ***Résultats et discussion***

Afin d'avoir une idée moins restrictive du contenu en ETs dans les génomes de Nototheniidae, nous avons recherché des éléments transposables dans les lectures de ddRAD-seq obtenues en vue de clarifier la phylogénie des « *Trematomus* » (chapitre III). Les ETs ont été identifiés par le filtre « éléments transposables » permettant d'éliminer les locus « ETs » partagés. Pour cette recherche, afin de maximiser le nombre de locus « ETs » obtenus, nous avons diminué le paramètre *p* en choisissant non plus 16 ou 20, mais seulement 6 individus pour former un locus partagé. Nous avons ainsi identifié 181 locus « ETs » (soit un total de 0,8 % des locus totaux polymorphes ou non) avec le logiciel PyRAD (Eaton and Ree 2013; Eaton 2014). La distribution de ces locus dans ces génomes a été classée par individu et par super-famille (Tableau IV.6).

Le nombre de locus « ETs » identifiés dans cette analyse pour chaque super-famille d'éléments dépend de certains paramètres, comme la taille de l'élément (transposons et LINEs plus courts que les autres types de rétrotransposons), la fréquence de sites de restriction EcoR1 et MspI présents sur les séquences, ou encore le paramètre *p* choisi. Cependant, nous avons choisi un seuil de détection non stringent (*e*-value de  $10^{-10}$ ) afin de pouvoir identifier un maximum de locus « ETs », sans risquer pour autant d'identifier des faux positifs. Si la composition et le nombre de locus « ETs » retrouvés ne reflètent donc pas de manière tout à fait exhaustive le répertoire d'ETs existant réellement dans les génomes de Nototheniidae, cette distribution permet de donner un aperçu du paysage global en ETs (à la fois pour les éléments étudiés précédemment, mais aussi pour d'autres super-familles d'éléments de classes I et II).

| ETs                           | classe II (transposons à ADN) |             |     |          |          |          |    |           |    |        | classe I (rétrotransposons) |        |       |         |       |       |       |          |     |  | Total locus par individu |      |  |     |
|-------------------------------|-------------------------------|-------------|-----|----------|----------|----------|----|-----------|----|--------|-----------------------------|--------|-------|---------|-------|-------|-------|----------|-----|--|--------------------------|------|--|-----|
|                               | TIR                           |             |     |          |          | Maverick |    |           |    |        | LINES                       |        |       |         |       | LTR   |       |          |     |  |                          | DIRS |  | PLE |
|                               | En-spm                        | Tc1-Mariner | hAT | PiggyBac | polinton | RTE      | L1 | Rex-Babar | L2 | I-Nimb | Jockey                      | Proto2 | Gypsy | Bel-Pao | Copia | DIRS1 | Ngaro | Penelope |     |  |                          |      |  |     |
| <i>I. cyanobrancha</i>        | 2                             | 1           | 4   | 2        | 5        | 2        | 7  | 10        | 6  | 1      |                             | 1      | 13    | 4       |       | 1     | 1     | 1        | 61  |  |                          |      |  |     |
| <i>T. scotti</i>              | 3                             | 3           | 5   | 2        | 1        | 3        | 24 | 16        | 7  | 3      |                             | 1      | 24    | 7       | 2     | 3     | 1     | 1        | 106 |  |                          |      |  |     |
|                               | 3                             | 2           | 4   | 2        | 3        | 3        | 19 | 13        | 8  | 3      |                             | 3      | 22    | 7       | 1     | 2     |       | 1        | 96  |  |                          |      |  |     |
| <i>T. nicolai</i>             | 3                             | 3           | 7   | 1        | 6        | 4        | 19 | 15        | 7  | 2      |                             | 1      | 24    | 9       |       | 3     |       |          | 104 |  |                          |      |  |     |
|                               | 1                             | 3           | 4   | 1        | 6        | 3        | 13 | 13        | 6  | 3      | 1                           | 3      | 29    | 7       |       |       | 1     |          | 94  |  |                          |      |  |     |
| <i>T. bernacchii</i>          | 2                             | 2           | 7   | 1        | 3        | 4        | 25 | 16        | 12 | 1      |                             | 3      | 34    | 11      | 2     | 2     | 2     | 1        | 128 |  |                          |      |  |     |
|                               | 2                             | 2           | 4   | 2        | 4        | 1        | 18 | 11        | 9  | 1      | 1                           | 1      | 26    | 9       |       | 2     | 1     |          | 94  |  |                          |      |  |     |
| <i>T. lepidorhinus</i>        | 1                             | 1           | 3   |          | 6        |          | 12 | 5         | 2  | 2      |                             | 1      | 18    | 3       |       | 1     |       |          | 55  |  |                          |      |  |     |
|                               | 2                             | 2           | 5   | 2        | 4        | 1        | 20 | 15        | 10 | 3      |                             | 2      | 25    | 9       | 2     | 3     | 2     | 1        | 108 |  |                          |      |  |     |
| <i>T. borchgrevinkii</i>      | 1                             | 2           | 4   | 1        | 6        | 2        | 12 | 7         | 8  |        |                             | 1      | 14    | 6       |       | 1     |       |          | 65  |  |                          |      |  |     |
| <i>T. hansonii</i>            | 2                             | 1           | 5   | 2        | 5        | 4        | 23 | 12        | 12 | 2      | 1                           | 1      | 29    | 9       | 1     | 1     |       | 1        | 111 |  |                          |      |  |     |
|                               | 2                             | 1           | 4   | 1        | 5        | 3        | 22 | 13        | 9  | 2      | 1                           | 1      | 24    | 6       | 2     |       |       |          | 96  |  |                          |      |  |     |
| <i>T. newnesi</i>             | 3                             | 2           | 4   | 1        | 4        |          | 14 | 9         | 3  | 1      | 1                           | 2      | 14    | 7       |       |       | 1     |          | 66  |  |                          |      |  |     |
|                               | 1                             | 3           | 4   | 1        | 5        | 1        | 8  | 9         | 5  | 1      |                             | 2      | 18    | 5       |       | 2     | 1     |          | 66  |  |                          |      |  |     |
| <i>T. loenbergii</i>          | 2                             | 3           | 6   |          | 4        | 1        | 14 | 9         | 5  |        | 1                           | 1      | 19    | 3       |       |       |       |          | 68  |  |                          |      |  |     |
|                               | 1                             | 3           | 4   |          | 5        | 1        | 12 | 5         | 7  |        |                             |        | 17    | 2       |       | 1     |       |          | 58  |  |                          |      |  |     |
| <i>T. pennellii</i>           | 2                             | 1           | 2   |          | 5        | 1        | 10 | 12        | 4  |        |                             |        | 10    | 4       | 1     |       |       |          | 52  |  |                          |      |  |     |
|                               | 4                             | 4           | 8   | 2        | 6        | 4        | 29 | 25        | 14 | 2      |                             | 3      | 45    | 14      | 2     | 4     | 1     |          | 167 |  |                          |      |  |     |
| <i>T. eulepidotus</i>         | 3                             | 2           | 5   |          | 5        | 1        | 15 | 14        | 3  | 1      | 1                           | 1      | 20    | 7       |       | 1     |       |          | 79  |  |                          |      |  |     |
|                               | 3                             | 1           | 7   |          | 5        | 2        | 14 | 12        | 7  | 1      | 1                           | 1      | 23    | 3       |       |       | 1     |          | 81  |  |                          |      |  |     |
| <i>T. tokarevi</i>            | 3                             |             | 3   |          | 5        | 1        | 10 | 9         | 6  | 1      |                             |        | 14    | 4       |       |       |       |          | 56  |  |                          |      |  |     |
| <i>L. squamifrons</i>         | 3                             |             | 4   | 2        | 3        | 2        | 15 | 16        | 6  | 1      | 1                           | 1      | 23    | 8       | 2     | 2     | 1     |          | 90  |  |                          |      |  |     |
| <i>P. antarctica</i>          |                               |             | 4   |          |          |          | 7  | 4         | 3  |        |                             | 2      | 6     | 3       |       |       |       |          | 30  |  |                          |      |  |     |
| Total locus par super-famille | 3                             | 5           | 8   | 2        | 6        | 4        | 33 | 28        | 17 | 3      | 1                           | 3      | 45    | 14      | 2     | 4     | 2     | 1        | 181 |  |                          |      |  |     |

**Tableau IV.6:** distribution des locus « ETs » de dd-RADseq partagés par un minimum de six individus et présents dans les génomes de vingt-et-un individus du groupe « *Trematomus* » et deux individus de *Nototheniidae* (*L. squamifrons* et *P. antarctica*). Les locus « ETs » ont été identifiés par BLASTx avec une base de données ETs générale, enrichie de nos consensus (e-value 10<sup>-10</sup>). Ils sont regroupés par super-familles. Le nombre indiqué dans chaque case correspond au nombre de locus identifiés chez cet individu pour une super-famille d'élément donnée. Les éléments MITEs sont probablement inclus dans les locus hAT et Tc1-Mariner, de même pour les SINES dans les locus LINES. Echelle couleur : jaune ; de 1 à 10 locus, vert ; de 11 à 20 locus, et rouge, plus de 21 locus.

Comme attendu, le nombre total de locus « ETs » pour un individu donné est proportionnel à la couverture de chaque banque. En effet, plus cette couverture est élevée, et plus la probabilité d'identifier des séquences de cette super-famille sera élevée et inversement. C'est par exemple le cas pour *P. antarctica*, dont la couverture est une des plus basses obtenues (627 297 lectures filtrées) et dont le nombre total de locus « ETs » est le plus faible (16,5 % des 181 locus ETs identifiés), ou encore pour l'individu *T. pennellii* CE3317, dont la couverture est une des plus hautes observées (996 239 lectures filtrées) et dont le nombre de locus « ETs » est le plus élevé (92,3 % des 181 locus ETs identifiés) (Tableau IV.6).

Un seuil d'un minimum de 85 % d'identité entre lectures a été requis pour former un locus au moyen de PyRAD. Nous savons par ailleurs que chez les « *Trematomus* », et même plus largement au sein de Nototheniidae, les ETs appartenant à une même famille partagent une très forte identité nucléotidique (plus de 90 % d'identité pour la plupart des familles constituant les trois éléments étudiés). Le nombre de locus « ETs » assignés à chaque super-famille devrait donc représenter le nombre de familles d'éléments existant dans ces génomes (et non le nombre de copies). Etant donné la différence de longueur entre les lectures Illumina analysées (140 pb) et la taille des ETs présents dans la base de données constituée (de un à plusieurs kb), une même famille d'élément peut être représentée par plusieurs locus (répartis sur toute la longueur de la séquence). Nous avons en effet analysés nos séquences nucléotidiques consensus d'ETs issus des « *Trematomus* » et identifié plusieurs sites de restriction EcoR1 (consensus d'éléments *Gypsy* et *Copia*) et MspI (consensus d'éléments *DIRS1*, *Gypsy* et *Copia*) ayant pu aboutir dans le jeu de données à plusieurs fragments de double digestion de ddRAD-seq pour une même famille. Par exemple, un total de quarante-cinq locus « *Gypsy* » ont été distribués dans vingt-cinq types de *Gypsy* différents : *GyNotoA* (3), *GyNotoB* (3), *GyNotoE* (3) et *GyNotoF* (1), *Gypsy-Danio rerio* (22 dans 14 groupes), *Gypsy-Gasterosteus aculeatus* (5), *Ronin-Takifugu rubripes* (4 dans 2 groupes), *Gypsy-Cigr Tetraodon nigroviridis* (2), *Gypsy-Alligator Mississippiensis* (1), *Gypsy-Strongylocentrotus Purpuratus* (1), et *Gypsy-Nematostella Vectensis* (1). Pour les *DIRS1*, un maximum de quatre locus a été identifié dans le génome *T. pennellii* (CE3317). Le même raisonnement s'applique ici également car, selon les résultats de BLAST, ces quatre locus ne correspondraient pas aux quatre familles identifiées dans les génomes des Nototheniidae (*YNotoJ*, *V*, *R* et *B*), mais seulement à deux familles différentes (*YNotoJ* (2), et *DIRS1\_TakifuguRubripes\_pol* (2)).

Nous retrouvons une cohérence entre cette étude et celle réalisée précédemment à l'échelle des « *Trematomus* ». De plus, ces résultats vont dans le même sens que ceux publiés par Capriglione *et al.* (2002), Ozouf-Costaz *et al.* (2004) et Detrich *et al.* (2010). Ainsi, le nombre total de locus « ETs » par individu semble plus important pour les éléments de classe I que de

classe II, toutes super-familles confondues (1609 vs 326). Ce résultat est toutefois à mesurer étant donné la différence de taille observée entre les différentes catégories d'ETs. Parmi les rétrotransposons, on note une représentation plus importante des LINEs de type *L1* et *Rex-Babar*, et rétrotransposons à LTR de type *Gypsy*. Pour ces trois dernières super-familles, la majorité des individus présentent en effet plus de dix à vingt locus « ETs », contrairement aux autres super-familles (Tableau IV.6). Il a été montré que les éléments *Gypsy* ont tendance à se diversifier dans les génomes, selon une stratégie évolutive appelée « théorie de la reine rouge » (Van Valen 1973). Au contraire les éléments *Copia* semblent présenter une autre dynamique. Ils sont en effet retrouvés plus rares et moins diversifiés dans les génomes des Nototheniidae examinés (Tableau IV.5 et IV.6). Ceci est illustré par une autre stratégie appelée « domino days spreading » (Piednoël et al. 2013).

Certains éléments, comme les rétrotransposons de type *Bel/Pao*, qui n'avaient jamais été identifiés par la méthode de clonage, ont pu être mis en évidence par cette recherche. Au contraire, d'autres éléments identifiés dans ces génomes par Detrich *et al.* (2010), comme les transposons à ADN de type *Kolobok* et *Harbinger*, n'ont pas été retrouvés ici. Cependant, une absence d'un locus « ET » dans une super-famille donnée ne signifie pas que ces éléments sont absents des génomes de Nototheniidae. Par exemple, des éléments SINEs pourraient avoir été confondus avec des éléments LINEs (partie commune en 5'UTR et en 3'UTR appelée « target site duplication » ou TSD de type L1), et des MITEs avec des ETs de classe II (partie commune avec les éléments *PIF/Harbinger*, *Piggybac* ou encore *Tc1-Mariner*).





## **V) Discussion générale : histoire évolutive des remaniements chromosomiques des « *Trematomus* »**

### **V.1 : les fusions chromosomiques, remaniements majoritaires survenus lors de la diversification des « *Trematomus* »**

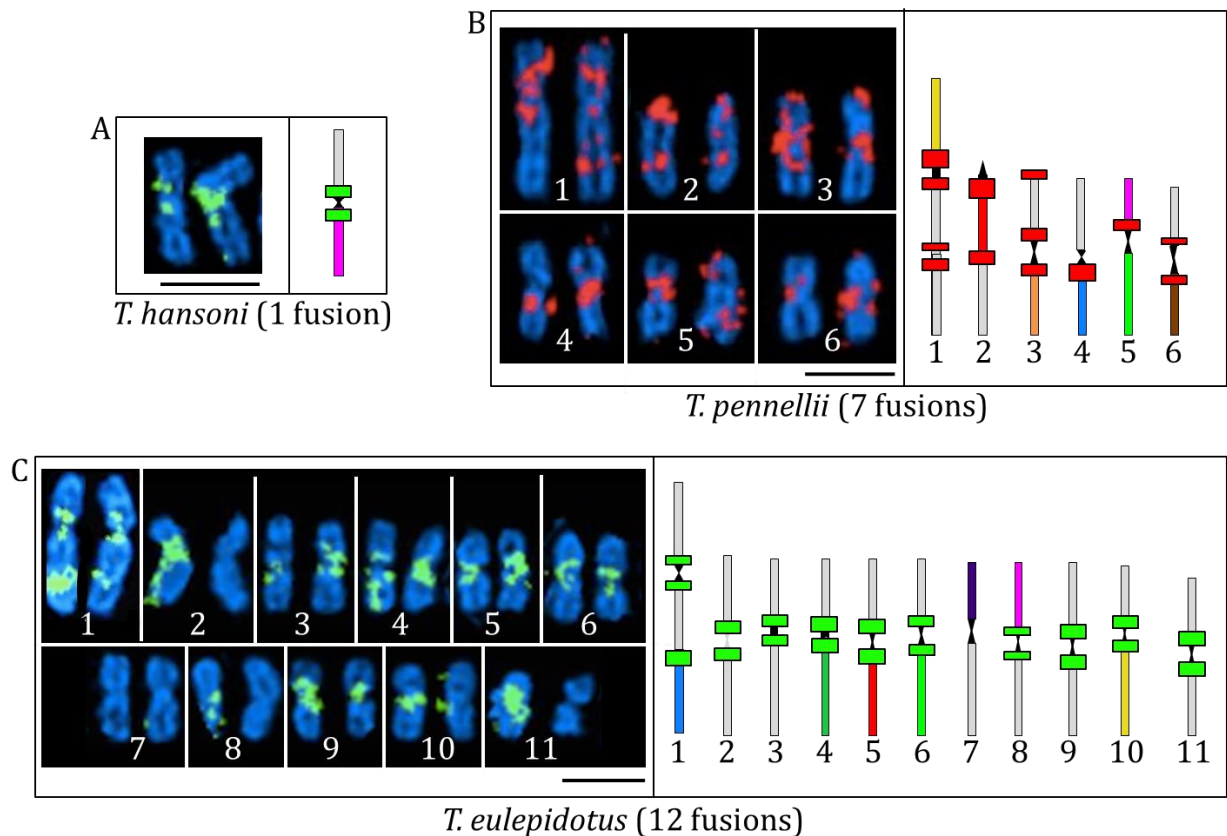
Parmi les remaniements chromosomiques, les fusions centriques ou Robertsoniennes (Robertson, 1916) font partie des événements les plus fréquemment observés (Bush 1975; Dancis and Holmquist 1979; Baker and Bickham 1986; Garagna et al. 2001), notamment chez les mammifères. Il s'agit de la fusion de deux chromosomes acrocentriques (télocentriques ou subtélocentriques, Figure II.2) au niveau de leur centromère pour former un chromosome à deux bras. Alors qu'elles avaient été évoquées dans la littérature (Morescalchi et al. 1992; Pisano and Ozouf-Costaz 2000; Pisano et al. 2003), nous avons montré au cours de cette étude que ces fusions centriques constituaient les remaniements chromosomiques majoritaires chez les « *Trematomus* » (caryotype ancestral inféré,  $2n = 48$  chromosomes acrocentriques, caryotypes actuels observés  $2n =$  ou  $< 48$ , à l'exception de *T. nicolai*,  $2n > 48$ ). Des fusions en tandem, entre le centromère et le télomère de deux chromosomes acrocentriques différents (*T. pennellii*, *T. eulepidotus*), ainsi que des inversions péricentriques et des fissions (*T. nicolai*) ont été également observées (Figure II.2, II.14). Elles semblent minoritaires dans ce groupe en comparaison aux fusions Robertsoniennes. Néanmoins, certains caryotypes et remaniements potentiels restent à découvrir pour certaines espèces de « *Trematomus* », comme *T. tokarevi*, *P. brachysoma*, *C. peninsulae* ou *C. amphitrete* (Ghigliotti et al. 2015).

Lors de la fusion centrique, des inactivations/raccourcissements/suppressions d'une partie des séquences centromériques et télomériques sont requises pour minimiser les problèmes de ségrégation mitotique ou méiotique liés au dicentrisme des chromosomes ou chromatides fusionnés (Schubert et al. 1992; Slijepcevic 1998; Garagna et al. 2001). Ces fusions peuvent survenir spontanément dans les lignées somatiques ou germinales par recombinaisons ectopiques (recombinaisons entre séquences homologues à des sites non homologues) (Garagna et al. 2001). Dans le cas d'une apparition dans les cellules germinales et avant leur fixation, l'essentiel du matériel génomique des chromosomes ayant fusionné est conservé, assurant une réversibilité de l'événement (fission chromosomique) (Schubert et al. 1992). L'association entre deux chromosomes acrocentriques différents s'opère entre des régions génomiques péricentromériques riches en ADN satellite (Garagna et al. 2001). Ces régions

hétérochromatiques sont également connues pour contenir des éléments transposables, accumulés à des points chauds d'insertion (Pimpinelli et al. 1995; Biémont and Vieira 2005; Biémont and Vieira 2006; Mazzuchelli and Martins 2009; Schneider et al. 2013).

Les éléments transposables ont été découverts par Barbara McClintock au milieu des années 1940 (McClintock 1948; McClintock 1950). Ce sera la première à étudier l'impact de l'insertion d'ETs dans les régions géniques chez le maïs, évoquant également leur implication potentielle dans des remaniements structuraux de type délétions, translocations et inversions (McClintock 1984). Par la suite, les éléments *BEL*, *roo*, *Doc*, *Hobo*, *P*, *FB* et *Galileo* ont été largement étudiés chez la drosophile (Lim and Simmons 1994; Casals 2003). Chez *D. melanogaster*, une corrélation a été établie entre des insertions de l'élément *Hobo* au cours des générations successives et de nouvelles inversions chromosomiques (dont 4 péricentriques) détectées sur les chromosomes polytènes (Aulard et al. 2004; Ladevèze et al. 2012). De même, le transposon *Galileo* est impliqué dans deux inversions naturelles détectées sur les chromosomes de *D. buzzatti* (Casals 2003). L'impact des ETs dans les restructurations génomiques a été depuis étudié chez de nombreux organismes présents dans différentes lignées comme les plantes (Bennetzen 2000; Parisod et al. 2010) ou encore les métazoaires (Lim and Simmons 1994; Miller and Capy 2004; Bourque 2009; Konkel and Batzer 2010). Leur rôle de médiateur des remaniements chromosomiques en lien avec la spéciation reste cependant le sujet de nombreux débats (Lim and Simmons 1994; Lönnig and Saedler 2002; Raskina et al. 2008; Bourque 2009; Chénais et al. 2012).

Suite à l'étude de trois types de rétrotransposons dans les génomes de « *Trematomus* » et autres Nototheniidae, nous avons conclu à un rôle possible de ces ETs en tant que facilitateurs des réarrangements survenus au cours des diversifications. Contrairement au cas des études sur la drosophile, nous n'avons pas la possibilité de suivre *in vivo* et en temps réel le mécanisme d'insertion des ETs dans ces génomes. Cependant nous avons observé la conséquence d'une mobilisation massive des éléments à tyrosine recombinase *DIRS1* dans leur distribution chromosomique et leur nombre chez les différentes espèces étudiées.

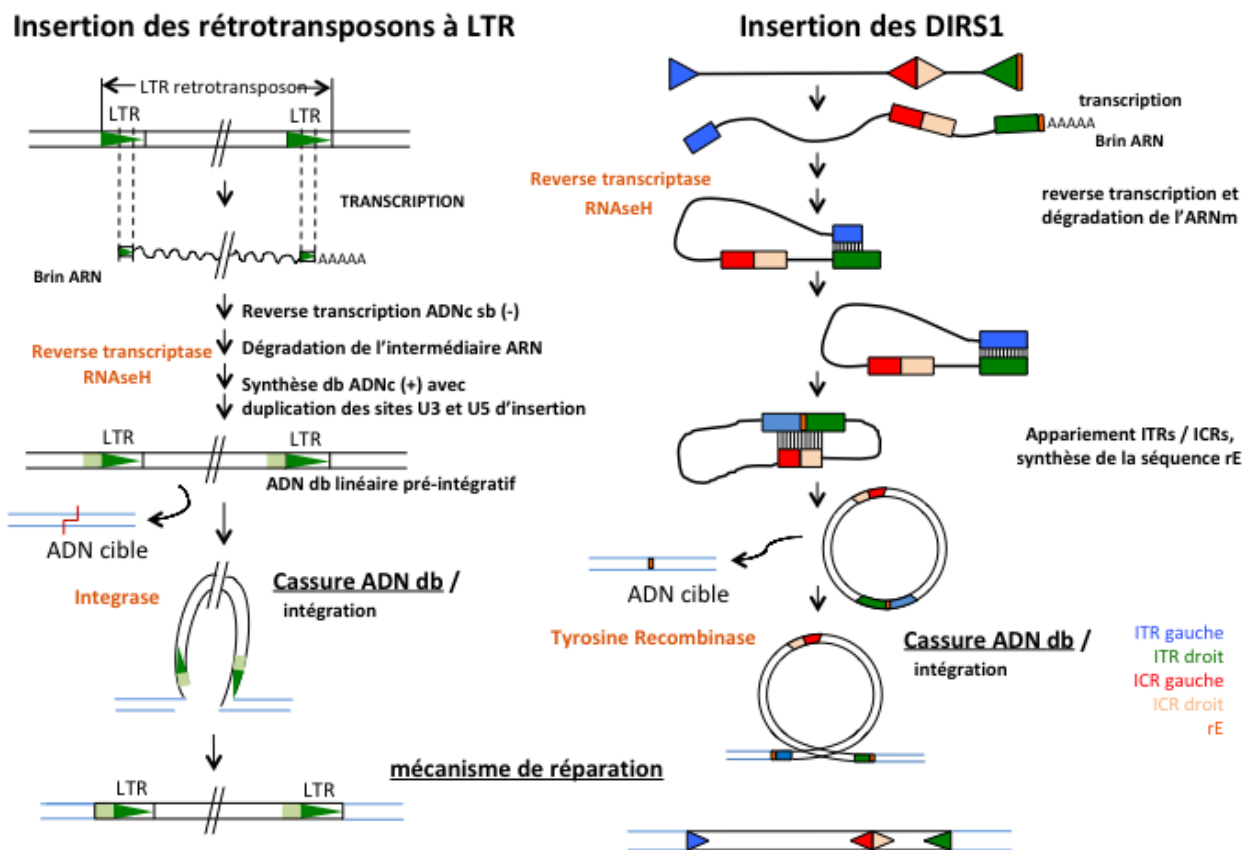


**Figure V.1 :** localisation des éléments *DIRS1* sur les chromosomes fusionnés de A) *T. hansonii*, B) *T. pennellii* et C) *T. eulepidotus* et leur correspondance sur les unités structurales reconstruites (lot haploïde). Les points chauds d'insertion des *DIRS1* observés par *FISH* sont représentés sur les différentes unités structurales par les plages colorées en vert clair (*T. hansonii*, *T. eulepidotus*) ou en rouge (*T. pennellii*). Ces métaphases ont été choisies pour représenter les patrons de localisation en points chauds d'insertion des éléments *DIRS1* sur les chromosomes. L'absence de *DIRS1* sur les paires 2, 7 et 11 chez *T. eulepidotus* est un artefact de cette *FISH*, ils sont présents sur ces paires pour d'autres métaphases. Les unités structurales mises en évidence par les BACs ou la peinture chromosomiques sont colorées selon leurs homologies interspécifiques (Figures II.13 et II.16). Les unités structurales communes impliquées dans des fusions indépendantes entre *T. eulepidotus* et *T. pennellii* sont : l'unité bleu (C4-F15) présente sur la paire numérotée 1 chez *T. eulepidotus* et 4 chez *T. pennellii*, l'unité rouge (25G6-G6-28D21) présente sur la paire numérotée 5 chez *T. eulepidotus* et 2 chez *T. pennellii*, l'unité verte claire (F7) présente sur la paire numérotée 6 chez *T. eulepidotus* et 5 chez *T. pennellii*, l'unité rose clair (M11-85H3) présente sur la paire numérotée 8 chez *T. eulepidotus* et 5 chez *T. pennellii*, et enfin l'unité jaune (P3) présente sur la paire numérotée 10 chez *T. eulepidotus* et 1 chez *T. pennellii*. Barre d'échelle : 10  $\mu$ m.

L'accumulation de ces multiples copies identifiées parmi les différentes familles (Figures IV.6, additionnelles IV.7 et IV.8) en points chauds d'insertions (patron d'insertion de type 1) dans les régions péricentromériques et parfois télomériques des chromosomes semble clairement correspondre à des zones privilégiées de cassures. Ces insertions localisées et répétées augmentent ainsi la probabilité de recombinaisons ectopiques lors des réparations. Les recombinaisons entre régions centromériques/péricentromériques et/ou télomériques de différentes unités structurales ont pu faciliter les fusions centriques ou en tandem, donnant lieu aux chromosomes fusionnés observés aujourd'hui dans le groupe des « *Trematomus* » (Figure V.1).

Les éléments *DIRS1* ont été localisés sur les chromosomes d'espèces proches des Nototheniidae (*E. maclovinus*, *B. diacanthus*, *P. fluviatilis*), mais n'ayant pas évolué dans cet environnement extrême. Ils ont été retrouvés dispersés le long des chromosomes étudiés, sans montrer d'accumulation préférentielle dans des régions particulières (Figure IV.7). Cette distribution rappelle plutôt le patron d'insertion de type 2 précédemment défini. Elle demanderait à être vérifiée par l'hybridation d'un mélange de sondes *DIRS1* homologues représentant les quatre familles identifiées jusqu'à présent dans les génomes de Nototheniidae.

D'autres propriétés des éléments *DIRS1* pourraient expliquer leur rôle possible dans les fusions chromosomiques dans le groupe des « *Trematomus* », et plus largement chez les Nototheniidae. Leur intégration dans les régions cibles est assurée non pas par une intégrase, comme chez les autres rétrotransposons mais par une tyrosine recombinase (YR) (Goodwin and Poulter 2001). Leur structure particulière (identification de la tyrosine recombinase et des régions répétées inversées et complémentaires : ITRs/ICRs, Figure V.2) a conduit au développement d'un modèle déductif décrivant leur mode d'intégration. Il s'appuie sur les connaissances des mécanismes de transposition de transposons à ADN (décrit chez les bactéries avec les *IS91* ou chez les eucaryotes avec les *Hélitrons* à YR par exemple) (Mendiola et al. 1994; Kapitonov and Jurka 2008). Elle s'effectue en « cercle roulant » (« rolling-circle ») au moyen d'un ADN double brin pré-intégratif circulaire sans duplication des sites d'intégration (Figure V.2) (Capello et al. 1985; Goodwin and Poulter 2001).



**Figure V.2 :** étapes décomposant la transposition et d'intégration (A) des rétrotransposons à LTR et (B) des rétrotransposons à tyrosine recombinase dans les régions cibles. D'après les mécanismes décrits par Eickbush et Malik (2002) et Wilhelm et Wilhelm (2001) pour les rétrotransposons à LTR, et Capello *et al.* (1985) et Goodwin et Poulter (2001) pour les rétrotransposons à tyrosine recombinase de type DIRS. Db : double brin, sb : simple brin, ITRs : inverted terminal repeats, ICRs : internal complementary regions.

D'autre part, tout comme l'élément *P*, les éléments *DIRS1* sont connus pour leur tendance à s'insérer dans des copies préexistantes (« homing ») (Cappello et al. 1984; Taillebourg and Dura 1999; Eickbush and Jamburuthugoda 2008). Enfin, le promoteur de l'élément *DIRS1* de l'amibe *Dictyostelium discoideum*, chez qui il a été découvert, présente des séquences très proches de motifs retrouvés dans des promoteurs « heat shock protein » (hsp) (Cappello et al. 1984). Le séquençage et la caractérisation des éléments *DIRS1* entiers chez les Nototheniidae seraient nécessaires pour confirmer notre hypothèse quant à la possible présence de ces motifs hsp dans leur promoteur. La présence de tels motifs renforceraient l'idée que ces éléments ont pu être fortement mobilisés dans le contexte de changements environnementaux caractérisant les vagues de diversifications des Nototheniidae dont font partie les « *Trematomus* ».

## **V.2 : spéciations chromosomiques multiples chez les « *Trematomus* »**

Dans le groupe « *Trematomus* », on trouve trois espèces (*T. hansonii*, *T. bernacchii*, *T. newnesi*) dont les caryotypes présentent une seule fusion chromosomique, et trois autres espèces (*T. pennellii*, *T. eulepidotus*, *T. loennbergii*) dont les caryotypes sont constitués de multiples chromosomes simplement ou doublement fusionnés. Ces chromosomes fusionnés peuvent être métacentriques ou submétacentriques (Figure II.5, V.1). Dans le premier cas, ce sont des associations de deux unités structurales (de taille équivalente). Dans le second cas, il s'agit d'associations entre deux unités (de taille différente) ou de trois unités structurales. Les unités structurales quant à elles semblent particulièrement conservées d'une espèce à une autre. Par analogie avec un jeu de construction, on pourrait comparer les unités structurales à des briques servant d'éléments de base. L'association de ces briques par deux ou par trois est différente entre les espèces. Les unités structurales ont dans ce sens été définies comme des unités de réarrangement, se combinant différemment pour former les caryomorphes actuellement observés.

Nous avons montré qu'au moins cinq unités structurales différentes ont été retrouvées impliquées dans des fusions indépendantes entre *T. eulepidotus* ( $2n = 24$ ) et *T. pennellii* ( $2n = 32$ ) (Figures II.13, II.16, V.1). On peut citer (1) l'unité structurale impliquée d'une part dans une fusion en tandem pour former la plus grande paire submétacentrique de *T. eulepidotus*, et d'autre part dans une fusion centrique pour former la deuxième paire métacentrique de *T. pennellii*; (2) l'unité structurale impliquée dans une fusion centrique pour former la quatrième paire métacentrique de *T. eulepidotus* est aussi impliquée dans une fusion en tandem pour former la plus grande paire acrocentrique de *T. pennellii*; et (3) l'unité structurale impliquée dans une fusion centrique pour former la cinquième paire métacentrique de

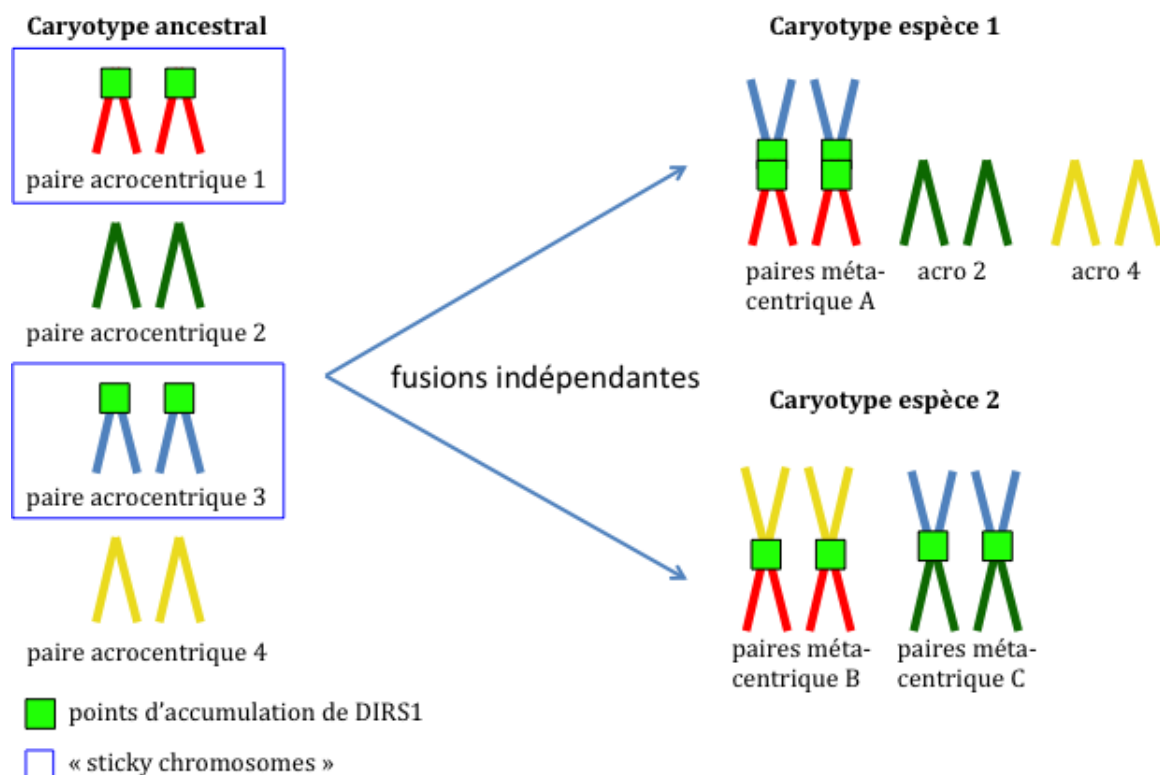
*T. eulepidotus* et également impliquée dans une fusion centrique pour former la troisième paire métacentrique de *T. pennellii*. Dans ce troisième et dernier cas, l'unité structurale ayant fusionné avec cette première unité chez *T. pennellii* est retrouvée sur une autre paire chromosomique chez *T. eulepidotus* (Figure II.13, V.1). Les fusions mises en évidence au cours de cette étude semblent donc multiples, c'est-à-dire apparues plusieurs fois séparément au cours de l'évolution des Nototheniidae. Cependant, il subsiste des fusions ou fissions (voire des inversions ou translocations) où nous n'avons pas encore identifié les unités structurales qui interviennent et qui pourraient être partagées entre espèces. Dans la littérature les deux hypothèses (remaniements multiples et indépendants ou partagés) ont été proposées chez les Nototheniidae. Des acquisitions indépendantes ont ainsi été évoquées chez les « *Trematomus* » pour la mise en place du chromosome sexuel hétéromorphe Y (Morescalchi et al. 1992; Ghigliotti et al. 2013). Les fusions donnant lieu à ce grand chromosome sexuel ne seraient pas partagées entre *T. nicolai* et *T. newnesi*, *T. lepidorhinus*, *T. hansonii*, *T. borchgrevinkii*. Au contraire, les réductions chromosomiques par fusions massives au sein de la sous-famille des Nototheniinae (groupe frère des Trematomiinae avec le reste du clade « antarctique », Figure I.1) auraient été acquises pour la plupart (onze sur un maximum de treize) chez l'ancêtre commun des *Notothenia* et des *Paranotothenia* et partagées par la descendance (*N. coriiceps*,  $2n = 22$  ; *N. rossii*,  $2n = 24$  ; *P. magellanica* et *N. angustata*,  $2n = 26$ ) (Amores et al. 2017). Ces résultats restent cependant des hypothèses et seraient à confirmer par marquage des unités structurales impliquées dans ces différentes fusions.

L'indépendance des événements de remaniements chromosomiques que nous avons observée chez les « *Trematomus* » est cohérente avec le modèle proposé dans lequel certains éléments transposables peuvent intervenir en facilitant les fusions. En effet, ces cassures double brins sont théoriquement aléatoires et peuvent s'opérer sur n'importe quelle unité structurale. Cependant, l'accumulation préférentielle de ces éléments dans des paires riches en hétérochromatine ainsi que dans des copies préexistantes (comme pour les éléments DIRS) pourrait conférer à certaines unités structurales une propension plus élevée à fusionner que d'autres unités (Figure V.3). C'est dans ce contexte qu'Amores *et al.* (2014) explique des fusions indépendantes mettant en jeu un chromosome commun (qualifié de « sticky ») chez deux espèces de téléostéens phylogénétiquement apparentés (*T. nigroviridis* et *G. aculeatus*).

Pour les remaniements chromosomiques indépendants mis en évidence entre *T. eulepidotus* et *T. pennellii*, il est à noter que des accumulations de *DIRS1* dans les régions péri-centromériques ont été localisées dans toutes les unités structurales impliquées dans les fusions (Figure V.1). D'autre part, la fusion chromosomique unique observée chez *T. hansonii* ( $2n = 46$ ) et chez *T. newnesi* ( $2n = 45/46$ ) met en jeu une même unité structurale (marquée par les BACs M11 et



85H3), héritée de l'ancêtre commun des « *Trematomus* » (Figure II.13 et II.16). Par chance, ces plus grandes paires fusionnées sont plutôt faciles à reconnaître chez ces deux espèces sur la base de la taille et de la morphologie. Chez *T. hansonii*, nous avons observé que cette paire est en effet riche en éléments de type *Gypsy* et *Copia* (Figure IV.6), et présente surtout des points chauds d'insertion d'éléments *DIRS1* (Figures IV.6 et V.1). Il conviendrait de localiser les ETs étudiés sur les chromosomes de *T. newnesi* pour observer la présence ou non d'un même patron d'insertion de ces éléments. En repositionnant ces fusions sur notre phylogénie des « *Trematomus* », il semble que ces deux fusions sont elles aussi indépendantes car *T. hansonii* est groupe frère de *T. bernacchii* et *T. newnesi*, groupe frère de *T. nicolai* (Figures III.6, V.6, annexe 3). Un marquage de la seconde unité structurale de chacune de ces deux paires chromosomiques métacentriques serait requis afin de confirmer ou d'invalider cette absence de partage. Selon notre hypothèse, ces deuxièmes unités structurales ne seraient pas homologues entre espèces (Figure V.3).



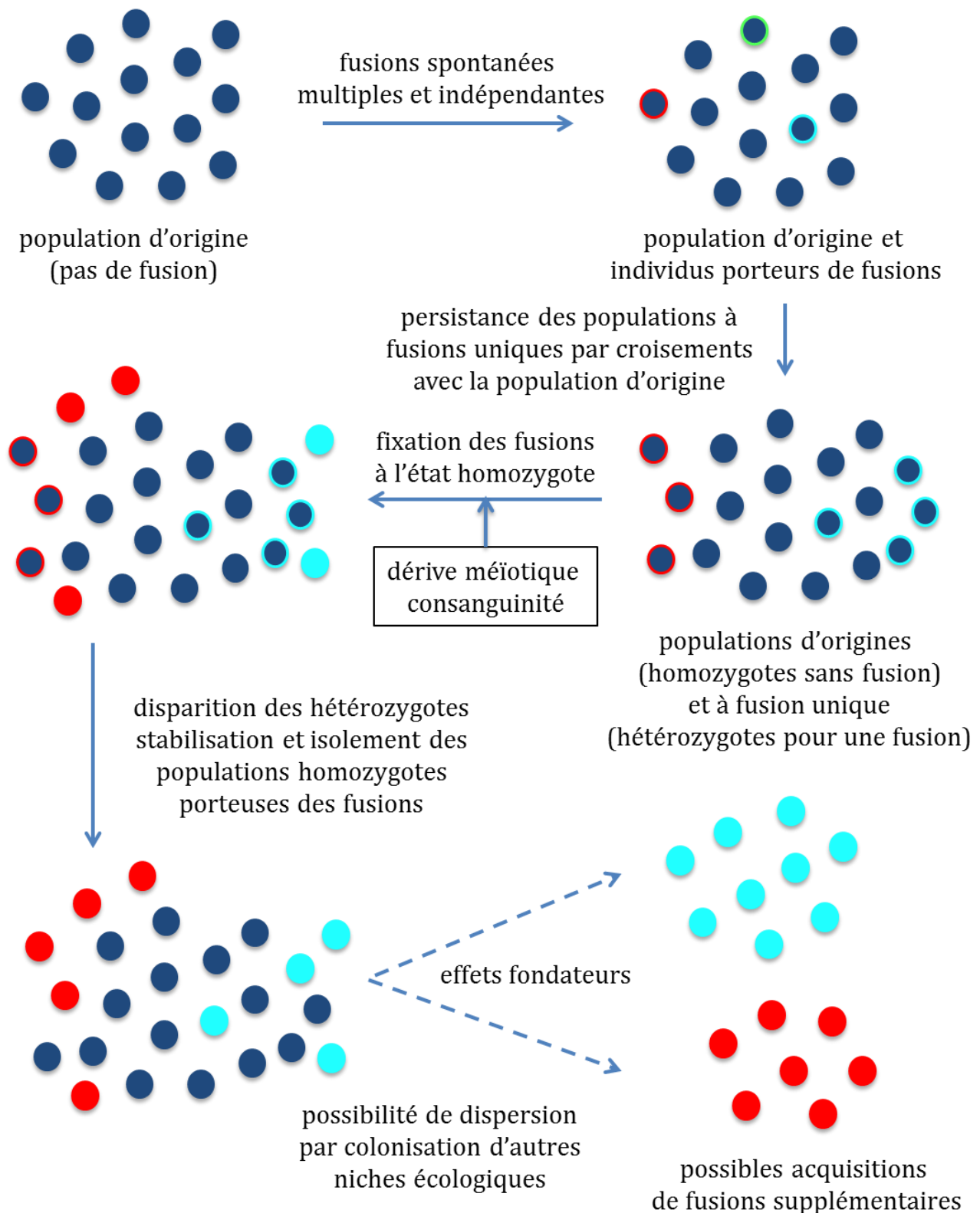
**Figure V.3 :** scénario de fusions chromosomiques indépendantes mettant en jeu une même unité structurale chez les « *Trematomus* ». Exemples entre *T. eulepidotus* et *T. pennellii* (Figure V.1), ainsi qu'entre *T. hansonii* et *T. newnesi* dans le cas de leur plus grande paire métacentrique, sous l'hypothèse que la deuxième unité structurale n'est pas homologue entre les deux paires.

Deux des trois radiations très récentes observées au sein de la famille des Nototheniidae présentent des caryotypes très peu variables :  $2n = 46$  pour la sous-famille des Artedidraconinae et  $2n = 48$  ou  $47/48$  en cas d'hétéromorphisme lié au sexe dans la sous-famille des Channichthyinae (Ghigliotti et al. 2013; Ghigliotti et al. 2015). Au contraire, la diversification rapide et récente des « *Trematomus* » s'est accompagnée d'une diversification chromosomique, la plus importante observée à l'échelle de la famille (on parle dans ce cas de spéciation chromosomique). Cependant, nous ne savons pas si ces fusions ou fissions chromosomiques multiples sont apparues spontanément avant ou après la divergence des espèces. En d'autres termes ont-elles été la cause de ces spéciations ou bien sont-elles intervenues comme éléments renforçant l'isolement reproducteur ?

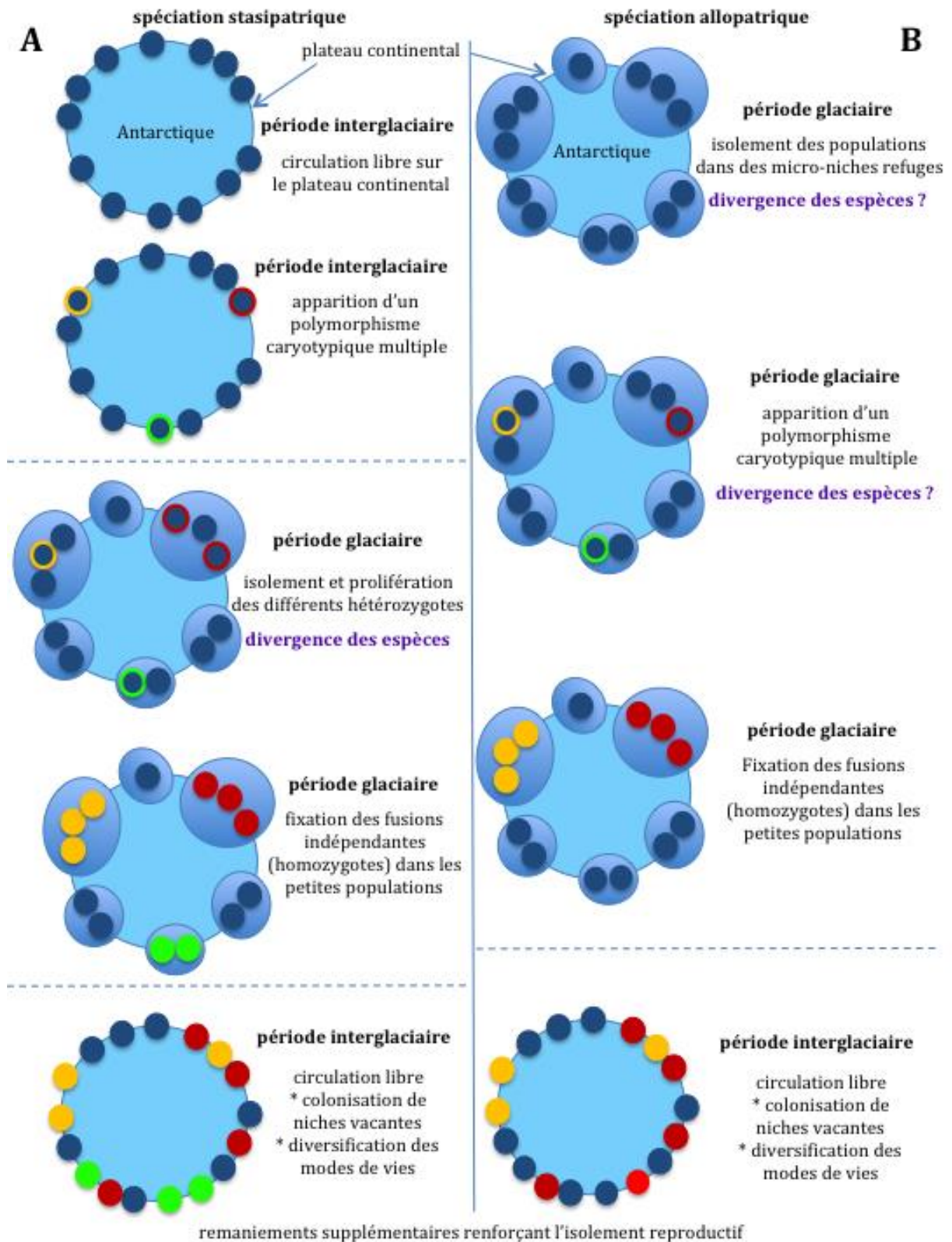
### **V.3 : vers une reconstruction des scénarios évolutifs des diversifications chromosomiques**

Si leur rôle dans la spéciation a longtemps été débattu, les réarrangements chromosomiques sont des événements importants au cours de l'évolution des génomes et peuvent accompagner la diversification des espèces (White 1970; Bush 1975; White 1978; White 1980; Baker and Bickham 1986; Rieseberg 2001). Plus qu'une simple mutation affectant une région génique donnée, ces réarrangements impliquent en effet des réorganisations génomiques majeures pouvant être source de nouveauté et d'innovation génétique (Bush 1975; Shaw et al. 1986). C'est dans cette perspective d'étude que des cas de polymorphismes caryotypiques ont été décrits dans la littérature, par exemple au sein de plusieurs populations ou espèces de plantes (Jones 1998), de mammifères (rongeurs (Auffray 1993; Britton-Davidian et al. 2000; Dobigny et al. 2003; Dobigny et al. 2005; Pialek et al. 2005); primates (Dutrillaux 1979)), d'insectes (orthoptères (White 1970; White 1978; Shaw et al. 1986); lépidoptères (Pringle et al. 2007)), ou encore de téléostéens (Characidae (Pansonato-Alves et al. 2011); Gastéroteidae (Kitano et al. 2009) et Gobiidae (Giles et al. 1985)). Lorsqu'elles sont transmises et fixées au sein des populations à l'état homozygote, les restructurations chromosomiques apparaissent en effet comme une barrière reproductrice majeure favorisant leur diversification (Baker and Bickham 1986; Barton and Rouhani 1991; Garagna et al. 2001). Cette barrière à l'échange de gènes entre individus s'effectue au moyen de la réduction de la valeur sélective des hybrides (hétérozygotes pour le réarrangement) ou encore le blocage de la recombinaison au niveau des sites fusionnés, où peuvent notamment se trouver des gènes dits de spéciation (Rieseberg 2001; Noor and Feder 2006).

Au moins sept modèles de spéciations chromosomiques ont été décrits sur la base de calculs théoriques ou d'observations expérimentales pour expliquer la fixation de ces événements dans les populations (White 1978; Templeton 1981; Baker and Bickham 1986). Un des modèles semble particulièrement adapté au contexte de diversification des « *Trematomus* ». Il s'agit de la spéciation stasipatrique proposée par White en 1969 à partir d'une étude sur des sauterelles australiennes, et repris par Auffray en 1993 avec les souris de la plaine de Pô en Italie (Figure V.4). Dans ce modèle, différentes populations présentent un caryotype réduit (fusions chromosomiques) par rapport à la population d'origine (pas de fusion) et se distinguent par la nature de leurs chromosomes fusionnés (indépendance des fusions). Les hybrides obtenus par croisement des populations à chromosomes fusionnés et non fusionnés (alors hétérozygotes pour les fusions) sont en général stériles par suite de difficultés d'appariements lors de méiose. Seules échappent à cette règle les populations ne différant que par une seule fusion (Auffray 2001). Aucun avantage sélectif des différentes fusions chromosomiques n'est mis en évidence dans ces populations. Par contre, il semble bien dans le cas des souris comme des sauterelles (White 1978; White 1980; Auffray 1993; Britton-Davidian et al. 2000; Auffray 2001), que ces fusions touchent essentiellement de petites populations isolées, pouvant être colonisées que par des groupes réduits d'immigrants. Sous l'effet de la dérive méiotique et de la consanguinité, le chromosome métacentrique passe à l'état homozygote (croisement des hétérozygotes entre eux). Les hybrides hétérozygotes étant moins fertiles du fait des problèmes d'appariements des chromosomes/chromatides homologues à la méiose, ils sont contre-sélectionnés (sous-dominance). La fixation de la fusion est facilitée d'une part par les effets fondateurs (non représentativité du premier groupe à l'origine d'une nouvelle population) et d'autre part par la dérive (plus de chance de fixer et plus rapidement). Plus l'effectif de la population est petit, plus la fixation est rapide. Les populations homozygotes pour les chromosomes fusionnés peuvent occuper l'aire de répartition de la population originelle ou bien peut coloniser une nouvelle niche écologique disponible dans l'environnement (dispersion). Ce processus est encore accentué par la fixation de nouvelles fusions après la divergence (White 1970; White 1978; White 1980; Auffray 1993; Britton-Davidian et al. 2000; Auffray 2001).



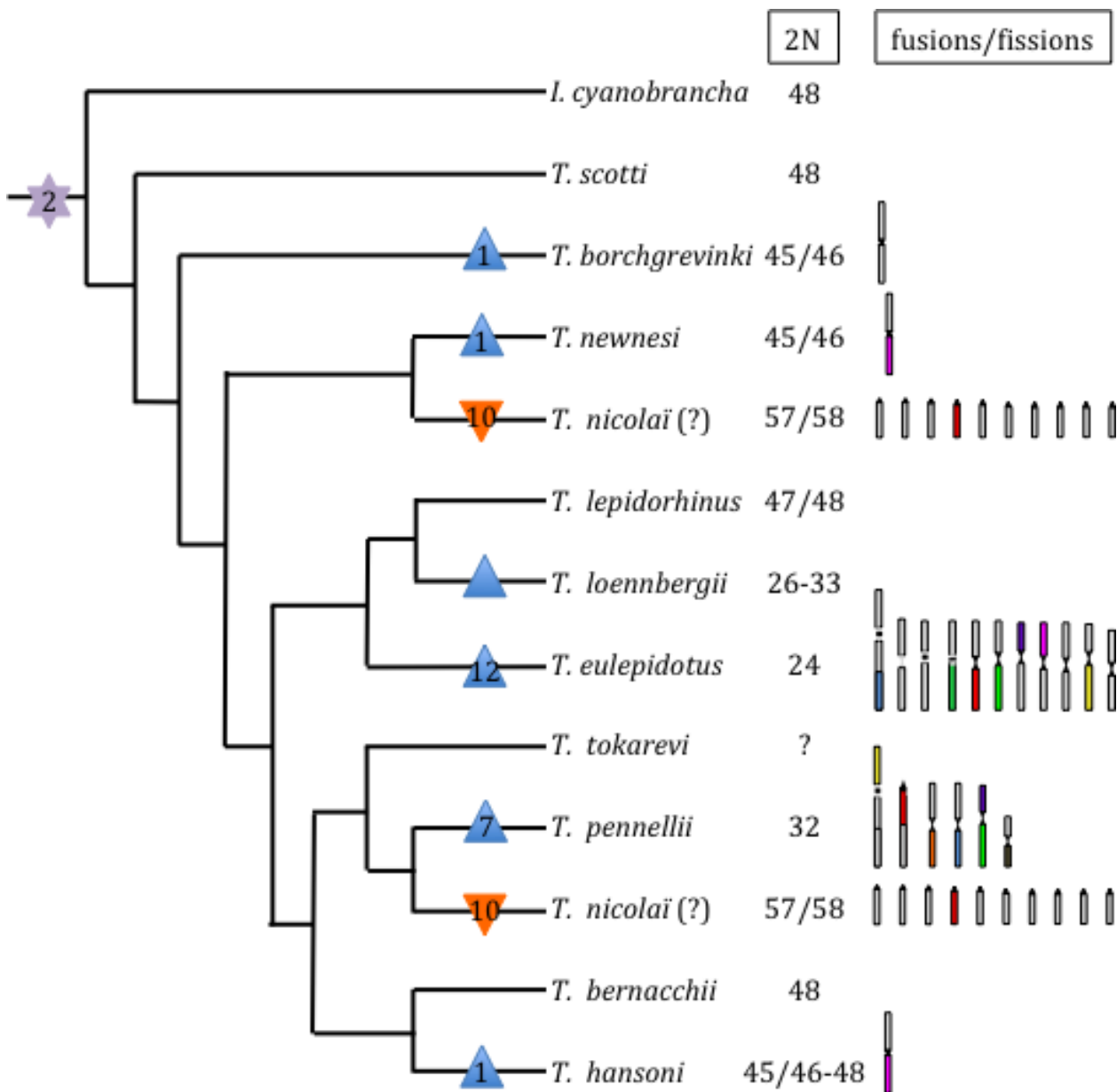
**Figure V.4 :** modèle de spéciation stasipatrique proposé par White (1969) et repris par Auffray (1993). Les populations rouges et bleu clair qui ont acquis une fusion centrique simple persistent. Leur hybridation avec la population d'origine (ne présentant pas de chromosomes fusionnés) est en effet viable et fertile, contrairement à l'individu vert chez qui plusieurs fusions sont apparues spontanément dans les gamètes.



**Figure V.5:** scénarios évolutifs des diversifications chez les « *Trematomus* » suivant l'hypothèse (A) de la spéciation stasipatrique ou (B) allopatrique. Dans les deux cas, les remaniements (principalement des fusions) apparaissent indépendamment au cours de l'évolution des « *Trematomus* ». Un événement apparaît, est fixé, et des événements secondaires peuvent venir renforcer la barrière reproductrice. Sous la première hypothèse (A), le polymorphisme caryotypique intervient avant la divergence des espèces. Sous la deuxième hypothèse (B), il peut intervenir avant ou après les spéciations (Figure V.6).

Chez les « *Trematomus* », nous proposons deux types de scénarios évolutifs pour retracer les remaniements chromosomiques ayant accompagné la radiation évolutive (Figure V.5). Ils s'inspirent fortement du modèle de la spéciation stasipatrique (Figures V.4) pour le premier, et allopatrique pour le deuxième. Dans le premier scénario, différentes fusions chromosomiques (ou fissions) apparaissent spontanément indépendamment dans la population d'origine présente au niveau du plateau continental antarctique, et non sujette à des colonisations de migrants (séparation de l'Océan Austral par la barrière géographique du courant circumpolaire antarctique). Les périodes de glaciations conduisent à une extension de la calotte glaciaire et à un isolement par partitionnement des populations suite à la fragmentation des habitats. Les différents remaniements (seulement s'ils sont uniques) (Auffray 2001) sont fixés dans chaque sous-population en passant à l'état homozygote. Ces fixations sont d'autant plus rapides que la taille de ces populations est petite. Les barrières de flux géniques entre les différentes populations isolées sont renforcées par la fixation d'autres fusions/fissions pouvant être acquises indépendamment après les divergences des espèces. Lors des périodes de réchauffement, la circulation libre sur le plateau continental est de nouveau permise et les différentes espèces sont de nouveau remises en contact (Figure V.5.A). La diversification peut être accentuée par des possibles colonisations de niches vacantes (extinctions massives des espèces non adaptées à des températures nécessitant des protéines antigels par exemple) combinées à la diversification importante des modes de vie (Eastman and McCune 2000; Eastman 2005; Lecoindre et al. 2013), par exemple cryopélagiques (Eastman and DeVries 1985; Clarke and Johnston 1996).

Dans le deuxième scénario, le polymorphisme intervient en période glaciaire lorsque les populations sont déjà isolées géographiquement dans des micro-niches refuges réparties tout autour du plateau continental. Les remaniements sont indépendants dans les différentes populations. Ils sont fixés dans ces petites populations qui deviennent homozygotes pour la/les fusion(s) donnée(s). En période de réchauffement, les espèces sont remises en contact (Figure V.5.B). Dans ce deuxième scénario, les remaniements indépendants peuvent apparaître dans les différentes lignées soit avant, soit après leur divergence (Figure V.6). Dans le cas où la fixation des différentes fusions/fissions uniques précède les spéciations (non partagés et donc non comptés comme événements au regard de la phylogénie), on compte un total de trente-neuf pas, tandis que dans le cas où tous les remaniements sont apparus après la divergence des espèces, on en compte quarante-cinq (Figure V.6).



**Figure V.6 :** positionnement des remaniements chromosomiques ayant accompagné les spéciations lors de la diversification des « *Trematomus* ». Le cladogramme reflète les résultats obtenus par nos reconstructions phylogénétiques. Des fusions centriques (ou fissions) indépendantes apparaissent dans différentes populations et sont fixées (si uniques) avant ou après (deuxième scénario explicité ici) les spéciations. Des remaniements chromosomiques supplémentaires (triangles bleus = fusion(s) Robertsonienne(s)-centriques ou en tandem, triangles orange = fission(s)) viennent renforcer l'isolement reproducteur après la divergence des différentes espèces. L'étoile violette désigne les deux inversions péricentriques acquises chez l'ancêtre commun du clade formé par les Pseudaphritidae, Eleginopsidae et les Nototheniidae (Figure I.2) (avec possible pertes chez *T. eulepidotus*, cercle violet). Sous cette hypothèse d'une acquisition des remaniements intégralement après les divergences, le nombre d'événements dans chaque branche est indiqué à l'intérieur de chaque triangle (quarante-cinq pas).

La divergence des Nototheniidae (ou clade antarctique) est considérée comme récente (23,9 Ma, Matschiner *et al.* (2011), 22,4 Ma, Near *et al.* (2012), 14,7 Ma, Colombo *et al.* (2015)). Bien que l'hybridation entre ces espèces semble un phénomène plutôt ponctuel en milieu naturel (isolements par le comportement, séparations des périodes et des lieux de ponte) (La Mesa *et al.* 2006; La Mesa *et al.* 2008), des traces d'introgessions passées et contemporaines ont été détectées dans les génomes de trois espèces de Channichthyinae (*C. myersi*, *C. rastrospinus* et *C. hamatus*) (Marino *et al.* 2013). Dans chacun de ces cas, les individus possèdent tous une même formule chromosomique de  $2n = 47$  ou  $48$ , (hétéromorphisme sexuel présent chez *C. myersi* et *C. hamatus*) (Ozouf-Costaz *et al.* 1991; Ghigliotti *et al.* 2015). La forte synténie des unités structurales montrée notamment chez les Channichthyinae (Figure II.15), et la divergence récente de cette radiation (-6,3 Ma, Near *et al.* (2012), -3,49 Ma, Colombo *et al.* (2015)) renforcent l'hypothèse de la possibilité de croisements. Au cours de nos reconstructions phylogénétiques, nous avons trouvé deux individus *T. nicolai* non monophylétiques. L'individu *T. nicolai* CE5684 est groupe frère des deux individus *T. newnesi*, tandis que l'individu *T. nicolai* CE5778 est groupe frère des deux individus *T. pennellii*. Comme mentionné auparavant, une étude détaillée des locus ayant servi à l'établissement de ces liens de parentés est requise pour pouvoir explorer ces génomes et expliquer ces positions. Cependant, la possibilité de génomes hybrides pour un ou les deux individus *T. nicolai* examinés n'est pas écartée pour le moment. La radiation des « *Trematomus* » est en effet récente et les unités structurales sont conservées (Figures II.10 et II.13). Cependant, *T. nicolai* ( $2n = 57/58$ ) est le seul cas de fissions massives (au moins dix paires chromosomiques) recensé dans ce groupe jusqu'à présent (Figures I.3 et II.5). La possibilité des croisements avec d'autres espèces de « *Trematomus* » dont les caryotypes présentent des chromosomes fusionnés de type  $2n = 45/46$  chromosomes comme *T. newnesi* ou bien  $2n = 32$  chromosomes comme *T. pennellii* reste donc à vérifier.





## VI) Conclusion et perspectives

L'étude de la reconstitution de l'histoire évolutive des réarrangements chromosomiques ayant accompagné les diversifications rapides chez le groupe « *Trematomus* » nous a permis de mettre en évidence des événements multiples et non partagés (majoritairement des fusions), acquis indépendamment au cours de l'évolution des génomes. Les fusions mises en jeu pour chacune de ces espèces combinent des unités structurales chromosomiques à la base des réarrangements observés. Ces unités ont été définies chez *I. cyanobrancha*, dont la position chez les « *Trematomus* » est basale. Nos résultats tendent à montrer que ces unités structurales auraient subi peu ou pas de modifications (remaniements secondaires de type inversions/translocations) depuis leur évolution à partir du caryotype de l'ancêtre commun de ce groupe.

Ces remaniements indépendants ont pu être facilités par la mobilisation massive de certains éléments transposables, notamment les rétrotransposons à tyrosine recombinase de type *DIRS1*. Il est probable que ces événements de réarrangements chromosomiques aient contribué, ou du moins renforcé l'isolement reproducteur entre les différentes populations dans un contexte d'alternance de périodes glaciaires et interglaciaires. Cependant, les barrières reproductives ne sont pas totales et des hybridations entre espèces différentes peuvent peut être encore s'opérer, même si elles restent très minoritaires en milieu naturel. Malgré des caryotypes parfois très variables, les génomes semblent en effet encore conservés, tant en terme d'identité en séquences qu'en synténies retrouvées entre unités structurales. Ceci est probablement dû à la divergence récente de ce groupe.

Le passage d'une étude de cytogénomique comparative à une étude de génomique comparative permettra d'obtenir des informations plus précises concernant ces phénomènes de diversifications chromosomiques à l'échelle de la famille. Les nouvelles technologies de séquençage (Nanopore pour générer des longs fragments, Illumina pour obtenir de courts fragments) permettront d'obtenir des génomes complets assemblés et annotés de diverses espèces de Nototheniidae. Nous serons alors en mesure de caractériser les ETs dans leur intégralité (structure, nombre de copies) et d'affiner les synténies chromosomiques interspécifiques. Idéalement, si l'assemblage est suffisant et que les contigs traversent de part en part les régions centromériques et péricentromériques, les points chauds d'insertion en ETs (comme les *DIRS1*) pourront également être localisés et étudiés de manière plus fine sur les chromosomes.



## VII) Bibliographie :

- Ahn D-H, Shin SC, Kim B-M, Kang S, Kim J-H, Ahn I, Park J, Park H. 2017. Draft genome of the Antarctic dragonfish, *Parachaenichthys charcoti*. *GigaScience* 6:1–6.
- Alkan C, Sajjadian S, Eichler EE. 2011. Limitations of next-generation genome sequence assembly. *Nature Methods* 8:61–65.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *Journal of molecular biology* 215:403–410.
- Alva V, Nam S-Z, Söding J, Lupas AN. 2016. The MPI bioinformatics Toolkit as an integrative platform for advanced protein sequence and structure analysis. *Nucleic Acids Research* 44:W410–W415.
- Amores A. 1998. Zebrafish hox Clusters and Vertebrate Genome Evolution. *Science* 282:1711–1714.
- Amores A, Bejar J, Alvarez MC. 1995. Replication, C and Ag-NOR chromosome-banding in two anguilliform fish species. *Marine Biology* 123:845–849.
- Amores A, Catchen J, Ferrara A, Fontenot Q, Postlethwait JH. 2011. Genome Evolution and Meiotic Maps by Massively Parallel DNA Sequencing: Spotted Gar, an Outgroup for the Teleost Genome Duplication. *Genetics* 188:799–808.
- Amores A, Catchen J, Nanda I, Warren W, Walter R, Scharl M, Postlethwait JH. 2014. A RAD-Tag Genetic Map for the Platyfish (*Xiphophorus maculatus*) Reveals Mechanisms of Karyotype Evolution Among Teleost Fish. *Genetics* 197:625–641.
- Amores A, Wilson CA, Allard CAH, Detrich HW, Postlethwait JH. 2017. Cold Fusion: Massive Karyotype Evolution in the Antarctic Bullhead Notothen *Notothenia coriiceps*. *G3 (Genes|Genomes|Genetics)* 7:2195–2207.
- Anderson BM, Thiele KR, Krauss SL, Barrett MD. 2017. Genotyping-by-Sequencing in a Species Complex of Australian Hummock Grasses (*Triodia*): Methodological Insights and Phylogenetic Resolution. Sun G, editor. *PLOS ONE* 12:e0171053.
- Anderson JB. 2000. *Antarctic Marine Geology*. Hardcover. New-york
- Anderson JB, Shipp SS, Lowe AL, Wellner JS, Mosola AB. 2002. The Antarctic Ice Sheet during the Last Glacial Maximum and its subsequent retreat history: a review. *Quaternary Science Reviews* 21:49–70.
- Anderson NC. 1984. Genera and subfamilies of the family Nototheniidae (Pisces, Perciformes) from the Antarctic and Subantarctic. *Steenstrupia* 10:1–34.
- Andrews S. 2016. *FastQC A Quality Control tool for High Throughput Sequence Data*. Babraham Bioinformatics.
- Andriashev AP, Jakubowski M. 1971. Morphological grounds for generic separation of the Antarctic broadhead-fishes (*Trematomus borchgrevinki* Boulenger and *T. brachysoma* Pappenheim) and a new status of the genus *Pagothenia* Nichols et Lamonte (Nototheniidae). *Zool. Zh.* 50:1041–1055.

- Aparicio S, Chapman J, Stupka E, Putman N, Chia J. 2002. Whole-Genome Shotgun Assembly and Analysis of the Genome of *Fugu rubripes*. *Science*, New Series 297:1301–1310.
- Arens YHJM, Engelen JJM, Govaerts LCP, van Ravenswaay CM, Loneus WH, van Lent-Albrechts JCM, van der Blij-Philipsen M, Hamers AJH, Schrande-Stumpel CTRM. 2004. Familial insertion (3;5)(q25.3;q22.1q31.3) with deletion or duplication of chromosome region 5q22.1-5q31.3 in ten unbalanced carriers. *American Journal of Medical Genetics* 130A:128–133.
- Arneson N, Hughes S, Houlston R, Done S. 2008. GenomePlex Whole-Genome Amplification. *Cold Spring Harbor Protocols* 2008:pdb.prot4920-pdb.prot4920.
- Auffray J-C. 1993. Chromosomal divergence in house mice in the light of palaeontology: A colonization-related event? *Quaternary International* 19:21–25.
- Auffray J-C. 2001. Developmental Stability in House Mice Heterozygous for Single Robertsonian Fusions. *Journal of Heredity* 92:23–29.
- Aulard S, Vaudin P, Ladevèze V, Chaminade N, Périquet G, Lemeunier F. 2004. Maintenance of a large pericentric inversion generated by the hobo transposable element in a transgenic line of *Drosophila melanogaster*. *Heredity* 92:151–155.
- Auvinet J, Graça P, Belkadi L, Petit L, Bonnivard E, Dettai A, Detrich WH, Ozouf-Costaz C, Higuët D. 2018. Mobilization of retrotransposons as a cause of chromosomal diversification and rapid speciation: the case for the Antarctic teleost genus *Trematomus*. *BMC Genomics* [Internet] 19. Available from: <https://bmcgenomics.biomedcentral.com/articles/10.1186/s12864-018-4714-x>
- Bailey JA. 2002. Recent Segmental Duplications in the Human Genome. *Science* 297:1003–1007.
- Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, Lewis ZA, Selker EU, Cresko WA, Johnson EA. 2008. Rapid SNP Discovery and Genetic Mapping Using Sequenced RAD Markers. *Fay JC, editor. PLoS ONE* 3:e3376.
- Baker RJ, Bickham JW. 1986. Speciation by monobrachial centric fusions. *Proceedings of the National Academy of Sciences* 83:8245–8248.
- Balushkin A. 1992. Classification, phylogenetic relationships, and origins of the families of the suborder Notothenioidei (Perciformes). :90–110.
- Balushkin A. 2000. Morphology, Classification, and Evolution of Notothenioid Fishes of the Southern Ocean (Notothenioidei, Perciformes). *Journal of Ichthyology* 40:S74–S109.
- Balushkin AV. 1982. Classification of Trematominae fishes of the Antarctic. In: *Biology of shelf zones of the world oceans 2*. Balushkin AV. Vladivostok.
- Bargelloni L, Lecointre G. 1998. Four years in Notothenioid systematics: a molecular perspective. In: *Fishes of Antarctica*. Di Prisco G., Pisano E., Andrew C. p. 259–273.
- Bargelloni L, Marcato S, Zane L, Patarnello T. 2000. Mitochondrial Phylogeny of Notothenioids: a molecular approach to Antarctic fish evolution and biogeography. *Systematic Biology* 49:114–129.
- Bargelloni L, Rithie PA, Patarnello T, Battaglia B, Lambert D., Meyer A. 1994. Molecular evolution at subzero temperatures: mitochondrial and nuclear phylogenies of fishes from

- Antarctica (suborder Notothenioidei), and the evolution of antifreeze glycopeptides. *Molecular Biology and Evolution* 6:854–863.
- Bartolomé C, Bello X, Maside X. 2009. Widespread evidence for horizontal transfer of transposable elements across *Drosophila* genomes. *Genome Biology* 10:R22.
- Barton NH, Rouhani S. 1991. THE PROBABILITY OF FIXATION OF A NEW KARYOTYPE IN A CONTINUOUS POPULATION. *Evolution* 45:499–517.
- Bates GP, Wainwright BJ, Williamson R, Brown SD. 1986. Microdissection of and microcloning from the short arm of human chromosome 2. *Molecular and cellular biology* 6:3826–3830.
- Bed'Hom B. 2000. Evolution of Karyotype Organisation in Accipitridae: a Translocation Model. In: Sankoff D, Nadeau JH, editors. *Comparative Genomics*. Vol. 1. Dordrecht: Springer Netherlands. p. 347–356. Available from: [http://www.springerlink.com/index/10.1007/978-94-011-4309-7\\_32](http://www.springerlink.com/index/10.1007/978-94-011-4309-7_32)
- Belyayev A. 2014. Bursts of transposable elements as an evolutionary driving force. *Journal of Evolutionary Biology* 27:2573–2584.
- Bennetzen JL. 2000. Transposable element contributions to plant gene and genome evolution. Doyle JJ, Gaut BS, editors. *Plant Molecular Evolution* 42:251–269.
- Betancur R, Wiley EO, Arratia G, Acero A, Bailly N, Miya M, Lecointre G, Orti G. 2017. Phylogenetic classification of bony fishes. *BMC Evolutionary Biology* 17:1–40.
- Biémont C, Vieira C. 2005. What transposable elements tell us about genome organization and evolution: the case of *Drosophila*. *Cytogenetic and genome research* 110:25–34.
- Biémont C, Vieira C. 2006. Junk DNA as an evolutionary force. *Nature* 443:4.
- Biltueva L, Vorobieva N, Perelman P, Trifonov V, Volobouev V, Panov V, Ilyashenko V, Onischenko S, O'Brien P, Yang F, et al. 2011. Karyotype Evolution of Eulipotyphla (Insectivora): The Genome Homology of Seven Sorex Species Revealed by Comparative Chromosome Painting and Banding Data. *Cytogenetic and Genome Research* 135:51–64.
- Blennow E. 2004. Reverse painting highlights the origin of chromosome aberrations. *Chromosome Research* 12:25–33.
- Böhne A, Brunet F, Galiana-Arnoux D, Schultheis C, Volff J-N. 2008. Transposable elements as drivers of genomic and biological diversity in vertebrates. *Chromosome Research* 16:203–215.
- Bonillo C, Coutanceau JP, D'Cotta H, Ghigliotti L, Ozouf-Costaz C, Pisano E. 2015. Standard Fluorescence in situ Hybridization Procedures. *Fish Cytogenetic Techniques: Ray-Fin Fishes and Chondrichthyans*:103.
- Bonnivard E, Higué D. 2009. Fluidity of eukaryotic genomes. *Comptes Rendus Biologies* 332:234–240.
- Bourque G. 2009. Transposable elements in gene regulation and in the evolution of vertebrate genomes. *Current Opinion in Genetics & Development* 19:607–612.

- Bowers JE, Chapman BA, Rong J, Paterson AH. 2003. Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. *Nature* 422:433–438.
- Braasch I, Gehrke AR, Smith JJ, Kawasaki K, Manousaki T, Pasquier J, Amores A, Desvignes T, Batzel P, Catchen J, et al. 2016. The spotted gar genome illuminates vertebrate evolution and facilitates human-teleost comparisons. *Nature Genetics* 48:427–437.
- Brawand D, Wagner CE, Li YI, Malinsky M, Keller I, Fan S, Simakov O, Ng AY, Lim ZW, Bezault E, et al. 2014. The genomic substrate for adaptive radiation in African cichlid fish. *Nature* 513:375–381.
- Breneman JW, Ramsey MJ, Lee DA, Eveleth GG, Minkler JL, Tucker JD. 1993. The development of chromosome-specific composite DNA probes for the mouse and their application to chromosome painting. *Chromoma* 102:591–598.
- Britton-Davidian J, Catalan J, da Graça Ramalhinho M, Ganem G, Auffray J-C, Capela R, Biscoito M, Searle JB, da Luz Mathias M. 2000. Rapid chromosomal evolution in island mice: Environmental genetics. *Nature* 403:158–158.
- Brookfield JFY. 2005. The ecology of the genome — mobile DNA elements and their hosts. *Nature Reviews Genetics* 6:128–136.
- Brosius J. 2003. The contribution of RNAs and retroposition to evolutionary novelties. *Genetica* 118:99–116.
- Buckley BA. 2004. Regulation of heat shock genes in isolated hepatocytes from an Antarctic fish, *Trematomus bernacchii*. *Journal of Experimental Biology* 207:3649–3656.
- Bush G. 1975. Modes of Animal Speciation. *Annual Review of Ecology and Systematics Annual Review of Ecology and Systematics*:339–364.
- Bustin SA, Benes V, Garson JA, Hellems J, Huggett J, Kubista M, Mueller R, Nolan T, Pfaffl MW, Shipley GL, et al. 2009. The MIQE Guidelines: Minimum Information for Publication of Quantitative Real-Time PCR Experiments. *Clinical Chemistry* 55:611–622.
- Caceres M. 2001. Molecular Characterization of Two Natural Hotspots in the *Drosophila buzzatii* Genome Induced by Transposon Insertions. *Genome Research* 11:1353–1364.
- Caceres ME, De Pace C, Mugnoz GTS, Kotsonis P, Ceccarelli M, Cionini PG. 1998. Genome size variations within *Dasypyrum villosum*: correlations with chromosomal traits, environmental factors and plant phenotypic characteristics and behaviour in reproduction. *TAG Theoretical and Applied Genetics* 96:559–567.
- Cappello J, Handelsman K, Lodish H F. 1985. Sequence of Dictyostelium DIRS-1: an apparent retrotransposon with inverted terminal repeats and an internal circle junction sequence. *Cell* 43:105–115.
- Cappello J, Zuker C, Lodish HF. 1984. Repetitive Dictyostelium heat-shock promotor functions in *Saccharomyces cerevisiae*. *Molecular and Cellular Biology* 4:591–598.
- Capriglione T, De Paolo S, Cocca E. 2014. Helinoto, a Helitron2 transposon from the icefish *Chionodraco hamatus*, contains a region with three deubiquitinase-like domains that exhibit transcriptional activity. *Comparative Biochemistry and Physiology Part D: Genomics and Proteomics* 11:49–58.

- Capriglione T, Odierna G, Caputo V, Canapa A, Olmo E. 2002. Characterization of a Tc1-like transposon in the Antarctic ice-fish, *Chionodraco hamatus*. *Gene* 295:193–198.
- Casals F. 2003. The Foldback-like Transposon Galileo Is Involved in the Generation of Two Different Natural Chromosomal Inversions of *Drosophila buzzatii*. *Molecular Biology and Evolution* 20:674–685.
- Castiglia R, Gornung E, Corti M. 2002. Cytogenetic analyses of chromosomal rearrangements in *Mus minutoides/musculoides* from North-West Zambia through mapping of the telomeric sequence (TTAGGG)<sub>n</sub> and banding techniques. *Chromosome Research* 10:399–406.
- Castresana J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular biology and evolution* 17:540–552.
- Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH. 2011. Stacks : Building and Genotyping Loci De Novo From Short-Read Sequences. *Genes|Genomes|Genetics* 1:171–182.
- Cavallini A, Natali L. 1991. Intraspecific variation of nuclear DNA content in plant species. *Caryologia* 44:93–107.
- Chalopin D, Naville M, Plard F, Galiana D, Volff J-N. 2015. Comparative Analysis of Transposable Elements Highlights Mobilome Diversity and Evolution in Vertebrates. *Genome Biology and Evolution* 7:567–580.
- Chaves R, Frönicke L, Guedes-Pinto H, Wienberg J. 2004. Multidirectional chromosome painting between the Hirola antelope (*Damaliscus hunteri*, Alcelaphini, Bovidae), sheep and human. *Chromosome Research* 12:495–503.
- Chaw C-C, Chen H-L, Li W-H, Chaw S-M. 2004. Dating the Monocot?Dicot Divergence and the Origin of Core Eudicots Using Whole Chloroplast Genomes. *Journal of Molecular Evolution* 58:424–441.
- Chen S, Yu M, Chu X, Li W, Yin X, Chen L. 2017. Cold-induced retrotransposition of fish LINES. *Journal of Genetics and Genomics* 44:385–394.
- Chen Z, Cheng C-HC, Zhang J, Cao L, Chen L, Zhou L, Jin Y, Ye H, Deng C, Dai Z, et al. 2008. Transcriptomic and genomic evolution under constant cold in Antarctic notothenioid fish. *Proceedings of the National Academy of Sciences* 105:12944–12949.
- Chénais B. 2015. Transposable elements in cancer and other human diseases. *Current cancer drug targets* 15:227–242.
- Chénais B, Caruso A, Hiard S, Casse N. 2012. The impact of transposable elements on eukaryotic genomes: From genome size increase to genetic adaptation to stressful environments. *Gene* 509:7–15.
- Cheng C, Detrich HW. 2007. Molecular ecophysiology of Antarctic notothenioid fishes. *Philosophical Transactions of the Royal Society B: Biological Sciences* 362:2215–2232.
- Cheng C-HC. 2003. Functional Antifreeze Glycoprotein Genes in Temperate-Water New Zealand Nototheniid Fish Infer an Antarctic Evolutionary Origin. *Molecular Biology and Evolution* 20:1897–1908.



- Chowdhary BP, Raudsepp T. 2001. Chromosome painting in farm, pet and wild animal species. *Methods in Cell Science* 23:37–55.
- Christian AT, Garcia HE, Tucker JD. 1999. PCR in situ followed by microdissection allows whole chromosome painting probes to be made from single microdissected chromosomes. *Mammalian Genome* 10:628–631.
- Clark MS, Clarke A, Cockell CS, Convey P, Detrich III HW, Fraser KPP, Johnston IA, Methe BA, Murray AE, Peck LS, et al. 2004. Antarctic Genomics. *Comparative and Functional Genomics* 5:230–238.
- Clarke A. 1990. Temperature and Evolution: Southern Ocean Cooling and the Antarctic Marine Fauna. In: Kerry KR, Hempel G, editors. *Antarctic Ecosystems*. Berlin, Heidelberg: Springer Berlin Heidelberg. p. 9–22. Available from: [http://www.springerlink.com/index/10.1007/978-3-642-84074-6\\_2](http://www.springerlink.com/index/10.1007/978-3-642-84074-6_2)
- Clarke A, Johnston I. 1996. Evolution and adaptive radiation of Antarctic fishes. *Tree* 11:212–218.
- Cocca E, Iorio SD, Capriglione T. 2011. Identification of a novel helitron transposon in the genome of Antarctic fish. *Molecular Phylogenetics and Evolution* 58:439–446.
- Cocca E, Petraccioli A, Morescalchi MA, Odierna G, Capriglione T. 2015. Laser microdissection-based analysis of the Y sex chromosome of the Antarctic fish *Chionodraco hamatus* (Notothenioidei, Channichthyidae). *Comparative Cytogenetics* 9:1–15.
- Cohen CJ, Lock WM, Mager DL. 2009. Endogenous retroviral LTRs as promoters for human genes: A critical assessment. *Gene* 448:105–114.
- Colombo M, Damerau M, Hanel R, Salzburger W, Matschiner M. 2015. Diversity and disparity through time in the adaptive radiation of Antarctic notothenioid fishes. *Journal of Evolutionary Biology* 28:376–394.
- Couturier J, Dutrillaux B. 1981. Conservation of replication chronology of homoeologous chromosome bands between four species of the genus *Cebus* and man. *Cytogenetics and Cell Genetics* 29:233–240.
- Crollius HR. 2000. Characterization and Repeat Analysis of the Compact Genome of the Freshwater Pufferfish *Tetraodon nigroviridis*. *Genome Research* 10:939–949.
- Cruaud A, Gautier M, Galan M, Foucaud J, Sauné L, Genson G, Dubois E, Nidelet S, Deuve T, Rasplus J-Y. 2014. Empirical Assessment of RAD Sequencing for Interspecific Phylogeny. *Molecular Biology and Evolution* 31:1272–1274.
- Curcio MJ, Derbyshire KM. 2003. The outs and ins of transposition: from Mu to Kangaroo. *Nature Reviews Molecular Cell Biology* 4:865–877.
- Cutter AD. 2013. Integrating phylogenetics, phylogeography and population genetics through genomes and evolutionary theory. *Molecular Phylogenetics and Evolution* 69:1172–1185.
- Cziko PA, Cheng C-HC. 2006. A New Species of Nototheniid (Perciformes: Notothenioidei) Fish from McMurdo Sound, Antarctica. Both D, editor. *Copeia* 2006:752–759.

- Dancis BM, Holmquist GP. 1979. Telomere replication and fusion in eukaryotes. *Journal of Theoretical Biology* 78:211–224.
- Davey JW, Hohenlohe PA, Etter PD, Boone JQ, Catchen JM, Blaxter ML. 2011. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics* 12:499–510.
- De Vries A. 1980. Biological antifreezes and survival in freezing environments. In: *Animals and environmental fitness*. Pergamon. Oxford: Gilles R. p. 583.
- De Vries AL. 1971. Glycoproteins as Biological Antifreeze Agents in Antarctic Fishes. *Science* 172:1152–1155.
- Deininger PL, Moran JV, Batzer MA, Kazazian HH. 2003. Mobile elements and mammalian genome evolution. *Current Opinion in Genetics & Development* 13:651–658.
- Deng H-X, Yoshiura K, Dirks R, Harada N, Hirota T, Tsukamoto K, Jinno Y, Niikawa N. 1992. Chromosome-band-specific painting: chromosome in situ suppression hybridization using PCR products from a microdissected chromosome band as a probe pool. *Human Genetics* 89:13–17.
- Denton JF, Lugo-Martinez J, Tucker AE, Schridder DR, Warren WC, Hahn MW. 2014. Extensive Error in the Number of Genes Inferred from Draft Genome Assemblies. Guigo R, editor. *PLoS Computational Biology* 10:e1003998.
- Detrich HW, Amemiya CT. 2010. Antarctic Notothenioid Fishes: Genomic Resources and Strategies for Analyzing an Adaptive Radiation. *Integrative and Comparative Biology* 50:1009–1017.
- Detrich HW, Parker SK, Williams RC, Nogales E, Downing KH. 2000. Cold Adaptation of Microtubule Assembly and Dynamics: STRUCTURAL INTERPRETATION OF PRIMARY SEQUENCE CHANGES PRESENT IN THE  $\alpha$ - AND  $\beta$ -TUBULINS OF ANTARCTIC FISHES. *Journal of Biological Chemistry* 275:37038–37047.
- Detrich HW, Stuart A, Schoenborn M, Parker SK, Methé BA, Amemiya CT. 2010. Genome enablement of the notothenioid: genome size estimates from 11 species and BAC libraries from 2 Representative Taxa. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution* 314B:369–381.
- Dettaï A, Berkani M, Lautredou A-C, Couloux A, Lecointre G, Ozouf-Costaz C, Gallut C. 2012. Tracking the elusive monophyly of nototheniid fishes (Teleostei) with multiple mitochondrial and nuclear markers. *Marine Genomics* 8:49–58.
- Dettaï A, Lecointre G. 2004. In search of notothenioid (Teleostei) relatives. *Antarctic Science* 16:71–85.
- Devos KM, Brown JK, Bennetzen JL. 2002. Genome size reduction through illegitimate recombination counteracts genome expansion in *Arabidopsis*. *Genome research* 12:1075–1079.
- DeWitt H, Heemstra PS, Gon O. 1990. Nototheniidae. In: *Fishes of the Southern Ocean*. Grahamstown: Gon, O, Heemstra, P.C. p. 279–331.

- Dion-Cote A-M, Renaut S, Normandeau E, Bernatchez L. 2014. RNA-seq Reveals Transcriptomic Shock Involving Transposable Elements Reactivation in Hybrids of Young Lake Whitefish Species. *Molecular Biology and Evolution* 31:1188–1199.
- Dobigny G, Aniskin V, Granjon L, Cornette R, Volobouev V. 2005. Recent radiation in West African *Taterillus* (Rodentia, Gerbillinae): the concerted role of chromosome and climatic changes. *Heredity* 95:358–368.
- Dobigny G, Ozouf-Costaz C, Bonillo C, Volobouev V. 2003. Evolution of rRNA gene clusters and telomeric repeats during explosive genome repatterning in *Taterillus* X (Rodentia, Gerbillinae). *Cytogenetic and Genome Research* 103:94–103.
- Donnart T, Piednoël M, Higuët D, Bonnivard É. 2017. Filamentous ascomycete genomes provide insights into Copia retrotransposon diversity in fungi. *BMC Genomics* [Internet] 18. Available from: <http://bmcgenomics.biomedcentral.com/articles/10.1186/s12864-017-3795-2>
- Duhamel G, Hulley P, Causse R, Koubbi P, Vacchi M, Pruvost P, Vigetta S, Irisson J, Mormède S, Belchier M, et al. 2014. Biogeographic atlas of the Southern Ocean. In: Scientific Committee on Antarctic Research. Cambridge: De Broyer C, Koubbi P, Griffiths HJ, Raymond B, Udekem d'Acoz C. p. 328–352.
- Dumas F, Bigoni F, Stone G, Sineo L, Stanyon R. 2005. Mapping genomic rearrangements in titi monkeys by chromosome flow sorting and multidirectional in-situ hybridization. *Chromosome Research* 13:85–96.
- Duncan L, Bouckaert K, Yeh F, Kirk DL. 2002. Kangaroo, a mobile element from *Volvox carteri*, is a member of a newly recognized third class of retrotransposons. *Genetics* 162:1617–1630.
- Dunn CW, Hejnal A, Matus DQ, Pang K, Browne WE, Smith SA, Seaver E, Rouse GW, Obst M, Edgecombe GD, et al. 2008. Broad phylogenomic sampling improves resolution of the animal tree of life. *Nature* 452:745–749.
- Dutrillaux B. 1979. Chromosomal evolution in Primates: Tentative phylogeny from *Microcebus murinus* (Prosimian) to man. *Human Genetics* 48:251–314.
- Dutrillaux B, Couturier J, Chauvier G. 1980. Chromosomal evolution of 19 species of sub-species of Cercopithecinae. *Annales de génétique* 23:133–143.
- Duvernell DD, Pryor SR, Adams SM. 2004. Teleost Fish Genomes Contain a Diverse Array of L1 Retrotransposon Lineages That Exhibit a Low Copy Number and High Rate of Turnover. *Journal of Molecular Evolution* 59:298–308.
- Eastman J, McCune AR. 2000. Fishes on the Antarctic continental shelf: evolution of a marine species flock? *Journal of Fish Biology* 57:84–102.
- Eastman JT. 2005. The nature of the diversity of Antarctic fishes. *Polar Biology* 28:93–107.
- Eastman JT. 2017. Bathymetric distributions of notothenioid fishes. *Polar Biology* 40:2077–2095.
- Eastman JT, DeVries AL. 1985. Adaptations for cryopelagic life in the antarctic notothenioid fish *Pagothenia borchgrevinki*. *Polar Biology* 4:45–52.

- Eastman JT, Devries AL. 1997. Biology and phenotypic plasticity of the Antarctic nototheniid fish *Trematomus newnesi* in McMurdo Sound. *Antarctic Science* [Internet] 9. Available from: [http://www.journals.cambridge.org/abstract\\_S0954102097000047](http://www.journals.cambridge.org/abstract_S0954102097000047)
- Eastman JT, Hubold G. 1999. The fish fauna of the Ross Sea, Antarctica. *Antarctic Science* [Internet] 11. Available from: [http://www.journals.cambridge.org/abstract\\_S0954102099000383](http://www.journals.cambridge.org/abstract_S0954102099000383)
- Eaton DAR. 2014. PyRAD: assembly of de novo RADseq loci for phylogenetic analyses. *Bioinformatics* 30:1844–1849.
- Eaton DAR, Ree RH. 2013. Inferring Phylogeny and Introgression using RADseq Data: An Example from Flowering Plants (Pedicularis: Orobanchaceae). *Systematic Biology* 62:689–706.
- Eickbush TH, Jamburuthugoda VK. 2008. The diversity of retrotransposons and the properties of their reverse transcriptases. *Virus Research* 134:221–234.
- Eickbush TH, Malik HS. 2002. Origins and Evolution of Retrotransposons. In: Craig NL, Lambowitz AM, Craigie R, Gellert M, editors. *Mobile DNA II*. American Society of Microbiology. p. 1111–1144. Available from: <http://www.asmscience.org/content/book/10.1128/9781555817954.chap49>
- Eklom R, Wolf JBW. 2014. A field guide to whole-genome sequencing, assembly and annotation. *Evolutionary Applications* 7:1026–1042.
- Elgar G. 1999. Generation and Analysis of 25 Mb of Genomic DNA from the Pufferfish *Fugu rubripes* by Sequence Scanning. *Genome Research* 9:960–971.
- Emanuel BS, Shaikh TH. 2001. Segmental duplications: an “expanding” role in genomic instability and disease. *Nature Reviews Genetics* 2:791–800.
- Emerson KJ, Merz CR, Catchen JM, Hohenlohe PA, Cresko WA, Bradshaw WE, Holzapfel CM. 2010. Resolving postglacial phylogeography using high-throughput sequencing. *Proceedings of the National Academy of Sciences* 107:16196–16200.
- Felsenstein J. 1985. Confidence Limits on Phylogenies: An Approach Using the Bootstrap. *Evolution* 39:783.
- Ferreira DC, Porto-Foresti F, Oliveira C, Foresti F. 2011. Transposable elements as a potential source for understanding the fish genome. *Mobile Genetic Elements* 1:112–117.
- Ferreira IA, Poletto AB, Kocher TD, Mota-Velasco JC, Penman DJ, Martins C. 2010. Chromosome Evolution in African Cichlid Fish: Contributions from the Physical Mapping of Repeated DNAs. *Cytogenetic and Genome Research* 129:314–322.
- Feschotte C. 2008. Transposable elements and the evolution of regulatory networks. *Nature Reviews Genetics* 9:397–405.
- Feuk L, Carson AR, Scherer SW. 2006. Structural variation in the human genome. *Nature Reviews Genetics* 7:85–97.
- Finnegan D. 1989. Eucaryotic transposable elements and genome evolution. *TIG* 5:103–107.

- Garagna S, Marziliano N, Zuccotti M, Searle JB, Capanna E, Redi CA. 2001. Pericentromeric organization at the fusion point of mouse Robertsonian translocation chromosomes. *Proceedings of the National Academy of Sciences* 98:171–175.
- Ghigliotti L, Cheng CC-H, Ozouf-Costaz C, Vacchi M, Pisano E. 2015. Cytogenetic diversity of notothenioid fish from the Ross sea: historical overview and updates. *Hydrobiologia* 761:373–396.
- Ghigliotti L, Cheng C-HC, Bonillo C, Coutanceau J-P, Pisano E. 2013. In Situ Gene Mapping of Two Genes Supports Independent Evolution of Sex Chromosomes in Cold-Adapted Antarctic Fish. *BioMed Research International* 2013:1–8.
- Ghigliotti L, Fevolden S-E, Cheng C-HC, Babiak I, Dettai A, Pisano E. 2012. Karyotyping and cytogenetic mapping of Atlantic cod (*Gadus morhua* Linnaeus, 1758). *Animal Genetics* 43:746–752.
- Giles V, Thode G, Alvarez MC. 1985. A new Robertsonian fusion in the multiple chromosome polymorphism of a mediterranean population of *Gobius paganellus* (Gobiidae, Perciformes). *Heredity* 55:255–260.
- Giraud T, Capy P. 1996. Somatic Activity of the mariner Transposable Element in Natural Populations of *Drosophila simulans*. *Proceedings of the Royal Society B: Biological Sciences* 263:1481–1486.
- Gold JR, Karel WJ, Strand MR. 1980. Chromosome Formulae of North American Fishes. *The Progressive Fish-Culturist* 42:10–23.
- Gon O, Heemstra PC. 1990. *Fishes of the Southern Ocean*. E. Institute of Ichthyologie. Grahamstown: O. Gon and P.C. Heemstra Available from: [http://www.journals.cambridge.org/abstract\\_S0954102092220381](http://www.journals.cambridge.org/abstract_S0954102092220381)
- Goodwin TJD. 2004. A New Group of Tyrosine Recombinase-Encoding Retrotransposons. *Molecular Biology and Evolution* 21:746–759.
- Goodwin TJD, Poulter RTM. 2001. The DIRS1 Group of Retrotransposons. *Molecular Biology and Evolution* 18:2067–2082.
- Goodwin TJD, Poulter RTM. 2004. A New Group of Tyrosine Recombinase-Encoding Retrotransposons. *Molecular Biology and Evolution* 21:746–759.
- Graphodatsky AS, Yang F, O'brien PCM, Serdukova N, Milne BS, Trifonov V, Ferguson-Smith MA. 2000. A comparative chromosome map of the Arctic fox, red fox and dog defined by chromosome painting and high resolution G-banding. *Chromosome Research* 8:253–263.
- Gray JW, Carrano AV, Moore DH, Steinmetz LL, Minkler J, Mayall BH, Mendelsohn ML, Van Dilla MA. 1975. High-speed quantitative karyotyping by flow microfluorometry. *Clinical chemistry* 21:1258–1262.
- Gray JW, Carrano AV, Steinmetz LL, Van Dilla MA, Moore DH, Mayall BH, Mendelsohn ML. 1975. Chromosome measurement and sorting by flow systems. *Proceedings of the National Academy of Sciences* 72:1231–1234.
- Green ED. 2001. Strategies for the systematic sequencing of complex genomes. *Nature Reviews Genetics* 2:573–583.

- Greilhuber J. 2005. The Origin, Evolution and Proposed Stabilization of the Terms “Genome Size” and “C-Value” to Describe Nuclear DNA Contents. *Annals of Botany* 95:255–260.
- Gribble S, Ng BL, Prigmore E, Burford DC, Carter NP. 2004. Chromosome paints from single copies of chromosomes. *Chromosome Research* 12:143–151.
- Griffin DK, Haberman F, Masabanda J, O’Brien P, Bagga M, Sazanov A, Smith J, Burt DW, Ferguson-Smith M, Wienberg J. 1999. Micro-and macrochromosome paints generated by flow cytometry and microdissection: tools for mapping the chicken genome. *Cytogenetic and Genome Research* 87:278–281.
- Grutzner F, LuÈtjens G, Rovira C, Barnes DW, Ropers H-H, Haaf T. 1999. Classical and molecular cytogenetics of the pufferfish *Tetraodon nigroviridis*. *Chromosome Research* 7:655–662.
- Guan XY, Meltzer PS, Trent JM. 1994. Rapid Generation of Whole Chromosome Painting Probes (WCPS) by Chromosome Microdissection. *Genomics* 22:101–107.
- Guan X-Y, Trent JM, Meltzer PS. 1993. Generation of band-specific painting probes from a single microdissected chromosome. *Human Molecular Genetics* 2:1117–1121.
- Hall T. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium series* 41:95–98.
- Hardie DC, Hebert PD. 2003. The nucleotypic effects of cellular DNA content in cartilaginous and ray-finned fishes. *Genome* 46:683–706.
- Hardie DC, Hebert PD. 2004. Genome-size evolution in fishes. *Canadian Journal of Fisheries and Aquatic Sciences* 61:1636–1646.
- Hartley SE. 1987. The chromosomes of salmonid fishes. *Biological Reviews* 62:197–214.
- Hartley SE, Horne MT. 1985. Cytogenetic techniques in fish genetics. *Journal of Fish Biology* 26:575–582.
- Hejnal A, Obst M, Stamatakis A, Ott M, Rouse GW, Edgecombe GD, Martinez P, Baguna J, Bailly X, Jondelius U, et al. 2009. Assessing the root of bilaterian animals with scalable phylogenomic methods. *Proceedings of the Royal Society B: Biological Sciences* 276:4261–4270.
- Hellsten U, Harland RM, Gilchrist MJ, Hendrix D, Jurka J, Kapitonov V, Ovcharenko I, Putnam NH, Shu S, Taher L, et al. 2010. The Genome of the Western Clawed Frog *Xenopus tropicalis*. *Science* 328:633–636.
- Henning F, Lee HJ, Franchini P, Meyer A. 2014. Genetic mapping of horizontal stripes in Lake Victoria cichlid fishes: benefits and pitfalls of using RAD markers for dense linkage mapping. *Molecular Ecology* 23:5224–5240.
- Henning F, Trifonov V, Almeida-Toledo LF de. 2008. Use of chromosome microdissection in fish molecular cytogenetics. *Genetics and Molecular Biology* 31:279–283.
- Herniou E, Martin J, Miller K, Cook J, Wilkinson M, Tristem M. 1998. Retroviral Diversity and Distribution in Vertebrates. *Journal of Virology* 72:5955–5966.
- Hill RJ, Watt F. 1977. Chromosomal proteins of *Drosophila melanogaster* and an approach for their localisation on polytene chromosomes. *Chromosoma* 63:57–78.

- Hirai H, Hirai Y, LoVerde PT. 2012. Evolution of sex chromosomes ZW of *Schistosoma mansoni* inferred from chromosome paint and BAC mapping analyses. *Parasitology International* 61:684–689.
- Hockner M, Erdel M, Spreiz A, Utermann G, Kotzot D. 2009. Whole Genome Amplification from Microdissected Chromosomes. *Cytogenetic and Genome Research* 125:98–102.
- Hoegg S, Brinkmann H, Taylor JS, Meyer A. 2004. Phylogenetic Timing of the Fish-Specific Genome Duplication Correlates with the Diversification of Teleost Fish. *Journal of Molecular Evolution* 59:190–203.
- Hoen DR, Hickey G, Bourque G, Casacuberta J, Cordaux R, Feschotte C, Fiston-Lavier A-S, Hua-Van A, Hubley R, Kapusta A, et al. 2015. A call for benchmarking transposable element annotation methods. *Mobile DNA* [Internet] 6. Available from: <http://www.mobilednajournal.com/content/6/1/13>
- Hofmann GE. 2000. Lack of heat-shock response in Antarctic fish. *The Journal of Experimental Biology* 203:2331–2339.
- Hohenlohe PA, Bassham S, Etter PD, Stiffler N, Johnson EA, Cresko WA. 2010. Population Genomics of Parallel Adaptation in Threespine Stickleback using Sequenced RAD Tags. Begun DJ, editor. *PLoS Genetics* 6:e1000862.
- Howe K, Clark MD, Torroja CF, Torrance J, Berthelot C, Muffato M, Collins JE, Humphray S, McLaren K, Matthews L, et al. 2013. The zebrafish reference genome sequence and its relationship to the human genome. *Nature* 496:498–503.
- Huang H, Knowles LL. 2016. Unforeseen Consequences of Excluding Missing Data from Next-Generation Sequences: Simulation Study of RAD Sequences. *Systematic Biology* 65:357–365.
- Hua-Van A, Le Rouzic A, Maisonhaute C, Capy P. 2005. Abundance, distribution and dynamics of retrotransposable elements and transposons: similarities and differences. *Cytogenetic and Genome Research* 110:426–440.
- Hureau JC. 1970. Biologie comparee de quelques poissons antarctiques (Nototheniidae). *Bull. Inst. Oceanog. Monaco* 68:1–244.
- Hurley I, Hale ME, Prince VE. 2005. Duplication events and the evolution of segmental identity. *Evolution Development* 7:556–567.
- Huybrechts P. 2002. Sea-level changes at the LGM from ice-dynamic reconstructions of the Greenland and Antarctic ice sheets during the glacial cycles. *Quaternary Science Reviews* 21:203–231.
- Inoue J, Sato Y, Sinclair R, Tsukamoto K, Nishida M. 2015. Rapid genome reshaping by multiple-gene loss after whole-genome duplication in teleost fish suggested by mathematical modeling. *Proceedings of the National Academy of Sciences* 112:14918–14923.
- Jaillon O, Aury J-M, Brunet F, Petit J-L, Stange-Thomann N, Mauceli E, Bouneau L, Fischer C, Ozouf-Costaz C, Bernot A, et al. 2004. Genome duplication in the teleost fish *Tetraodon nigroviridis* reveals the early vertebrate proto-karyotype. *Nature* 431:946–957.
- Janko K, Marshall C, Musilová Z, Houdt JV, Couloux A, Cruaud C, Lecointre G. 2011. Multilocus analyses of an Antarctic fish species flock (Teleostei, Notothenioidei, Trematominae):

- Phylogenetic approach and test of the early-radiation event. *Molecular Phylogenetics and Evolution* 60:305–316.
- Jones DT, Taylor WR, Thornton JM. 1992. The rapid generation of mutation data matrices from protein sequences. *Computer applications in the biosciences: CABIOS* 8:275–282.
- Jones JC, Fan S, Franchini P, Schartl M, Meyer A. 2013. The evolutionary history of Xiphophorus fish and their sexually selected sword: a genome-wide approach using restriction site-associated DNA sequencing. *Molecular Ecology* 22:2986–3001.
- Jones K. 1998. Robertsonian fusion and centric fission in karyotype evolution of higher plants. *The Botanical Review* 64:273–289.
- Kai W, Nomura K, Fujiwara A, Nakamura Y, Yasuike M, Ojima N, Masaoka T, Ozaki A, Kazeto Y, Gen K, et al. 2014. A ddRAD-based genetic map and its integration with the genome assembly of Japanese eel (*Anguilla japonica*) provides insights into genome evolution after the teleost-specific genome duplication. *BMC Genomics* 15:233.
- Kao F-T. 1993. Genes and genomes: Microdissection and microcloning of human chromosome regions in genome and genetic disease analysis. *BioEssays* 15:141–146.
- Kapitonov VV, Jurka J. 2008. A universal classification of eukaryotic transposable elements implemented in Repbase. *Nature Reviews Genetics*:2.
- Kasahara M, Naruse K, Sasaki S, Nakatani Y, Qu W, Ahsan B, Yamada T, Nagayasu Y, Doi K, Kasai Y, et al. 2007. The medaka draft genome and insights into vertebrate genome evolution. *Nature* 447:714–719.
- Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular biology and evolution* 30:772–780.
- Kaus H. 2005. The Antarctic Notothenioids as a model for macro-adaptation and molecular evolution.
- Kazazian HH. 2004. Mobile Elements: Drivers of Genome Evolution. *Science* 303:1626–1632.
- Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, et al. 2012. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28:1647–1649.
- Kehrer-Sawatzki H, Cooper DN. 2007. Understanding the recent evolution of the human genome: insights from human-chimpanzee genome comparisons. *Human Mutation* 28:99–130.
- Kellis M, Birren BW, Lander ES. 2004. Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*. *Nature* 428:617–624.
- Kennett J. 1986. Miocene Paleoceanography and Plankton Evolution. In: *Mesozoic and Cenozoic Oceans*. Vol. 15. Kenneth J. Hsü. p. 199–122. Available from: <https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/GD015p0119>
- Kennett J, Barker P. 1990. Latest Cretaceous to Cenozoic climate and oceanographic development in the Weddell Sea: an ocean drilling perspective. *Proc Ocean Drilling Project Sci Results* 113:937–960.



- Kennett JP. 1977. Cenozoic evolution of Antarctic glaciation, the circum-Antarctic Ocean, and their impact on global paleoceanography. *Journal of Geophysical Research* 82:3843–3860.
- Kennett JP, Keller G, Srinivasan MS. 1985. Miocene planktonic foraminiferal biogeography and paleoceanographic development of the Indo-Pacific region. In: *Geological Society of America Memoirs*. Vol. 163. Geological Society of America. p. 197–236. Available from: <https://pubs.geoscienceworld.org/books/book/172/chapter/3792458/>
- Kerk NM. 2003. Laser Capture Microdissection of Cells from Plant Tissues. *PLANT PHYSIOLOGY* 132:27–35.
- Kidwell MG. 2002. Transposable elements and the evolution of genome size in eukaryotes. *Genetica* 115:49–63.
- Kim JM, Vanguri S, Boeke JD, Gabriel A, Voytas DF. 1998. Transposable Elements and Genome Organization: A Comprehensive Survey of Retrotransposons Revealed by the Complete *Saccharomyces cerevisiae* Genome Sequence. *Genome Research* 8:464–478.
- Kiss AJ. 2004. Cold-stable eye lens crystallins of the Antarctic nototheniid toothfish *Dissostichus mawsoni* Norman. *Journal of Experimental Biology* 207:4633–4649.
- Kitano J, Ross JA, Mori S, Kume M, Jones FC, Chan YF, Absher DM, Grimwood J, Schmutz J, Myers RM, et al. 2009. A role for a neo-sex chromosome in stickleback speciation. *Nature* 461:1079–1083.
- Koga A, Higashide I, Hori H, Wakamatsu Y, Kyono-Hamaguchi Y, Hamaguchi S. 2007. The Toll element of medaka fish is transposed with only terminal regions and can deliver large DNA fragments into the chromosomes. *Journal of Human Genetics* 52:1026–1030.
- Kohn M, Hogel J, Vogel W, Minich P, Kehrerawatzki H, Graves J, Hameister H. 2006. Reconstruction of a 450-My-old ancestral vertebrate protokaryotype. *Trends in Genetics* 22:203–210.
- de Koning APJ, Gu W, Castoe TA, Batzer MA, Pollock DD. 2011. Repetitive Elements May Comprise Over Two-Thirds of the Human Genome. *Copenhaver GP, editor. PLoS Genetics* 7:e1002384.
- Konkel MK, Batzer MA. 2010. A mobile threat to genome stability: The impact of non-LTR retrotransposons upon the human genome. *Seminars in Cancer Biology* 20:211–221.
- Korbel JO, Urban AE, Affourtit JP, Godwin B, Grubert F, Simons JF, Kim PM, Palejev D, Carriero NJ, Du L, et al. 2007. Paired-End Mapping Reveals Extensive Structural Variation in the Human Genome. *Science* 318:420–426.
- Kraaijeveld K. 2010. Genome Size and Species Diversification. *Evolutionary Biology* 37:227–233.
- Kraaijeveld K, Zwanenburg B, Hubert B, Vieira C, De PATER S, Van ALPHEN JJM, Den DUNNEN JT, De KNIJFF P. 2012. Transposon proliferation in an asexual parasitoid: TRANSPOSON PROLIFERATION. *Molecular Ecology* 21:3898–3906.
- Kuhn KL, Near TJ. 2009. Phylogeny of *Trematomus* (Notothenioidei: Nototheniidae) inferred from mitochondrial and nuclear gene sequences. *Antarctic Science* 21:565.

- La Mesa M, Ascoli L, Caputo V. 2006. Gametogenesis and reproductive strategies in some species of the Antarctic fish genus *Trematomus* (Nototheniidae) from Terra Nova Bay, Ross Sea. *Polar Biology* 29:963–970.
- La Mesa M, Caputo V, Eastman JT. 2008. The reproductive biology of two epibenthic species of Antarctic nototheniid fish of the genus *Trematomus*. *Antarctic Science* [Internet] 20. Available from: [http://www.journals.cambridge.org/abstract\\_S095410200800103X](http://www.journals.cambridge.org/abstract_S095410200800103X)
- Ladevèze V, Chaminade N, Lemeunier F, Periquet G, Aulard S. 2012. General survey of hAT transposon superfamily with highlight on hobo element in *Drosophila*. *Genetica* 140:375–392.
- van de Lagemaat LN, Landry J-R, Mager DL, Medstrand P. 2003. Transposable elements in mammals promote regulatory variation and diversification of genes with specialized functions. *Trends in Genetics* 19:530–536.
- Laudicina A, Mühlmann M. 2015. Fish Chromosome Microdissection and its Utilization in the Synthesis of Fluorescence in situ Hybridization Probes. In: Ozouf-Costaz C, Pisano E, Foresti F, de Almeida L, editors. *Fish Cytogenetic Techniques*. CRC Press. p. 132–143. Available from: <http://www.crcnetbase.com/doi/10.1201/b18534-18>
- Lautrédou A-C, Bonillo C, Denys G, Cruaud C, Ozouf-Costaz C, Lecointre G, Dettai A. 2010. Molecular taxonomy and identification within the Antarctic genus *Trematomus* (Notothenioidei, Teleostei): How valuable is barcoding with COI? *Polar Science* 4:333–352.
- Lautrédou A-C, Hinsinger DD, Gallut C, Cheng C-HC, Berkani M, Ozouf-Costaz C, Cruaud C, Lecointre G, Dettai A. 2012. Phylogenetic footprints of an Antarctic radiation: The Trematominae (Notothenioidei, Teleostei). *Molecular Phylogenetics and Evolution* 65:87–101.
- Lecointre G, Améziane N, Boisselier M-C, Bonillo C, Busson F, Causse R, Chenuil A, Couloux A, Coutanceau J-P, Cruaud C, et al. 2013. Is the Species Flock Concept Operational? The Antarctic Shelf Case. Ropert-Coudert Y, editor. *PLoS ONE* 8:e68787.
- Lecointre G, Bonillo C, Ozouf-Costaz C. 1997. Molecular evidence for the origins of Antarctic fishes: paraphyly of the Bovichtidae and no indication for the monophyly of the Notothenioidei (Teleostei). *Polar Biology* 18:193–208.
- Levan A, Fredga K, Sandberg AA. 1964. Nomenclature for centromeric position on chromosomes. *Hereditas* 52:201–220.
- Levis RW, Ganesan R, Houtchens K, Tolar LA, Sheen F. 1993. Transposons in place of telomeric repeats at a *Drosophila* telomere. *Cell* 75:1083–1093.
- Liehr T ed. 2009. *Fluorescence in situ hybridization (FISH) - application guide*. Berlin: Springer
- Lim JK, Simmons MJ. 1994. Gross chromosome rearrangements mediated by transposable elements in *Drosophila melanogaster*. *Bioessays* 16:269–275.
- Lipnerova I, Bures P, Horova L, Smarda P. 2013. Evolution of genome size in *Carex* (Cyperaceae) in relation to chromosome number and genomic base composition. *Annals of Botany* 111:79–94.

- Lisch D. 2009. Epigenetic Regulation of Transposable Elements in Plants. *Annual Review of Plant Biology* 60:43–66.
- Livermore R, Nankivell A, Eagles G, Morris P. 2005. Paleogene opening of Drake Passage. *Earth and Planetary Science Letters* 236:459–470.
- Llorens C, Muñoz-Pomer A, Bernad L, Botella H, Moya A. 2009. Network dynamics of eukaryotic LTR retroelements beyond phylogenetic trees. *Biology Direct* 4:41.
- Logue LA, De Vries AL, Fodor E, Cossins AR. 2000. LIPID COMPOSITIONAL CORRELATES OF TEMPERATURE-ADAPTIVE INTERSPECIFIC DIFFERENCES IN MEMBRANE PHYSICAL STRUCTURE. *The Journal of Experimental Biology* 203:2105–2115.
- Lönnig W-E, Saedler H. 2002. Chromosome Rearrangements and Transposable Elements. *Annual Review of Genetics* 36:389–410.
- Machado TC, Pansonato-Alves JC, Pucci MB, Nogaroto V, Almeida MC, Oliveira C, Foresti F, Bertollo LA, Moreira-Filho O, Artoni RF. 2011. Chromosomal painting and ZW sex chromosomes differentiation in *Characidium* (Characiformes, Crenuchidae). *BMC genetics* 12:65.
- Mahé F. 2009. Phylogénie, éléments transposables et évolution de la taille des génomes chez les lupins. Available from: <https://tel.archives-ouvertes.fr/tel-00494607/>
- Mandáková T, Hloušková P, German DA, Lysak MA. 2017. Monophyletic Origin and Evolution of the Largest Crucifer Genomes. *Plant Physiology* 174:2062–2071.
- Mandáková T, Pouch M, Harmanová K, Zhan SH, Mayrose I, Lysak MA. 2017. Multispeed genome diploidization and diversification after an ancient allopolyploidization. *Molecular Ecology* 26:6445–6462.
- Mansour A. 2007. Epigenetic activation of genomic retrotransposons. *Journal of Cell and Molecular Biology* 6:99–107.
- Mardis ER. 2008. Next-Generation DNA Sequencing Methods. *Annual Review of Genomics and Human Genetics* 9:387–402.
- Marino IAM, Benazzo A, Agostini C, Mezzavilla M, Hoban SM, Patarnello T, Zane L, Bertorelle G. 2013. Evidence for past and present hybridization in three Antarctic icefish species provides new perspectives on an evolutionary radiation. *Molecular Ecology* 22:5148–5161.
- Martens K. 1997. Speciation in ancient lakes. *TREE* 12:177–182.
- Matschiner M, Hanel R, Salzburger W. 2011. On the Origin and Trigger of the Notothenioid Adaptive Radiation. Escrivá H, editor. *PLoS ONE* 6:1–9.
- Matsunaga S, Kawano S, Michimoto T, Higashiyama T, Nakao S, Sakai A, Kuroiwa T. 1999. Semi-automatic laser beam microdissection of the Y chromosome and analysis of Y chromosome DNA in a dioecious plant, *Silene latifolia*. *Plant and Cell Physiology* 40:60–68.
- Mazzei F, Ghigliotti L, Coutanceau J-P, Detrich HW, Prirodina V, Ozouf-Costaz C, Pisano E. 2008. Chromosomal characteristics of the temperate notothenioid fish *Eleginops maclovinus* (Cuvier). *Polar Biology* 31:629–634.

- Mazzei F, Ghigliotti L, Lecointre G, Ozouf-Costaz C, Coutanceau J-P, Detrich W, Pisano E. 2006. Karyotypes of basal lineages in notothenioid fishes: the genus *Bovichtus*. *Polar Biology* 29:1071–1076.
- Mazzuchelli J, Martins C. 2009. Genomic organization of repetitive DNAs in the cichlid fish *Astronotus ocellatus*. *Genetica* 136:461–469.
- McClintock B. 1948. Mutable loci on Maize. Carnegie institution of Washington:157–167.
- McClintock B. 1950. The origin and behavior of mutable loci in maize. *Proceedings of the National Academy of Sciences* 36:344–355.
- McClintock B. 1984. The Significance of Responses of the Genome to Challenge. *Science* 226:792–801.
- McCluskey BM, Postlethwait JH. 2015. Phylogeny of Zebrafish, a “Model Species,” within *Danio*, a “Model Genus.” *Molecular Biology and Evolution* 32:635–652.
- Medrano L, Bernardi Giacomo, Couturier J, Dutrillaux B, Bernardi Giorgio. 1988. Chromosome banding and genome compartmentalization in fishes. *Chromosoma* 96:178–183.
- Mendiola MV, Bernales I, de la Cruz F. 1994. Differential roles of the transposon termini in IS91 transposition. *Proceedings of the National Academy of Sciences* 91:1922–1926.
- Miller WJ, Capy P. 2004. *Mobile Genetic Elements as Natural Tools for Genome Evolution*. Humana Press. Totowa, NJ.: Miller W.J, Capy P.
- Milne I, Wright F, Rowe G, Marshall DF, Husmeier D, McGuire G. 2004. TOPALi: software for automatic identification of recombinant sequences within DNA multiple alignments. *Bioinformatics* 20:1806–1807.
- Mirarab S, Reaz R, Bayzid MS, Zimmermann T, Swenson MS, Warnow T. 2014. ASTRAL: genome-scale coalescent-based species tree estimation. *Bioinformatics* 30:i541–i548.
- Mirarab S, Warnow T. 2015. ASTRAL-II: coalescent-based species tree estimation with many hundreds of taxa and thousands of genes. *Bioinformatics* 31:i44–i52.
- Miyaki K, Tabeta O, Kayano H. 1995. Karyotypes in Six Species of Pufferfishes Genus *Takifugu* (Tetraodontidae, Tetraodontiformes). *Fisheries Science* 61:594–598.
- Molina WF. 2006. Chromosomal changes and stasis in marine fish groups. In: *Fish cytogenetics*. Science Publishers. Enfield: Pisano E, Ozouf-Costaz C, Forest F and Kapoor BG. p. 69–110.
- Molina WF, de Freitas Bacurau TO. 2006. Structural and numerical chromosome diversification in marine Perciformes (Priacanthidae and Gerreidae). *Cytologia* 71:237–242.
- Monard SP. 1998. Chromosome Sorting and Analysis by FACS. In: *Flow Cytometry Protocols*. Vol. 91. Humana Press. Totowa, NJ.: Jaroszeski M.J., Heller R. p. 239–254. Available from: <https://doi.org/10.1385/0896033546>
- Montgomery JC, Foster BA, Milton RC, Carr E. 1993. Spatial and temporal variations in the diet of nototheniid fish in McMurdo Sound, Antarctica. *Polar Biology* 13:429–431.

- Morescalchi A, Morescalchi MA, Odierna G, Sitingo V, Capriglione T. 1996. Karyotype and genome size of zoarcids and notothenioids (Teleostei, Perciformes) from the Ross Sea: cytotaxonomic implications. *Polar Biology* 16:559–564.
- Morescalchi A, Pisano E, Stanyon R, Morescalchi MA. 1992. Cytotaxonomy of antarctic teleosts of the Pagothenia/Trematomus complex (Nototheniidae, Perciformes). *Polar Biology* 12:553–558.
- Muleris M, Dutrillaux B, Chauvier G. 1983. Mise en évidence d'une fission centromérique hétérozygote chez un mâle Theropithecus gelada et comparaison chromosomique avec les autres Papioninae. *Genetics Selection Evolution* 15:177–184.
- Müller S, Stanyon R, O'Brien PCM, Ferguson-Smith MA, Plesker R, Wienberg J. 1999. Defining the ancestral karyotype of all primates by multidirectional chromosome painting between tree shrews, lemurs and humans. *Chromosoma* 108:393–400.
- Muotri AR, Marchetto MCN, Coufal NG, Gage FH. 2007. The necessary junk: new functions for transposable elements. *Human Molecular Genetics* 16:R159–R167.
- Murphy KR. 2015. Investigating the larval/juvenile Notothenioid fish species assemblage in McMurdo sound, Antarctica.
- Nadeau NJ, Martin SH, Kozak KM, Salazar C, Dasmahapatra KK, Davey JW, Baxter SW, Blaxter ML, Mallet J, Jiggins CD. 2013. Genome-wide patterns of divergence and gene flow across a butterfly radiation. *Molecular Ecology* 22:814–826.
- Nakatani Y, Takeda H, Kohara Y, Morishita S. 2007. Reconstruction of the vertebrate ancestral genome reveals dynamic genome reorganization in early vertebrates. *Genome Research* 17:1254–1265.
- Nanda I, Benisch P, Fetting D, Haaf T, Schmid M. 2011. Synteny Conservation of Chicken Macrochromosomes 1–10 in Different Avian Lineages Revealed by Cross-Species Chromosome Painting. *Cytogenetic and Genome Research* 132:165–181.
- Naruse K, Hori H, Shimizu N, Kohara Y, Takeda H. 2004. Medaka genomics: a bridge between mutant phenotype and gene function. *Mechanisms of Development* 121:619–628.
- Navarro A, Barton NH. 2003. Chromosomal Speciation and Molecular Divergence: Accelerated Evolution in Rearranged Chromosomes. *Science* 300:321–324.
- Near TJ. 2004. Estimating divergence times of notothenioid fishes using a fossil-calibrated molecular clock. *Antarctic Science* 16:37–44.
- Near TJ, Cheng C-HC. 2008. Phylogenetics of notothenioid fishes (Teleostei: Acanthomorpha): Inferences from mitochondrial and nuclear gene sequences. *Molecular Phylogenetics and Evolution* 47:832–840.
- Near TJ, Dornburg A, Harrington RC, Oliveira C, Pietsch TW, Thacker CE, Satoh TP, Katayama E, Wainwright PC, Eastman JT, et al. 2015. Identification of the notothenioid sister lineage illuminates the biogeographic history of an Antarctic adaptive radiation. *BMC Evolutionary Biology* [Internet] 15. Available from: <http://www.biomedcentral.com/1471-2148/15/109>

- Near TJ, Dornburg A, Kuhn KL, Eastman JT, Pennington JN, Patarnello T, Zane L, Fernandez DA, Jones CD. 2012. Ancient climate change, antifreeze, and the evolutionary diversification of Antarctic fishes. *Proceedings of the National Academy of Sciences* 109:3434–3439.
- Near TJ, Kendrick BJ, William Detrich H, Jones CD. 2007. Confirmation of neutral buoyancy in *Aethotaxis mitopteryx* DeWitt (Notothenioidei: Nototheniidae). *Polar Biology* 30:443–447.
- Noor MAF, Feder JL. 2006. Speciation genetics: evolving approaches. *Nature Reviews Genetics* 7:851–861.
- Nygren A, Edlund P, Hirsch U, \AAhsgren L. 1968. CYTOLOGICAL STUDIES IN PERCH (*PERCA FLUVIATILIS* L.), PIKE (*ESOX LUCIUS* L.), PIKE-PERCH (*LUCIOPERCA LUCIOPERCA* L.), AND RUFF (*ACERINA CERNUA* L.). *Hereditas* 59:518–524.
- Obbard DJ, Gordon KH, Buck AH, Jiggins FM. 2009. The evolution of RNAi as a defence against viruses and transposable elements. *Philosophical Transactions of the Royal Society B: Biological Sciences* 364:99–115.
- Ocalewicz K. 2005. Cytogenetic analysis of platyfish (*Xiphophorus maculatus*) shows location of major and minor rDNA on chromosomes: Cytogenetic analysis of platyfish. *Hereditas* 141:333–337.
- Ohno S. 1974. *Evolution by gene duplication*. Springer-Verlag. New-york: Ohno S Available from: <http://doi.wiley.com/10.1002/tera.1420090224>
- Oliver KR, Greene WK. 2009. Transposable elements: powerful facilitators of evolution. *BioEssays* 31:703–714.
- Orsoly K, Virág K, Anita S, Zsuzsa E, Ágota A, György V, Lajos M, Zsuzsanna I, Zoltán I, Balázs S, et al. 2011. Reliable transgene-independent method for determining Sleeping Beauty transposon copy numbers. Available from: <https://dea.lib.unideb.hu/dea/handle/2437/110559>
- Ozouf-Costaz C, Brandt J, KöRting C, Pisano E, Bonillo C, Coutanceau J-P, Volff J-N. 2004. Genome dynamics and chromosomal localization of the non-LTR retrotransposons Rex1 and Rex3 in Antarctic fish. *Antarctic Science* 16:51–57.
- Ozouf-Costaz C, Hureau J-C, Beaunier M. 1991. Chromosome studies on fish of the suborder Notothenioidei collected in the Weddell sea during EPOS 3 cruise. *Cybium* 15:271–289.
- Ozouf-Costaz C, Pisano E, Foresti F, de Almeida LF. 2015. *Fish cytogenetic techniques: Ray-Fin fishes and chondrichthyans*. CRC Press Available from: <http://www.crcnetbase.com/doi/pdf/10.1201/b18534-1>
- Ozouf-Costaz C, Pisano E, Thaeron C, Hureau J-C. 1997. Antarctic fish chromosome banding: significance for evolutionary studies. *Cybium* 21:399–409.
- Pace JK, Feschotte C. 2007. The evolutionary history of human DNA transposons: Evidence for intense activity in the primate lineage. *Genome Research* 17:422–432.
- Pansonato-Alves JC, Vicari MR, Oliveira C, Foresti F. 2011. Chromosomal diversification in populations of *Characidium cf. gomesi* (Teleostei, Crenuchidae). *Journal of Fish Biology* 78:183–194.

- Parisod C, Alix K, Just J, Petit M, Sarilar V, Mhiri C, Ainouche M, Chalhoub B, Grandbastien M-A. 2010. Impact of transposable elements on the organization and function of allopolyploid genomes: Research review. *New Phytologist* 186:37–45.
- Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE. 2012. Double Digest RADseq: An Inexpensive Method for De Novo SNP Discovery and Genotyping in Model and Non-Model Species. Orlando L, editor. *PLoS ONE* 7:e37135.
- Pfender WF, Saha MC, Johnson EA, Slabaugh MB. 2011. Mapping with RAD (restriction-site associated DNA) markers to rapidly identify QTL for stem rust resistance in *Lolium perenne*. *Theoretical and Applied Genetics* 122:1467–1480.
- Pialek J, Hauffe HC, Searle JB. 2005. Chromosomal variation in the house mouse: CHROMOSOMAL VARIATION IN THE HOUSE MOUSE. *Biological Journal of the Linnean Society* 84:535–563.
- Picone B, Masters J, Silvestro D, Sineo L, DelPero M. 2011. A Phylogenetic Analysis of Human Syntenies Revealed by Chromosome Painting in Euarchontoglires Orders. *Journal of Mammalian Evolution* 18:131–146.
- Piednoël M, Bonnivard E. 2009. DIRS1-like retrotransposons are widely distributed among Decapoda and are particularly present in hydrothermal vent organisms. *BMC Evolutionary Biology* 9:86.
- Piednoël M, Donnart T, Esnault C, Graça P, Higuete D, Bonnivard E. 2013. LTR-Retrotransposons in *R. exoculata* and Other Crustaceans: The Outstanding Success of GalEa-Like Copia Elements. Kashkush K, editor. *PLoS ONE* 8:e57675.
- Piednoël M, Gonçalves IR, Higuete D, Bonnivard E. 2011. Eukaryote DIRS1-like retrotransposons: an overview. *BMC genomics* 12:1.
- Pimpinelli S, Berloco M, Fanti L, Dimitri P, Bonaccorsi S, Marchetti E, Caizzi R, Caggese C, Gatti M. 1995. Transposable elements are stable structural components of *Drosophila melanogaster* heterochromatin. *Proceedings of the National Academy of Sciences* 92:3804–3808.
- Pinton A, Ducos A, Yerle M. 2003. Chromosomal rearrangements in cattle and pigs revealed by chromosome microdissection and chromosome painting. *Genetics Selection Evolution* 35:685–696.
- Pisano E, Angelini C, Mazzei F, Stanyon R. 2000. Adaptive radiation in Antarctic notothenioid fish: Studies of genomic change at chromosomal level. *Italian Journal of Zoology* 67:115–121.
- Pisano E, Cocca E, Mazzei F, Ghigliotti L, Di Prisco G, Detrich III HW, Ozouf-Costaz C. 2003. Mapping of  $\alpha$ - and  $\beta$ -globin genes on Antarctic fish chromosomes by fluorescence in-situ hybridization. *Chromosome Research* 11:633–640.
- Pisano E, Ozouf-Costaz C. 2000. Chromosome change and the evolution in the Antarctic fish suborder Notothenioidei. *Antarctic Science* [Internet] 12. Available from: [http://www.journals.cambridge.org/abstract\\_S0954102000000390](http://www.journals.cambridge.org/abstract_S0954102000000390)
- Pisano E, Ozouf-Costaz C, Hureau JC, Williams R. 1995. Chromosome differentiation in the subantarctic Bovichtidae species *Cottoperca gobio* (Günther, 1861) and *Pseudaphritis urvillii* (Valenciennes, 1832) (Pisces, Perciformes). *Antarctic Science* 7:381–386.

- Platt RN, Blanco-Berdugo L, Ray DA. 2016. Accurate Transposable Element Annotation Is Vital When Analyzing New Genome Assemblies. *Genome Biology and Evolution* 8:403–410.
- Pokorná M, Giovannotti M, Kratochvíl L, Caputo V, Olmo E, Ferguson-Smith MA, Rens W. 2012. Conservation of chromosomes syntenic with avian autosomes in squamate reptiles revealed by comparative chromosome painting. *Chromosoma* 121:409–418.
- Pontecorvo G, De Felice B, Carfagna M. 2000. A novel repeated sequence DNA originated from a Tc1-like transposon in water green frog *Rana esculenta*. *Gene* 261:205–210.
- Poulter RT, Butler MI. 2015. Tyrosine Recombinase Retrotransposons and Transposons. *Microbiology Spectrum* 3.
- Poulter RT, Goodwin T. 2005. DIRS-1 and the other tyrosine recombinase retrotransposons. *Cytogenetic and Genome Research* 110:575–588.
- Pringle EG, Baxter SW, Webster CL, Papanicolaou A, Lee SF, Jiggins CD. 2007. Synteny and Chromosome Evolution in the Lepidoptera: Evidence From Mapping in *Heliconius melpomene*. *Genetics* 177:417–426.
- di Prisco G, Maresca B, Tota B. 1991. *Biology of Antarctic Fish*. Springer. Berlin, Heidelberg  
Available from: <https://doi.org/10.1007/978-3-642-76217-8>
- Pukk L, Ahmad F, Hasan S, Kisand V, Gross R, Vasemägi A. 2015. Less is more: extreme genome complexity reduction with ddRAD using Ion Torrent semiconductor technology. *Molecular Ecology Resources* 15:1145–1152.
- Ráb P, Rábová M, Pereira CS, Collares-Pereira MJ, Pelikánová Š. 2008. Chromosome studies of European cyprinid fishes: interspecific homology of leuciscine cytotaxonomic marker—the largest subtelocentric chromosome pair as revealed by cross-species painting. *Chromosome Research* 16:863–873.
- Radice AD, Bugaj B, Fitch DH, Emmons SW. 1994. Widespread occurrence of the Tc1 transposon family: Tc1-like transposons from teleost fish. *Molecular and General Genetics MGG* 244:606–612.
- Rajeev L, Malanowska K, Gardner JF. 2009. Challenging a Paradigm: the Role of DNA Homology in Tyrosine Recombinase Reactions. *Microbiology and Molecular Biology Reviews* 73:300–309.
- Raskina O, Barber JC, Nevo E, Belyayev A. 2008. Repetitive DNA and chromosomal rearrangements: speciation-related events in plant genomes. *Cytogenetic and Genome Research* 120:351–357.
- Rebollo R, Horard B, Hubert B, Vieira C. 2010. Jumping genes and epigenetics: Towards new species. *Gene* 454:1–7.
- Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, Fiegler H, Shapero MH, Carson AR, Chen W, et al. 2006. Global variation in copy number in the human genome. *Nature* 444:444–454.
- Rey O, Danchin E, Mirouze M, Loot C, Blanchet S. 2016. Adaptation to Global Change: A Transposable Element–Epigenetics Perspective. *Trends in Ecology & Evolution* 31:514–526.



- Rey O, d'Hont A, Coutanceau J, Pisano E, Chilmonczyk S, Ozouf-Costaz C. 2015. Cephalic Kidney and Spleen Cell Culture in Antarctic Teleosts. In: Ozouf-Costaz Catherine, Pisano Eva, Foresti F, de Almeida L, editors. *Fish Cytogenetic Techniques*. CRC Press. p. 74–81. Available from: <http://www.crcnetbase.com/doi/10.1201/b18534-12>
- Rho M, Schaack S, Gao X, Kim S, Lynch M, Tang H. 2010. Research article LTR retroelements in the genome of *Daphnia pulex*. Available from: <http://ai2-s2-pdfs.s3.amazonaws.com/4293/39a6ca462df8d91dc9405dd0cd236568f6ba.pdf>
- Ried T. 1998. Chromosome painting: a useful art. *Human Molecular Genetics* 7:1619–1626.
- Rieseberg LH. 2001. Chromosomal rearrangements and speciation. *Trends in Ecology & Evolution* 16:351–358.
- Ritchie PA, Bargelloni L, Meyer A, Taylor JA, Macdonald JA, Lambert DM. 1996. Mitochondrial Phylogeny of Trematomid Fishes (Nototheniidae, Perciformes) and the Evolution of Antarctic Fish. *Molecular Phylogenetics and Evolution* 5:383–390.
- Ritchie PA, Lavoué S, Lecointre G. 1997. Molecular phylogenetics and the evolution of Antarctic notothenioid fishes. *Comparative Biochemistry and Physiology Part A: Physiology* 118:1009–1025.
- Röhme D, Fox H, Herrmann B, Frischauf A-M, Edström J-E, Mains P, Silver LM, Lehrach H. 1984. Molecular clones of the mouse t complex derived from microdissected metaphase chromosomes. *Cell* 36:783–788.
- Romanenko SA, Volobouev VT, Perelman PL, Lebedev VS, Serdukova NA, Trifonov VA, Biltueva LS, Nie W, O'Brien PCM, Bulatova NS, et al. 2007. Karyotype evolution and phylogenetic relationships of hamsters (Cricetidae, Muroidea, Rodentia) inferred from chromosomal painting and banding comparison. *Chromosome Research* [Internet]. Available from: <http://link.springer.com/10.1007/s10577-007-1124-3>
- Ross JA, Urton JR, Boland J, Shapiro MD, Peichel CL. 2009. Turnover of Sex Chromosomes in the Stickleback Fishes (Gasterosteidae). Mank JE, editor. *PLoS Genetics* 5:e1000391.
- Rubin BER, Ree RH, Moreau CS. 2012. Inferring Phylogenies from RAD Sequence Data. Kolokotronis S-O, editor. *PLoS ONE* 7:e33394.
- Rutschmann S, Matschiner M, Damerau M, Muschick M, Lehmann MF, Hanel R, Salzburger W. 2011. Parallel ecological diversification in Antarctic notothenioid fishes as evidence for adaptive radiation: ECOLOGICAL DIVERSIFICATION IN NOTOTHENIOIDS. *Molecular Ecology* 20:4707–4721.
- Sadofsky MJ. 2001. The RAG proteins in V(D)J recombination: more than just a nuclease. *Nucleic Acids Research* 29:1399–1409.
- Šafář J, Bartoš J, Janda J, Bellec A, Kubaláková M, Valárik M, Pateyron S, Weiserová J, Tušková R, Číhalíková J, et al. 2004. Dissecting large and complex genomes: flow sorting and BAC cloning of individual chromosomes from bread wheat. *The Plant Journal* 39:960–968.
- Sanchez S, Dettai A, Bonillo C, Ozouf-Costaz C, Detrich HW, Lecointre G. 2007. Molecular and morphological phylogenies of the Antarctic teleostean family Nototheniidae, with emphasis on the Trematominae. *Polar Biology* 30:155–166.

- Sánchez-Gracia A, Maside X, Charlesworth B. 2005. High rate of horizontal transfer of transposable elements in *Drosophila*. *Trends in Genetics* 21:200–203.
- Sanmiguel P. 1998. Evidence that a Recent Increase in Maize Genome Size was Caused by the Massive Amplification of Intergene Retrotransposons. *Annals of Botany* 82:37–44.
- Santini F, Harmon LJ, Carnevale G, Alfaro ME. 2009. Did genome duplication drive the origin of teleosts? A comparative study of diversification in ray-finned fishes. *BMC Evolutionary Biology* 9:194.
- Scalenghe F, Turco E, Edström JE, Pirrotta V, Melli M. 1981. Microdissection and cloning of DNA from a specific region of *Drosophila melanogaster* polytene chromosomes. *Chromosoma* 82:205–216.
- Scannell DR, Butler G, Wolfe KH. 2007. Yeast genome evolution—the origin of the species. *Yeast* 24:929–942.
- Schaack S, Gilbert C, Feschotte C. 2010. Promiscuous DNA: horizontal transfer of transposable elements and why it matters for eukaryotic evolution. *Trends in Ecology & Evolution* 25:537–546.
- Scher HD, Martin EE. 2006. Timing and Climatic Consequences of the Opening of Drake Passage. *Science* 312:428–430.
- Scherthan H, Cremer T, Arnason U, Weier H-U, Lima-de-Faria A, Fronicke L. 1994. Comparative chromosome painting discloses homologous segments in distantly related mammals. *Nature Genetics* 6:342–347.
- Schluter D. 2000. Ecological Character Displacement in Adaptive Radiation. *The American Naturalist* 156:S4–S16.
- Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S, Liang C, Zhang J, Fulton L, Graves TA, et al. 2009. The B73 Maize Genome: Complexity, Diversity, and Dynamics. *Science* 326:1112–1115.
- Schneider C, Gross M, Terencio M, do Carmo E, Martins C, Feldberg E. 2013. Evolutionary dynamics of retrotransposable elements Rex1, Rex3 and Rex6 in neotropical cichlid genomes. *BMC Evolutionary Biology* 13:152.
- Schubert I, Schriever-Schwemmer G, Werner T, Adler ID. 1992. Telomeric signals in Robertsonian fusion and fission chromosomes: implications for the origin of pseudoaneuploidy. *Cytogenetics and Cell Genetics* 59:6–9.
- Schwartz D, Cantor C. 1984. Separation of yeast chromosome-sized DNAs by pulsed field gradient gel electrophoresis. *Cell* 37:67–75.
- Seifertova E, Zimmerman LB, Gilchrist MJ, Macha J, Kubickova S, Cernohorska H, Zarsky V, Owens ND, Sesay AK, Tlapakova T, et al. 2013. Efficient high-throughput sequencing of a laser microdissected chromosome arm. *BMC Genomics* 14:357.
- Sela N, Kim E, Ast G. 2010. The role of transposable elements in the evolution of non-mammalian vertebrates and invertebrates. *Genome biology* 11:13.

- Shaw DD, Coates DJ, Wilkinson P. 1986. Estimating the genic and chromosomal components of reproductive isolation within and between subspecies of the grasshopper *Caledia captiva*. *Canadian Journal of Genetics and Cytology* 28:686–695.
- Shen C-H, Steiner LA. 2004. Genome Structure and Thymic Expression of an Endogenous Retrovirus in Zebrafish. *Journal of Virology* 78:899–911.
- Shin SC, Ahn DH, Kim SJ, Pyo CW, Lee H, Kim M-K, Lee J, Lee JE, Detrich HW, Postlethwait JH. 2014. The genome sequence of the Antarctic bullhead notothen reveals evolutionary adaptations to a cold environment. *Genome biology* 15:468.
- Sineo L. 1990. The Banded Karyotype of *Cercopithecus mitis maesi* Compared with the Karyotypes of *C. albogularis samango* and *C. nictitans stampflii*. *International Journal of Primatology* 11:12.
- Slijepcevic P. 1998. Telomeres and mechanisms of Robertsonian fusion. *Chromosoma* 107:136–140.
- Slotkin RK, Martienssen R. 2007. Transposable elements and the epigenetic regulation of the genome. *Nature Reviews Genetics* 8:272–285.
- Smale D, Barnes D, Fraser K, Peck L. 2008. Benthic community response to iceberg scouring at an intensely disturbed shallow water site at Adelaide Island, Antarctica. *Marine Ecology Progress Series* 355:85–94.
- Sola L, Gornung E. 2001. Classical and molecular cytogenetics of the zebrafish, *Danio rerio* (Cyprinidae, Cypriniformes): an overview. *Genetica* 111:397–412.
- Sorek R. 2007. The birth of new exons: Mechanisms and evolutionary consequences. *RNA* 13:1603–1608.
- Sotero-Caio CG, Platt RN, Suh A, Ray DA. 2017. Evolution and Diversity of Transposable Elements in Vertebrate Genomes. *Genome Biology and Evolution* 9:161–177.
- Splendore de Borba R, Lourenço da Silva E, Parise-Maltempi PP. 2013. Chromosome mapping of retrotransposable elements *Rex1* and *Rex3* in *Leporinus* Spix, 1829 species (Characiformes: Anostomidae) and its relationships among heterochromatic segments and W sex chromosome. *Mobile Genetic Elements* 3:e27460.
- Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30:1312–1313.
- Stankovic A, Spalik K, Kmalar E, Borsuk P, Weglenski P. 2002. Recent origin of sub-Antarctic notothenioids. *Polar Bio* 25:203–205.
- Stanyon R, Stone G. 2008. Phylogenomic Analysis by Chromosome Sorting and Painting. *Methods Molecular Biology* 422:13–29.
- Stein, Ponelies, Musket, McMullen, Weber. 1998. Chromosome micro-dissection and region-specific libraries from pachytene chromosomes of maize (*Zea mays* L.). *The Plant Journal* 13:281–289.
- Steward N, Ito M, Yamaguchi Y, Koizumi N, Sano H. 2002. Periodic DNA Methylation in Maize Nucleosomes and Demethylation by Environmental Stress. *Journal of Biological Chemistry* 277:37741–37746.

- Stock AD, Mengden G. 1975. Chromosome banding pattern conservatism in birds and nonhomology of chromosome banding patterns between birds, turtles, snakes and amphibians. *Chromosoma* [Internet] 50. Available from: <http://link.springer.com/10.1007/BF00284963>
- Stott LD, Kennett JP, Shackleton NJ, Corfield RM. 1990. 48. THE EVOLUTION OF ANTARCTIC SURFACE WATERS DURING THE PALEOGENE: INFERENCES FROM THE STABLE ISOTOPIC COMPOSITION OF PLANKTONIC FORAMINIFERS, ODP LEG 1131. *Proc Ocean Drilling Project Sci Results* 113:849–863.
- Swennenhuis JF, Foulk B, Coumans FAW, Terstappen LWMM. 2012. Construction of repeat-free fluorescence in situ hybridization probes. *Nucleic Acids Research* 40:e20–e20.
- Symonová R, Majtánová Z, Sember A, Staaks GB, Bohlen J, Freyhof J, Rábová M, Ráb P. 2013. Genome differentiation in a species pair of coregonine fishes: an extremely rapid speciation driven by stress-activated retrotransposons mediating extensive ribosomal DNA multiplications. *BMC Evolutionary Biology* 13:42.
- Taillebourg E, Dura J-M. 1999. A novel mechanism for P element homing in *Drosophila*. *Proceedings of the National Academy of Sciences* 96:6856–6861.
- Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. 2011. MEGA5: Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Molecular Biology and Evolution* 28:2731–2739.
- Taylor JS. 2003. Genome Duplication, a Trait Shared by 22,000 Species of Ray-Finned Fish. *Genome Research* 13:382–390.
- Templeton AR. 1981. Mechanisms of Speciation - A Population Genetic Approach. *Annual Review of Ecology and Systematics* 12:23–48.
- Teng L, Firpi HA, Tan K. 2011. Enhancers in embryonic stem cells are enriched for transposable elements and genetic variations associated with cancers. *Nucleic Acids Research* 39:7371–7379.
- Terrat Y, Bonnivard E, Higuete D. 2008. GalEa retrotransposons from galatheid squat lobsters (Decapoda, Anomura) define a new clade of Ty1/copia-like elements restricted to aquatic species. *Molecular Genetics and Genomics* 279:63–73.
- Teruel M, Cabrero J, Montiel EE, Acosta MJ, Sánchez A, Camacho JPM. 2009. Microdissection and chromosome painting of X and B chromosomes in *Locusta migratoria*. *Chromosome Research* 17:11–18.
- Thatje S, Hillenbrand C-D, Larter R. 2005. On the origin of Antarctic marine benthic community structure. *Trends in Ecology & Evolution* 20:534–540.
- The 1000 Genomes Project Consortium. 2012. An integrated map of genetic variation from 1,092 human genomes. *Nature* 491:56–65.
- The Arabidopsis Genome Initiative. 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408:796–816.
- The *C. elegans* Sequencing Consortium. 1998. Genome Sequence of the Nematode *C. elegans*: A Platform for Investigating Biology. *Science* 282:2012–2018.

- The Chimpanzee Sequencing and Analysis Consortium. 2005. Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature* 437:69–87.
- Tomaszkiewicz M, HauTecoœur M, Coutanceau JP, Bonillo C, Dettai A, Mazzei F, Ghigliotti L, Pisano E, Couloux A, Chanet B, et al. 2011. Comparative cytogenetic studies of the Nototheniidae (Teleostei: Acanthomorpha) from the Indian (Kerguelen-Heard Plateau) and Atlantic (South Georgia, South Sandwich, Falkland/Malvinas, Bouvet Islands) sectors of the Southern Ocean. *Cytogenetics of nototheniids from the Indian and Atlantic Southern Ocean*:109–121.
- Trifonov VA, Perelman PL, Kawada S-I, Iwasa MA, Oda S-I, Graphodatsky AS. 2002. Complex structure of B-chromosomes in two mammalian species: *Apodemus peninsulae* (Rodentia) and *Nyctereutes procyonoides* (Carnivora). *Chromosome Research* 10:109–116.
- Trifonov VA, Vorobieva NV, Serdyukova NA, Rens W. 2017. FISH with and Without COT1 DNA. In: Liehr T, editor. *Fluorescence In Situ Hybridization (FISH)*. Berlin, Heidelberg: Springer Berlin Heidelberg. p. 123–133. Available from: [http://link.springer.com/10.1007/978-3-662-52959-1\\_11](http://link.springer.com/10.1007/978-3-662-52959-1_11)
- Turcotte K, Srinivasan S, Bureau T. 2008. Survey of transposable elements from rice genomic sequences: Transposable elements in rice. *The Plant Journal* 25:169–179.
- Ulsh BA, Mühlmann-Díaz MC, Whicker FW, Hinton TG, Congdon JD, S. Bedford J. 2000. Chromosome Translocations in Turtles: A Biomarker in a Sentinel Animal for Ecological Dosimetry. *Radiation Research* 153:752–759.
- Urton JR, McCann SR, Peichel CL. 2011. Karyotype Differentiation between Two Stickleback Species (Gasterosteidae). *Cytogenetic and Genome Research* 135:150–159.
- Valente GT, Mazzuchelli J, Ferreira IA, Poletto AB, Fantinatti BEA, Martins C. 2011. Cytogenetic Mapping of the Retroelements Rex1, Rex3 and Rex6 among Cichlid Fish: New Insights on the Chromosomal Distribution of Transposable Elements. *Cytogenetic and Genome Research* 133:34–42.
- Van Valen L. 1973. A new evolutionary law.
- Ventura K, O'Brien PCM, Yonenaga-Yassuda Y, Ferguson-Smith MA. 2009. Chromosome homologies of the highly rearranged karyotypes of four Akodon species (Rodentia, Cricetidae) resolved by reciprocal chromosome painting: the evolution of the lowest diploid number in rodents. *Chromosome Research* 17:1063–1078.
- Vicari M, Pansonato-Alves J, Foresti F. 2015. Microdissection and Chromosome Painting. In: Ozouf-Costaz C, Pisano E, Foresti Fausto, de Almeida L, editors. *Fish Cytogenetic Techniques*. CRC Press. p. 144–165. Available from: <http://www.crcnetbase.com/doi/10.1201/b18534-19>
- Vicari MR, de Mello Pistune HF, Castro JP, de Almeida MC, Bertollo LAC, Moreira-Filho O, Camacho JPM, Artoni RF. 2011. New insights on the origin of B chromosomes in *Astyanax scabripinnis* obtained by chromosome painting and FISH. *Genetica* 139:1073–1081.
- Viersbach R, Schwanitz G, Nöthen MM. 1994. Delineation of marker chromosomes by reverse chromosome painting using only a small number of DOP-PCR amplified microdissected chromosomes. *Human genetics* 93:663–667.

- Volff J-N. 2005. Genome evolution and biodiversity in teleost fish. *Heredity* 94:280–294.
- Volff J-N, Bouneau L, Ozouf-Costaz C, Fischer C. 2003. Diversity of retrotransposable elements in compact pufferfish genomes. *Trends in Genetics* 19:674–678.
- Volff J-N, Körting C, Meyer A, Schartl M. 2001. Evolution and Discontinuous Distribution of Rex3 Retrotransposons in Fish. *Molecular Biology and Evolution* 18:427–431.
- Voltolin TA, Laudicina A, Senhorini JA, Bortolozzi J, Oliveira C, Foresti F, Porto-Foresti F. 2010. Origin and molecular organization of supernumerary chromosomes of *Prochilodus lineatus* (Characiformes, Prochilodontidae) obtained by DNA probes. *Genetica* 138:1133–1139.
- Voskoboynikova O. 1993. Evolution of the Visceral Skeleton and Phylogeny of Nototheniidae. *Journal of Ichthyology* 33:23–47.
- Wagner CE, Keller I, Wittwer S, Selz OM, Mwaiko S, Greuter L, Sivasundar A, Seehausen O. 2013. Genome-wide RAD sequence data provide unprecedented resolution of species boundaries and relationships in the Lake Victoria cichlid adaptive radiation. *Molecular Ecology* 22:787–798.
- Wang H, Liu J-S. 2008. LTR retrotransposon landscape in *Medicago truncatula*: more rapid removal than in rice. *BMC Genomics* 9:382.
- Warren IA, Naville M, Chalopin D, Levin P, Berger CS, Galiana D, Volff J-N. 2015. Evolutionary impact of transposable elements on genomic diversity and lineage-specific innovation in vertebrates. *Chromosome Research* 23:505–531.
- Watanabe K, Koga H, Nakamura K, Fujita A, Hattori A, Matsuda M, Koga A. 2014. Spontaneous germline excision of Tol1, a DNA-based transposable element naturally occurring in the medaka fish genome. Danzmann R, editor. *Genome* 57:193–199.
- Weimer J, Kiechle M, Senger G, Wiedemann U, Ovens-Raeder A, Schuierer S, Kautza M, Siebert R, Arnold N. 1999. An easy and reliable procedure of microdissection technique for the analysis of chromosomal breakpoints and marker chromosomes. *Chromosome Research* 7:355–362.
- Wessler S. 1995. LTR-retrotransposons and MITEs: important players in the evolution of plant genomes. *Current Opinion in Genetics & Development* 5:814–821.
- White MJD. 1970. Cytogenetics of speciation. *Australian Journal of Entomology* 9:1–6.
- White MJD. 1978. Chain Processes in Chromosomal Speciation. *Systematic Zoology* 27:285.
- White MJD. 1980. Modes of speciation in orthopteroid insects. *Bolletino di zoologia* 47:83–94.
- Wicker T, Keller B. 2007. Genome-wide comparative analysis of copia retrotransposons in Triticeae, rice, and *Arabidopsis* reveals conserved ancient evolutionary lineages and distinct dynamics of individual copia families. *Genome Research* 17:1072–1081.
- Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, Flavell A, Leroy P, Morgante M, Panaud O, et al. 2007. A unified classification system for eukaryotic transposable elements. *Nature Reviews Genetics* 8:973–982.

- Wickett NJ, Mirarab S, Nguyen N, Warnow T, Carpenter E, Matasci N, Ayyampalayam S, Barker MS, Burleigh JG, Gitzendanner MA, et al. 2014. Phylotranscriptomic analysis of the origin and early diversification of land plants. *Proceedings of the National Academy of Sciences* 111:E4859–E4868.
- Wienberg J, Stanyon R, Jauch A, Cremer T. 1992. Homologies in human and *Macaca fuscata* chromosomes revealed by in situ suppression hybridization with human chromosome specific DNA libraries. *Chromosoma* 101:265–270.
- Wilhelm M, Wilhelm F. 2001. Reverse transcription of retroviruses and LTR retrotransposons. *Cell Mol Life Sci* 58:1246–1262.
- Williams RC, Correia JJ, De Vries AL. 1985. Formation of microtubules at low temperature by tubulin from Antarctic fish. *Biochemistry* 24:2790–2798.
- Winnepennincks B, Backeljau T, De Wachter R. 1993. Extraction of high molecular weight DNA from molluscs. *Trends in Genetics* 9:407.
- Woods IG. 2005. The zebrafish gene map defines ancestral vertebrate chromosomes. *Genome Research* 15:1307–1314.
- Xi Z, Liu L, Davis CC. 2016. The Impact of Missing Data on Species Tree Estimation. *Molecular Biology and Evolution* 33:838–860.
- Xi Z, Liu L, Rest JS, Davis CC. 2014. Coalescent versus Concatenation Methods and the Placement of *Amborella* as Sister to Water Lilies. *Systematic Biology* 63:919–932.
- Yang F, Fu B, O'Brien PCM, Nie W, Ryder OA, Ferguson-Smith MA. 2004. Refined genome-wide comparative map of the domestic horse, donkey and human based on cross-species chromosome painting: insight into the occasional fertility of mules. *Chromosome Research* 12:65–76.
- Yu J. 2002. A Draft Sequence of the Rice Genome (*Oryza sativa* L. ssp. *indica*). *Science* 296:79–92.
- Zhang L, Rong YS. 2012. Retrotransposons at *Drosophila* telomeres: Host domestication of a selfish element for the maintenance of genome integrity. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* 1819:771–775.
- Zhou R-N, Hu Z-M. 2007. The Development of Chromosome Microdissection and Microcloning Technique and its Applications in Genomic Research. *Current Genomics* 8:67–72.
- Zhou Y, Hu Z, Dang B, Wang H, Deng X, Wang L, Chen Z. 1999. Microdissection and microcloning of rye (*Secale cereale* L.) chromosome 1R. *Chromosoma* 108:250–255.

## VIII) Annexes

### Annexe 1: ensemble des spécimens utilisés pour les analyses de la thèse.

\* matériel généreusement fourni par Eva Pisano (\*) ou Chris Cheng (\*\*).

| Sous-famille | Genre             | Espèce             | code terrain | type d'analyse                 |
|--------------|-------------------|--------------------|--------------|--------------------------------|
| Tremataminae | <i>Trematomus</i> | <i>pennellii</i>   | CE 4313      | peinture + BACs                |
|              |                   |                    | CE4320       | BACs                           |
|              |                   |                    | TA647        | ETs ( <i>FISH</i> )            |
|              |                   |                    | CE1434       | cytométrie de flux             |
|              |                   |                    | TA364TRPE1   | ETs                            |
|              |                   |                    | si98         | ddRAD-seq                      |
|              |                   |                    | CE3317       | ddRAD-seq                      |
| Tremataminae | <i>Trematomus</i> | <i>newnesi</i>     | R6TA495      | BACs                           |
|              |                   |                    | TA401TRNE3   | ETs                            |
|              |                   |                    | CE4319       | cytométrie de flux             |
|              |                   |                    | TA456        | cytométrie de flux             |
|              |                   |                    | TRNE6        | ddRAD-seq                      |
|              |                   |                    | REVO002      | ddRAD-seq                      |
| Tremataminae | <i>Trematomus</i> | <i>hansoni</i>     | TA646TRHA2   | peinture + ETs ( <i>FISH</i> ) |
|              |                   |                    | CE4004       | BACs                           |
|              |                   |                    | CE4002       | cytométrie de flux             |
|              |                   |                    | CE4003       | cytométrie de flux             |
|              |                   |                    | CE4004       | cytométrie de flux             |
|              |                   |                    | CE4316       | cytométrie de flux             |
|              |                   |                    | CE5660       | cytométrie de flux             |
|              |                   |                    | CE5664       | cytométrie de flux             |
|              |                   |                    | CE5708       | cytométrie de flux             |
|              |                   |                    | CE5711       | cytométrie de flux             |
|              |                   |                    | CE5738       | cytométrie de flux             |
|              |                   |                    | CE5856       | cytométrie de flux             |
|              |                   |                    | CE5880       | cytométrie de flux             |
|              |                   |                    | CE5661       | cytométrie de flux             |
|              |                   |                    | TA440TRHA2   | ETs                            |
|              |                   |                    | CE5659       | ddRAD-seq                      |
| CE5781       | ddRAD-seq         |                    |              |                                |
| Tremataminae | <i>Trematomus</i> | <i>eulepidotus</i> | M2S1(*)      | BACs + ETs ( <i>FISH</i> )     |
|              |                   |                    | CE23815      | ETs ( <i>FISH</i> )            |
|              |                   |                    | 3152         | ETs                            |
|              |                   |                    | si487n31524  | ETs + ddRAD-seq                |
|              |                   |                    | CE4011       | cytométrie de flux             |



|              |                   |                       | si218                   | ddRAD-seq          |
|--------------|-------------------|-----------------------|-------------------------|--------------------|
| Trematominae | <i>Trematomus</i> | <i>bernacchii</i>     | REVO2009-10<br>214-2738 | BACs               |
|              |                   |                       | Ich1098                 | cytométrie de flux |
|              |                   |                       | Ich1099                 | cytométrie de flux |
|              |                   |                       | Ich1100                 | cytométrie de flux |
|              |                   |                       | Ich1101                 | cytométrie de flux |
|              |                   |                       | Ich1102                 | cytométrie de flux |
|              |                   |                       | Ich1114                 | cytométrie de flux |
|              |                   |                       | CE4314                  | cytométrie de flux |
|              |                   |                       | CE4315                  | cytométrie de flux |
|              |                   |                       | CE5658                  | cytométrie de flux |
|              |                   |                       | CE5659                  | cytométrie de flux |
|              |                   |                       | CE5685                  | cytométrie de flux |
|              |                   |                       | CE5709                  | cytométrie de flux |
|              |                   |                       | CE5737                  | cytométrie de flux |
|              |                   |                       | CE6042                  | cytométrie de flux |
|              |                   |                       | CE6043                  | cytométrie de flux |
|              |                   |                       | CE6091                  | cytométrie de flux |
|              |                   |                       | CE6092                  | cytométrie de flux |
|              |                   |                       | si352n2561              | ETs                |
|              |                   |                       | TA337TRBE7              | ETs                |
| CE5791       | ddRAD-seq         |                       |                         |                    |
| CE5659       | ddRAD-seq         |                       |                         |                    |
| Trematominae | <i>Trematomus</i> | <i>borchgrevinkii</i> | CE3588                  | BACs               |
|              |                   |                       | CE5560                  | cytométrie de flux |
|              |                   |                       | CE5561                  | cytométrie de flux |
|              |                   |                       | CE5562                  | cytométrie de flux |
|              |                   |                       | CE5655                  | cytométrie de flux |
|              |                   |                       | CE5656                  | cytométrie de flux |
|              |                   |                       | CE5662                  | cytométrie de flux |
|              |                   |                       | CE5661                  | cytométrie de flux |
|              |                   |                       | TA263PAB01              | ETs                |
|              |                   |                       | CE5655                  | ddRAD-seq          |
| CE5466       | ddRAD-seq         |                       |                         |                    |
| Trematominae | <i>Trematomus</i> | <i>loennbergii</i>    | si259n923               | ETs                |
|              |                   |                       | si170n1265              | ETs                |
|              |                   |                       | ST50AID220              | ETs                |
|              |                   |                       | CE5916                  | cytométrie de flux |
|              |                   |                       | si174                   | ddRAD-seq          |
| si205        | ddRAD-seq         |                       |                         |                    |
| Trematominae | <i>Trematomus</i> | <i>lepidorhinus</i>   | TNB238                  | ETs                |
|              |                   |                       | si453                   | ddRAD-seq          |

|                 |                       |                     |                   |                           |
|-----------------|-----------------------|---------------------|-------------------|---------------------------|
|                 |                       |                     | si26              | ddRAD-seq                 |
| Trematomiinae   | <i>Trematomus</i>     | <i>scotti</i>       | si541n6317        | ETs                       |
|                 |                       |                     | NLT3              | ETs                       |
|                 |                       |                     | si287             | ddRAD-seq                 |
|                 |                       |                     | si394             | ddRAD-seq                 |
| Trematomiinae   | <i>Trematomus</i>     | <i>nicolai</i>      | CE5684REVO856     | BACs + cytométrie de flux |
|                 |                       |                     | TNB214            | ETs                       |
|                 |                       |                     | Ich1115           | ETs                       |
|                 |                       |                     | CE5684REVO856     | ddRAD-seq                 |
|                 |                       |                     | CE5778            | ddRAD-seq                 |
| Trematomiinae   | <i>Trematomus</i>     | <i>tokarevi</i>     | si171             | ETs + ddRAD-seq           |
| Trematomiinae   | <i>Indonotothenia</i> | <i>cyanobrancha</i> | PK9343            | BACs                      |
|                 |                       |                     | Aus30             | ETs                       |
|                 |                       |                     | PK9344            | ddRAD-seq                 |
| Trematomiinae   | <i>Lepidonotothen</i> | <i>larseni</i>      | LL12              | BACs                      |
|                 |                       |                     | ICEFISH2004SR     | ETs                       |
| Trematomiinae   | <i>Lepidonotothen</i> | <i>squamifrons</i>  | PK9224            | ddRAD-seq                 |
|                 |                       |                     | ST59N°1249        | ETs                       |
| Trematomiinae   | <i>Lepidonotothen</i> | <i>nudifrons</i>    | ICEFISH2005SR     | ETs                       |
| Trematomiinae   | <i>Patagonotothen</i> | <i>ramsayi</i>      | PR1               | BACs                      |
| Trematomiinae   | <i>Patagonotothen</i> | <i>guntheri</i>     | ICEFISH2004       | ETs                       |
| Pleuragramminae | <i>Pleuragramma</i>   | <i>antarctica</i>   | si394             | ETs + ddRAD-seq           |
| Nototheniinae   | <i>Notothenia</i>     | <i>coriiceps</i>    | TA415             | peinture                  |
|                 |                       |                     | REVO2009-10lot316 | BACs                      |
|                 |                       |                     | TA25              | BACs                      |
|                 |                       |                     | TA15              | BACs                      |
|                 |                       |                     | TA335             | BACs                      |
|                 |                       |                     | NC5R2             | BACs                      |
|                 |                       |                     | TA731             | BACs                      |
|                 |                       |                     | TA449             | ETs (FISH)                |
|                 |                       |                     | NC4               | ETs (FISH)                |
|                 |                       |                     | N°10-2001         | ETs                       |
|                 |                       |                     | TAA421NOCO8       | ETs                       |
|                 |                       |                     | Ich1095           | cytométrie de flux        |
|                 |                       |                     | Ich1096           | cytométrie de flux        |
|                 |                       |                     | Ich1097           | cytométrie de flux        |
| Nototheniinae   | <i>Notothenia</i>     | <i>rossii</i>       | NR1 (**)          | BACs                      |
|                 |                       |                     | ICEFISH2004SG     | ETs                       |
| Nototheniinae   | <i>Notothenia</i>     | <i>angustata</i>    | Nang1 (**)        | BACs + ETs                |
| Nototheniinae   | <i>Paranotothenia</i> | <i>magellanica</i>  | Aus29             | ETs                       |

|                         |                       |                     |                |                                      |
|-------------------------|-----------------------|---------------------|----------------|--------------------------------------|
| Dissostichinae          | <i>Dissostichus</i>   | <i>mawsoni</i>      | DM1 (**)       | peinture + BACs+ ETs ( <i>FISH</i> ) |
|                         |                       |                     | REVO759        | ETs                                  |
|                         |                       |                     | CE4054         | cytométrie de flux                   |
|                         |                       |                     | CE4055         | cytométrie de flux                   |
|                         |                       |                     | CE4056         | cytométrie de flux                   |
| Channichthyinae         | <i>Chionodraco</i>    | <i>hamatus</i>      | R60-II         | BACs                                 |
|                         |                       |                     | TNB235         | ETs                                  |
|                         |                       |                     | CE5342         | cytométrie de flux                   |
| Channichthyinae         | <i>Champscephalus</i> | <i>gunnari</i>      | ICEFISH2004SR  | ETs                                  |
| Channichthyinae         | <i>Champscephalus</i> | <i>esox</i>         | ICEFISH2004F   | ETs                                  |
| Channichthyinae         | <i>Cryodraco</i>      | <i>antarcticus</i>  | TNB246         | ETs                                  |
| Cygnodraconinae         | <i>Cygnodraco</i>     | <i>C. mawsoni</i>   | TA373          | BACs                                 |
|                         |                       |                     | TA53CYMA1      | ETs                                  |
| Gymnodraconinae         | <i>Gymnodraco</i>     | <i>G. acuticeps</i> | TA736          | BACs                                 |
|                         |                       |                     | TA109          | ETs                                  |
|                         |                       |                     | CE5883         | cytométrie de flux                   |
| Artedidraconinae        | <i>Pogonophryne</i>   | <i>scotti</i>       | si238n1819     | ETs                                  |
| Artedidraconinae        | <i>Histiodraco</i>    | <i>velifer</i>      | si283n978      | ETs                                  |
| Gobionototheniinae      | <i>Gobionotothen</i>  | <i>gibberifrons</i> | 1229-5R        | ETs                                  |
|                         |                       |                     | ICEFISH2004SR  | ETs                                  |
| Eleginopsidae (famille) | <i>Eleginops</i>      | <i>maclovinus</i>   | EM2            | BACs                                 |
|                         |                       |                     | RG             | ETs ( <i>FISH</i> )                  |
|                         |                       |                     | EM4            | ETs                                  |
| Bovichtidae (famille)   | <i>Bovictus</i>       | <i>diancanthus</i>  | BD1            | BACs + ETs ( <i>FISH</i> )           |
|                         |                       |                     | BD4            | ETs ( <i>FISH</i> )                  |
|                         |                       |                     | I2004(TDC)1206 | ETs                                  |

**Annexe 2: protocole détaillé de préparation des bibliothèques de ddRAD-seq en vue d'un séquençage Illumina (Miseq) mis au point par l'équipe du Service de Systématique Moléculaire (SSM) du MNHN (UMS 2700).** Les paramètres utilisés sont indiqués en gras.

| étape    | Réactifs  |
|----------|---|
| 1.       | EB Buffer (10 mM Tris-Cl, Ph7.5-8)  |
| 1.       | kit Qubit BR  |
| 1.       | digestion ER<br>- 10X buffer 4<br>- EcoRI<br>- MspI<br>- BSA (100X/10X)                                     |
| 2.       | oligos adaptateurs RAD<br>- Top Ax / Bot Ax<br>- Top P1 /Bot P1<br>- tampon 10X AB (10X TE avec 500mM NaCl) |
| 3.       | réactifs de ligation<br>- 10X T4 ligase<br>- T4 DNA ligase (400,000 U / ml)                                 |
| 1. et 4. | purification magnétique<br>- billes Nucleomag (Macherey-Nagel)<br>- EtOH 80 % frais                         |
| 4.       | TET buffer (0.1X TE + 0,05 % Tween-20)  |

| Consommables                                 |
|--|
| tubes DNA lo-bind Millipore 0.5 ml et 1.5 ml |
| Plaque low profile semi-jupée (Biorad)       |
| strips de bouchons pour plaque PCR           |
| film adhésif pour qPCR optique               |

## 0. Calcul de la concentration d'adaptateurs nécessaire

### 0.1 Test de digestion par chaque enzyme séparément

a) Préparer sur glace pour chaque enzyme d'intérêt

| Premix                     | [final] | / puits |
|----------------------------|---------|---------|
| Volume de réaction (ul)    |         | 20      |
| 10X CUTsmart               | 1X      | 2       |
| BSA 10X                    | 1X      | 2       |
| MspI ou EcoRI (20000 U/ml) | 20 U    | 1       |
| ADN à digérer              |         | 15      |

b) Lancer le programme : 3 heures à 37°, refroidissement progressif (1°/min) jusque 4°C, stock à 4°.

## 0.2 Purification magnétique

- Diluer chaque réaction dans volume final de 40  $\mu$ l (ajouter 20  $\mu$ l eau MilliQ)
- SPRI simple de facteur 1.8X (72 $\mu$ l de billes)
- Ré-éluer dans 22  $\mu$ l TET. Récupérer 20  $\mu$ l (TE 0.1X) dans un tube lo-bind 0.5 ml.

## 0.3 Profil bioanalyzer

- Faire migrer les produits de digestion sur une puce bioanalyzer HS.
- Déterminer la fréquence de coupe pour chaque enzyme.
- Utiliser le [ligation molarity calculator](#) de Peterson *et al* (2012) pour déterminer la concentration nécessaire de chaque adaptateur.

# 1. Préparation des adaptateurs

En fonction des résultats du test bioanalyzer ou des analyses in silico, préparer les dilutions des oligos d'adaptateurs

## 1.1 Dilution initiale des oligos d'adaptateurs

- Préparer du tampon 1X AB à partir du stock de 10X AB et d'eau nuclease-free
- Réaliser un mélange de chaque paire d'oligos (top/bot) à 10  $\mu$ M par exemple (attention! ces volumes doivent être ajustés pour chaque projet)

| réactif                             | pour les P1 EcoRI | pour le P2 MspI |
|-------------------------------------|-------------------|-----------------|
| tampon 1X AB                        | 8 $\mu$ l         | 8 $\mu$ l       |
| oligo Ax/P1 top (stock 100 $\mu$ M) | 1 $\mu$ l         | 1 $\mu$ l       |
| oligo Ax/P1 bot (stock 100 $\mu$ M) | 1 $\mu$ l         | 1 $\mu$ l       |

## 1.2 Hybridation des oligos

Lancer le programme : 2.5 min à 97.5°C, refroidissement de 1°/min jusqu'à 20.5°C

## 1.3 Dilution fonctionnelle des adaptateurs P1 barcodés et P2

En fonction de la paire d'enzymes utilisée, du génome ciblé et de la quantité d'ADN digéré (étape 0.3c), ajuster la concentration fonctionnelle des adaptateurs en préparant un volume total (x+Y) suffisant pour chaque type d'adaptateur.

| réactif                 | pour P1 EcoRI                               | pour P2 MspI                                 |
|-------------------------|---|--|
| tampon 1X AB            | x1 $\mu$ l / <b>58,97 <math>\mu</math>l</b> | x2 $\mu$ l / <b>55,895 <math>\mu</math>l</b> |
| adaptateur à 10 $\mu$ M | y1 $\mu$ l / <b>1,03 <math>\mu</math>l</b>  | y2 $\mu$ l / <b>4,105 <math>\mu</math>l</b>  |

$x1+y1 > 2 \times N$  lots d'échantillons

$x2+y2 > 2 \times$  nombre total d'échantillons

## 2. Double Digestion par EcoRI et MspI

### 2.1 Normalisation des ADNs génomiques

- a) Quantification Qubit en plaque de tous les individus
- b) Réaliser une dilution équimolaire de chaque individu dans du EB pour en utiliser 20 ul. Distribuer 10 ul de chaque extrait dans le puits assigné. Quantité minimum requise d'ADN génomique: 420 ng total par extrait. Possibilité de préparer des lots avec des quantités d'ADN total différentes.

### 2.2 Digestion enzymatique (25 ul par puits)

Les deux enzymes (MspI et EcoRI-HF) fonctionnent à 100 % dans le tampon CUTsmart. MspI ne peut pas être dénaturée par la chaleur. Il faudra donc purifier le digest en SPRI.

- a) Préparer le premix suivant sur glace en fonction du nombre de banques:

| Premix                   | [final] | Puits (µl) |
|--------------------------|---------|------------|
| Volume de premix (ul)    |         | 7          |
| 10X CUTsmart             | 1X      | 2.5        |
| BSA 10X                  | 1X      | 2.5        |
| MspI<br>(20000 U/ml)     | 20 U    | 1          |
| EcoRI-HF<br>(20000 U/ml) | 20 U    | 1          |

- b) Vortexer, quick spin  
Répartir 15 ul du prémix par puits de banque
- c) Lancer le programme : 2 heures à 37°, refroidissement progressif (1°/min) jusque 4°C, stock à 4°.

### 2.3 Purification magnétique

SPRI simple avec facteur 1.8X de billes.

## 3. Ligation des adaptateurs

### 3.1 Ligation

- a) Ajouter à chaque échantillon de 25ul (donc volume de réaction final de 40 ul)

|                       |             |
|-----------------------|-------------|
| Mix                   | 1 puit (µl) |
| Volume de premix (ul) | (15)        |

|                              |     |
|------------------------------|-----|
| 10X T4 ligase buffer         | 4   |
| H2O                          | 6.5 |
| Adaptateur P1 dilué          | 2   |
| Adaptateur P2 dilué          | 2   |
| T4 DNA ligase<br>(2 MU / ml) | 0.5 |

**b)** Lancer le programme : 2h à température ambiante, 20' à 65°C, refroidissement de 1.5°C par min

#### Liste des adaptateurs EcoR1 utilisés :

**1 : GCATG, 2 : AACCA, 3 : CGATC, 4 : TCGAT, 5 : TGCAT, 6 : CAACC, 7 : GGTTG, 8 : AAGGA, 9 : AGCTA, 10 : ACACA, 11 : AATTA, 12 : ACGGT, 13 : ACTGG, 14 : ACTTC, 15 : ATACG, 16 : ATGAG, 17 : ATTAC, 18 : CATAT, 19 : CGAAT, 20 : CGGCT, 21 : CGGTA, 22 : CGTAC, 23 : CGTCG et 24 : CTGAT.**

### 3.2 Migration sur gel des produits ligués

Préparer un gel à 2 %

Faire migrer 2ul de chaque produit pour vérifier l'aspect du smear quantif

### 3.3 Pool des échantillons par lot de barcodes

Mélanger le contenu des tubes de chaque lot

### 3.4 Première purification magnétique des banques

Réaliser une purification à 1.8X de chaque lot.

Prendre chaque lot dans 50 ul de TET.

## 4. Sélection de taille par double SPRI

**a)** Faire une double purification Nucleomag de chaque lot avec les conditions de ration billes/ADN identifiés pour garder la gamme de taille voulue (à adapter à chaque projet)

| pour chaque lot (50 µl) | ratio billes/ADN | volume à utiliser                      |
|-------------------------|------------------|--|
| 1ère sélection          | <b>0,55X</b>     | <b>50,05 µl pool1, 36,85 µl pool 2</b> |
| 2ème sélection          | <b>0,35X</b>     | <b>31,85 µl pool1, 23,4 µl pool 2</b>  |

**b)** Ajouter le premier facteur de billes à chaque tube lo-bind

Vortexer à vitesse moyenne pour homogénéiser

**c)** Laisser s'hybrider sur la paillasse 3' et préparer pendant ce temps les tubes lo-binds contenant le deuxième facteur de billes.

**d)** Passer chaque lot sur le magnet et transférer le surnageant dans le tube contenant le deuxième facteur de billes.

Vortexer à vitesse moyenne pour homogénéiser

Laisser s'hybrider sur la paillasse pendant 3'.

**e)** Passer sur le magnet,

Supprimer le surnageant.

Rincer 2 fois avec 200 ul d'EtOH 80 %.

Supprimer l'alcool

Laisser sécher les billes jusqu'à ce que leur surface devienne opaque (pas craquelée)

**f)** Reprendre les billes hors du magnet dans 20 ul de EB et vortexer

Patienter 3 minutes, puis repasser sur le magnet et récupérer le surnageant dans un tube lo-bind 1.5 ml.

## 5. Amplification de chaque lot et indexage

### 5.1 Quantification des banques pré-amplification

**a)** Préparer deux dilutions du pool des banques (dans tubes lo-bind 0.5 ml) pour les quantifier en qPCR:

Banque au 1/10: 9 ul TET + 1 ul banque brute

Banque au 1/50: 4 ul de TET + 1 ul de banque au 1:10

**b)** Stocker les banques brutes à -20°C jusqu'à utilisation

**c)** Réaliser une qPCR des banques avec un couple quelconque d'amorces P1/P2 indexé

En fonction du résultat, identifier le nombre de cycles d'écart entre les différents lots de banques pour les compenser lors de la PCR d'indexage ou un dosage en Qubit (HS).

### 5.2 PCR d'indexage

**a)** En fonction du nombre de lots d'échantillons à traiter, identifier une combinaison favorable d'index Illumina (6nuc). **Pool 1 : index 5 (ACAGTG) et pool 2 : index 19 (GTGAAA).**

**b)** Réaliser une PCR d'indexage sur au plus ½ des banques brutes, en assortissant le nombre de cycles de telle sorte que le nombre de cycles maximum (càd pour le lot le moins concentré) ne dépasse pas 15-16 cycles.

**c)** Faire une purification 1X à 1.2X des banques amplifiées

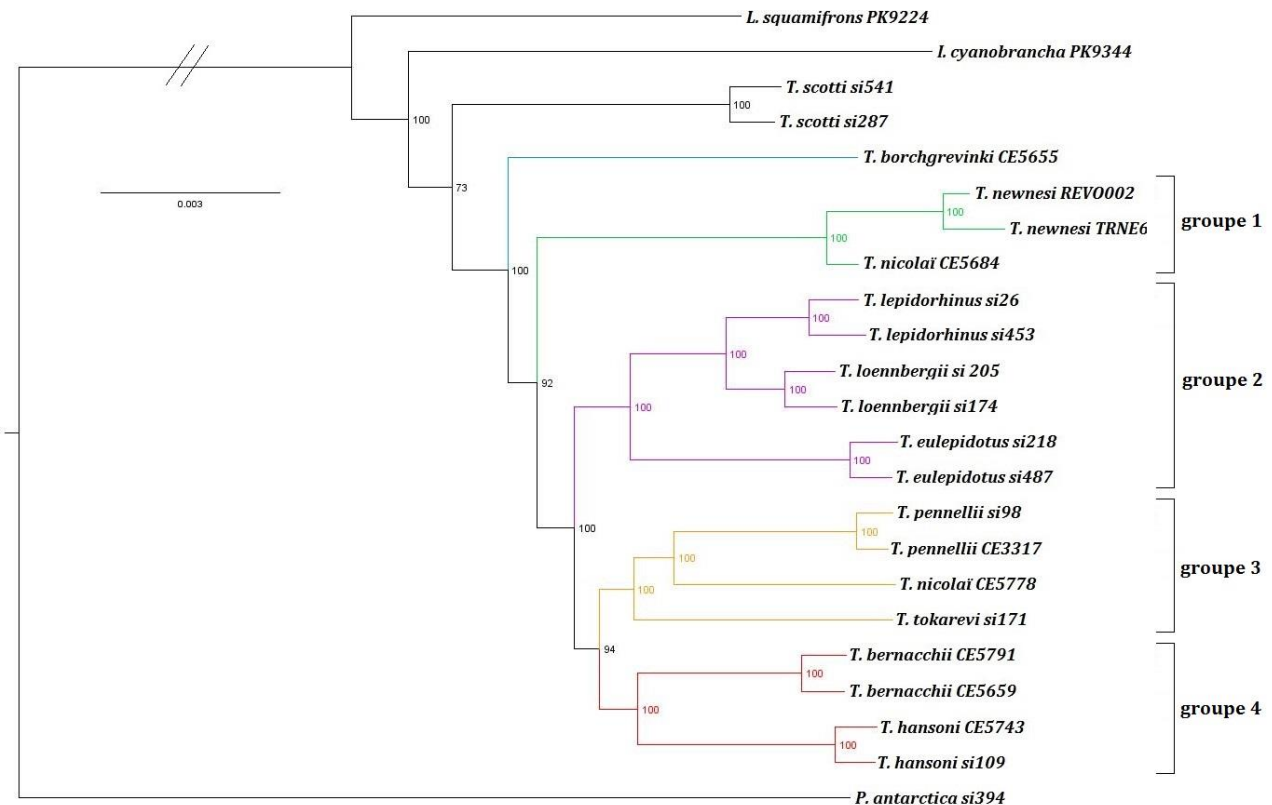
### 5.3 Caractérisation des banques post-amplification

**a)** Faire une quantification qubit HS des banques

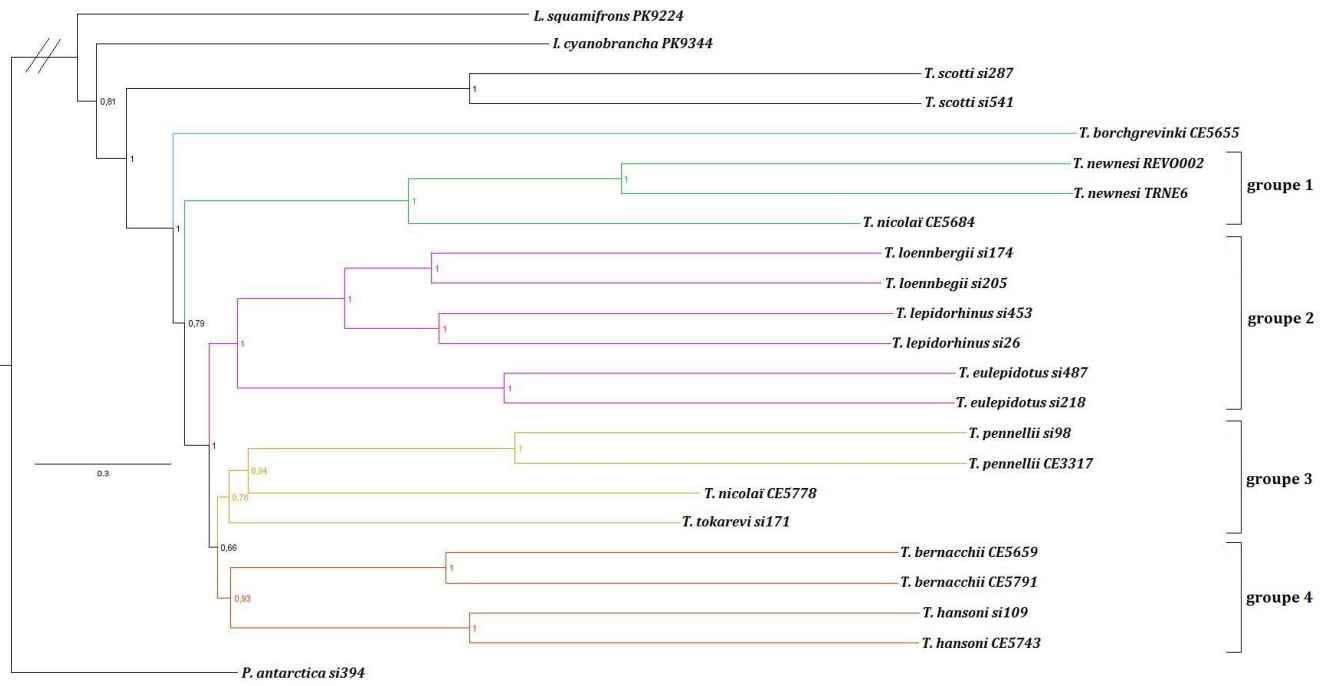
**b)** En fonction du rendement des banques amplifiées constaté en qubit, diluer ces banques pour les passer sur une puce bioanalyzer HS avant envoi pour séquençage



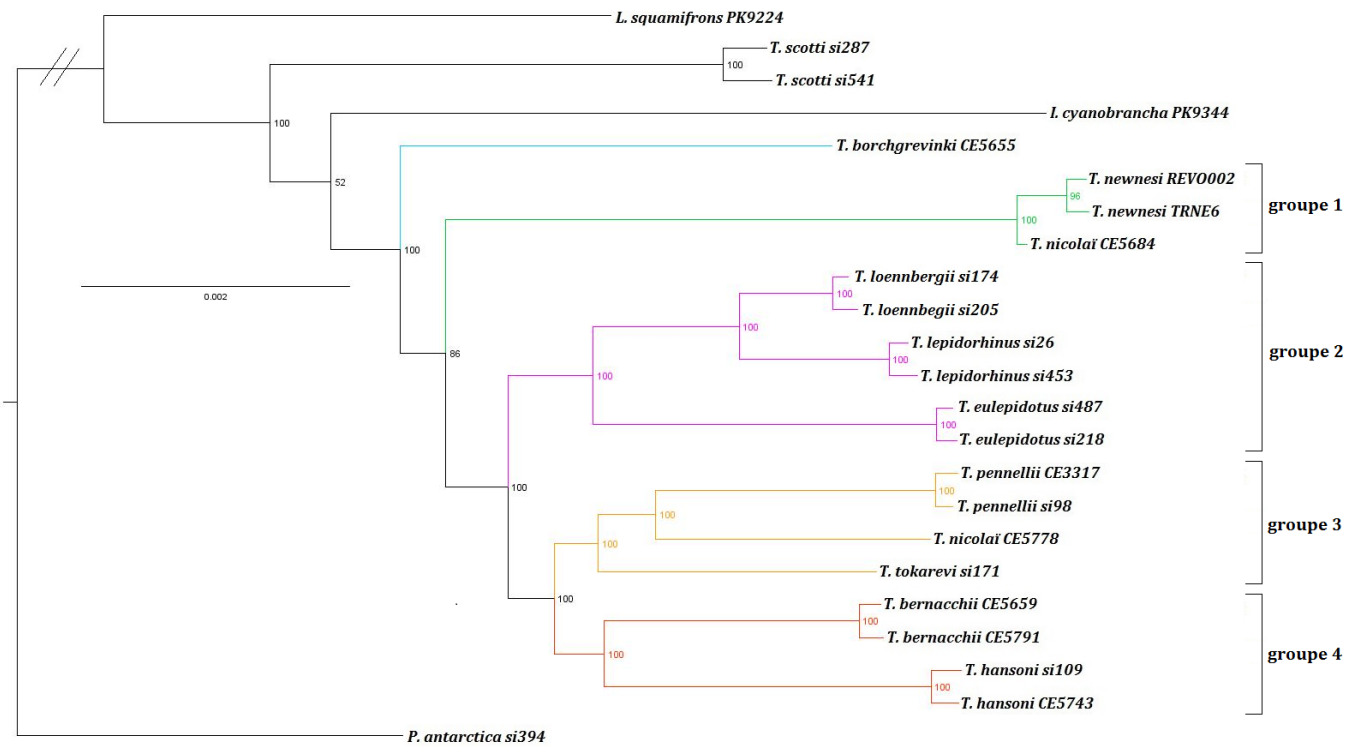
**Annexe 3 : topologies obtenues avec les jeux de données issus de PyRAD ou Stacks et un minimum de 16 individus pour former un locus polymorphe partagé (MinDepth ou  $p = 16$ ).** Les valeurs de soutien des nœuds (en pourcentage ou en décimal) sont indiqués par les chiffres. La barre d'échelle correspondant au nombre de substitutions par unités de temps. Pour faciliter la lecture, *T. borchgrevinki* est indiqué en bleu, le groupe un (*T. newnesi*/*T. nicolai* CE5684) en vert, le groupe deux (*T. eulepidotus*,*T. loennbergii*/*T. lepidorhinus*) en rose, le groupe trois (*T. pennellii*/*T. nicolai* CE 5778,*T. tokarevi*) en orange et groupe quatre (*T. bernacchii* et *T. hansonii*) en rouge.



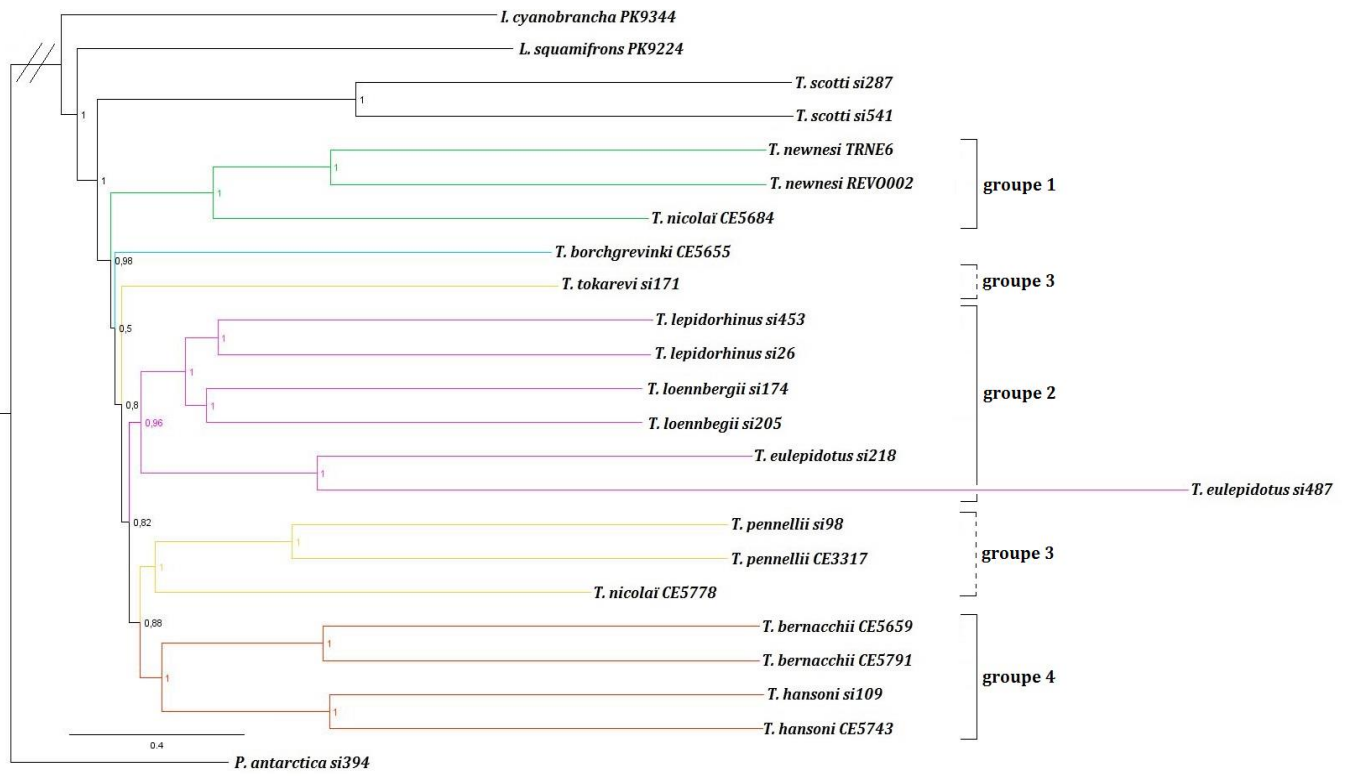
PyRAD, méthode de concaténation



### PyRAD, méthode de congruence



### Stacks, méthode de concaténation



Stacks, méthode de congruence

## Résumé / Abstract

L'alternance de périodes glaciaires et interglaciaires durant les 20 derniers Ma a mené à des changements environnementaux répétés au niveau du plateau continental antarctique. C'est dans ce contexte que les téléostéens de la famille des Nototheniidae se sont adaptés et diversifiés à travers plusieurs vagues de radiations (dont les Trematominae), dominant l'ichtyofaune australe. Parmi les Nototheniidae, le groupe « *Trematomus* » (genres *Cryothernia*, *Pagothenia*, *Trematomus* et *Indonotothenia*) est celui où l'on observe la plus grande diversité chromosomique, avec des nombres diploïdes de chromosomes allant de 24 à 58, impliquant de nombreux réarrangements ayant accompagné les spéciations.

Nous avons cherché à caractériser ces remaniements chromosomiques. Avec un caryotype ancestral inféré de  $2n = 48$ , une conservation des unités chromosomiques entre espèces, et une constance des tailles de génome, l'hypothèse de réarrangements structuraux sans polyploïdisation préalable est la plus probable. Afin de reconstruire l'histoire évolutive de ces événements, nous avons recherché les homologues chromosomiques interspécifiques. Ceci nous a permis de reconstituer les remaniements (majoritairement des fusions) que nous avons repositionnés sur la phylogénie résolue des « *Trematomus* ». Contrairement à ce qui a été publié pour le genre *Notothenia*, nos résultats suggèrent des acquisitions multiples et indépendantes.

Les éléments transposables (ETs) peuvent être impliqués dans les remaniements chromosomiques par le biais de recombinaisons ectopiques. Ils participent alors à la diversification des lignées au cours de l'évolution. En raison de leur régulation épigénétique, leur mobilisation massive peut être induite en cas de variations environnementales importantes. Nous nous sommes intéressés à trois super-familles d'ETs (*DIRS*, *Gypsy* and *Copia*) dans ces génomes. Les *DIRS1* ont montré des patrons d'insertions en points chauds dans les régions centromériques et péri-centromériques. Étant donné leur mode de transposition décrit et leur propension à s'insérer dans des copies préexistantes, nous proposons un rôle des éléments *DIRS1* comme facilitateurs des fusions observées lors de la diversification des « *Trematomus* ».

In the last 20 My, multiple glacial-interglacial cycles led to strong and repeated environmental changes on the Antarctic continental shelf. In this changing environment, nototheniid fishes diversified through several rounds of species radiation (one of which within Trematominae), and now constitute the dominant group in Antarctic teleosts. Among Nototheniidae, the group « *Trematomus* » (genera *Cryothernia*, *Pagothenia*, *Trematomus* and *Indonotothenia*) exhibits the highest chromosomal diversity, with diploid chromosome numbers ranging between 24 and 58, involving many rearrangements probably linked to speciation.

We characterized the nature of these chromosomal repatternings. With an inferred ancestral state of  $2n = 48$  acrocentric chromosomes, a conserved number of chromosomal structural units, and a constancy of the genomes sizes we measured; the hypothesis of structural modifications is favored rather than a whole genome duplication associated to drastic reductions.

In order to reconstruct an evolutionary scenario of such chromosomal rearrangements accompanying the trematomine diversification, we identified interspecific chromosomal homologies. This allowed us to reconstruct the rearrangements events (mostly centric and tandem fusions). We plotted them on a phylogeny we reconstructed based on our own ddRAD-seq data. Contrary to what was reported for the *Notothenia*, our results are in favor of independent acquisitions.

Transposable elements (TEs) can lead to chromosomal rearrangements through ectopic recombination events, hinting at a role as drivers of specific-lineage diversification. Moreover, due to their epigenetic regulation, TEs can be mobilized when thermic changes occur. We focused on three retrotransposon superfamilies (*DIRS*, *Gypsy* and *Copia*) in nototheniid genomes. The *DIRS1* showed unexpected accumulation patterns of insertion in the centromeric and pericentromeric regions. Given the mechanism of *DIRS1* transposition and their tendency to sometimes insert on pre-existing copies (homing), we suggest a role of *DIRS1* elements as facilitators of the fusions that occurred during the trematomine radiation.