UNIL | Université de Lausanne

Unicentre
CH-1015 Lausanne
http://serval.unil.ch

Year : 2016

Long term evolutionary responses to whole genome duplication

Sacha Laurent

Sacha Laurent, 2016, Long term evolutionary responses to whole genome duplication

Originally published at : Thesis, University of Lausanne

Posted at the University of Lausanne Open Archive http://serval.unil.ch
Document URN : urn:nbn:ch:serval-BIB_8E5536BAC6042

Droits d'auteur
L'Université de Lausanne attire expressément l'attention des utilisateurs sur le fait que tous les documents publiés dans l'Archive SERVAL sont protégés par le droit d'auteur, conformément à la loi fédérale sur le droit d'auteur et les droits voisins (LDA). A ce titre, il est indispensable d'obtenir le consentement préalable de l'auteur et/ou de l'éditeur avant toute utilisation d'une oeuvre ou d'une partie d'une oeuvre ne relevant pas d'une utilisation à des fins personnelles au sens de la LDA (art. 19, al. 1 lettre a). A défaut, tout contrevenant s'expose aux sanctions prévues par cette loi. Nous déclinons toute responsabilité en la matière.

Copyright
The University of Lausanne expressly draws the attention of users to the fact that all documents published in the SERVAL Archive are protected by copyright in accordance with federal law on copyright and similar rights (LDA). Accordingly it is indispensable to obtain prior consent from the author and/or publisher before any use of a work or part of a work for purposes other than personal use within the meaning of LDA (art. 19, para. 1 letter a). Failure to do so will expose offenders to the sanctions laid down by this law. We accept no liability in this respect.

UNIL | Université de Lausanne

Faculté de biologie
et de médecine

Département d'Écologie et d'Évolution

Long term evolutionary responses to whole genome duplication

Thèse de doctorat ès sciences de la vie (PhD)
Écologie et Évolution

présentée à la
Faculté de Biologie et de Médecine
de l'Université de Lausanne
par

Sacha LAURENT

Master en Bioinformatique et Modélisation de l'Institut National des Sciences Appliquées de Lyon, France

Jury

Dirk Fasshauer, président
Marc Robinson-Rechavi, co-directeur
Nicolas Salamin, co-directeur
Nils Arrigo, expert
Christian Parisod, expert

Lausanne 2016

Unil

UNIL | Université de Lausanne

Faculté de biologie et de médecine

Ecole Doctorale

Doctorat ès sciences de la vie

Imprimatur

Vu le rapport présenté par le jury d'examen, composé de

Président · e	Monsieur	Prof.	Dirk	**Fasshauer**
Directeur · rice de thèse	Monsieur	Prof.	Marc	**Robinson-Rechavi**
Co-directeur · rice	Monsieur	Prof.	Nicolas	**Salamin**
Experts · es	Monsieur	Dr	Nils	**Arrigo**
	Monsieur	Dr	Christian	**Parisod**

le Conseil de Faculté autorise l'impression de la thèse de

Monsieur Sacha Laurent

Master en bioinformatique et modélisation de l' INSA de Lyon, France

intitulée

Long term evolutionary responses to whole genome duplication

Lausanne, le 26 août 2016

pour le Doyen
de la Faculté de biologie et de médecine

Prof. Dirk Fasshauer

Abstract

The study of species appearance and disappearance at the geological scales is the main topic of my doctoral work. The joint use of reconstructed phylogenies and of methods identifying changes in extinction and speciation rates through the ages deepens our knowledge of the mechanisms explaining current day biodiversity. During my project, I use simulations and different methods for quantifying biodiversity changes to try to address the question of the influence of the genome onto the apparition of new species. In particular, I focus on the possibility for lineages to be more likely to form new species after a duplication of their genome. To do so, I first develop an algorithm simulating phylogenies, incorporating different types of macro-evolutionary events, such as a sudden increase of the number of species appearing, or their abrupt extinction, to test the power and specificity of commonly used methods set to detect such events. Then, I use biological data, from ray finned fishes and plant lineages, where genome duplications had been identified, to determine how these events shape the patterns of the apparition of species. In the first part of my work, I show that methods used are sufficiently robust to correctly identify changes in diversification tempos, even when simulations included noise. In the second part, I find a number of evidences indicating that if genome duplication in ray finned fishes had an influence on the appearance of species, it could not be detected a long time after the genomic event, as was proposed by some authors. In the last part of my work, I show using data from around sixty plant genera, that genome duplication is concomitant to the apparition of new species in around half of the cases. In general, this work builds on the rich debate regarding the study of long term consequences of genome duplication and brings some proof indicating its limited repercussion upon the success of species at the geological scale.

Résumé

L'étude de l'apparition et de la disparition des espèces à l'échelle des temps géologiques est au centre de mon travail de doctorat. L'utilisation conjointe de reconstructions d'arbres phylogéniques et de méthodes identifiant les changements de taux d'extinction et de spéciation à travers les âges permet d'améliorer notre compréhension des mécanismes expliquant la biodiversité observée de nos jours. Dans mon projet, j'ai utilisé d'une part des simulations et d'autres part différentes méthodes de quantification des changements de la biodiversité afin d'essayer d'adresser la question de l'influence du génome sur l'apparition de nouvelles espèces. Notamment, je me suis intéressé à la possibilité pour les lignées d'être plus aptes à former de nombreuses espèces après la duplication de leur génome. Pour cela, j'ai dans un premier temps développé un algorithme de simulation d'arbre phylogénique incorporant différents types d'événements, tels que la soudaine augmentation du nombre d'espèces qui apparaissent, ou leur extinction subite, afin de tester la puissance et la spécificité des méthodes couramment employées pour les détecter. Dans un second temps, j'ai utilisé des données biologiques, provenant d'espèces de poissons à nageoires rayonnées ainsi que de plantes, chez lesquelles des duplications de génome avaient été identifiées, afin de chercher les conséquences de ces événements sur le processus d'apparition des espèces. Dans la première partie de mon travail, j'ai trouvé que les méthodes que nous utilisions étaient suffisamment robustes pour identifier correctement les changements de tempo d'apparition d'espèces, même lorsque les simulations avaient été réalisées avec du bruit. Dans la seconde partie, j'ai trouvé un certain nombre d'indices laissant penser que si la duplication de génome chez les poissons à nageoires rayonnés avait une influence sur l'apparition des espèces, cet effet n'était pas observé longtemps après l'événement génomique, comme certains auteurs l'avaient proposé. Dans la dernière partie, je montre que, d'après les données d'une soixantaine de genres de plantes, la duplication de génome arrive simultanément avec l'apparition d'une nouvelle espèce dans environ la moitié des cas. Dans l'ensemble, ce travail s'inscrit dans le foisonnant thème de l'étude des conséquénces évolutives de la duplication de génome, et amène quelques preuves laissant penser que la duplication de génome a des répercussions limitées sur le succès des espèces à l'échelle des temps géologique.

Acknowledgements

First I would like to warmly thank both my supervisers, Marc Robinson-Rechavi and Nicolas Salamin who supported me during my five years of work. Without their guidance, help and trust I would never have been able to get through my PhD. Thank you for giving me the chance to go through my work and try to learn how to be a scientist.

The two members of my jury, Nils Arrigo and Christian Parisod, have given a large amount of time to the improvement and the evaluation of my work and for this I thank them very much. All their comments and ideas tremendously improved this manuscript.

Thanks to all the people I have shared time with during those five years, in particular the members of both my groups (and others) in the Department of Ecology and Evolution, in particular to Elza, Valentine, Fred, Julien, Martha, Chloé, Nicolas. Thanks for the fun times together and the presence during good and bad moments.

I would like to thank all the people from Toulouse and INSA who came to visit or with whom we regularly gathered: Hugo, Alice, Greg, Elsa, Bérénice, PL, Étienne, Vèl, Camille, Raph, Sam, Romain, Audrey, Mélaine, Julien, Léa. Thanks for showing me the way.

My family has always been supportive of my choices (and my multiple changes thereof). Thanks in particular to my parents and my sister Chloé for their love no matter what. Thanks also to all my cousins, for acting as models and pushing me towards improvement. All my aunts and uncles have always been supported and offered help when needed. Thanks for coming so numerous at my defence.

Lots of friend have made these five years in Lausanne really fun, in particular Clémence, who was an awesome roommate for a few months but is now a long lasting friend, and all the cool people I have met through her. Thanks to all the Swiss people I have met here with whom I discovered the Swiss way of life and this beautiful country.

Of course I'll be always indebted to Yoann, who helped me go through the final years of my PhD as smoothly as possible. Thank you for being so curious and open minded, for always being so amazed at the wonders of life and nature.

Contents

Contents **vii**

1 Introduction **1**

1.1 Duplication mechanisms 1
1.2 Identification of polyploidy 3
1.3 Evolutionary fates of gene duplicates 8
1.4 Fates of populations and species after polyploidization 12
1.5 Testing the link between polyploidy and diversification 15

2 Detecting patterns of species diversification **19**

2.1 Abstract 19
2.2 Background 20
2.3 Methods 22
2.4 Results and discussion 28
2.4.1 Baseline performances 28
2.4.2 Mixed scenarios of diversification 29
2.4.2.1 Tree size influence 30
2.4.2.2 Impact of events violating the model 31
2.4.2.3 Impact of patterns of extinction 34
2.5 Conclusion 35
2.6 Authors contributions 38
2.7 Acknowledgements 38

3 Teleostei and radiation time lag model after polyploidy **39**

3.1 Abstract 39

3.2 Background 40
3.3 Methods 43
3.3.1 Phylogeny-based diversification analysis 43
3.3.2 Fossil-based method 46
3.4 Results 47
3.4.1 Salmonids, cyprinids and botiids whole genome duplications studied with phylogeny methods 47
3.4.2 Teleost whole genome duplication studied with fossil data . 52
3.5 Discussion 54

4 Anagenetic and cladogenetic polyploidization as likely in land plants 59
4.1 Abstract 59
4.2 Background 60
4.3 Materials and Methods 63
4.3.1 Phylogenetic reconstruction and aggregation 63
4.3.2 Chromosome number evolution reconstruction 64
4.3.3 Macro-evolutionary rates estimation 65
4.4 Results 67
4.5 Discussion 71
4.5.1 Origin of polyploid individuals 71
4.5.2 Methodological considerations 74
4.6 Conclusion 75

5 Perspectives 77
5.1 Outline of my work 77
5.2 Linking diversification and polyploidy 78
5.3 Divergences and convergences between plant and animal polyploidies 80
5.4 Short and long term consequences of polyploidy 83

A Discussion regarding "Polyploidy can drive rapid adaptation in yeast" by Selmecki et al. (2015) 87

B Looking for other events of polyploidization in Teleostei **95**
B.1 Identification of the clades of interest 95
B.2 Polyploidy event determination in Callichthyidae 97

References **101**

Chapter 1

Introduction

The aim of my work will be to investigate the link between whole genome duplications and the evolutionary success of lineages. In particular, I will study how whole genome duplications have been related to speciation processes, and after introducing the molecular and cellular effects of gene and genome duplications, I will endeavour to look for the large scale impacts of whole genome duplications, at the level of species and lineages, using simulated and real biological data.

1.1 Duplication mechanisms

Whether evolution is mostly governed by neutral or selective processes is still debated today (Nei, 2005; Wagner, 2008), but both mechanisms can only be seen as mere choosing algorithms, without any creative power. The appearance of novelties requires other phenomena: mutation and recombination have been recognized as having the potential to bring new functionalities to the host that experiences them. Among mutations, duplication has been studied from the beginning of the 20th century (reviewed in Taylor and Raes (2004)). Similarly to human progress and innovations, that seem more likely by progressive tinkering and building upon previous findings, evolving new functions would naively appear to us more probable by using material that has already been selected to perform some other function, rather than *ex nihilo*. At the same time, throwing away

something that has been working for millions of years might not seem a sensitive move, as new functionalities can always be built using the simplest blocks, akin to using the thousand-year old wheel invention to create bikes during the 19th century.

Gene duplications in a host can happen through a variety of mechanisms. Appearance of retrocopies can occur when a mature mRNA is recruited by the retrotranposition molecular machinery encoded by transposable elements littering most eukaryotic genomes, and inserted back into the genome (Zhang, 2003). As the mRNA lacks introns and flanking promoter sequence, the new gene can be distinctively recognized as retrocopy and is usually assumed to be nonfunctional (Hurles, 2004). Nevertheless, examples of functional retroposed genes have now been identified in most organisms, notably in mammals and *Drosophila* (Kaessmann et al., 2009). Moreover, correlation between expression of parental and retroposed gene copies has been observed in zebrafish (Zhong et al., 2015), suggesting the possibility of co-retrotransposition of the flanking sequences along with the gene. The origin of the mRNA from which the gene was copied can help explaining this finding: because retrotransposable elements have weak transcription stop sites, their mRNA intermediates can sometimes incorporate neighbouring downstream genes, along with their promoters (Lynch, 2007). Upon retrotransposition of the mRNA, the passenger gene will also include its promoter sequences.

Segmental duplications, resulting in the addition of one to a few genes in the host, can happen as well through unequal crossing-over or gene capture during DNA repairs (Lynch, 2007). These duplications typically include neighbouring elements promoting the expression of the gene, thus the additional product is usually expected to be functional from its apparition. All these mechanisms usually result in the duplication of a small part of the genome, but the amplification of a large number of genes, through whole genome duplication in particular, is also known to happen.

Indeed, wider disturbances in the genome, such as chromosomal aberrations,

can drive the addition of an elevated number of gene copies. Abnormal number of homologous chromosome copies, known as aneuploidy, are frequently found in tumor cells or involved in different human syndromes. Nevertheless, the duplication of all chromosome copies, known as polyploidization or whole genome duplication — that can arise through a variety of mechanisms (figure 1.1) — do not necessarily lead to immediate adverse effect in every host. As a matter of fact, Hugo de Vries, an early 20th century botanist, when studying the primrose genus *Oenothera*, accidentally observed the non-detrimental effect of such chromosome number variation, when he proposed his "Mutation" theory, whose name lived on, although not its essence (discussed by Endersby (2013) and Nei and Nozawa (2011)). I will now focus onto the historical studies of the identification of polyploidy to present how our current view of the presence of whole genome duplications across the tree of life has emerged.

1.2 Identification of polyploidy

When de Vries discovered, at the same locality, individuals of widely diverging phenotypes belonging to what he thought to be the *O. lamarckiana* complex, de Vries (1904) put forward that saltationism led to the — almost instantaneous — apparition of new species, not the gradual accumulation of slight variations from one generation to another, generally accepted by Darwin and his followers. Among the varieties of *O. lamarckia* he identified, the ones he named *gigas*, because of their unusual density of leaves and robustness, and *nanella*, because of their dwarfish aspect, are of particular interest. On subsequent studies (de Vries and Boedijn, 1923), he identified *O. nanella* as carrying 14 chromosomes, whereas *O. gigas* had 28. Following experiments in breeding the organisms and creating hybrids, he also added to the list *O. semigigas*, of intermediate phenotype, carrying 21 chromosomes (de Vries, 1915). Those changes in chromosome number were not due to independent duplications of particular chromosomes, but to a change in ploidy — the number of homologous chromosome sets present in the organism — occurring through mitotic or meiotic error leading to the duplication of the entire chromosomal collection (figure 1.1). Although de Vries granted us

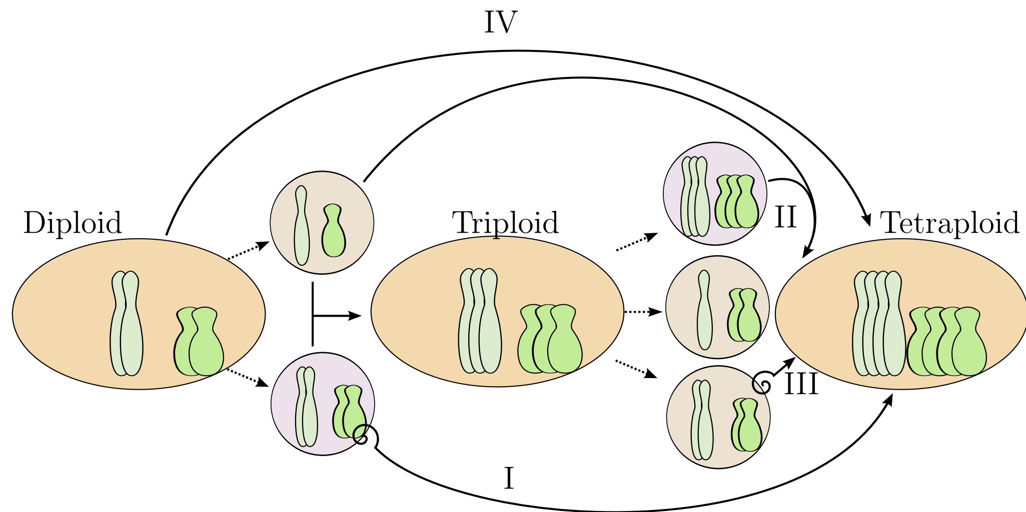


Figure 1.1: Different pathways towards polyploidy. Orange ovals are individuals, light violet disks unreduced gametes carrying the same number of chromosome as the individuals that generated them, brown disks other types of gametes, balanced or unbalanced. Full lines represent fertilization and dotted lines gametogenesis. At a low rate, diploids individuals will give rise to unreduced gametes, carrying 2n chromosomes. If, by chance, fertilization of two unreduced gametes occur, a tetraploid individual will be born (I). However, as unreduced gametes are rare, it is more likely that fertilization will happen with a haploid gamete, leading to the formation of a triploid individual. This individual will produce unbalanced gametes (center, not exhaustive), balanced diploid (bottom) or haploid gametes, or unreduced gametes (top). Tetraploids can thus arise through fertilization of this unreduced triploid gamete with a haploid one (II) or through selfing, if possible in the species under consideration, of diploid gametes (III). Somatic mutations, spontaneous doubling of the chromosome number before the origination of reproductive organs are also theoretically possible (IV) (Mason and Pires, 2015).

with one of the first example of how ploidy levels had an effect on the phenotype of an organism, his theory of “mutationism” was discarded and he is nowadays mostly remembered as one of the scientists having rediscovered Mendel’s inheritance laws.

Polyploidy, the state of carrying more than two copies for every homologous chromosomes, was extensively studied from the twenties (discussed by Ramsey and Ramsey (2014)). Biologists realized early on that many angiosperms species or even genera showed variation in their ploidy levels (Müntzing, 1927), acquired

through natural evolution or human selection. Wheat varieties, providing overall almost 20% of the worldwide calorie intake (Food and Agriculture Organisation, 2011), have been recognized as different polyploid forms almost a century ago (discussed by Feldman et al. (2012)). Gametes of the wild members of the *Triticum* and *Aegilops* genera carry 7 chromosomes (Feldman et al., 2012), whereas *Triticum turgidum durum*, used for pasta production, is an allotetraploid, a polyploid hybrid between two species of the different genera and carries 14 chromosomes (The International Wheat Genome Sequencing Consortium, 2014). The most cultivated wheat, *Triticum aestivum*, used for bread preparation, is an hexaploid hybrid between *Triticum turgidum durum* and another *Aegilops* species (figure 1.2) with 21 chromosomes.

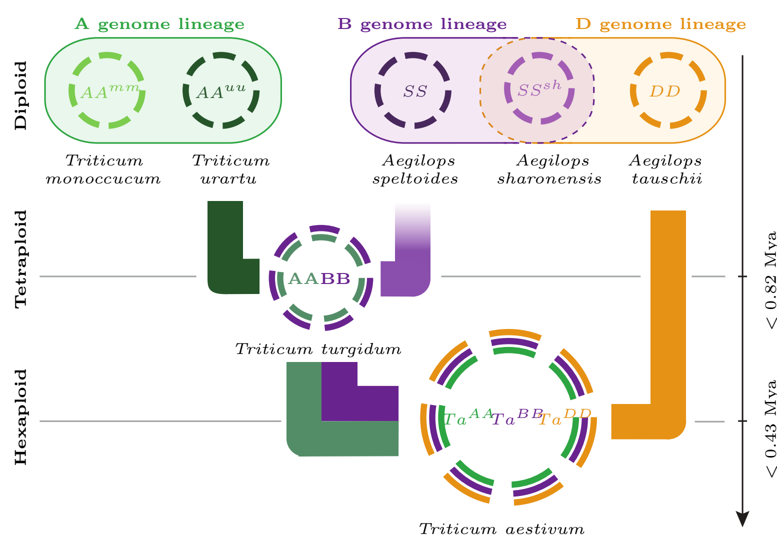


Figure 1.2: Genomic relationships of cultivated wheats with respect to wild members of the *Triticum* and *Aegilops* genera. Chromosome numbers are schematically represented for the gametophytic state. Adapted from The International Wheat Genome Sequencing Consortium (2014)

Counting the number of chromosomes with a microscope was feasible at the beginning of the genetics era, but determining the origin of each chromosome

in the polyploid complex of *Triticum*, for instance, was out of the reach of the early 20th century scientists. Moreover, even though they suspected that polyploidization had an effect on land plant evolution given its present-day prevalence (Müntzing, 1927), they had no tools to determine experimentally if lineages had experienced and fixed whole genome duplications, an event sometimes called paleopolyploidization (see for instance Renny-Byfield et al. (2015)). With the advent of the molecular sequencing techniques, a variety of methods can be used to identify and date the occurrence of past whole genome duplications, such as the analysis of synonymous mutations (K_s or d_S) between duplicated pairs of genes — paralogs — inside a genome (Barker et al., 2009; Vanneste et al., 2013). The massive number of paralogs created after polyploidization — sometimes called ohnologs in honor of Susumu Ohno, who hypothesized early on about the evolutionary consequences of whole genome duplications (Ohno, 1970) — leaves distinctive peaks in the distribution of d_S, and if the substitution rate of the species under consideration is known, the timing of the event can be approximately determined (Blanc and Wolfe, 2004). For whole genome duplications happening far in the evolutionary past, divergence preventing the recognition of paralogs or genetic saturation impedes good results of this technique.

A more formal evidence of paleopolyploidy can be given by studying syntenic regions inside genomes (Abrouk et al., 2010). Syntenic regions are traditionally defined as portions of chromosome carrying a number of different genes loci. The identification, in a single genome or between genomes of various species, of synteny regions exhibiting homologous gene loci, can tell us about the evolutionary past of duplications: if two syntenic regions of a species are systematically homologous to a single region in another species, whole genome duplication is likely to have occurred in the lineage leading to one but not to the other. The Brassicaceae family, comprising more than 3000 plant species, provides a fascinating example of polyploidization and chromosomal rearrangements (figure 1.3). It thus appeared that, even *Arabidopsis thaliana* with as few as 5 chromosomes, can be a paleopolyploid (The Arabidopsis Genome Initiative, 2000): given enough time, diploidization, the phenomenon of going back to a diploid state through chromosomal fusion, loss, or fractionation (Leitch et al., 2004), prevents the iden-

tification of polyploids by mere counting. Both methods, relying on the study of substitutions between ohnologs or simply their recognition, depend on the ability to identify the gene duplicates that are maintained into the genome after whole genome duplications.

As my work will focus onto the long term dynamics of lineages after whole genome duplications, and as the processes happening at the lineage scale will partly hinge on the genomic substrate that will be preserved after polyploidization, I will now mention the current state of the literature regarding how genes are kept following duplications.

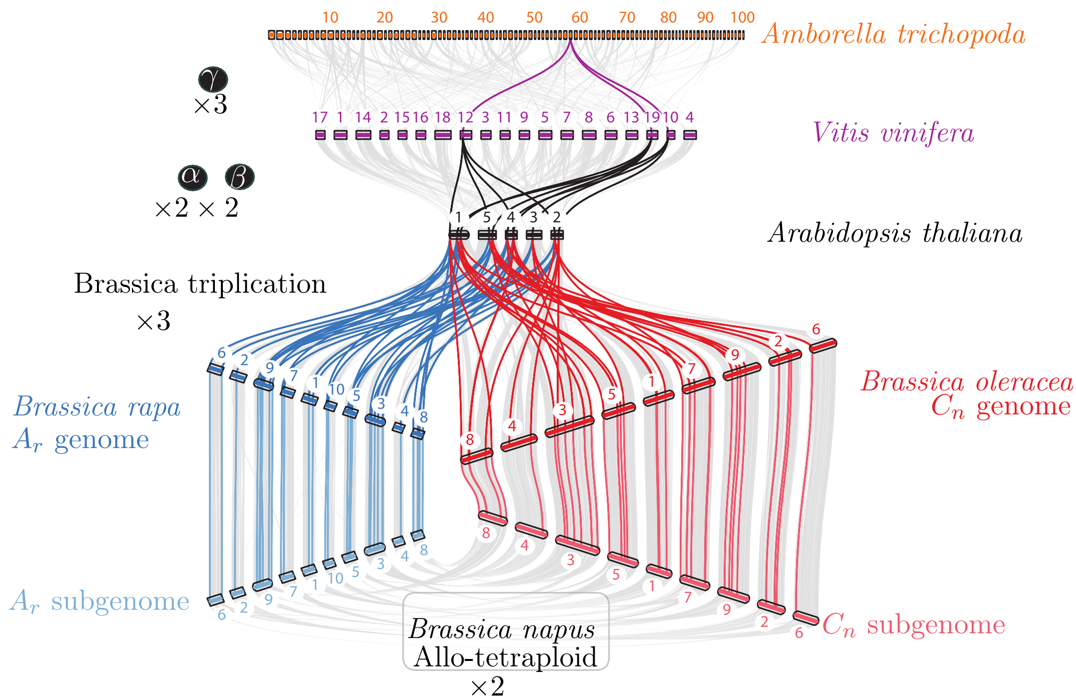


Figure 1.3: (Previous page) Representation of the syntenic regions between 6 land plants species. *Amborella trichopoda* is an endemic New Caledonian species which is assumed to be the sister species of every other angiosperm plant. The gamma triplication is one of the oldest duplication currently identified in angiosperms and is thought to have occurred after the divergence with the lineage leading to *Amborella trichopoda*, hence the 1:3 ratio between syntenic regions of this species and the eudicot *Vitis vinifera* (Amborella Genome Project, 2013). Two additional paleopolyploidizations, termed alpha and beta, have been recognized somewhere along the branch leading to the ancestor of all Brassicaceae, after its divergence with the ancestor of *Carica papaya* (Soltis et al., 2009). Thus most *Vitis vinifera* genomic regions have 4 homologs in *Arabidopsis thaliana*. Extensive chromosomic rearrangements in this latter species led to a karyotype of only 5 chromosomes, whereas *Arabidopsis lyrata* retained the ancestral Brassicaceae karyotype of 8 chromosomes (Murat et al., 2015). Another hexapolyploidy event has been found in the tribe Brassiceae (Lysak et al., 2007), including *Brassica oleracea* (cabbage and allies) and *Brassica rapa*, even though their gametes only carry 9 and 10 chromosomes, respectively. Finally, *Brassica napus* (rapeseed) is an hybrid species dating around ten thousand years ago, whose 19 chromosomes originate from the sum of its parents *Brassica rapa* and *Brassica oleracea* (Chalhoub et al., 2014). Adapted from Chalhoub et al. (2014).

1.3 Evolutionary fates of gene duplicates

A number of models have been proposed to explain fixation of changes in gene dosage — the number of gene copies in a genome. One of the first proposed, neofunctionalization, denoting the retention of the ancestral gene function by a duplicate whereas the release of selective constraints on the second duplicate enables the evolution of a new function, has been frequently used in the literature (Taylor and Raes, 2004). A recently identified example of gene duplicate evolving a new function after whole genome duplication includes a new cellular location and a higher activation capacity for the androgen receptor after the teleost specific polyploidy (Ogino et al., 2016). Similarly, the evolution of a new transcription repressing factor involved in the nitric oxide detoxification pathway was linked to the paleopolyploidization found in the common ancestor of *Candida glabrata* and *Saccharomyces cerevisiae*, baker's yeast (Merhej et al., 2015).

Another process leading to gene pair retention could also be subfunctionalization, where both duplicates together recapitulate the function of the ancestral

gene (Force et al., 1999; Stoltzfus, 1999). Subfunctionalization can occur at different phenotypic levels, as either duplicates can carry on different molecular functions that were all carried by the ancestral gene, or by summarizing its expression patterns across tissues, developmental stages or differentiated sex and caste individuals. Specific instances of this phenomenon include the rapidly evolving binding-GTPase gene pair showing complementarity of expression in different carp tissues, having originated after the cyprinid specific whole genome duplication (Zhao et al., 2015). The evolution of the large MADS-box regulation factors family responsible for flower development in angiosperms has also been recurrently associated with subfunctionalization (Ng and Yanofsky, 2001). Finally, using *Saccharomyces cerevisiae*, Soria et al. (2014) showed that the phenotypic space where subfunctionalization occurs can indeed be very wide: at the level of the function of the proteins, of gene expression or of organismic growth. This phenotypic space might even be wider for multicellular organisms, as subfunctionalization can also happen at the tissue-specific level.

Nowadays, our theoretical corpus has grown from these two models to a decade of them (Innan and Kondrashov, 2010), to resolve the preservation of gene copies by means that can be neutral, adaptive or a mixture of both. For instance, when the availability of the product of a particular gene is decisive for the host, increasing its copy number or modifying its expression level will be selected for. The case of the weedy *Amaranthus palmari* plant provides an extreme example of such positive gene dosage effect: a 160-fold increase in the copy number of its gene producing a necessary enzyme inhibited by an herbicide was found in response to the herbicide pressure (Gaines et al., 2010). Another model, the "escape from adaptive conflict" (Conant and Wolfe, 2008) case, was proposed to explain duplicate maintenance when the ancestral gene product bears different molecular processes. If the generalist condition of the gene product prevents one or more of its functions to be fully optimised, the appearance of a another gene copy enabling the independent evolution of more efficient reactions by the two products will be advantageous. It represents a subcase of subfunctionalization where the fixation of the new gene is favored by selection and is likely to be the mode of action explaining the post whole genome duplication evolution of the

GAL1 and GAL3 genes, involved in galactose metabolism in yeast, for instance (Conant and Wolfe, 2008).

The patterns governing conservation or loss of excess gene copies have been investigated for the first time over the genomes of six eukaryotic species in a seminal study by Lynch and Conery (2000). The authors showed that even if duplicated genes arise at high rates, most of them are subsequently lost, making nonfunctionalization largely dominant over models predicting retention. But they also argued that the same rules might not apply if an elevated number of genes were to duplicate at the same time: as gene products often work in association between themselves, duplication of all the genes working together would keep their outputs at the same relative levels. Conversely, the duplication or the change in expression of a single gene coding a subunit of a protein complex could prove deleterious, by disturbing the assembly of the complex or by directly competing for binding of the catalytic target (Papp et al., 2003). These effects, summarized in the gene balance hypothesis (Birchler and Veitia, 2010), predict that additional gene copy retention is strongly dependent on the function of the encoded product as well as on the processes leading to the duplication (Conant et al., 2014). Genes involved in particular molecular functions that are more likely to be carried by large complexes will be preferentially retained after whole genome duplications, such as expression regulation, ribosomal structures or signaling proteins (McGrath et al., 2014b).

Observing that genes belonging to duplicate pairs evolve significantly slower than singletons in eukaryotes and prokaryotes species, Jordan et al. (2004) and Davis and Petrov (2004) showed that the probability of retention of a duplicate gene was highly correlated to the ancestral function: the more constrained the protein encoded by a gene, the more likely a duplicate of this gene is to persist over time. The globin or opsin families provide fitting examples to this finding: all the different proteins belonging to those families identified nowadays in vertebrates evolved through sequential duplications of their genes in their ancestors, and have clearly defined function in the organism expressing them. Beyond its function, the level of expression of a gene has also been shown to be positively cor-

related to the probablity of one of its additional copy to be kept in the genome of the host, proving that highly expressed genes are more sensitive to changes in gene copies (Gout et al., 2010), resulting in the high retention of metabolic genes after whole genome duplication, that are usually more expressed than non metabolic genes (Gout et al., 2009).

Another process explaining the non-random loss of gene or the expression changes observed after allopolyploidy is the phenomenon of genome dominance. It can happen simply at the expression change level, as in the case of the hexaploid wheat, where each of the three genomes has been found to be dominant over certain cell types or development stages (The International Wheat Genome Sequencing Consortium, 2014), but it could also favour the maintenance of only one of the ohnologs over longer evolutionary timescales. Because the reunited genomes might have different expression levels of their ohnologous genes, due to different methylation levels controlling transposable elements (Woodhouse et al., 2014) for instance, loss of the most highly expressed gene will be more deleterious than loss of its ohnologs (Freeling et al., 2012). Genome dominance could then lead to loss of the genes belonging to the recessive genome over the long evolutionary term. Most studies indicate that after whole genome duplication, gene loss occurs via DNA deletion through non homologous recombination in plants (Freeling et al., 2015), but whole genome duplications are also known to occur in other organisms.

After studying what processes are involved in the fixation or the loss of gene duplicates, I will now introduce how these various processes might affect the evolutionary fates of the organisms experiencing them after duplication of their whole gene repertoire.

1.4 Fates of populations and species after polyploidization

The fact that most of actinopterygians diversity is actually included in the teleost clade, whose ancestor arose after a paleopolyploidy event (Taylor et al., 2003), or that angiosperms, where polyploidy is rampant, is a highly successful group, led researchers to hypothesize on the possible link between the events of genome duplication and the evolutionary success of the lineages that arose after such events (Hoegg et al., 2004; Santini et al., 2009; Soltis et al., 2009). Moreover, the ancestor to all vertebrates also experienced paleopolyploidy, once (Smith and Keinath, 2015) or maybe twice (Dehal and Boore, 2005). Paleopolyploidy is responsible for the architecture of the *Hox* genes loci in vertebrates, controlling the development of the correct body plan and that are found in a single locus in most animals' genome, but in up to four copies in vertebrates or even seven in teleostei (Amores et al., 1998). As the evolution of the vertebrate body plan led to the appearance of a large species-rich clade compared to its closest relatives, Tunicates and Cephalocordates (Cañestro et al., 2013), this paleopolyploidization has also been linked to higher diversification (Crow and Wagner, 2006).

Conceptually, the occurrence of polyploidization could almost instantaneously lead to the appearance of a new species, as triploid hybrids between polyploids and diploids are very likely to suffer from decreased fertility (Ramsey and Schemske, 1998). Alternatively, polyploid lineages could be more inclined to form new species because of intrinsic properties of polyploidy, such as the elevated gene content. Tentatively, the extra-duplicated material could facilitate the apparition of new functions, according to the neofunctionalization model, and the filling of new ecological niches. Nevertheless, adaptive radiations, defined as the rapid appearance of new phenotypes through ecological speciation (Rundell and Price, 2009) and classically illustrated by the cichlids (Brawand et al., 2014), the anolis (Pincheira-Donoso et al., 2015) or Darwin's finches species, are traditionnaly not associated with whole genome duplications. Another mechanism that could favour the apparition of new species after whole genome duplication has been called reciprocal gene loss (Lynch et al., 2000) (figure 1.4). Although initial evi-

dence of this mechanism was found in different organisms (Scannell et al., 2006; Sémon and Wolfe, 2006), recent theoretical modelling cast some doubts upon the significance of reciprocal gene loss in building species barriers (Muir and Hahn, 2015).

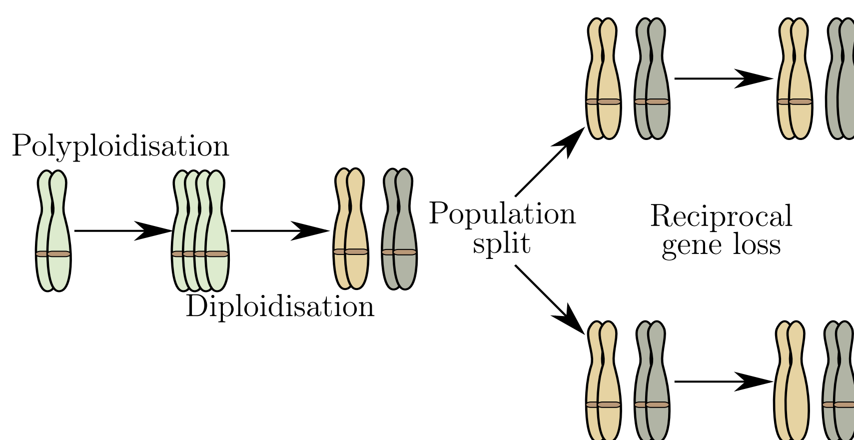


Figure 1.4: Reciprocal gene loss, a model to explain barriers to the gene flow after whole genome duplication. This models builds on the first genetic model to explain the appearance of new species, the Bateson-Dobzhanksi-Muller model (Orr, 1996). This model originally posits that when populations get isolated mutations appearing in different members of a pair of interacting genes could rise to fixation in each population, for adaptive reasons or simply by drift. Upon secondary contact of the two populations, the new alleles of each pair member never interacted before and could lead to incompatibilities in the organism carrying them, thus leading to gene flow barriers between the populations. Similarly, when genome duplication occurs, a likely outcome for duplicated genes is their loss. If, once disomic inheritance is restored, different populations lose one of the two paralogs at different loci, crosses between those two populations would give birth to a significant number of offsprings lacking this gene. Given the high number of paralogs generated during whole genome duplication, this could theoretically favour the apparition of new species.

Actually, debates regarding the evolutionary consequences of polyploidy started early in the scientific community: Stebbins (1950), acknowledging that around 30 to 35% of angiosperms species are polyploids, stated that most long term evolution happened at the diploid level, even though he admitted that some families of plants had a polyploid origin. Although the idea is pervasive in the present-day literature (see for instance te Beest et al. (2012); Madlung (2013); Vanneste et al.

(2014); Kagale et al. (2014); Dodsworth et al. (2015)), Stebbins never deemed polyploidy as an "evolutionary dead-end" (discussed by Soltis et al. (2014b)). In opposition to this idea of minor effect of polyploidy in the long run, Levin (1982) strongly emphasized polyploidy as a powerful source of new species and innovations, and a facilitator of colonization of new environments. More recently, people have argued that polyploidy in plant lineages have increased their chances to survive through the mass extinction that occurred at the Cretaceous-Paleogene boundary (Fawcett et al., 2009; Vanneste et al., 2014) or that it can drive adaptation to new environment in yeast (Selmecki et al., 2015) (but see Appendix A for a discussion of this finding).

Another hypothesis proposed for explaining the link between diversification and polyploidization is the "radiation time lag" model (Schranz et al., 2012). Recognizing that in land plants, many paleopolyploidization are associated with the appearance of very large group of species, whose sister clades are usually species poor — such as Asteraceae, Brassicaceae, Fabaceae, Poaceae, Solanaceae or even angiosperms themselves — this model postulates that contrary to adaptive radiations, the effect on diversification will only be seen in the distant future after whole genome duplication and some required steps. It posits that the whole genome duplication is necessary for the evolution of new key traits, but that the radiation will only be triggered millions of years later, after a dispersion event that will displace polyploids in a new environment (Schranz et al., 2012). Similarly, Dodsworth et al. (2015) hypothesized that it is the diploidization process that is responsible for the delay in time between the paleopolyploidy and the diversification. Underlying that polyploids at birth face many difficulties — meiosis defects (Grandont et al., 2013; Stenberg and Saura, 2013), shock associated to the merging of diverged genomes (Parisod et al., 2010), chromosomal rearrangements (Leitch and Leitch, 2008) or minority cytotype exclusion (Husband, 2000) — polyploids must first survive this depression phase before they can thrive. Return to a normal disomic inheritance, higher genomic and transcriptomic variation can then supposedly favor the success of paleopolyploids (Dodsworth et al., 2015). Similarly, reciprocal gene loss can only occur after diploidization and could neutrally lead to the appearance of reproductive isolation.

Studying how polyploidy affects evolutionary success implies defining and identifying formally its variation in the tree of life. I will introduce how I plan to use lineages' diversification — their propensity to create new species — as a proxy for evolutionary success and how it can be estimated through the ages. Using these methods, I will present the work I performed to deepen our knowledge between polyploidy and diversification.

1.5 Testing the link between polyploidy and diversification

The testing of differential evolutionary successes between lineages requires the development of appropriate methods. One of the most natural approaches aiming at reconstructing historical species diversity trends is studying the fossil record. Nevertheless, the sparsity of the fossil record for plants, harbouring a significant part of ploidy variation in the tree of life, and the difficulty of recognizing polyploidy in the record — except by using cell size as a proxy for ploidy level, see Scott et al. (2015a) for instance — prevents researchers from systematically investigating this relation using such an approach. More generally, the fossil record being incomplete, models incorporating other kinds of data can also be used to understand biodiversity patterns.

The first model predicting species growth was developed by Yule (1925). Discarding the effect of species extinction, he devised how monotypic genera grew based on the probability of a new species appearing. Based on the distribution of species number in genera of a few animal families, he was one of the first to estimate macro-evolutionary rates — parameters governing trends at or above the species-level — such as the time of species number doubling in a genus or the ratio of "specific mutation" — species appearance — to "generic mutation" — genus appearance (Yule, 1925). Nowadays, macro-evolutionary models discarding the effect of extinction are named in his honor. Kendall (1948) solved

the problem of any population growth given the birth (λ) and death (μ) rates of the individuals of the given population. Kendall's work is the basis of subsequent methodological breakthroughs in a variety of fields, such as modelling the waiting time in any given queue or the estimation of speciation and extinction rates using phylogenies (Nee et al., 1994b), which is of particular interest in the present context as identifying changes in these rates will allow us to find elevated evolutionary success in a phylogeny. Indeed, Nee et al. (1994b) devised the likelihood equations governing the distribution of speciation times based on constant λ and μ, and thus constituting the null model of how phylogenies grow with the apparition and disparition of species. Since this important finding, a large number of methods have been developped to account for more and more complicated modes of diversification (for a full review, see Morlon (2014)).

Since then, researchers were given a handy tool to investigate biodiversity patterns without exclusively using fossil data. By reconstructing phylogenies of extant species using molecular data and possibly dating them using fossil knowledge, one is able to estimate the times of appearance of new species and thus infer macro-evolutionary rates. Rather than working with speciation and extinction rates, researchers have taken the habit of using diversification, the difference between speciation and extinction, usually termed r, and turn-over, the ratio of extinction to speciation, termed ϵ, for convenience. Indeed, diversification represents the number of species created per units of time and thus can be used as a proxy for the effective evolutionary success of lineages. From these estimations a surprising observation springs: extinction rates can be inferred from molecular data alone, by definition including at best only very recently (on geological scale) extinct species. This apparent paradox has been hotly debated in the literature (Nee et al., 1994a; Paradis, 2004; Rabosky, 2010; Beaulieu and O'Meara, 2015) and the idea behind is that the shape itself of the tree will be modified by the extinction rate of the lineages. For instance, at constant diversification rate, the higher the extinction, the closer to the present the speciation nodes, because old species will tend to be replaced by new ones more often.

Using these variety of models, we are now capable of explicitly testing the

association between characters and diversity. Thus, the debate regarding the potential link between diversification and polyploidy in angiosperms has been fueled in the past years by the publication of Mayrose et al. (2011) which has prompted extensive arguments on the philosophical, technical and experimental implications of this debate (Soltis et al., 2014a; Mayrose et al., 2014). Other publications, using different methods or metrics that are somewhat related to polyploidy, such as genome size, have been issued in the perspective of this debate (Knight et al., 2005; Soltis et al., 2009; Kraaijeveld, 2010; Pandit et al., 2014; Scarpino et al., 2014; Bromham et al., 2015; Puttick et al., 2015), with contrasting results. This debate has many implications, in particular on how to design computational experiments to test the effect of young polyploidization events.

During my work, I will use a combination of simulated and publicly available data to further our understanding about diversification models and their impact of polyploidy. From the wealth of tools now available for reconstructing diversity trends through phylogenies, we can tentatively design experiments to test specific hypothesis, such as "Are diversification patterns altered after polyploidy and in which way?" or "Did polyploidy lower extinction rates in plants at the Cretaceous-Paleogene boundary?". In chapter 2, I will explore methods commonly used when looking for patterns of diversification using simulations and assess the power of our current means of detecting ancient events. Specifically, I will study how the joint occurrences of different macro-evolutionary processes, such as mass extinctions or abrupt changes in speciation or extinction rates between lineages affect the shape of reconstructed phylogenies and the outcomes of our methods. This work was motivated by the finding that many paleopolyploidy events in plant lineages cluster around 66 Million years ago (Vanneste et al., 2014): in order to test for the correlation between trait (diploidy or polyploidy) and events (mass extinctions), one has first to be able to check if these mass extinctions are recognizable in the molecular data. I developped a simulation algorithm to study the influence of many parameters, such as number of extant species, number of events occurring during the evolution of the lineages, values of the macro-evolutionary parameters, to conclude on some of the methods used to detect mass extinction events and changes in diversification, building upon, and extending a bit, the

flourishing literature discussing diversification methods.

Then, I will move on to explicitly test one assumption of the radiation time lag model in teleost fishes. Both reciprocal gene loss and radiation time lag will predict that the diversification increase should be delayed after whole genome duplication. Such signal of diversification was found in angiosperms (Tank et al., 2015), although using methods estimating changes in macro-evolutionary rates in a common ancestor to living species, whose accuracy has been questioned (May and Moore, 2016). In chapter 3, I will try to find evidence of such pattern in Teleostei, by using methods that can recover changes in macro-evolutionary rates through time and thus potentially find the delayed responses assumed by the radiation time lag. I will study the changes in speciation and extinction following the few old and more recent whole genome duplications occurring in teleosts using a variety of data collected from the literature, and discuss the implications of differences in age of paleopolyploidies.

In chapter 4 of this manuscript, I will present another angle of the debate regarding polyploidy and diversification. I will focus on the nature of the processes leading to the apparition of polyploid species and investigate it using phylogenies of plant genus showing ploidy variation. I will use diversification methods that estimate macro-evolutionary parameters, and compare models using different assumptions on how character changes arise. Namely, I will try to find out to what extent polyploid lineages arise from cladogenetic or anagenetic processes. For this I will use already published methods and data, as well as newly computed data.

In general, the aim of my manuscript is to provide technical advancements and insights on the polyploidy and diversification conundrum. Chapter 2 of my thesis will focus primarily onto methodological considerations, whereas chapters 3 and 4 will use biological data to test concrete models of diversification.

Chapter 2

Detecting patterns of species diversification in the presence of both rate shifts and mass extinctions

This chapter was published as a research article in BMC Evolutionary Biology.

2.1 Abstract

Recent methodological advances allow better examination of speciation and extinction processes and patterns. A major open question is the origin of large discrepancies in species number between groups of the same age. Existing frameworks to model this diversity either focus on changes between lineages, neglecting global effects such as mass extinctions, or focus on changes over time which would affect all lineages. Yet it seems probable that both lineages differences and mass extinctions affect the same groups. Here we used simulations to test the performance of two widely used methods under complex scenarios of diversification. We report good performances, although with a tendency to over-predict events with increasing complexity of the scenario. Overall, we find that lineage shifts are

better detected than mass extinctions. This work has significance to assess the methods currently used to estimate changes in diversification using phylogenetic trees. Our results also point toward the need to develop new models of diversification to expand our capabilities to analyse realistic and complex evolutionary scenarios.

2.2 Background

The estimation of the rates of speciation and extinction provides important information on the macro-evolutionary processes shaping biodiversity through time (Ricklefs, 2007). Since the seminal paper by Nee et al. (1994a), much work has been done to extend the applicability of the birth-death process, which now allows us to test a wide range of hypotheses on the dynamics of the diversification process.

Several approaches have been developed to identify the changes in rates of diversification occurring along a phylogenetic tree. Among them, we can distinguish between lineage-dependent, trait-dependent, time-dependent and diversity-dependent changes. Lineage specific methods identify changes in macro-evolutionary rates — speciation and extinction rates, denoted as λ and μ, respectively — at inner nodes of a phylogenetic tree (Rabosky et al., 2007; Alfaro et al., 2009; Silvestro et al., 2011). We can also identify trait-dependence in speciation and extinction rates if the states of the particular trait of interest are known for the species under study (Maddison et al., 2007; FitzJohn et al., 2009; Mayrose et al., 2011). It is also possible to look for concerted changes in rates on independent branches of the phylogenetic tree by dividing it into time slices (Stadler, 2011a). Finally, diversity-dependent effects can be detected when changes of diversification are correlated with overall species number (Etienne et al., 2012). Most methods can correct for incomplete taxon sampling, by assigning species numbers at tips of the phylogeny (Alfaro et al., 2009; Stadler and Bokma, 2013), or by introducing a sampling parameter (Nee et al., 1994a). By taking into account this sampling parameter at time points in the past, it is also possible to look for

events of mass extinction (Stadler, 2011a).

These methods provide insights into the dynamics of species diversification and it is now well accepted that differences in lineage-specific rates exist (Jetz et al., 2012; Barker et al., 2013). However, it seems unlikely that both lineage specific shifts and mass extinction events would not have occurred, especially when studying large phylogenetic trees covering hundreds of million years of evolution. For example, several global crises, which caused the extinction of a high proportion of species (Raup and Sepkoski, 1982), have occurred since the appearance of the last common ancestor of vertebrates. Among them, the Cretaceaous-Paleogene (K-Pg) boundary and the Permian-Triassic events, which happened 65 million years ago (Mya) and 251 Mya, respectively, induced the most dramatic losses of biodiversity (Erwin, 2006). Moreover, other less extensive events have also occurred in the past hundred million years (Benton, 1995).

Alternative models have been proposed for mass extinctions. They could be represented as a high number of species disappearing at the same time (single-pulse model), or as an increase of the background rate of extinction during an extended period of time (time-slice model) (Condamine et al., 2013). They could also impact biodiversity in different ways. Three main hypotheses, corresponding to different patterns of extinction, have been proposed (Raup, 1992). First, the event could affect all lineages equally and terminate any extant lineage with the same probability. This "field of bullets" scenario is often used as a null model (Nee, 1997; Faller et al., 2008). Second, in the "fair game" scenario, some form of lineage selection would occur, where the most successful species — in our case, the most diversifying species — before the event would be the most likely to survive. This could, for instance, happen if the probability of survival depends on a specific trait varying across the lineages of the phylogeny (Faller and Steel, 2012). Finally, in the "wanton destruction" scenario (Eble, 1999), the event could induce such changes in the environmental conditions that the probability of extinction of the species and their post-event diversification rate would be uncorrelated to their initial speciation and extinction rates.

Although lineage-dependent differences in macro-evolutionary rates and mass extinctions are known to happen, the performances of the existing methods to identify both lineage-specific rate shifts when mass extinctions have occurred, and mass extinctions when lineage-specific rate shifts have occurred has not, to our knowledge, been investigated. The aim of this study was thus to assess the performance of current methods to estimate the rates of diversification using complex scenarios involving both mass extinctions and lineage shifts. We used simulations to assess the impact of varying number and magnitude of rate shifts and mass extinction events.

2.3 Methods

Figure 2.1 gives an overview of the simulation design. We used a backward algorithm to simulate phylogenetic trees as implemented in the function *sim.rateshift.taxa* from the R (R Core Team, 2013) package TreeSim (Stadler, 2011b), since a direct forward approach to simulate trees using a birth-death process can lead to bias when conditioning on the number of tips (Hartmann et al., 2010). Forward simulations can be used to simulate trees when conditioning on the total amount of time of the process. However, this approach is not practical in this context as the procedure would result in trees with highly variable numbers of taxa, in particular when adding mass extinction events. A backward simulation procedure is therefore the best solution to simulate the different diversification scenarios of interest for our study. This procedure enables both single-pulse or time-slice modelling of mass extinctions, but we chose to represent them only using the single-pulse model because paleontological data indicates very high species loss at major mass extinction events in a limited amount of time. For instance, a 52% decrease in marine families was observed at the Permian-Triassic boundary (Raup and Sepkoski, 1982). Further, there is currently no approach available to simulate continuous birth-death process that incorporates rate variation among lineages. We thus designed a new backward algorithm that we detail below.

Our algorithm takes as input the number of extant species, the evolutionary

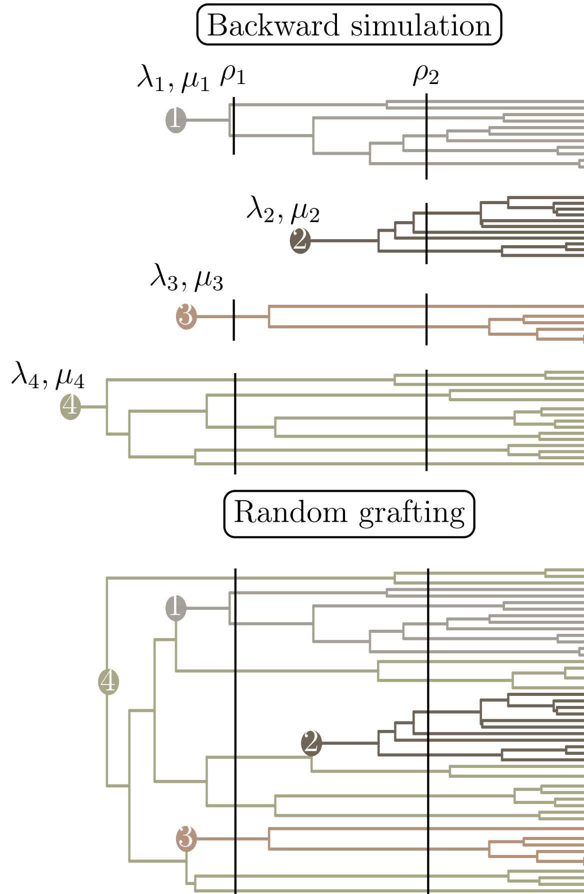


Figure 2.1: Workflow of the simulation process. Hypothetic case of 50 species tree, 3 lineages shifts and 2 mass extinctions. The number of species in each lineage is randomly drawn first. Each tree is grown separately with different (λ, μ) but with identical survival rates (ρ) at each mass extinction events. The four trees are then successively joined at branches ensuring ultrametricity. Vertical continuous lines: simulated mass extinction events, full circles: ancestor where diversification change occurred.

rates λ and μ, and the time of occurrence and survival rate ρ for mass extinction events. We assumed in the first part of our simulations that these events happened according to the field of bullet scenario (step 1). We randomly grafted different trees having experienced the same mass extinction events but different evolutionary rates to account for rate shifts in diversification (step 2; see Table 2.1). First, we ran as many backward simulations as the number of lineages shifts in our tree. We defined the number of species in each backward simulation

by drawing samples from a Dirichlet distribution to keep the total sum equal to the overall number of species. We then ranked the trees by decreasing order of their total age, which included the stem branch length provided by TreeSim. We selected from the oldest tree (referred to as acceptor tree) the branches that overlapped in time with the age of the stem branch of the second oldest tree (referred to as donor tree). Thus, the branches considered for possible grafting were the ones that included the age of the donor tree between the timing of the two speciation events defining them in the acceptor tree. We randomly chose one of those branches to graft the donor tree onto the acceptor. This ensures ultrametricity of the newly created tree and leaves the branch lengths of each separate tree unmodified once the lineage having experienced the diversification shift is removed. We iterated over this protocol until all donor trees, whose number varied in our simulations between 0 and 5 (Table 2.1), were grafted. Finally, we ran Medusa (Alfaro et al., 2009) and TreePar (Stadler, 2011a) analyses on each simulated tree to investigate our capacity to recover the signal of mass extinctions and diversification shifts (Fig. 2.2). We simulated trees with different numbers of lineages and extinction events to assess the influence of these factors. Table 2.1 summarizes the parameter space explored for the $16,371$ trees that we simulated. For the values of λ and μ, we targeted distributions similar to the estimates calculated on a mammalian phylogeny (Bininda-Emonds et al., 2007).

Medusa is a maximum likelihood-based framework to detect shifts in diversification by iteratively adding breakpoints on inner branches of the tree with different rates of speciation and extinction. It uses ΔAIC to discriminate between models with an increasing number of parameters (Alfaro et al., 2009). Rabosky (2014) also recently presented a new method (BAMM) to estimate the number of possible rate changes along a phylogenetic tree and to fit exponential responses in macroevolutionary rates to time or to species number. Unlike Medusa, BAMM uses a Bayesian framework, with reversible jump Markov chain Monte Carlo to estimate the number of shifts in diversification in the phylogeny. In our design, we chose not to simulate varying speciation and extinction rates except at speciation nodes, thus using higher complexity models is not necessary. Comparisons between BAMM and Medusa have been performed, but only

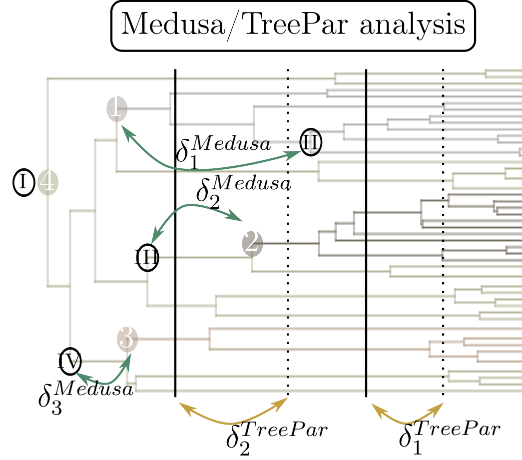


Figure 2.2: Exemple output of the analyzes. We run the Medusa and TreePar analysis, and group the pairs of simulated/estimated events by minimizing the sum of the distance separating the events in each pair ($\sum_i \delta_i^{Medusa}$ and $\sum_i \delta_i^{TreePar}$). Vertical dotted lines: estimated mass extinction events by TreePar, dotted circles with roman letters: estimated diversification rate shift by Medusa, by decreasing significance, other: as in figure 2.1. The first estimated shift is always at the root of the tree.

on simulations involving either time-dependent or diversity-dependent rates (Rabosky, 2014). This framework led to a clear bias in favor of BAMM as Medusa can not evaluate such models, and resulted in Medusa estimating a lower number of events than what was actually simulated (Rabosky, 2014). The numbers of estimated shifts obtained with Medusa can therefore be considered as conservative. Finally, we do not expect a different behavior for Medusa and BAMM regarding the identification of mass extinction events, as neither method incorporates them in their model. Those reasons, as well as the large computational burden to run Bayesian analyses on over $16,000$ trees, led us to favor the simpler Medusa framework for the rest of the study. Medusa was run until a more complex model was not supported by the ΔAIC. We did not extract the macro evolutionary rate estimations from Medusa as we were only interested in testing the ability of the method to detect the events, and not the accuracy of the parameter estimation.

TreePar uses the birth-death process to identify changes in λ and μ through

time. This is done by estimating the probability of a change in parameter values within small time intervals, which can be extended to test for the occurrence of mass extinction events (Stadler, 2011a). The parameters of the rate shifts might be correlated with those related to mass extinction (Stadler, 2011a), which will be a problem for our simulations. We therefore restricted our analysis to the identification of mass extinction events to avoid this issue. The number of iterations of TreePar was set to the simulated number of mass extinction events plus one to test for the appearance of false positive events. A standard Likelihood Ratio Test (LRT) is used to extract the most likely models from TreePar and more complex models were favored when their p-value was less than 0.01, following the standard approach for this framework (Stadler, 2011a). Similarly to what was done with Medusa, we did not analyze estimations of survival rates at mass extinctions events given by this framework.

To verify that our simulation design had no effects on the methods evaluated, we tested the influence of the subtree grafting approach with a constant rate of diversification. We simulated trees with 200 species using both the standard procedures implemented in TreeSim and by grafting two subtrees of 150 and 50 species having evolved under the same λ and μ values. We then compared the results obtained by TreePar and Medusa. We ran 250 pairs of simulations and we observed no significant differences in the number of false positive found between the groups with and without artificial grafting (7 and 13 for Medusa respectively, and none in both cases for TreePar), showing that our simulation design does not bias the estimation of the rate shifts by the two methods used.

We used a slightly different framework to study the impact of the different types of mass extinction events. We simulated a scenario that aimed at testing for the presence of the K-Pg mass extinction event using high order phylogenetic trees. We therefore simulated trees with a large number of extant species ($5,000$ tips, similar to the number of mammalian species) and a large number of lineage shifts (5), but only one event of mass extinction. The other parameters were still drawn at random from the ranges specified in Table 2.1, except for the survival rate ρ that was modified according to the models of mass extinction. For the fair

game hypothesis, we randomly drew λ and μ for the 5 different lineage shifts, but the survival rate ρ was modified for each lineage based on its diversification rate (r, $\lambda - \mu$). We thus considered that the trait influencing the probability of extinction for each species was its diversification rate. For the wanton destruction hypothesis, the mass extinction event induced a change in rates for each lineage, again drawn according to the distribution stated in Table 2.1, and their survival rate ρ was then based on their new diversification value. For the wanton destruction, our simulations included both a global rate shift and a mass extinction and we ran TreePar twice in order to detect both events. For the two latter cases, we chose to linearly parametrize ρ with regards to diversification. As diversification could range between 0 and 0.25 and ρ between 0 and 1, we applied a factor four to the diversification to obtain the survival rates of the lineages. We also ran Medusa on the three sets of simulations to assess the potential impact of the three extinction hypotheses on the detection of lineage shifts. For this second part, we generated over 700 trees for each model of mass extinction event, for a total of 2289 simulations.

Parameter	**Possible values**
λ	$Unif(0.05, 0.25)$
μ	$Unif(0, 0.05)$
ρ	$Unif(0.2, 0.9)$
Number of tips	$200, 500, 1000, 2500, 5000$
Mass extinction event number	0 to 5
Rateshift event number	0 to 5
Mass extinction event time	$Unif(0, min(\frac{Log(N_i)}{\lambda_i - \mu_i}))$

Table 2.1: Universe explored for parameters values. $Unif$: Uniform distribution, i: lineage identifier

2.4 Results and discussion

2.4.1 Baseline performances

To estimate the baseline behavior of both frameworks, we first tested the performance of the methods on the simplest scenarios. We thus selected simulations that included a single rate shift for Medusa, or a single mass extinction for TreePar. Figure 2.3 represents the fraction of shifts detected by Medusa relative to the absolute difference between the new and the old diversification values (Figure 2.3A) and to the number of species in the lineage (Figure 2.3B). More than 80% of the changes in diversification larger than 0.05 are detected by Medusa, which shows a good performance in assessing strong shifts. Further, Figure 2.3B shows that the overall tree size has no influence on the detection, since lineages of the same size are as likely to be detected in small or larger trees.

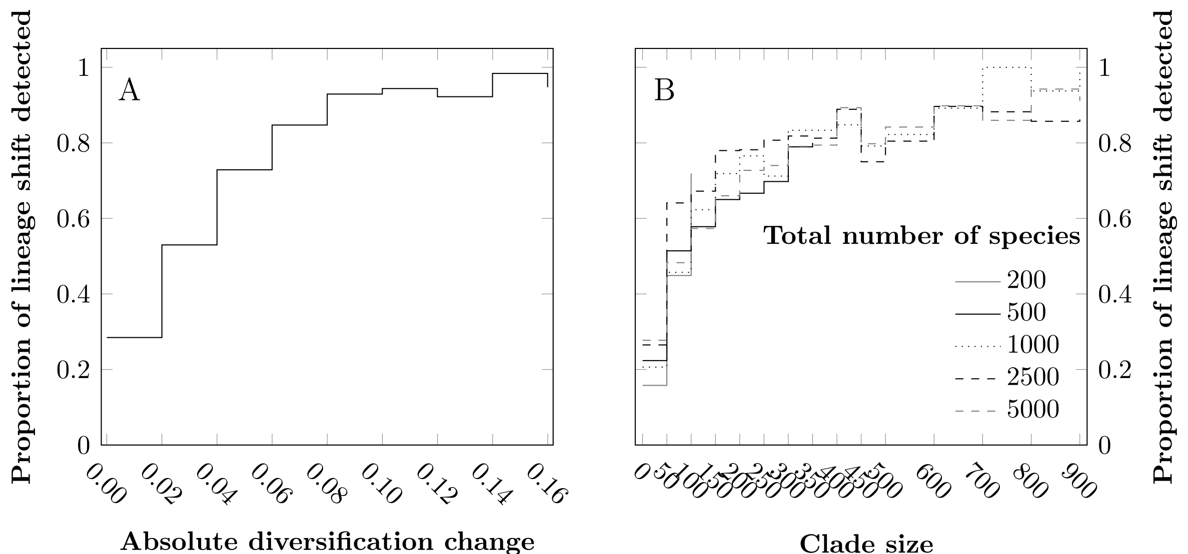


Figure 2.3: Baseline detection level for Medusa, for simulations with one rate shift and no mass extinction event. A: Proportion of detected events for ranges of values of diversification, B: Proportion of detected events for ranges of extant species number in lineages.

We then checked the ability of TreePar to detect mass extinction as a function of the survival rate, ρ, as well as of the number of ancestral species predating this

event in the reconstructed tree. We also used first the simplest simulation to limit the effect of other parameters. Figure 2.4A shows that the signal of mass extinction in the phylogenetic tree is very weak when less than 100 ancestral species are present before the event. This has implications for our ability to find evidence for the K-Pg boundary using phylogenetic trees of vertebrates, for example. We can only reach more than a hundred ancestral species older than 65 My by considering phylogenetic trees encompassing distantly related lineages of tetrapods (see Bininda-Emonds et al. (2007) or Meredith et al. (2011)). Besides, as detection drops with increasing survival rate (Fig. 2.4B), the signal is even less likely to be picked as the ancestors of the extant species might have experienced the mildest extinction rates.

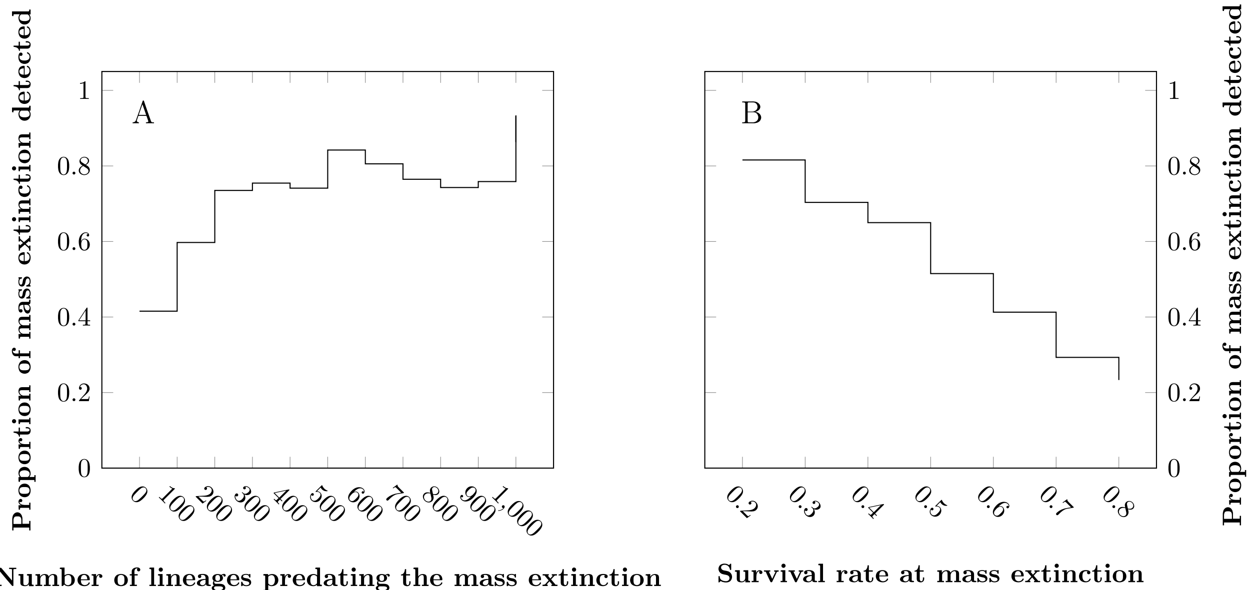


Figure 2.4: Baseline detection level for TreePar, for simulations with one mass extinction and no diversification shift. A: Number of lineages predating the mass extinction event influence, B: Survival rate influence.

2.4.2 Mixed scenarios of diversification

In a second stage, we analyzed simulations with more events and a mix of different types of events. We evaluated the performance of rate shift detection by Medusa,

or of mass extinction events by TreePar, by comparing the events detected to the relevant simulated events. To perform the assignment between detected and simulated events (see Fig. 2.2), we chose to minimize the sum of the distances between each potential pairing of events ($\sum_i \delta_i^{Medusa}$ and $\sum_i \delta_i^{TreePar}$). The distance metric used for Medusa was the sum of the branch lengths along the shortest path separating the two nodes, whereas we used the time between the estimated and simulated pairs of mass extinction events for TreePar (see caption of Figure 2.2 for details).

The simulations incorporated several factors and we tested the effect on the framework of three categorical parameters: total number of tips, number of mass extinctions and number of shifts in diversification rate (see Table 2.1 for their possible values). To ensure that the effects observed were related to the parameter of interest, we designed a reshuffling scheme for each parameter. First, we randomly selected an equal number of simulations for each combination of every possible value of the other two parameters. As an example, to study the outputs for trees of 200 tips, we randomly drew an equal number of simulations with (i) no lineage shift, no mass extinction and 200 tips; (ii) one lineage shift, no mass extinction and 200 tips; (iii) one lineage shift, one mass extinction and 200 tips; etc. This draw was repeated a hundred times and we determined, for each bin created, the proportion of simulations for which each method favored the model with the correct number of relevant events it was looking for, and the proportion of simulations for which they favored a model with too many events. Finally, we report the median and 95% intervals of those proportions based on our hundred bins.

2.4.2.1 Tree size influence

Both Medusa and TreePar perform better in assessing the correct number of events they are set to detect with an increasing number of tips (Fig. 2.5). The median proportion of simulations correctly assessed reaches 60% for Medusa and 32% for TreePar with $5,000$ tips. The increase in the number of tips also leads

to an increased acceptance by TreePar of models with too many mass extinctions (28% for $5,000$ tips). However, the number of tips in the tree has no effect on the error of the estimated time of mass extinction (Fig. 2.6), even though more events are predicted. We only see a slight effect of tree size for Medusa, which is probably due to the fact that the method only detects lineage related events and does not depend on the total number of tips. We also investigated the effect of lineage size on the outputs of Medusa. We first compared the variance of lineage sizes relative to the overall tree size, contrasting the simulations with false positives to those with the correct number of rate shifts found. To remove the effect of lineage number, we compared groups of trees with the same number of diversification shifts. To account for a potential effect of tree imbalance, we compared the variance in lineage sizes inside trees, with or without false positives. There is no effect in most cases, except in the simulations with 4 or 5 rate shifts (p-values: 0.01 and $3.6 \cdot 10^{-3}$, respectively, Mann-Whitney test). Thus, simulations with lineages of similar size are more likely to yield false positives only when they include more than 4 rate shifts. We also compared the variance in lineage sizes between simulations for which we recovered the correct number of events against those for which we recovered too few events. For every possible number of lineages, we find significantly lower variance for simulations that were correctly assessed. Thus, we only see a slight effect of the lineage size on the occurrence of false positives, whereas high variance in lineage size significantly increases false negatives. This indicates on the one hand, a tendency to overestimate the number of shifts when lineages are comparable in size, and on the other hand, problems with Medusa for identifying diversification shifts specific to a low number of species, as showed in the first part.

2.4.2.2 Impact of events violating the model

We tested the robustness of the methods by studying the behavior of (1) Medusa to detect rate shifts with an increasing number of mass extinctions, and (2) TreePar to detect mass extinction events with an increasing number of lineages shifts. The results of Medusa are unaffected by the number of mass extinctions

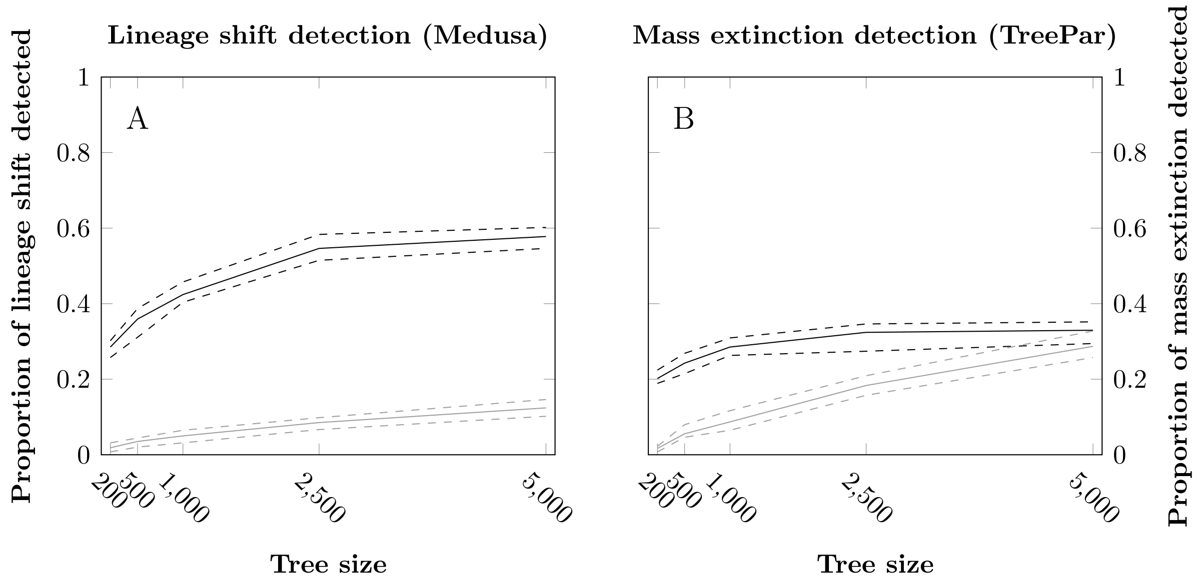


Figure 2.5: Influence of tree size on the detection of lineage shifts (A) and mass extinction events (B). Continuous lines correspond to median proportions of simulations and dotted lines correspond to 95% confidence interval, both based on resampling. Dark lines represent the proportion of simulations where the model with the correct number of events was the most favored, and light lines where a model with too many events was favored.

in the simulations (Fig. 2.7). In contrast, an increase in the number of lineage shifts results in an increase of the proportion of false positives for TreePar (2% with no lineage shift vs. 20% with five; Fig. 2.7). However, the accuracy of the estimate of the timing of the event is not affected (Fig. 2.8). The number of lineage shifts has almost no impact on the probability of detecting a true mass extinction event, *i.e.* on false negatives.

We note that false positive rates remain very low throughout all cases for Medusa, less than 10% overall and around 5% when dealing with simulations without mass extinctions (Fig. 2.7A). Recently, May *et al.* May and Moore (2016) have also studied the performances of Medusa but with a different focus. Medusa also enables the characterization of diversification changes on incomplete phylogenies by letting the user assign species diversity at each tips of the tree. Two different equations are then used to calculate the likelihood function. One of them incorporates the likelihood of getting a specific number of species given

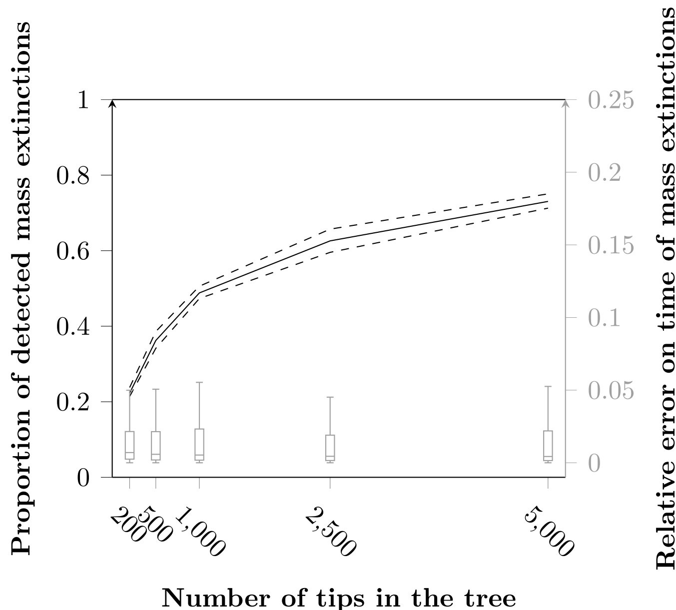


Figure 2.6: Influence of tree size on the detection of mass extinctions by TreePar. Line: proportion of detected mass extinctions; boxplots: distribution of the errors on their timing relative to the time of the first speciation event of the tree.

a pair of λ and μ after a certain amount of time, and is now used to account for the terminally input species numbers. May *et al*. simulated complete phylogenies before introducing uncertainties by sequentially collapsing some of the tips, and tested the different flavors of the three different Medusa algorithms ever made available. They found high Type I errors in every algorithm and biased parameter estimates. We note that in our study, we did not consider the estimation of the macro evolutionary parameters, and did not use unresolved trees, that can be used in Medusa to account uncertainties in the phylogeny. Interestingly, May *et al*. also tested the algorithm that we used in this study ($turboMedusa$, defined as $tMEDUSA$ in their study) on completely resolved trees, and found about the same rate of Type I errors as we did in the comparable trees (Figure S.20 of their study). Thus even though the focus of the two studies differs, they are in agreement in the few common analysis.

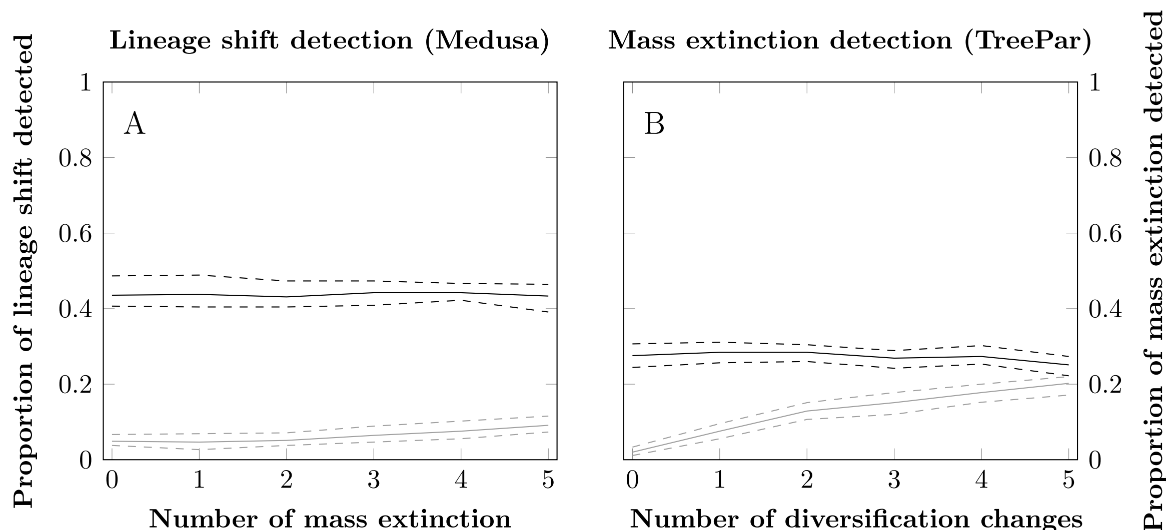


Figure 2.7: Influence of increasing model violations on the tests. A: Lineage shift detection against an increasing number of mass extinctions; B: Mass extinction event detection against an increasing number of lineage shifts. Dark lines: simulations where the correct number of events was found, light lines: simulations where too many events was favoured.

2.4.2.3 Impact of patterns of extinction

The effect of different scenarios of mass extinction on the results of Medusa and TreePar are presented in Figure 2.9. First, as expected, no effect of the extinction scenarios is observed on the detection of lineage rate shifts detected by Medusa (Fig. 2.9A). In contrast, the fair game and wanton destruction scenarios impact the estimation made by TreePar. They produce, for comparable levels of detection, more false positives than the field of bullets which was used in the previous simulations (73% and 74% for fair and wanton against 58% for field of bullets, Fig. 2.9B). Irrespective of the type of mass extinction simulated, there are very few false negatives, *i.e.* at least one extinction was detected in almost every tree. The error on the timing of this event was kept under 5% of the root age. We also performed a search for global rate shifts in the case of wanton destruction (Fig. 2.9B, dashed background). Regarding this scenario, we also compared simulations where all lineages undergo an increase of diversification after the mass extinction event against those who undergo a decrease and observe no difference between the outcomes of the two frameworks. Even though the shifts are differ-

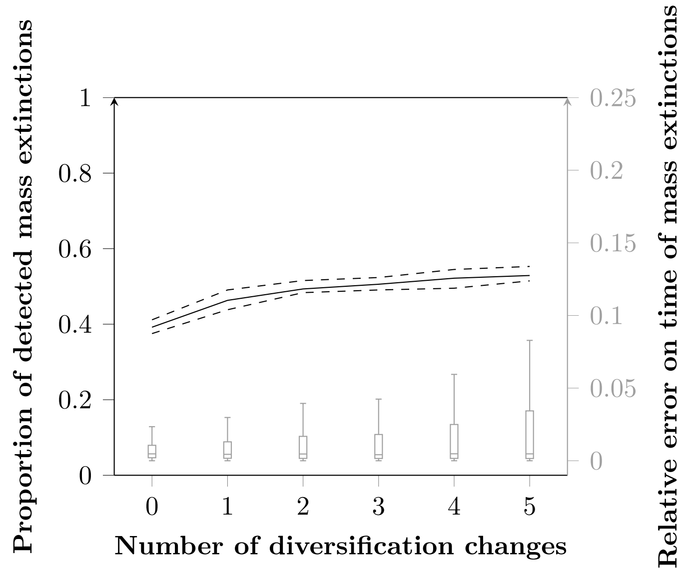


Figure 2.8: Influence of the number of lineage shifts in a simulation upon the detection of mass extinctions. Line: proportion of detected mass extinctions; boxplots: distribution of the errors on their timing relative to the time of the first speciation event of the tree.

ent between lineages (*i.e.*, increase of diversification in some lineages, decrease in others), TreePar detects the period of this shift with more power than for the detection of the associated mass extinction (34% and 21% correctly assessed simulations, respectively). Overall, these results show that departure from the simplest model of mass extinction should not affect our ability to detect these events in phylogenetic trees (*i.e.* no increase in false negatives rate). But it should lead to an increase of false positive detections.

2.5 Conclusion

Previous studies involving mass extinctions and changes in macro-evolutionary rates have only focused on their effect on lineage through time plots (Crisp and Cook, 2009). This lead to the identification of a possible mass extinction event in some plants lineages around 32 Mya, which was further suggested to be linked

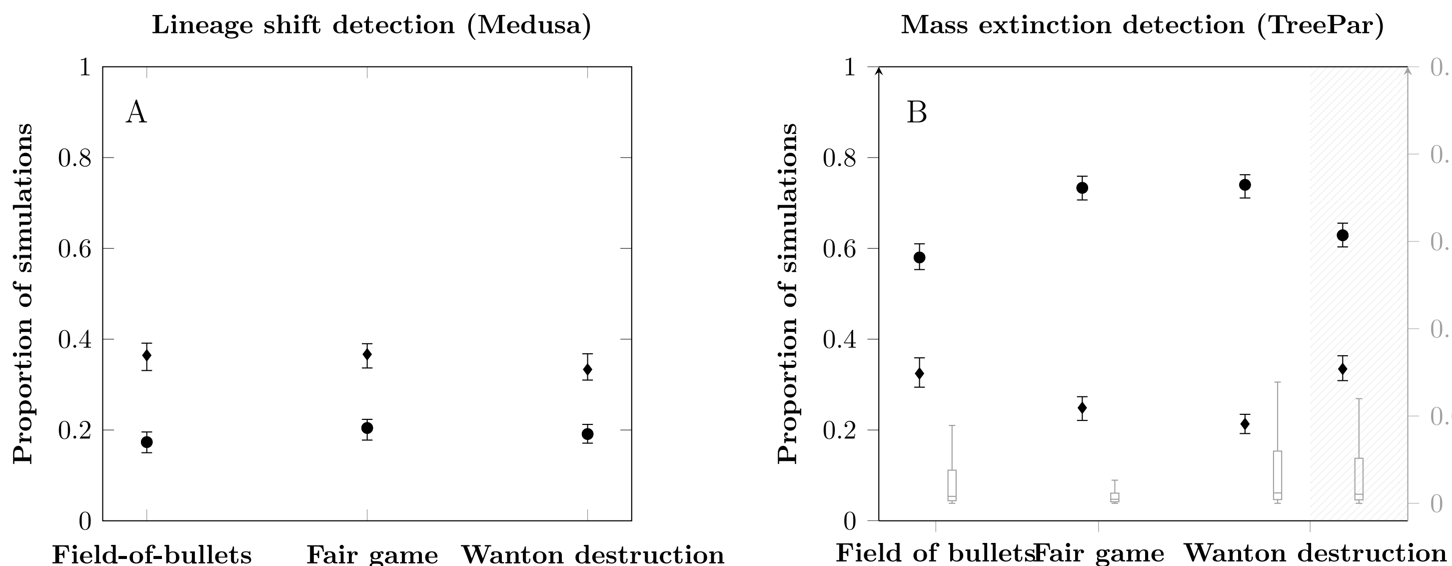


Figure 2.9: Influence of distinct extinction scenarios on Medusa and TreePar predictions. A: Medusa outcome; diamonds: proportion of simulations where the model with the correct number of events is chosen; circles: proportion of simulations where a model with too many events is chosen; there are less correctly assessed simulations for Medusa because of the high number of lineage shifts in these simulations (5). B: TreePar outcome and error on the timing of events: boxplots: error on the timing of the estimated extinction relative to the first speciation event; blank background: detection of mass extinctions; dashed background: detection of global rate shifts; other symbols as in A.

with changes in climate. Recently, Höhna (2013) developed a new algorithm to perform simulations with varying macro-evolutionary rates, allowing for mass extinction events. Other ongoing work aims at studying and simulating increasingly complex scenarios of diversification (Hartmann et al., 2010; Morlon et al., 2010), but we would like to emphasize that no method allows the simultaneous discovery of both time-specific or lineage-specific rate changes and mass extinction events.

The study of diversification rates has become a standard part of the analysis of large phylogenetic trees (Meredith et al., 2011; Jetz et al., 2012; Near et al., 2013), and recent efforts have also assessed the methods used when their assumptions are violated (Rabosky, 2014). We have shown that departure from the assumption of consistency in rates across lineages causes a large increase in false positives when looking for mass extinction events. This can be problematic as we

know that rate consistency rarely holds (Rabosky et al., 2007; Jetz et al., 2012; Barker et al., 2013), and casts doubts on our ability to reliably find such events using only phylogenetic trees. Nevertheless, an increasing number of disparities between lineages caused neither a decrease in the probability of detecting an event nor an increase in the error on its timing. As we observed the same pattern under more complex scenarios of extinction, the difficulty in detecting the K-Pg event in mammals is therefore probably not due to biases in the methods used. We might be limited by the power of TreePar to detect mass extinction events, although in simulations we reach 60% of true events detected for a tree size similar to that of mammals.

Recent efforts aim to reach a better agreement between paleontological and molecular data (Morlon et al., 2011), including looking for mass extinctions in molecular phylogenies. For instance, there is much debate on whether the K-Pg extinction event triggered the mammalian diversification (Bininda-Emonds et al., 2007; Meredith et al., 2011; Stadler, 2011a; Dos Reis et al., 2012; O'Leary et al., 2013). The fossil record also indicates higher extinction rates of mammalians species around 65 Mya (Wilson, 2005). In this work, we have shown that for phylogenetic trees similar in size to that of mammals (i.e. *ca.* 5000 species), the signal for mass extinctions was usually recovered in the tree, even though lineage discrepancies in macro-evolutionary rates had a tendency to yield more false positives. Thus, if the ancestor lineages of the extant mammal families did experience a mass extinction at the K-Pg boundary, we should theoretically be able to identify it using phylogenetic trees. The underlying assumption about the mass extinction made when using TreePar is that lineages are terminated randomly with a fixed ρ value everywhere in the tree, *i.e.* a field of bullets type of mass extinction. But other models of extinction seem to increase false positives but not false negatives, not explaining difficulties in finding a K-Pg signal in real phylogenetic trees.

Recent studies have used Markov processes to account for the effect of specific traits upon the probability of extinction of a species, thus extending models of mass extinction beyond the field of bullets scenario (Faller and Steel, 2012).

Such models can be used for instance to estimate the loss of phylogenetic diversity after a mass extinction event (Lambert and Steel, 2013). Our simulations can be seen as a special case of such models, where the trait influencing survival probabilities is the diversification value of the species. We have shown that more complex models of mass extinction cause more false positive detection than the simple field of bullets, as well as a decrease in the error for the fair game scenario. Choosing a specific model of extinction (field of bullets, wanton destruction, fair game) might require the incorporation of fossil information into the phylogenetic tree, and thus the further development of methods capable of dealing with both molecular and fossil data.

2.6 Authors contributions

SL, MRR and NS designed the study, SL performed the simulations, SL, MRR and NS analyzed the results and wrote the manuscript.

2.7 Acknowledgements

This work was supported by the ProDoc grant number 134931 of the Swiss National Science Fondation; and État de Vaud. The computations were performed at the Vital-IT (http://www.vital-it.ch) Center for high-performance computing of the SIB Swiss Institute of Bioinformatics. We thank Tanja Stadler, Daniele Silvestro and four anonymous reviewers for helpful discussions.

Chapter 3

No evidence for the radiation time lag model after whole genome duplications in Teleostei

3.1 Abstract

The effect of polyploidization on lineages fates over the short or long term is the focus of intensive research. First recognized in land plants, the full sequencing of distantly related species has lead to the recognition of its conspicuousness across the tree of life, in particular at the origin of vertebrates and teleost fishes. Many hypotheses have been proposed to explain the potential link between evolutionary success and whole genome duplication. For instance, the radiation time lag model posits that the whole genome duplication would favour the apparition of key innovations but the evolutionary success would not become apparent after a later diversification event. Some preliminary results might indicate that this model may be observed during land plant evolution. In this work, we test the prediction of the radiation time lag model using both fossil and phylogenies in old and more recent teleost whole genome duplication. We fail to find any evidence of delayed evolutionary success after any of these events and conclude that paleopolyploidization still remains to be unambiguously linked to evolutionary success in this group.

3.2 Background

The understanding of how biodiversity changes and is maintained on earth has long fascinated naturalists. Studying historical biodiversity trends was originally performed by looking at the fossil record observed on successive geological layers (Benton, 1995) and using models to explain the appearance of new clades (Raup, 1985). Nowadays, the joint use of molecular clocks and fossil calibrations in a maximum likelihood (reviewed by Paradis (2013)) or Bayesian (Drummond et al., 2012a; Ronquist et al., 2012) framework on the one hand, and the flourishing methods enabling evolutionary inferences based on the reconstructed phylogenies (O'Meara, 2011; Morlon, 2014) on the other hand now enables alternative studies about how lineages evolve through time. As the development of complex methodologies for retracing evolutionary trends using fossil instances was not left apart either (Silvestro et al., 2014; Wagner and Estabrook, 2014), researchers are now left with the choice of using data from molecules, fossils, or a mix of both when retracing biodiversity (Givnish et al., 2015; Smith and Marcot, 2015; Condamine et al., 2016).

Many methods assessing diversification based on different assumptions were developed to test macro-evolutionary hypothesis. By comparing which model fits the best our data, whether it be molecular or fossilized, one is able to judge under which scenario their lineage of interest is evolving. Thus, authors now routinely test whether diversity depends on particular states, such as a particular diet (Wiens, 2015) or key ornamental displays (Maia et al., 2013), on species age (Hagen et al., 2015; Alexander et al., 2015), on environments (Cantalapiedra et al., 2015; Wiens, 2015) or how it responds to mass extinction events (Feduccia, 2014; Bronzati et al., 2015). Moreover, one can address more general questions regarding species diversification, such as whether their appearance is bounded by the environment and the number of niches available or not, as was recently debated in the literature (Harmon and Harrison, 2015; Rabosky and Hurlbert, 2015). In this case, authors can compare density-dependent diversification frameworks to other models (Etienne et al., 2012; Slater, 2015), or argue that slowdowns in diversification can be explained by other factors (Moen and Morlon, 2014). Exhaustive

testing with different models as well as considering a variety of distantly related organisms is usually performed to bring contributions to these debates.

Among intrinsic properties of the lineages that have been scrutinized for potential effect upon evolutionary success, polyploidy has been the object of intense publications for the last decade, with researchers hypothesizing about how whole genome duplication could impact species fate (Taylor and Raes, 2004; Comai, 2005; Donoghue and Purnell, 2005; Crow and Wagner, 2006). Various scenarios have been proposed to explain how diversification rates could react to polyploidy. One of them is the radiation lag-time model, where the increase of diversification would occur only after a substantial amount of time of evolution (Schranz et al., 2012). The expansion of the lineages would only occur after the evolution of a key innovative trait, thanks to the duplicated gene material, and a subsequent dispersion event. Similarly, Dodsworth et al. (2015) have argued that the diploidization process is responsible for the lag between polyploidisation and radiation of the lineages. Another scenario, where the increase in diversification is mediated by reciprocal gene loss leading to reproductive isolation (Scannell et al., 2006), would predict that diversification is highest after the polyploidy event, once the differential fixation of the duplicated pairs has started in the subpopulations resulting from the polyploid ancestor. Diversification is then predicted to slowly decay as the process of reproductive isolation is complete between all subpopulations. However, recent modelling work showed that reciprocal gene loss was not likely to be sufficient for explaining potential diversity increases (Muir and Hahn, 2015), although some initial evidence for this model had been found in yeast (Maclean and Greig, 2011).

In plants, recent findings tend to support the radiation time lag model when considering nine paleopolyploidization events (Tank et al., 2015). Using a phylogenetic tree resolved at the plant family level, Tank et al. (2015) estimated diversification rates and changes using species richness associated for each family. They identified increases in diversification in ancestors of the current plant species and showed that these increases were preferentially clustered after paleopolyploidization. Nevertheless, the method used for estimating diversification

based on family richness has been shown to yield a very high rate of false positives (May and Moore, 2016). Moreover, it does not explicitly model the changes of diversification as a function of time but rather the global rate at which children of a particular inner node of the phylogeny diversify. Finally, this work needs to be expended once every paleopolyploidy in land plants will be identified and precisely dated, as new events are still currently being found (Barker et al., 2009; Li et al., 2015; Scott et al., 2015a; Shi et al., 2015).

Whole genome duplications have been identified as well in vertebrates, in particular in the ancestor of all present-day teleost species (Taylor et al., 2003). Moreover, subsequent polyploidy has also been found in few teleost genera (Mable et al., 2011), as well as in non-teleostei species, such as sturgeons (Havelka et al., 2011). The Salmoniformes-specific genome duplication has recently been thoroughly studied (Berthelot et al., 2014) and dated as occurring at least 88 million years ago (Macqueen and Johnston, 2014). Some authors tentatively linked it with anadromy — leaving in a marine environment and migrating to freshwaters to mate — in some salmonid species (Alexandrou et al., 2013), hypothesizing that the duplicated genomic material was co-opted for the evolution of this particular behaviour. Other events of more recent polyploidization have also been investigated in the Cyprinidae family, as identified in the genome of *Cyprinus carpio* (Xu et al., 2014) or in *Squalius alburnoides* (Collares-Pereira et al., 2013), in Botiidae (clown loaches and allies) (Slechtová et al., 2006) and Callichthyidae (Otto and Whitton, 2000), among others.

Patterns of diversification after paleopolyploidization have already been investigated in actinopterygians. Although the occurrence of the genome duplication in actinopterygians coincides with the ancestor to all teleost fishes (Hoegg et al., 2004), Santini et al. (2009) determined that the teleost-specific genome duplication was responsible for only part of the diversity of this group observed today, using the same methods as Tank et al. (2015). By reconstructing diversification patterns, Zhan et al. (2014) studied the differences between polyploids and diploids in four groups of actinopterygians and concluded that polyploidy had inconsistent consequences across these clades, namely that it had a posi-

tive impact on cyprinids evolutionary success but not in salmonids, Botiidae or sturgeons. Similarly, Macqueen and Johnston (2014) found a major decoupling in time between the whole genome duplication of salmonids and their diversification, although without detailing diversification patterns, hinting at potential evidence for the radiation time lag model of diversification in Salmoniformes.

In our work, we propose to further test the assumption of the radiation lag-time scenario by studying the responses in diversification in Teleostei experiencing whole genome duplication and comparing them to their sister clades, under the expectation that we should see a surge in diversification some time after the whole genome duplication event. For this we will explicitly model the changes in diversification through time, using both molecular and fossil based knowledge extracted from the literature, on ancient and more recent whole genome duplications.

3.3 Methods

3.3.1 Phylogeny-based diversification analysis

We used the data generated from the study of Zhan et al. (2014). They reconstructed the phylogenies of three actinopterygian families: Acipenseridae (sturgeons), Botiidae, Cyprinidae (carp, goldfish and allies) and the monophyletic group formed by the sister orders Salmoniformes and Esociformes (pike and allies). Paleopolyploidization has been identified in the lineage leading to the ancestor Salmoniformes species whereas Esociformes are diploids (Berthelot et al., 2014; Macqueen and Johnston, 2014). Similarly, Botiidae is a freshwater family belonging to the Cypriniformes order originating from Southeast Asia including two subfamilies, one of them being Botiinae, whose members are all tetraploids (Slechtová et al., 2006). Sturgeons (Acipenseridae) are also a well known-case of non-Teleostei group where polyploidy has been recurrently identified (Havelka et al., 2011).

Using explicit modelling of chromosome number evolution (Glick and May-

rose, 2014), Zhan et al. (2014) recovered 12 polyploidiziation events in Cyprinids evolutionary history. They estimated the diversification rates of polyploid and diploids using the widely used Binary State Speciation and Extinction framework (BiSSE) (Maddison et al., 2007) that discriminates between ploidy levels but does not reconstruct changes through time in diversification.

To identify the potential effects of radiation time lag, we aimed to compare diversification rates through time of polyploid clades with their sister diploid clade, to find changes in those rates that would be consistent with this model. The identification of changes in diversification through time necessitates a sufficient number of data point to draw meaningful inferences, thus we discarded for our analysis Acipenseridae and 7 events from Cyprinids, because they involved either diploid or polyploid clades of less than 5 species. From the set of polyploidy events under consideration, 5 for cyprinids, 1 for salmonids and 1 for Botiidae, we extracted every polyploid clade and its sister diploid clade. We reconstructed the diversification pattern for each clade independently. We used two different diversification methods enabling macro-evolutionary rate fitting as a function of time to compare the dynamics of diversification and highlight potential signs of delayed rise after the paleopolyploidy compared to species that did not experience such event.

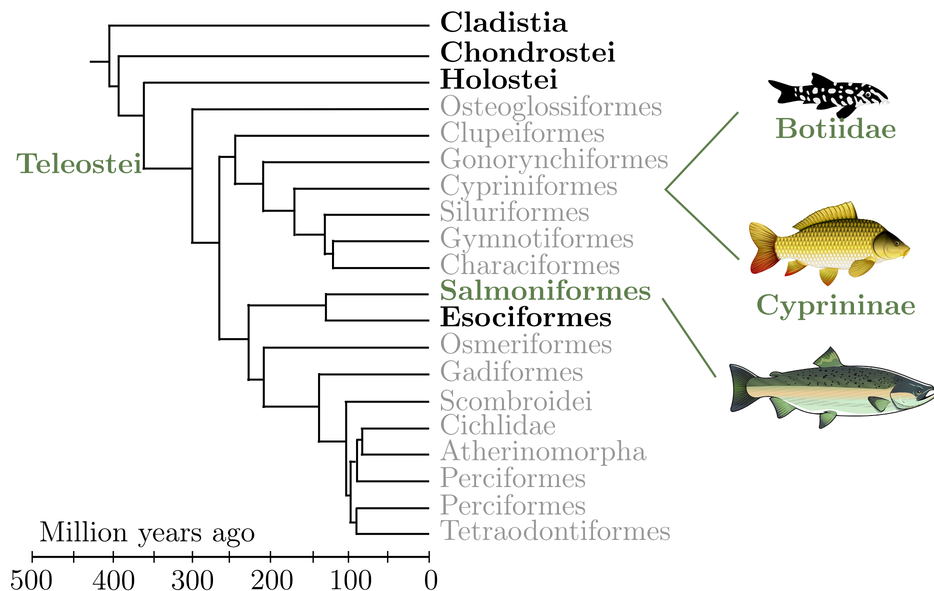


Figure 3.1: Polyploidization events considered in this study. Green names represents either lineages where the common ancestor experienced paleopolyploidy (Teleostei, Salmoniformes) or groups where polyploidy has been found, which will be studied here. Black names represents groups of species that will be studied in comparison to polyploid groups. Gray names represent other teleost orders or other taxonomic ranks. The Teleostei paleopolyploidy event will be studied using fossil data whereas the other events using phylogeny-based methods. Backbone tree modified from Berthelot et al. (2014). Although pictured paraphyletic, the monophyly of Perciformes has been recovered in other studies (Betancur-R et al., 2013; Near et al., 2012).

The first method we used was TreePar (Stadler, 2011a). TreePar can estimate diversification rates across time ranges for a phylogeny. The method is sequentially run, from a model where only one speciation and extinction rate governs the entirety of the tree, from its root to extant species. Then, phylogenies are split across a range of time point and macro-evolutionary rates are estimated for each slice created. The best estimates and time break are chosen by their associated log likelihood values. To determine if a model with an additional break was preferred over the simpler model, we used a likelihood ratio test at a 0.01 significance threshold, according to standard procedure for this method (Stadler, 2011a). If the likelihood ratio test favours the simpler model, the computation is stopped and the diversification pattern will be set for the considered tree. If

the more complex model is preferred, the computation is run once more with one additional time break and this process will continue as long as the more complex models are chosen. Using this procedure thus enables us to find abrupt changes in diversification at certain time points.

The second method used was developed by Morlon et al. (2011). It enables the fitting of any function to either macro-evolutionary rate with time as the variable. We chose to fit either constant, linear or exponential response of rates. We fit each type of function for each rate, summing to 9 models in total and chose the one that fitted best our data using ΔAIC, as employed by Morlon et al. (2011). This method will be referred as the function-fitting method.

We ran both analysis on the whole distribution of trees that we had acquired, to check if all phylogenetic trees agreed onto a similar diversification scenario. We performed the diversification analysis on all the 500 trees constituting the distribution we extracted from the literature, on trees of Botiidae and salmonids species, and on five subclades of cyprinids showing differential ploidy level between sister clades.

3.3.2 Fossil-based method

We downloaded from the Paleobiology database all teleost and actinopterygians non-teleost fossil occurrences identified at the species level. For the non-teleost analysis, we selected the occurrences matched to Chondrostei, Cladistia or Holostei taxonomic groups. We obtained 1239 and 1515 occurrences matched to species names for teleosts and non-teleosts respectively. Every species name present in our datasets that matched accepted names extracted from FishBase (Froese and Pauly, 2015) was deemed as extant.

We used PyRates (https://github.com/dsilvestro/PyRate) (Silvestro et al., 2014), a probabilistic framework to estimate diversification rates from fossil occurrences of extinct and extant species without prior knowledge of relationships

between the considered species. Additionally to speciation and extinction, preservation and sampling rates for the lineages are also estimated, so that the differences between organisms of chances to be effectively observed into the fossil record does not bias the macro-evolutionary rate estimations. On top of this probabilistic framework, a bayesian procedure is used to explore models with different number of changes in macro-evolutionary rates. Thus, from the total distribution of the dating of the occurrences, we are able to reconstruct the changes in diversification rates through geological times, with both extinct and extant species fossils.

3.4 Results

3.4.1 Salmonids, cyprinids and botiids whole genome duplications studied with phylogeny methods

In most of the cases, both phylogeny-based diversification methods found consistent results across trees. For the cases where diversification was determined as constant for each clades (figure 3.2 and figure 3.3 panels B, C, D and E), both methods led to the same conclusions when comparing diversification between polyploids and diploids in all cases except for Salmonids. Indeed, in this group, TreePar overwhelmingly led to conclude that polyploids diversified faster, whereas when fitting varying-through-time functions, the range of estimated diversification values for polyploids and diploids overlapped. However, for the subtrees of Cyprinids where constant diversification was favoured in general, both methods concluded faster diversification in every tree for polyploids, except for panel E where no difference was observed between the two groups. Interestingly, this subtree represents the youngest whole genome duplication event studied in this work. We note that in some cases, individual trees led to apparent incoherent or wrong inferences when using the function-fitting method: some trees (1 in figure 3.2 and 1 in figure 3.3, panel D) were fit with exponentially decaying diversification functions, disagreeing with the rest of the distribution.

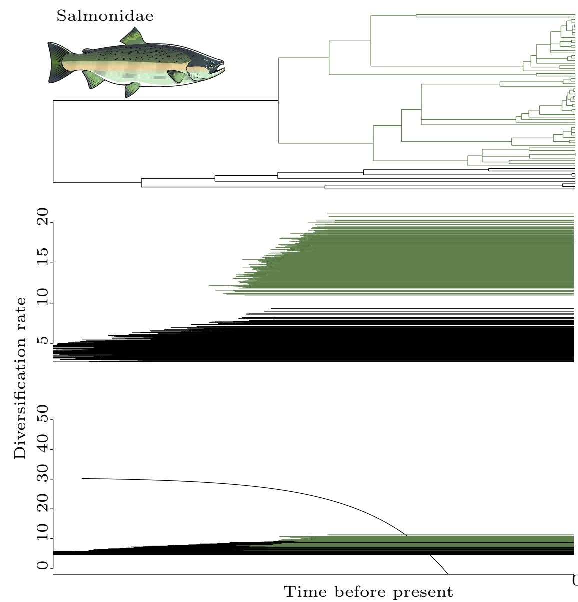


Figure 3.2: Results for the Salmoniformes and Esociformes phylogeny. The black clade represents the diploid Esociformes, the dark green clades the tetraploid Salmoniformes. Only the consensus tree is represented but analysis were run on a set of 500 phylogenies, separately for each diploid and tetraploid clade, and the results of each run is represented with one line. TreePar analysis is on the middle plot, function-fitting scenario on the bottom plot. One line represents the result for one diploid or polyploid clade extracted from one the 500 hundred trees distribution of the phylogeny. Constant diversification models were preferred with every analysis, as no TreePar result includes a break in diversification value, and as only constant functions were chosen by the function-fitting method except for one outlier diploid clade. The analysis were run with the sampling ratios extracted from Zhan et al. (2014): 0.69 for the polyploid clade and 0.28 for the diploid one.

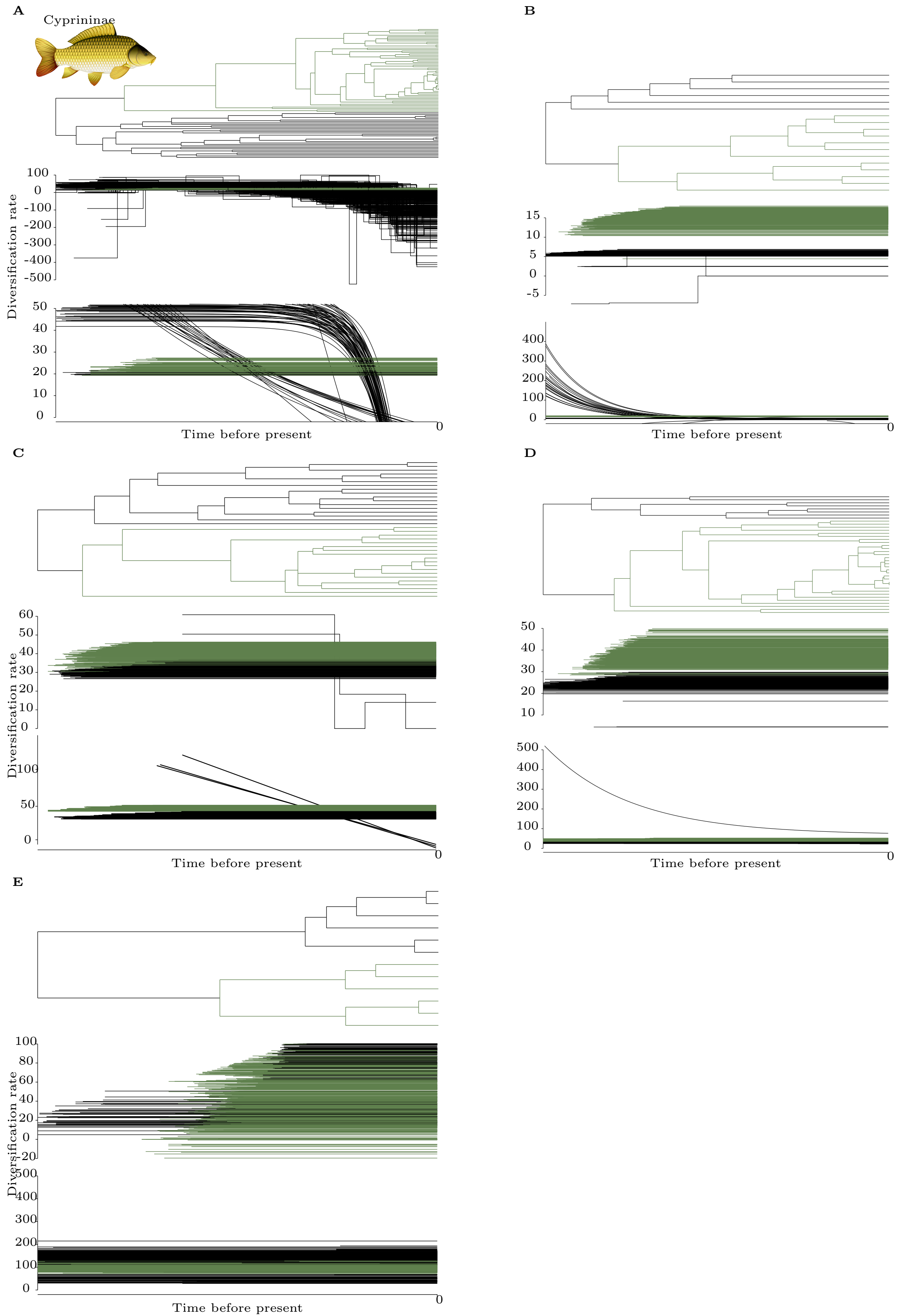

A
Cyprininae
100
0
-100
-200
-300
-400
-500
Diversification rate
50
40
30
20
10
0
Time before present
0
B
15
10
5
0
-5
400
300
200
100
0
Time before present
0
C
60
50
40
30
20
10
0
Diversification rate
100
50
0
Time before present
0
D
50
40
30
20
10
500
400
300
200
100
0
Time before present
0
E
100
80
60
40
20
0
-20
Diversification rate
500
400
300
200
100
0
Time before present
0

Figure 3.3: (Previous page) Results for five Cyprininae subtrees, ordered from oldest to most recent event, presented as in figure 3.2. No fossil calibration was performed thus the absolute timing of the events cannot be estimated. Most genera were recovered as paraphyletic in Zhan et al. (2014), except *Capoeta, Pseudobarbus, Schizothorax* and *Sinocyclocheilus*, genera for which monophyly had been documented before. Lack of monophyly prevents the estimation of the sampling ratio of each subtrees, thus the analysis were run without sampling information. Panel A: results comparing a diploid clade (black), encompassing *Gara* and *Labeo* genera, with a tetraploid one (dark green) encompassing the *Barbus, Labeobarbus, Neolissochilus, Varicorhinus* and *Tor* genera. Panel B: subtrees encompassing *Pseudobarbus* and *Barbus* species. Panel C: subtrees with hexaploid *Capoeta* genus members and tetraploid *Barbus* and *Luciobarbus* species. Panel D: results from a subtree of panel A, comprising *Casobarbus* genera and some *Barbus* species in the tetraploid subclade (black), and hexaploid species of *Labeobarbus* and *Barbus* genera (dark green). Panel E: comparison inside the *Schizothorax* genus.

When diversification was determined to be changing through time, we observe that identical scenarios were reconstructed by both methods (figures 3.4 and 3.3 panel A). For the oldest duplication in cyprinids, TreePar results led to the scenario where initially diploids diversify faster than their polyploid counterpart but, as their diversification rates decreases nearer to the present, and the polyploid one stays constant, most of the members of the polyploid clades appear to have a higher diversification rates in recent time scales. When fitting varying-through-time function responses, a similar scenario is observed for most of the trees of the distribution, with some variability identified for the diploid clades. Indeed, this method apparently favours alternatively one of the following scenarios for diploids: most of them show exponential decrease in diversification, either gradually or more sharply and closer to the present, and a small portion of the trees show constant diversification, at a lower rate than the polyploids. In all cases, polyploids are consistently diversifying faster than diploids near the present.

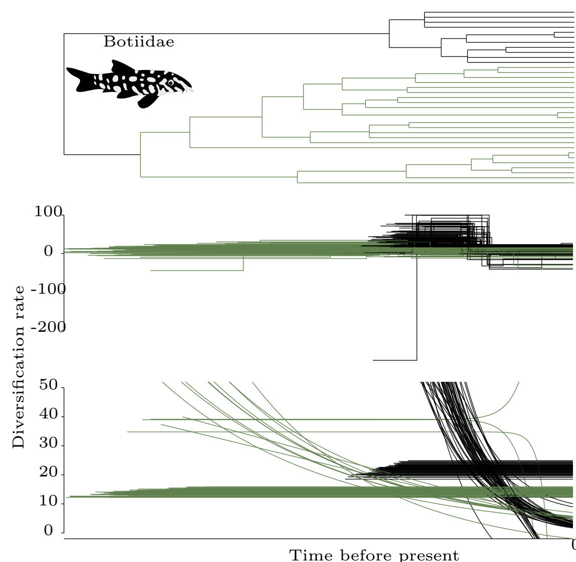


Figure 3.4: Results for the Botiidae (clown loaches and allies) species, presented as in figure 3.2. Middle plot: results for TreePar. Bottom plot: results when fitting varying through time functions.

For Botiidae, a somewhat similar pattern was observed, where diploids initially diversified faster and then reached the approximate levels of diversification of polyploids. For TreePar, this was materialized for the vast majority of trees as a two-phase diversification values for the diploid clade only, with a switch in rates near the present at roughly the same time for every tree, and a constant diversification rate for polyploids. When fitting function-time-responses, it is apparent by the exponential decay being favoured for a significant part of the distribution of the diploid clade, and a constant value of diversification for most of the remainder of the distribution of the diploid clade. Still, we note that diploid clades for which constant values were favoured retained higher diversification than polyploids towards the present. We also observe in both Botiidae and Cyprinidae that both methods output some degree of error: 3 polyploid clades in Botiidae apparently follow incoherent exponential behavior, whereas some degrees of divergence from the mostly observed patterns of TreePar is seen for both Botiidae and cyprinids (figure 3.3, panel A).

3.4.2 Teleost whole genome duplication studied with fossil data

From the origin of non-teleost actinopterygians till the Permian-Triassic extinction event, the diversification rate for these species stayed constant and at a relatively high level (figure 3.5, black line), level that were never reached again in subsequent 250 My of evolution. Their diversification rate then plunged below zero, around the Permian-Triassic boundary (250 Mya), denoting massive loss of species and stabilized over the long run near null diversification, explaining the relative rarity of those species in the present day.

Although teleost fishes appeared around the boundary of the Carboniferous and Permian periods (298 Mya) (Near et al., 2012), we are not able to estimate diversification rates before 250 Mya because of lack of data. From 250 Mya, teleost fishes diversified at a constant rate for more than 50 My, faster than their non-teleostean counterparts around the same time but at lower levels than the latter experienced before Teleostei appeared, during the Carboniferous and Permian periods. They experienced a sudden drop in their diversification in the middle of the Jurassic (around 175 Mya), consistent with the reported loss of 53 genera of ray finned fishes around this boundary (Guinot and Cavin, 2015). They went on to a steady increase in their diversification during the Early Cretaceous but around 100 Mya, their diversification once again sharply decreased, decrease possibly driven by marine specific groups (Guinot and Cavin, 2015). Interestingly, most of the ray finned fish families disappearing at the Cretaceous-Paleogene boundary were exclusively marine, whereas no fully freshwater family disappeared (Guinot and Cavin, 2015). After this last drop, Teleostei diversification rate stayed constant at a positive value until the present.

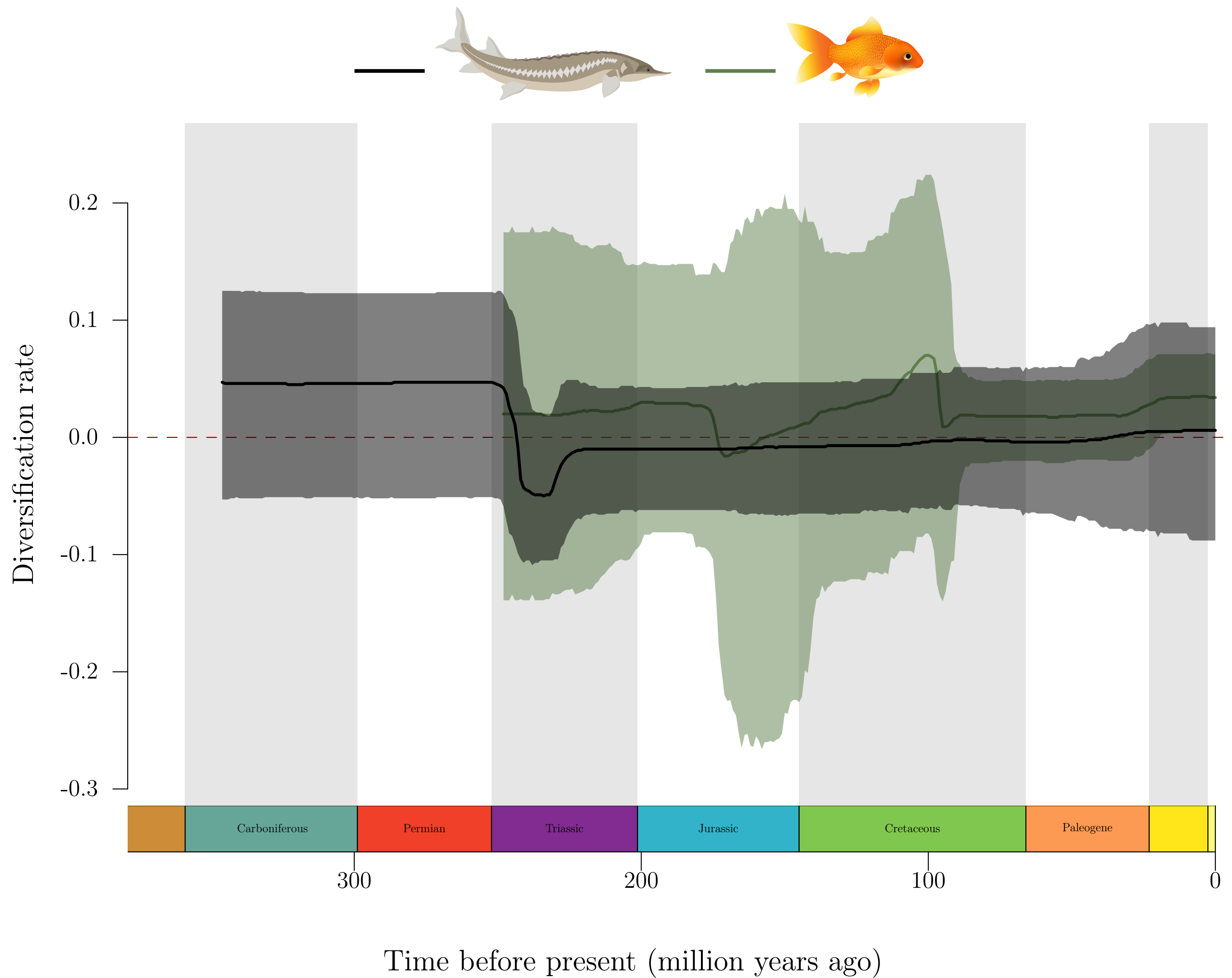


Figure 3.5: Results of the fossil analysis, for species-matched fossils. In black is the reconstructed diversification mean (thick line) and 95% highest density interval (transparent area) from fossil occurrences of the Chondrostei, Cladistia, or Holostei groups, including as extant species bichirs, bowfins, gars or sturgeons (pictured) and allies, which did not experience paleopolyploidy. In dark green is the results of the same analysis for the fossil attributed to Teleostei, encompassing most of the approximate 30000 extant species, among them cyprinids and goldfishes (pictured), whose last ancestor experienced polyploidization.

3.5 Discussion

In this study, we tested the radiation time lag model after the whole genome duplication instances in teleost fishes. Assuming that such model should be verified, an increase in diversification rate should be seen some time after the whole genome duplication and not exactly from the time of the event. For this we have compared some recent and older events in ray-finned fishes, using both fossil and phylogenetic data, and compared the fates of polyploid clades to their sister diploid clades. We did not find evidence for such a model in any cases we investigated. We also used methods with different initial assumption in order to test their consistency across datasets.

Both methods used usually led to similar patterns of diversification. We note, that although our study does not give evidence for the radiation time-lag model, some clades appear to follow a somewhat related scenario for diversification. Indeed, cyprinids show a pattern where diploids diversification is initially higher than polyploids but then decays until it reaches lower values. The radiation time lag model posits that key innovations occur because of the whole genome duplication, but that the radiation and the appearance of new species happen only once the lineage carrying the key innovation is dispersed to some new environment (Schranz et al., 2012). This hypothesis can not be fit to patterns of diversification observed for Botiidae and cyprinids here, as it should lead to an increase in diversification for polyploids, not a decrease in diploids. Nevertheless, the results presented here show the potential case where the evolutionary advantage of polyploids upon diploids is not apparent before a significant amount of time, as in the radiation time lag model.

Paleopolyploidization events in cyprinids or Botiidae have not been precisely dated, thus it remains hard to estimate if sufficient time has elapsed so that the hypothetical effects of the radiation time lag can be observed. Nevertheless, in every cyprinid and Botiidae case we investigated, constant diversification was observed after the polyploidy event. Even though in some case this polyploid event led to higher diversification compared to diploids (figure 3.3 panel B, C and D),

this pattern does not give support for the radiation time lag model because the increase in diversification appears concomitantly with the polyploidization.

The second oldest event concerns the salmonids species. Contrary to what Macqueen and Johnston (2014) reported when studying the diversification of Salmoniformes, we did not find increasing diversification rates with time, but a constant rate of diversification for both polyploid salmonids and diploids esociformes, using both TreePar and the function-fitting method. The salmonid specific genome duplication has been dated according to different sources no later than 88 Mya (Macqueen and Johnston, 2014) or around 96 Mya (90.5-101.5) (Berthelot et al., 2014), whereas the common ancestor to every extant Salmonidae species is usually dated around 55 Mya (52.2-58.0 (Near et al., 2012), 52.1-59.5 (Campbell et al., 2013)). Based on a lineage through time analysis, Macqueen and Johnston (2014) concluded that the pattern of diversification was more consistent with an increased diversification correlated to the cooling down of the ocean. Nevertheless, they did not estimate diversification rates and thus cannot conclude that diversification increased through time in this group. Overall, we do not find evidence of delayed increase of diversification after the salmonid specific duplication.

The importance of the anadromous behaviour in salmonids has already been reported (Macqueen and Johnston, 2014) and tentatively linked with the whole genome duplication (Alexandrou et al., 2013). Macqueen and Johnston (2014) performed a BiSSE analysis on the phylogeny of the Salmoniformes, enabling discriminated calculation of the diversification rates between anadromous and fully freshwater species, potentially testing a correlation between evolutionary success and this behaviour. Nevertheless, they only reported the estimation of the speciation rate of anadromous and freshwater species, and although anadromous species had higher speciation, a higher extinction in these species could still lead to a non significant or deleterious effect upon diversification. Alexandrou et al. (2013) found that anadromy evolved multiple times and no sooner than 55 My after the event of whole genome duplication, a pattern that may theoretically fit the prediction of the radiation time lag, if the evolution of such behaviour would lead to increase in diversification. Nevertheless, in our analysis, we found no dif-

ferences in diversification rates even after the appearance of anadromy, and with one of the two method used, we even found similar diversification rates between Salmoniformes and Esociformes (figure 3.2, bottom panel), highlighting the lack of evidence for such mode of diversification.

Most of the results found here are in partial agreement with previous studies (Zhan et al., 2014), where differences were looked for by studying the trees as a whole and discriminating between ploidy levels using Binary states Speciation en Extinction (BiSSE) (Maddison et al., 2007), rather than separately analysing each clades by ploidy-level. Indeed, their results also indicated higher speciation rates of salmonids. Nevertheless, they reported higher diversification for Botiidae polyploids whereas we do not find significant difference between the two clades. But one has to bear in mind that comparison is hard, as the methods presented here allow to model more complex evolutionary scenarios than the BiSSE models.

The teleost whole genome duplication has been alternatively dated around 350 Mya (Christoffels et al., 2004), between 300 and 450 Mya (Volff, 2005), or between 226 and 316 Mya (Hurley et al., 2007). The most recent common ancestor to all teleosts was also recently dated, ranging from 307 Mya (285-333) (Near et al., 2012), to 283 Mya (255-305) (Betancur-R et al., 2013). There is thus a minimum lag of 50 My between the occurrence of the genome duplication and the appearance of the common ancestor of all extant teleosts. In our results, the first estimations of diversification are available from around 250 Mya, meaning that fossil occurrences are not frequent enough between this date and the whole genome duplication to precisely estimate its value.

Effect of the whole genome duplication had already been studied by looking a the fossil record (Crow and Wagner, 2006), but without explicit modelling of the processes of origination of families, by studying the probability of survival of families emerging before and after whole genome duplications and hypothesized that whole genome duplication acted as a protection against extinction of such families. In our study, we explicitly modelled the diversification rates and found no signal indicating a protection against extinction after whole genome duplica-

tion over the long run in teleost fishes. Recently, the evolution of new organs, such as the bulbus arteriosus which is specific to the teleost heart, has been linked to their specific paleopolyploidization (Moriyama et al., 2016). Both sub— and neofunctionalization of a duplicate gene arisen after the whole genome duplication is thought to have enabled the apparition of the bulbus arteriosus. Although this organ had tentatively been linked to the success of Teleostei, we fail to find such a signal in their diversification rate.

In the oldest event in our dataset, common to all teleosts, there was no increase in diversification found before at least 50 My after the appearance of the first teleost fossil, and only after a sharp decrease in diversification. Moreover, their diversification rate reached the value of their initial diversification after more than 150 My after their origination and never quite reached the values at which non-Teleostei species were diversifying across the Permian and Carboniferous. This seems to indicate that the tremendous diversity observed today in Teleostei is rather the result of a steady diversification process through the geological times rather than an increased diversification promoted by the whole genome duplication some time after its occurrence. Overall, through the 8 different events studied here, none showed support for the radiation time lag model, yet 3 events supported higher diversification in polyploids (all belonging to Cyprininae) and one showed inconsistent results between the two methods (Salmonidae).

Chapter 4

Anagenesis at least as likely as cladogenesis for recent ploidy variation in 67 land plant groups

4.1 Abstract

Present day angiosperms species show significant variation at their ploidy level. Moreover, through genome sequencing, most of them have been found to be derived from ancestors that also went through multiple polyploidization. It has been traditionally assumed that polyploidization would lead to direct reproductive isolation from the parental diploids, as triploid crosses between them would suffer from decreased fitness. However, how frequently polyploidization directly leads to the formation of new species has not been addressed globally. Using a model that can reconstruct the apparition and the loss of species together with rates of polyploidization, we can infer the relevant parameters using phylogenetic trees. We compare models incorporating only anagenetic polyploidization, the change of ploidy level in the same species, to models accounting in addition for cladogenetic polyploidization, the co-occurrence of polyploidization and speciation. Using data both from the literature and newly generated, we study 67 plant groups and found that 27 of them did not show support for a cladogenetic process of polyploidization. Moreover, in the genera where evidence for both

cladogenesis and anagenetic was found, both processes contributed equally to the creation of polyploids. These results indicate that recent ploidy variation is as likely to originate from cladogenesis as from anagenesis in the few groups that have been investigated and need to be confirmed in other groups. Nevertheless, they emphasize the need of understanding which life history traits of species can influence the relative ratio of one process over the other.

4.2 Background

Polyploidizations, alternatively viewed as powerful events mediating evolutionary success (Levin, 1982) or complicating factors of evolution (Stebbins, 1950), have been recognized as an important process shaping plants genomes (Wendel, 2015). Polyploidy at the intraspecific level is indeed rampant in plant species: Rice et al. (2014) estimated that up to 69% of the intraspecific changes in chromosome numbers were due to polyploidy and that around 16% of plant species harbour variation at their ploidy levels. These events, mediated by gametic or somatic errors leading to the addition of one or more chromosome sets (Rieseberg and Willis, 2007), are thought to be at the origin of up to 15% and 31% of speciation events in angiosperms and ferns, respectively (Wood et al., 2009).

Moreover, ancient whole genome duplications, or paleopolyploidizations, have been identified in multiple instances in the ancestry of seed plants and angiosperms (Jiao et al., 2012). Dating at least many tens or hundreds of millions years ago, they have been frequently linked to the evolution of increased organismal complexity (Freeling and Thomas, 2006) and of key innovations (Soltis et al., 2014c) in plants. These events have also been found at the origin of the most speciose families (Soltis et al., 2009), prompting the hypothesis of a positive impact onto the evolutionary fates of lineages. Additionally, Fawcett et al. (2009) and Vanneste et al. (2014) proposed that genome duplications protected against extinction during periods of elevated environmental stress, for instance during the Cretaceous-Paleogene boundary.

Polyploidization is generally assumed to occur through aberrant meiotic processes (Grandont et al., 2013), whereas the cases of somatic doubling are assumed to be much rarer. Barriers to the gene flow between polyploids and diploids can originate from the decreased fitness of triploids offspring resulting from the fertilization of an haploid and a diploid gametes. Ramsey and Schemske (1998) studied the frequency of auto— and allopolyploid formation by estimating triploid fertility. They found the triploid individuals produced, along with the unbalanced gametes, a minority of haploid, diploid and triploid ones that could permit interbreeding between different ploidy levels at low levels. The minority exclusion principle (Levin, 1975), states that the success of a cytotype depends on its relative frequency in the population. Indeed, in absence of assortative mating, most of the offspring of the least abundant cytotypes will have parents from the other cytotype of the population and will suffer from fitness loss.

Stemming from this reproductive isolation between diploids and polyploids, Scarpino et al. (2014) estimated that diploids had a tendency to create more species just because they could polyploidize more often than their polyploid counterparts. The establishment of polyploid lineages can thus be explained without needs of adaptive arguments. Similarly, Oswald and Nuismer (2011) showed using simulations that under the assumption that polyploids display assortative matings, coexistence with diploid species is expected and demonstrates that the patterns observed at the Cretaceous-Paleogene boundary are not necessarily due to selection. Moreover, studying the whole genome duplications in Andropogoneae, a grass tribe, Estep et al. (2014) found no support for higher success of polyploid species although allopolyploidy was an important mechanism for speciation. Under this emerging picture, polyploidy helps building barriers to the gene flow during the establishment of a new species, differing at the ploidy level from its closest relative, but does not subsequently bring elevated evolutionary success.

The study of evolutionary success can be performed by considering diversification, the difference between speciation (λ) and extinction (μ) rates. It has been widely studied through the tree of life since the influencial publication of Nee et al. (1994b), enabling its estimation from phylogenetic trees. Denoting the net num-

ber of species created per unit of time, diversification is widely used as a proxy for evolutionary success and a number of methodologies has been developed in order to explain its changes in the tree of life (Morlon, 2014). The variety of models estimating diversification across time, environments or traits makes possible testing macro-evolutionary hypotheses regarding the processes leading to species appearance or loss. Paleopolyploidy in plants has recently been investigated using these methods, and deemed to positively influence diversification after a time delay (Tank et al., 2015). Another variety of these methods estimating diversification builds on the Binary States Speciation and Extinction (Maddison et al., 2007) (BiSSE) framework, that enables to test association between speciation or extinction rates and qualitative or quantitative traits of the extant species of a phylogeny (Maia et al., 2013; Rolland et al., 2014a,b; Spriggs et al., 2014; Weber and Agrawal, 2014).

Using these phylogeny-based methods, Mayrose et al. (2011) set on to study the effect of recent polyploidies upon diversification. By selecting 63 genera where ploidy variation was known and analyzing them separately, Mayrose et al. (2011) focused on the recurrent polyploidization events, dating recently on the evolutionary time-scale. Based on a BiSSE analysis, the authors observed that diploids seemed to create more species than their recent polyploid counterparts. This initial result has led to discussions regarding the philosophical, technical, and data implications regarding the study of the effect of recent ploidy changes on evolutionary fate (Soltis et al., 2014a; Mayrose et al., 2014). Studying diversification rates after recent polyploidization comes with a variety of hurdles that have been addressed in the discussion: the reticulation pattern of allopolyploid speciation, frequent in plants (Majure et al., 2012; Mason-Gamer, 2013; Díaz-Pérez et al., 2014), that can not be represented using phylogenetic tree, the low sample size and power of the methods used in those cases (Davis et al., 2013), the difficulty of classifying plants either as polyploids or diploids, since most of them stem from paleopolyploidy somewhere in their ancestry (Soltis et al., 2009) or the taxonomic uncertainty surrounding polyploid accessions.

In this work, we aim to study the contribution of cladogenesis and anage-

nesis to the formation of polyploids using phylogenetic data for the first time. Anagenesis refers to the process of character change occurring inside a species, whereas cladogenetic change occurs simultaneously to speciation. Anagenetic polyploidization would indicate a shift from diploidy to polyploidy along the same species whereas cladogenetic polyploidization would trigger the apparition of a new species because of the polyploidy change. We propose, by using a method simultaneously estimating all macro-evolutionary rates relevant to the speciation and character change processes considered here, to quantify how much much recent polyploidy variation stems from anagenesis and cladogenesis in a few tens of vascular plant genera. We will compare different models including and excluding cladogenetic processes to find out whether speciation usually occurs with polyploidization.

4.3 Materials and Methods

4.3.1 Phylogenetic reconstruction and aggregation

From Soltis et al. (2014a), we identified four additional groups of interest to study anagenetic and cladogenetic polyploidization processes: the genera *Draba, Medicago* and *Opuntia* and the subtribe Loliinae. To reconstruct the phylogeny of our four groups of interest, we used orthoMCL (Fischer et al., 2011) to cluster orthologous genes of species belonging to theses groups extracted from GenBank (Benson et al., 2005). Table 4.1 lists the genes used for each group.

We aligned the clustered genes using MAFFT (Katoh and Toh, 2008) and ran jModelTest (Darriba et al., 2012) to identify the most likely model of substitution. We used BEAST (v1.7.1) and its associated software (Drummond et al., 2012b) to reconstruct the trees, using lognormal relaxed clock, birth-death process for the tree prior and the closest substitution model configurable by this software (table 4.1). We did not calibrate our trees using fossil knowledge. We let the MCMC run for 200 million generations and visually checked with Tracer that convergence was reached. We used TreeAnnotator to produce a single consensus

Group	Loci name	Substitution model
Draba	ITS1+ITS2	GTR+G+I
	trnL	TN93+G
Loliinae	ITS1+ITS2	GTR+G+I
	trnL	GTR+G+I
Medicago	ITS1+ITS2	GTR+G+I
	ETS	GTR+G
	trnL+matK	GTR+G
	GA3ox1	GTR+G+I
	Bcop	GTR+G
	rpS14 intergenic spacer	GTR+I
	CNGC5-like protein gene	GTR+I
Opuntia	ITS1 + ITS2	GTR+G+I
	trnL + trnF	GTR+G+I
	matK	GTR+G+I
	psbJ spacer	GTR+G+I
	atpB-rbcL intergenic spacer	GTR+G+I
	ndhF / ndhF-rpl32 intergenic spacer	GTR+G+I
	phosphoenolpyruvate carboxylase gene	GTR+G+I
	ycf1	GTR+G+I

Table 4.1: List of genes used for reconstruction

tree, discarding as a burnin the first 10000 trees of the MCMC distribution.

We carried on a reanalysis of the 63 genera presented in Mayrose et al. (2011). We downloaded the phylogenies and the ploidy calls of the extant species from their dryad repository (`http://datadryad.org/resource/doi:10.5061/dryad.6hf21`). These 63 genera with the four groups we reconstructed in this analysis constitutes the bunk of our dataset in the present study.

4.3.2 Chromosome number evolution reconstruction

For our four newly assembled trees (*Draba, Medicago, Opuntia* and Loliinae), we mapped the chromosome counts extracted from the Chromosome Counts Database (CCDB), listing chromosome number for plant accessions (Rice et al., 2014), and used chromEvol (Glick and Mayrose, 2014) to identify past polyploidisation based solely on those counts. ChromEvol implements different models of chromosome number evolution and estimates rates of various events, such as chro-

mosome gain, loss, or whole genome duplication.

We used the associated Polyploidy Inference Pipeline (PIP), following the instruction provided by the authors (`http://www.tau.ac.il/~itaymay/cp/chromEvol/chromEvol_v2.0_manual.pdf`). The pipeline uses ΔAIC to choose the optimal model of chromosome number evolution. If the favoured model included polyploidization, the pipeline can call the species as either diploid or polyploid. Using the phylogenetic tree distribution and simulations of chromosome number evolution using parameter estimates from the model (rate of chromosome loss, gain, and of whole genome duplication), PIP estimates duplication thresholds, above or below which species can be determined as diploid or polyploid.

4.3.3 Macro-evolutionary rates estimation

To estimate the relative contribution of anagenetic and cladogenetic processes of polyploidization, we used the Binary State Speciation and Extinction (BiSSE) (Maddison et al., 2007) and its alternative node enhanced state shift (ness) model (Magnuson-Ford and Otto, 2012) coupled with Bayesian Model Averaging (BMA) (Silvestro et al., 2013). The classic BiSSE model incorporates 6 different parameters: λ_0 and λ_1, respectively speciation rates of diploids and polyploids, μ_0 and μ_1, extinction rates of diploids and polyploid, and q_{01} and q_{10}, the rate of becoming a polyploid for a diploid species and vice-versa. q_{01} and q_{10} represent the anagenetic rates, modeling the change in ploidy level without a speciation event. As our dataset represents only very recently formed polyploids, we constrained q_{10}, the rate of polyploids going back to the diploid state, or diploidiziation, to 0. The anagenetic rate of polyploidization will then be estimated by q_{01}.

To incorporate cladogenesis, we used the BiSSE-ness model (Magnuson-Ford and Otto, 2012). It incorporates four additional parameters: p_{0c}, p_{1c}, p_{0a}, p_{1a}. The first two parameters are the probability of a diploid or a polyploid lineage undergoing ploidy increase (for diploids) or decrease (for polyploids) when speciating, and the last two are the probability of asymetric character inheritance

after a simultaneous speciation and character change event. p_{0a} is the probability of having one daughter lineage being diploid and the other polyploid when the ancestral lineage is speciating and undergoing polyploidization at the same time, the opposite event being having both daughter lineages polyploids after a diploid ancestral lineage speciating. p_{1a} refers to the equivalent event when the ancestral lineage is polyploid and undergoes diploidization and speciation at the same time.

As we constrained the rate of diploidization to 0, we also constrained p_{1c} and p_{1a} to 0 in all our calculations. Indeed, as polyploid species are never going back to the diploid level, the probability of such an event occurring simultaneously with speciation (p_{1c}) or being simultaneous to speciation and leading to asymmetric character change in the daughter species (p_{1a}) both remains irrelevant for our study. As we consider that, in the case of simultaneous polyploidization and speciation (p_{0c}), the reproductive isolation between the two daughter species is a direct consequence of the ploidy difference, we set the probability of having asymmetric character change after simultaneous polyploidization and speciation (p_{0a}) to 1. From the four parameters added by BiSSE-ness, only one will be effectively estimated, p_{0c}, which is the probability of a speciation event in a diploid species occurring at the same time as polyploidization. Magnuson-Ford and Otto (2012) showed using simulations that this parameters can be reliably estimated with BiSSE-ness. Thus, the net cladogenetic rate will be estimated by $\lambda_0 \cdot p_{0c}$, the product of the speciation rate of diploid with the probability of such an event being simultaneous with polyploidization.

To estimate if models incorporating cladogenetic character change significantly explained better the data than pure anagenetic character change, we used Bayesian Model Averaging (BMA) following Silvestro et al. (2014). We run preliminary MCMC runs for each of the models listed in Table 4.2, with the variants of the models incorporating or excluding cladogenetic rates of polyploidization using the consensus tree for each group. We used thermodynamic integration to calculate the marginal likelihood and the relative probability of each 20 models. We selected all models whose relative probability was higher than 5% and ran a full MCMC chain for each, on a distribution of 100 trees. At the end, each

MCMC run was weighted by the relative probability of the associated model. For the phylogenies collected from the literature, we used the sampling values provided by the authors (Mayrose et al., 2011). We estimated the sampling of the newly reconstructed groups using species number estimates from the plant list (The Plant List, 2013).

Model number	Condition on speciation	Condition on extinction	Description
I	$\lambda_0 = \lambda_1$	$\mu_0 = \mu_1 = 0$	Equal speciation, no extinction
II	$\lambda_0 \neq \lambda_1$	$\mu_0 = \mu_1 = 0$	Different speciation, no extinction
III	$\lambda_0 = \lambda_1$	$\mu_0 = 0$	Equal speciation, no extinction in diploids
IV	$\lambda_0 = \lambda_1$	$\mu_1 = 0$	Equal speciation, no extinction in polyploids
V	$\lambda_0 = \lambda_1$	$\mu_0 = \mu_1 \neq 0$	Equal speciation, equal extinction
VI	$\lambda_0 \neq \lambda_1$	$\mu_0 = 0$	Different speciation, no extinction in diploids
VII	$\lambda_0 \neq \lambda_1$	$\mu_1 = 0$	Different speciation, no extinction in polyploids
VIII	$\lambda_0 \neq \lambda_1$	$\mu_0 = \mu_1 \neq 0$	Different speciation, equal extinction
IX	$\lambda_0 = \lambda_1$	$\mu_0 \neq \mu_1 \neq 0$	Equal speciation, different extinction
X	$\lambda_0 \neq \lambda_1$	$\mu_0 \neq \mu_1 \neq 0$	Different speciation, different extinction

Table 4.2: Summary of the 10 different standard BiSSE Models. We tested this 10 models twice, once enabling only for anagenetic polyploidization, once enabling for both anagenetic and cladogenetic polyploidization.

4.4 Results

Out of the 67 groups investigated, models excluding cladogenesis were favoured for 27 of them. Thus, the adjunction of cladogenetic rate of polyploidization did not help to explain better the patterns of diversification based on phylogenetic data of those groups. Anagenesis was deemed to be the most likely process of origination for polyploids in those groups. Figure 4.1 shows the repartition of the groups studied over the global vascular plant phylogeny. The groups showing purely anagenetic polyploidization do not seem to show any pattern of occurrence across plants (in brown). No part of the tree of land plant seems to be particularly affected or depleted from purely anagenetic clades. Nevertheless, only three groups of ferns out of 11 are purely anagenetic. Similarly, closely related species do not appear to evolve necesarily under the same mode, as for instance *Primula* shows cladogenetic polyploidization but *Dodecatheon* does not, or as *Aeonium* and *Aichryson* also do but *Graptopetalum* does not. The appearance or the dis-

appearance of cladogenetic behaviour thus appear to be under rapid turnover.

Table 4.2 shows the estimates of anagenetic (q_{01}) and cladogenetic ($\lambda_0 \cdot p_{0c}$) rates for the groups of species where models using cladogenesis were favoured by the thermodynamic integration method. No groups showed significantly higher cladogenetic rate over anagenetic polyploidization and only one group (*Asplenium* species of New Zealand) showed higher anagenetic over cladogenetic polyploidization. Overall, in every other group where a mixture of model was favoured, no difference between the two processes was found.

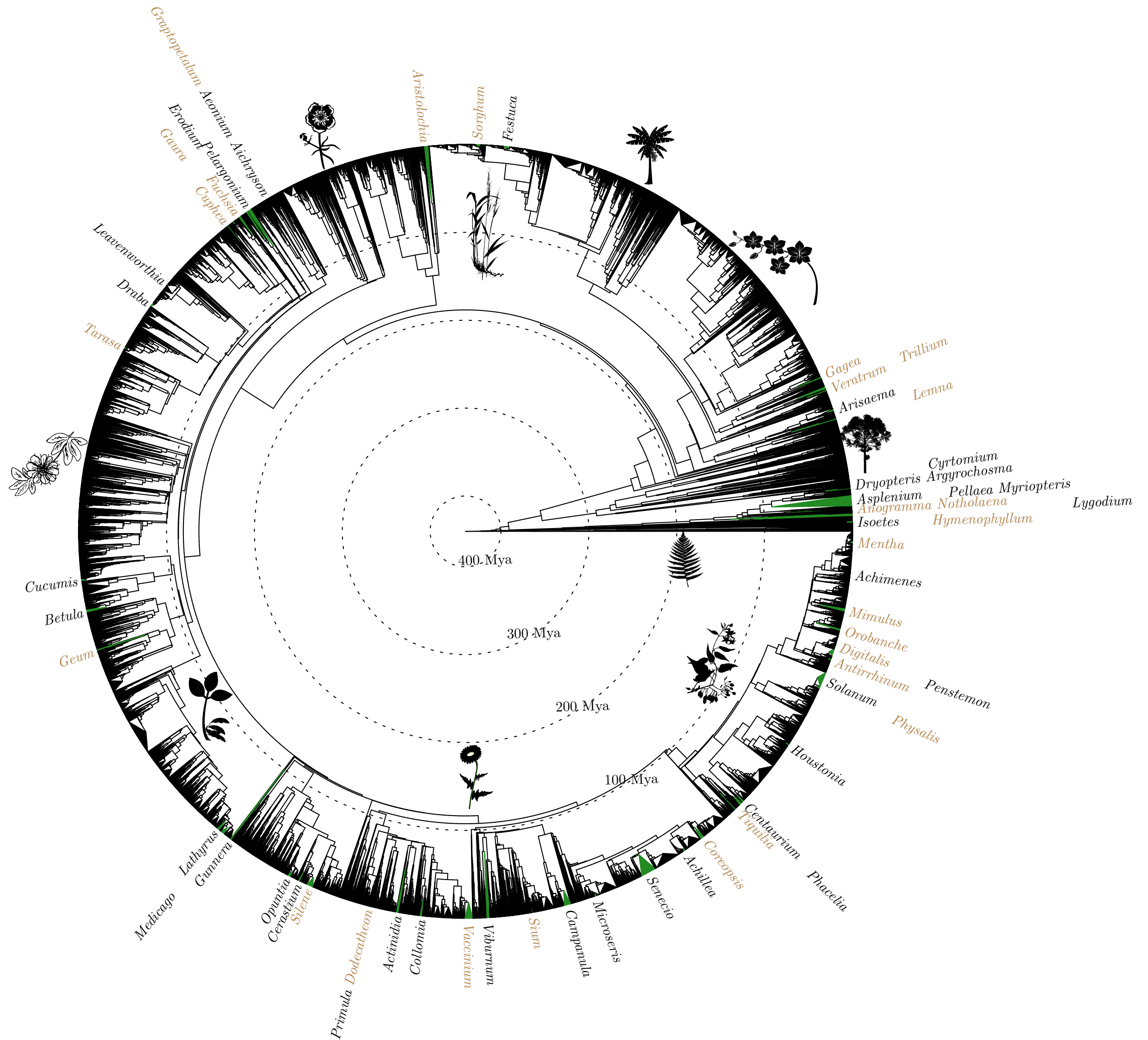


Figure 4.1: Overview of the groups studied in this study in a phylogeny of 30000 vascular plant species produced by Zanne et al. (2014). Dark green clades are the groups that were studied in the present work, with their associated genus names. Clades were collapsed to represent their diversity in the sampled tree, not their overall diversity. Brown labeled genera represent groups were cladogenetic rate was estimated as zero. Two groups, *Asplenium* and *Dryopteris*, represent respectively 2 and 3 different monophyletic clades that were analysed separately, but that all supported models with cladogenetic polyploidization.

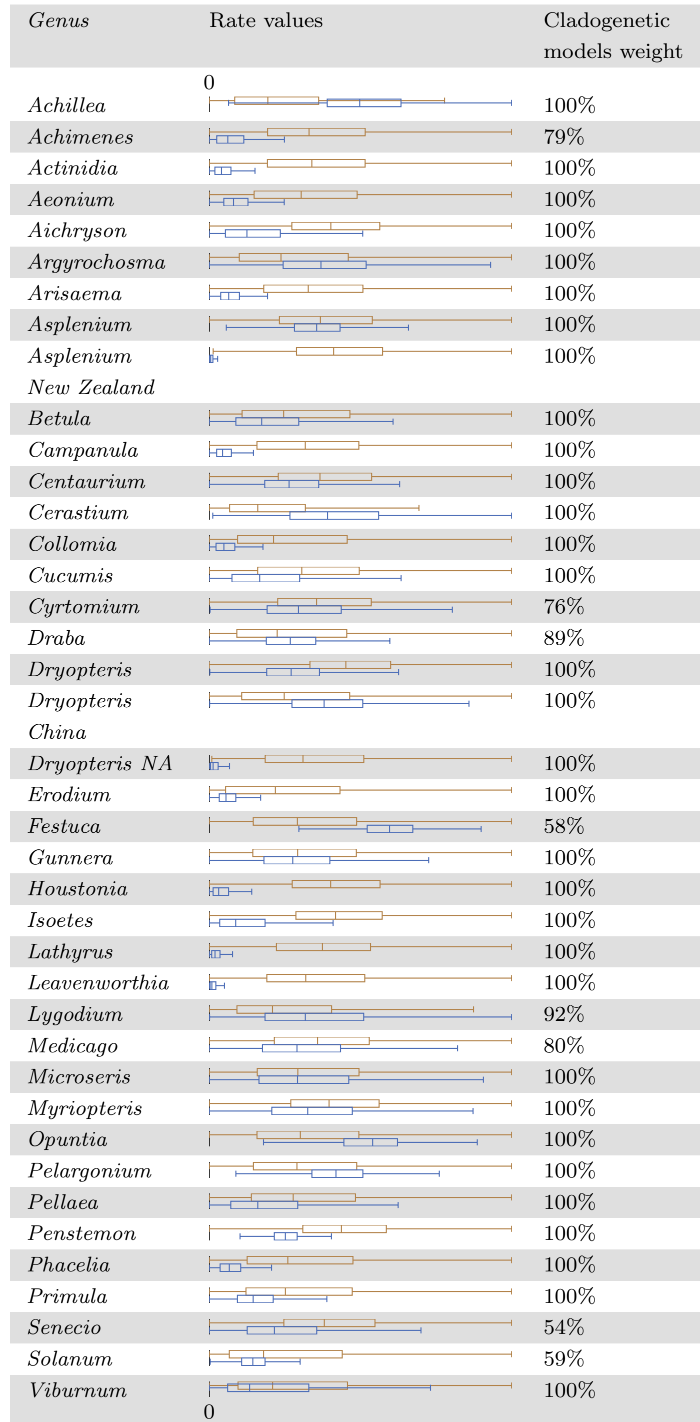


Figure 4.2: Boxplots of the estimates of cladogenetic (blue) and anagenetic (brown) rates of polyploidization from groups where models with cladogenetic processes were favoured, with the total weight explained by models with both cladogenesis and anagenesis.

4.5 Discussion

4.5.1 Origin of polyploid individuals

In our work, using phylogenies and macro-evolutionnary rate estimations, we have shown that the anagenetic polyploidization rate is at least as high as the cladogenetic polyploidization rate in the land plant genera we investigated. Anagenetic polyploidization leads to the entirety of a species to switch to polyploidy. It would occur through the formation of polyploids and the subsequent disappearance of diploids individuals. On the other hand, cladogenetic polyploidization denotes the appearance of a different polyploid species, without the loss of the parental diploid. Both processes could happen with or without gene flow occurring between the two populations during the early stages of isolation.

Recent discussion arose regarding the relevance of the anagenesis and cladogenesis notions (Vaux et al., 2015). Vaux et al. (2015) argued that anagenesis and cladogenesis were unecessary terms, that have been used loosely and to refer to different processes. Originally, Gould and Eldredge (1977) proposed that phenotypic evolution observed in the fossil record was punctuational, and that most changes occurred at speciation, as opposed to phyletic gradualism. Pennell et al. (2014) tried to untangle the search for punctuated equilibrium in macroevolutionary studies by dividing it in different questions: whether evolution is saltational or gradual, whether it happens at speciation or not, among others. Although chromosomal number can evolve gradually, by fussion or fission (Escudero et al., 2014), change in ploidy level is a purely saltational process. What we tested presently is how likely such a change can occur simultaneously to a speciation event, for which most authors agree that it is synonymous to testing cladogenetic processes (Pennell et al., 2014).

One of the key question remains on how much gene flow is preserved between the cytotypes. However, such processes, happening at the level of populations, cannot be captured by the phylogenetic models that we used presently. A variety of factors might influence the relative chance of each of the processes occurring,

but we have shown in the present work that polyploid species in general emerge as much from one of the process as from the other.

Ecological differentiation between ploidy levels could theoretically favour cladogenetic polyploidization (Parisod, 2012). It has been identified in many instances in the building up of reproductive isolation in plants (Baack et al., 2015). As phenotypic variation has frequently been associated with ploidy variation (Chen, 2007), specialization into new niches of neopolyploids might be likely and was indeed found in some instances (Hijmans et al., 2007). Polyploids would then escape competition from their parental, better adapted diploid counterparts and both populations could remain isolated and persist over time, eventually leading to cladogenetic polyploidization. Nevertheless, recent findings showed significant ecological niches overlap between polyploids and their diploid parents (Glennon et al., 2014). By studying diploid-polyploid pairs of 20 different species, Glennon et al. (2014) found that both members of the pair occupied similar climatic environments. This study casts doubt onto whether niche shifting can explain the process of cladogenetic polyploidization, although other reports on individual species reported climatic differentiation across ploidy levels (Theodoridis et al., 2013).

The origin of the neopolyploids might also influence how likely they are to displace or complement their parental diploid species. Polyploids have traditionally been classified as either allopolyploids or autopolyploids. Although sometimes distinguished cytogically, based on the chromosomal segregation pattern (discussed by Soltis et al. (2010)), these terms have been mainly used in a taxonomic context. Autopolyploids then refers to polyploids whose parents belong to the same species whereas allopolyploids are hybrids between two different species. Outcomes of allopolyploidy events have been frequently studied in the lab (Liu, 2003; Petit et al., 2010) or over evolutionary timescales (Hegarty et al., 2013; Senerchia et al., 2014). For instance, non additive gene expression in polyploid hybrids with respect to the diploid parents is expected because of complex interactions between the two different genomes (Cox et al., 2014; Yoo et al., 2014). Besides, allopolyploids could benefit from hybrid vigor without suffering potential

defects in chromosomal pairing of the diploid hybrids at meiosis (Otto, 2007). Allopolyploids also exhibit extensive gene expression variation (Buggs et al., 2014), epigenetic remodelling (Liu, 2003), differential mobilization of repeat sequences (Renny-Byfield et al., 2013), compared to their progenitors. All mechanisms possibly favour phenotypic alteration. In their study, Glennon et al. (2014) did not differentiate between autopolyploids and allopolyploids, so it might still be possible that allopolyploid formation favours cladogenesis through occupancy of different climatic and ecological niche. If most allopolyploidizations lead to cladogenesis and autopolyploidization to anagenesis, then our results are in agreement with the recent finding of autopolyploids being as frequent as allopolyploids (Barker et al., 2015).

Anagenetic polyploidization could occur through the complete replacement of the diploid population by the polyploid one, that can be mediated by neutral or adaptive processes. If anagenesis is to occur, both diploids and polyploids must be sufficiently related so that they form one single species complex. Thus, some degree of coexistence between the two populations is expected after polyploidization. The mechanisms governing the evolution and coexistence of polyploid and diploid populations have been heavily investigated. One of the first processes described is the minority cytotype exclusion, whereby the individuals of the rarest cytotype are driven to extinction (Levin, 1975). Using populations of *Chamerion angustifolium*, Husband (2000) found indeed that the fitness of tetraploid individuals was correlated with their prevalence in the population. The role of triploids in the fate of diploid and polyploid coexistence in particular has been emphasized (Petit et al., 1999). They can act either as blocks, preventing gene flow to occur between the cytotypes, or as bridges, in this case enabling interbreeding. For instance, Husband (2004), in the same system, discovered that a significant portion of new tetraploids at each generation originated from triploid individuals. Similarly, Slotte et al. (2008) found that in the genus *Capsella*, polyploidization did not lead to instantaneous isolation from the diploid parents, possibly through gene flow *via* the triploid hybrids. On the other hand, many other examples where triploid act as blocks have been found in nature (Moyle et al., 2004; Baack, 2005; Köhler et al., 2010). However, using mathematical modelling of gametogenesis

in mixed cytotypes systems, Suda and Herben (2013) showed that parameters related to the even-ploid organisms were the most important for determining the subsistence of polyploids in a population. They emphasized that simple models were sufficient to explain the coexistence of mixed cytotypes populations and that the prerequisites were less stringent than previously thought.

4.5.2 Methodological considerations

In this work, we have used BiSSE and BiSSE-ness to infer macro-evolutionnary parameters governing the formation of polyploid species. A note of caution has recently been issued regarding the use of this suite of methods (Rabosky and Goldberg, 2015), emphasizing a high risk of Type I errors. The authors showed that if a trait was indeed influencing diversification in a phylogeny but one was to study another trait without effect, spurious correlation between diversification and this trait was likelier to be found for the slowest evolving traits. Following this, they advised caution when studying traits that changed only once in a potential ancestor and was inherited by a single clade on the tree, rather than independently changing in multiple places of the phylogeny. In our study, we used model choosing to infer which evolutionary scenario was more likely in 67 plant groups. Selecting an appropriate number of groups to study the trait of interest is one of the advice given by Rabosky and Goldberg (2015) to avoid the pitfalls of the BiSSE methods.

Magnuson-Ford and Otto (2012) have investigated the precision of the BiSSE-ness method to estimate anagenetic and cladogenetic rates. They found good performance for the parameters studied presently (p_{0c}), which is the cladogenetic speciation probability, but poor estimates for p_{0a} and p_{1a}, which are the probability of asymetric character changes, that we did not consider here. Overall, Magnuson-Ford and Otto (2012) found similar estimates of speciation and extinction between BiSSE and BiSSE-ness, and, consistently with other methods, better speciation estimates compared to extinction. This result shows that we can draw reliable conclusions regarding the anagenetic and cladogenetic processes

using BiSSE-ness.

Davis et al. (2013) also explored the power of the BiSSE methods, focusing more on Type II errors. Simulating phylogenies with character states influencing speciation and extinction, they found a very low chance of recovering significant results when the phylogenies show extreme character state bias or low number of extant species. They advised caution for authors drawing conclusions based on analysis on trees having less than 300 species (Davis et al., 2013). In the present study, all trees used included less than 300 species, and the biggest ones had 195 (Loliinae subtribe), 162 (*Draba*), 144 (*Opuntia*) and 80 (*Medicago*) species. Thus, the phylogeny we used presently are generally undersized for testing the association between one trait and diversification. However, the power of BiSSE and BiSSE-ness methods to test for cladogenesis and anagenesis, with respect to the size of the tree, has not yet been investigated.

4.6 Conclusion

We showed that polyploids are as likely to emerge from anagenesis as from cladogenesis. Nevertheless, the influence of hybridization, reproductive mode and life history traits over the contribution of each process still need to be clarified. This could be performed once data regarding the groups considered in this study have been systematically collected, enabling the correlation between those traits and the likeliness of anagenesis and cladogenesis polyploidization.

Chapter 5

Perspectives

5.1 Outline of my work

In the first part of my work, chapter 2, I have simulated how one could expect phylogenies to look like under a wide variety of models of diversification, going from the standard "nothing interesting is happening" scenario to scenarios mimicking some of the hypothesis linking diversification and polyploidy (Laurent et al., 2015). I have used those simulations to learn how methods can be tricked into wrongly favouring one model over another. Still, I have shown that in general the methods tested performed well at detecting the events they were set to identify, even in the presence of substantial amount of noise. Moreover, as a multiple number of factors are constantly influencing the fate of lineages, I have shown that aiming at a global understanding of biodiversity trends will probably require unification or comparison of most diversification methods. Finally, I have emphasized that taking into account and identifying mass extinction events was an important part of the diversification pipeline, and that developing new tools for this task is important (May et al., 2015), in particular if one is to show that some traits allow escape from such events.

In chapter 3, I have studied how scenarios with delayed rise in biodiversity can be identified using a variety of methods. I have shown that even using frameworks that make different assumptions about how diversity changes, one can reconstruct

the same biological histories. I have also brought some evidence and discussion of how actinopterygians respond to polyploidy, in particular how we did not detect signal for the radiation time lag model using both phylogenetic and fossil data. We tested a few relatively old paleopolyploidization events, at the origins of Teleostei and Salmoniformes, as well as more recent or undated events, those occurring in cyprinids and botiids. Luckily, the fossil record of actinopterygians enables us to test the effect on diversification on old events.

The final part of my project, chapter 4, is related to the link between polyploidy and evolutionary success, but I explored an alternative angle of the debate. I used models incorporating anagenesis and cladogenesis to explain polyploidy fixation and quantified the contribution of each process. Using phylogenies, I showed that in a significant number of cases, the addition of cladogenesis does not help to explain the data we are observing. For the cases where cladogenesis did help, it was estimated that polyploids emerged as much from this process as from anagenesis. Nevertheless, the actual biological traits that influence the favour on process over the other still need to be thoroughly investigated. I have also provided a discussion of which botanical traits might be at play in the balance of these two processes.

5.2 Linking diversification and polyploidy

Replication is key to the scientific approach but carrying it out is cumbersome in macro-evolutionary studies (Bromham et al., 2015). Redoing an experiment in the lab is feasible, but we will never be able to do controlled and replicable experiments for determining whether the dinosaur extinction led to the diversification of mammals (O'Leary et al., 2013), whether the acquisition of novel microRNAs families is key to the increased morphological complexity of vertebrates (Heimberg et al., 2008) or whether genome size correlates with fitness in some beetles species (Arnqvist et al., 2015). Thus, the need of designing another kind of "replication procedure" is paramount to hypothesis-testing in macro-evolution. It can be done by repeatedly identifying our trait of interest in the tree of life and running our

methodology of choice in each of those instances, hoping that given enough data points, some signal will emerge out of the noise induced by all the aside variables that cannot be controlled in such settings. Doing so for polyploidization requires the study of their instances across every groups in the tree of life.

Settling the debate concerning the link between polyploidization and diversification will require extensive additional computational experiment and data collection. Moreover, it will also include testing a variety of models of diversification to encompass the many hypothesis presented, as was performed in the present work. Uncertainties regarding every step of the endeavour can be addressed in a variety of ways. First, phylogenetic uncertainties can be incorporated by integrating many sources of information. Right now this correction is mostly carried on by using a distribution of trees for the diversification computations. Nevertheless, querying resources storing phylogenetic knowledge, such as TreeBase (Sanderson et al., 1994) or Open Tree of Life (Hinchliff et al., 2015), for all phylogenies concerning our organisms of choice can also be perfomed in order to incorporate uncertainties about relationships between species.

Next, the development and the use of appropriate diversification methods are also paramount in this search. For instance, as BiSSE-like methods deal with character states at the tips, one must recognize that they can only bring answers regarding recent events of polyploidization. Thus, paleopolyploidy must be investigated through other tools, as was performed in this work with fossil data and methods reconstructing diversification rates through time. Doing so will permit us to determine if different patterns of diversification are associated with old or recent polyploidizations.

To test polyploidy effects on diversification, one should be able to disentangle the effects of polyploidy and hybridization. For instance, Marcet-Houben and Gabaldón (2015) determined that the paleopolyploidization event in the lineage of yeast was actually masking an hybridization event. They showed that old hybridization events were identifiable using molecular data and that polyploidization was a possible means of fertility restoration for the hybrids. Thus,

the possibility exists that some of the effects seen after polyploidization can be attributed to hybridization. Therefore, before giving a definite answer regarding this question, formally distinguishing between allopolyploidy and autopolyploid is necessary. Recent work aiming for better identification of homoeologs (Bertrand et al., 2015; Glover et al., 2016) or assessment of the polyploidy origin with whole genome data (Roux and Pannell, 2015) might for instance enable us to perform such discrimination.

5.3 Divergences and convergences between plant and animal polyploidies

Through my work (chapters 3 and 4), I have used data from two of the groups where polyploidy has been the most investigated, namely actinopterygians and plants. Although I have not tested the same hypothesis in both cases, one interesting open question remains whether polyploidy has different outcomes between plants and this group of animals. One could test for instance if recent polyploidizations in actinopterygians also originate equally from anagenesis and cladogenesis. Nevertheless, the sparsity of recent polyploidy in actinopterygians, contrary to plants, would enable us to test the hypothesis solely on a reduced number of groups. Regarding the radiation time lag hypothesis, testing the effect of old events in angiosperms is hard because of lack of fossil data. Apart from using a backbone tree and families richness as a proxy for evolutionary success (Tank et al., 2015), an alternative test could be performed using the time-varying diversification estimations of chapter 3. A potential experiment could be designed by identifying on the land plant phylogeny the ancestral nodes where paleopolyploidy is likely to have occurred. Then, one could reconstruct the subsequent diversification trends, to check if an increase consistent with radiation time lag is detected. However, the signal in favour or against an increase in diversification rate would mostly rely on the few speciation nodes occurring after the paleopolyploidization. The uncertainty on such results might therefore be very wide.

The most apparent discrepancy between plants and animal regarding polyploidy relies in the fact that it is much more widespread in angiosperms (Jiao et al., 2011) than in animals. Additionally to a few actinopterygians, variation at the species level in vertebrates appears to be limited to amphibians (Mable et al., 2011) and to few squamates (Bogart, 1980), although the potential case of a tetraploid red vizcacha rat has been debated for more than a decade (Gallardo et al., 1999; Svartman et al., 2005; Gallardo et al., 2006; Teta et al., 2014). Muller (1925) argued that the reason for this discrepancy was not lack of viability of polyploid animals but because of their sex chromosomes. This assumption has been revisited first by Orr (1990), who emphasized the role of dosage compensation between sexes as major deterrent for polyploidization, and then by Mable (2004), who chose to focus on the traits that would favour polyploidy rather than prevent it, such as elevated number of unreduced gametes or mechanisms favouring intra-cytotype mating. Mable (2004) also emphasized the link between reproductive mode and polyploidy, whether it be for plants or animals.

Cases of polyploid asexuals have been identified in insects and in squamates (Otto and Whitton, 2000) as well as in molluscs (Larkin et al., 2015). This latter case demonstrated that under lab conditions, triploid and tetraploid asexuals outgrew sexual diploids (Larkin et al., 2015), raising the question of how sexual diploids could survive in the wild. One possible explanation could reside in the limited availability of resources in the environment. In plants, Guignard et al. (2016) showed that growth of polyploids under nitrogen and phosphorus limitation was significantly lower than diploid growth. Similarly, Neiman et al. (2013) argued that polyploidy costs could provide a way for sexual individuals to persist when asexuals are associated with higher ploidy levels.

In plants, the conversion from hermaphroditism to dioecy through polyploidization has been found to be quite frequent (Ashman et al., 2013) although the reverse case has also been found, in *Mercurialis annua* for instance (Buggs and Pannell, 2006). In both plants and animals, polyploidy and sexual systems are suspected to be intertwined, but determining which is cause and consequences is hard. For instance, polyploidy could lead to disruption of gene expression

responsible for sex-determination, but unreduced gametes as the results of a modified reproductive system could favour the apparition of polyploids (Ashman et al., 2013). Gonochorism is much more frequent in animals than dioecy in plants, as well as sex chromosomes and dosage compensation (Ashman et al., 2013; Wertheim et al., 2013). Thus, one of the key component for concluding whether polyploidy has different outcomes between plants and animals would be to compare across the two groups events occurring on the similar biological background of the organisms.

Polyploidy does not have the same phenotypic impact on plants and animals. Although cell size is usually doubled with the doubling of the genome, total organ size is controlled during the establishment of the animal body plan (Orr-Weaver, 2015), hence the overall size of a polyploid animal is expected to be similar to a diploid one. On the other hand, polyploid plants are usually bigger than their closest diploid relatives (Mable, 2003). Plant polyploids are expected to be more phenotypically different compared to their parents than animal polyploids, thus the evolutionary perspective could diverge between the two orders. As anagenesis was as likely than cladogenesis in plants and if the probability of cladogenesis does depend on the phenotypic divergence between cytotypes, then the same results, or an even greater contribution of anagenesis, should be expected in actinopterygians.

Historically the study of the link between diversification and polyploidy has rarely been performed on other organisms, whether they be animals or not. For instance, polyploidy in yeast (Scannell et al., 2006) and in plasmodium (McGrath et al., 2014a,b), was identified early on and heavily studied at the molecular level, has not been investigated through its link with diversity or success, probably because these groups are species-poor compared to plants and animals. Evidence of paleopolyploidy in other groups of animals has also been found, in one of the chelicerata ancestor for instance, as the genome of the present day horse-shoe crab showed (Kenny et al., 2015). Some other parts of the plant tree of life have recently been identified as paleopolyploids, such as conifers (Li et al., 2015; Scott et al., 2015b) or horsetails (Vanneste et al., 2015).

5.4 Short and long term consequences of polyploidy

Polyploids could be no more likely than diploids to survive when they originate but once species or lineages underwent polyploidization, their evolutionary potential could be higher, as the radiation time lag model posits (Schranz et al., 2012). Some authors went as far as suggesting that the mechanisms leading to polyploid formation through unreduced gametes were not molecular errors but selected for because generating polyploids was adaptive. Mason and Pires (2015), in an opinion article pushing for polyploidy as an adaptive phenomena, concluded with:

> "Taking these observations together, we propose that unreduced gametes are maintained across widely disparate lineages because the ability to produce unreduced gametes facilitates lineage survival by allowing polyploid speciation, particularly in response to stress."

For this to be true would require natural selection to be effective on a trait that could eventually be adaptive in a few generations. It would suggest that natural selection is strong enough to favour the fitness loss induced by unreduced gametes that would eventually lead to an increase in fitness in some undetermined time in the future. It is akin to thinking that transposable elements litter eukaryotic genomes because they provide variation to be selected for (Oliver and Greene, 2009) or that mutations rate is not nil for the same reason. To the contrary, Lynch (2011) argued that natural selection is not strong enough to lower mutation rates below a certain threshold, that depend on the effective population size of the species under consideration. Subsequent enhancement to the replication machinery would then be lost because of drift.

Regarding the hypothesis of Mason and Pires (2015), a few points particularly disprove their idea, first specifically concerning the process of gametogenesis, and

then more generally applicable to any hypothesis that prompts natural selection to favour a trait that will be adaptive only in a few generations.

For gametogenesis, Mason and Pires (2015) argue that since unreduced gametes are created everywhere in the tree of life and through many mechanisms, their appearance must be selected for. Moreover, they hypothesize that unreduced gametes are more likely to appear when environmental conditions are changing so that the polyploid offspring can adapt. However, meiosis is made of an important number of phases and indeed any mistake in one of those phases could lead to defects, whether the result be aneuploidy or unreduced gametes. As meiosis is one of the most fundamental cellular process for sexual species, upon which natural selection is the most likely to act, perfect gametogenesis would be highly adaptive: the prospect of out-competing any gamete of their peers and having the most fit progeny would be very favoured. The selection on genes associated with meiosis during the first generations of tetraploid *Arabidopsis* (Le Comber et al., 2010; Yant et al., 2013; Wright et al., 2014) shows this phase is strongly submitted to natural selection. Thus, it is very unlikely that "meiotic mishap" (Mason and Pires, 2015) during gametogenesis could be favoured by natural selection, in the presence of better variants.

Finally, the unit of selection traditionally recognized in classic darwinism is the individual. However, for the supposed adaptation in this case, polyploid speciation, the advantage will be granted at a much higher lever than the individual. This selection pressure is assumed to happen at the level of the lineage, as stated by the authors, and the timescale at which it will unfold will also be much larger. Thus the negative effect affecting the individual, over its life span, must be counterbalanced by the positive effect at the species level, that will not be apparent before generations. Unrealistically large population size, for multicellular organisms, would be needed for the evolution of such "feedback loop" from the species to the individual. Individuals belonging to any species, however fit as a group it is to exploit resources or pass their genes to the next generation, will be out-competed by more fit individuals of other less collaborative species, as Mayr (1992) put it when he proposed "species turnover" as a better term than "species

selection". The same goes for species that would be more efficient at creating new species. For these few reasons I do not think that unreduced gametes production can be selected for as a way to produce more polyploid speciations.

Appendix A

Discussion regarding "Polyploidy can drive rapid adaptation in yeast" by Selmecki et al. (2015)

Recently was published by Selmecki et al. (2015) a study that is highly relevant to the subject of the present manuscript. In this paper, the authors propose to test the differences in adaptation to a new medium of *Saccharomyces cerevisae*. They derived isogenic lines, stained with two different markers (Cyan and Yellow) and constructed clones with different ploidy levels: 1N & 2N are usual states in yeast, but they also constructed 4N cells. The new medium they wanted to test adaptation to was chosen as raffinose, a poor-carbon source medium for yeast, because it is constituted of one molecule of fructose, one of galactose and one of glucose, which have to be broken up before they can be processed.

To test adaptation, they performed the following competition experiment: inside a number of well, they put the same number of cyan and yellow stained clones of the same ploidy level, with the raffinose medium. They used plates of 12*8 wells and left some of the wells empty to check for cross contamination, which means there are approximately 80-90 experiment going on on each plate. Each 24h, they performed flow cytometry to count in each well the number of cyan and yellow marked cells. Departure from a 50/50 proportion inside a well

means one of the individual acquired beneficial mutation and its progeny starts outcompeting every other individual inside this well. They let the first experiment run for 250 generations (Figure A.1).

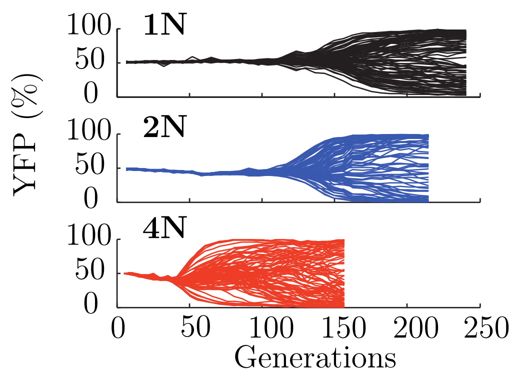


Figure A.1: Percent of yellow stained cells in each experiment through generations. Adapted from Selmecki et al. (2015)

We can see that individuals in the 4N experiments indeed adapt faster than in the other ploidy levels: the proportion of cyan/yellow marked cells start diverging from 50/50 at the 50th generation for the 4N, whereas divergence does not occur before the 100/120th generation for 2N and 1N cells. They also show the rate of adaptation for each ploidy state and each replicate, which is almost always higher in any 4N experiment (Figure A.2).

They have, according to Figure A.2, 2 replicates (plates) for 2N and 3 replicates for 4N. The number of points in this figure is, 82 (1N), 92 (2N, A), 68 (2N B), 85 (4N A), 86 (4N B), 92 (4N C). This fits an approximate number of 80 wells per plates. The problem being that in the supplementary material, the author mention 2 plates for the 1N and 3 plates for both 2N and 4N, making it respectively 173, 264 and 265 parallel evolution experiments. The 3 plates for the 4N are present in the main text analysis, but one plate is missing for each 1N and 2N, and are never mentioned again. The 2N states are always stated as having 160 experiments, differing from the supplementary material information.

They performed Illumina and SOLiD sequencing in order to identify SNPs

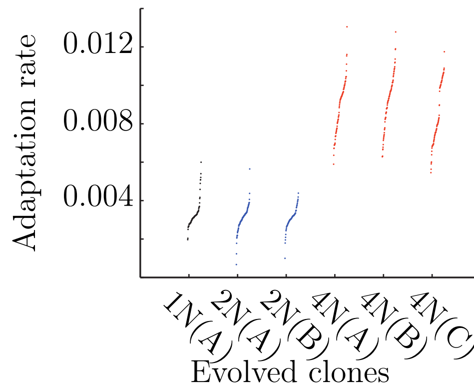


Figure A.2: Rate of adaptation for every ploidy level, for each replicate experiment. Adapted from Selmecki et al. (2015)

after 250 generations, for 22, 24 and 28 clones for the 1N, 2N and 4N states. They also performed a number of array of Comparative Genomic Hybridization (aCGH) to estimate copy number variation, in particular in chromosome copy numbers. In total, they are able to estimate the number of chromosome for 30 4N evolved clones, and they realized that most of them kept the chromosome XIII, so they went on to do a fitness analysis. They wanted to check the effect of the addition of another chromosome XIII to the 4N ancestor, and observed a much better fitness than the 4N ancestor without additional chromosome XIII (Figure A.3). Nevertheless, we can also see from the plot that the 4N ancestor has a much lower fitness than the 2N ancestor on the new medium. The fitness difference is pretty important (-0.5) and had the 2N and 4N ancestors competed against each other in the new raffinose medium, the 4N would probably have disappeared in every case.

When reading this article and doing some background bibliography search, I came up with this highly relevant paper to this study: Allelic variation, aneuploidy, and nongenetic mechanisms suppress a monogenic trait in yeast, Sirr et al. (2015). This paper is more medically focused, but basically looks at the adaptation of yeast strains to galactose, in the context of a human disease (classic galactosemia). They removed a gene on chromosome II of yeast, GAL7, that prevents the organism to grow on galactose medium because galacticol and galactose-1-

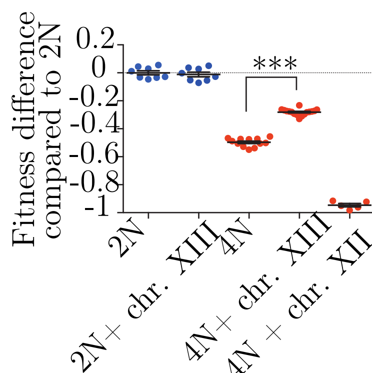


Figure A.3: Fitness values of each expriment with respect to the mean diploid fitness value, for diploid and tetraploid cells, with or without aneuploidy for chromosomes XII and XIII. Adapted from Selmecki et al. (2015)

phosphate accumulates in the yeast cells, leading to toxicity. It is interesting because in the case of human classic galactosemia, it is thought that toxicity comes from the same metabolites. To do this, they constructed different yeast crosses missing this GAL7 gene and grew them on a medium composed of 0.5% galactose and 2% raffinose. At the end, they selected 247 strains that overcame the toxicity induced by the missing GAL7 and sent them to perform some genotyping/sequencing. The strains considered here are haploid. They found that around half (122) of the strains had kept one copy for every chromosome, but for the other half whose chromosome number had varied, 92 of them were disomic at chromosome XIII. So although in Selmecki et al. (2015), no difference in fitness for 2N strains on raffinose medium was observed when you add a chromosome XIII, it is clear that there is a fitness effect for the 1N strains (Sirr et al. provide some statistical assessment in their paper). It appears that chromosome XIII is playing an important role in galactose pathways for yeast, and it might be ploidy dependent. For instance, it might be that the GAL7 gene is disrupted at the 4N level, and that the galactose pathway can be rescued by duplicating chromosome XIII. Anyway I think this study is pretty relevant to Selmecki et al. (2015), predates it, but it is not cited in this latter paper.

Selmecki et al. (2015) went on to perform fitness analysis by competing the evolved clones with the 2N ancestors. They wanted to check if the faster adaptation capacity of 4N was still apparent after the 250 initial generations. So they

chose 48 of the evolved 4N clones that had reached the same fitness as the 2N ancestors. So over 265 clones, after 250 generations, only 48 of them had reached the level of 2N ancestors. The fitness values of the other evolved 4N clones are not given, so it could be that the 210 left had better or worse fitness than the 2N ancestor. Was it that the fitness values of the 210 4N evolved individuals were not computed, or they only chose to mention those who had the same fitness as the 2N ancestors? Anyway those 48 clones are left evolving for another 500 generations and are compared to the two replicate experiments of the 2N strains (Figure A.4). The 4N after 500 generations reached a better fitness than the 2N, so they crossed a bigger fitness gap during the same amount of generation (250), as they started with the same fitness as the 2N ancestor.

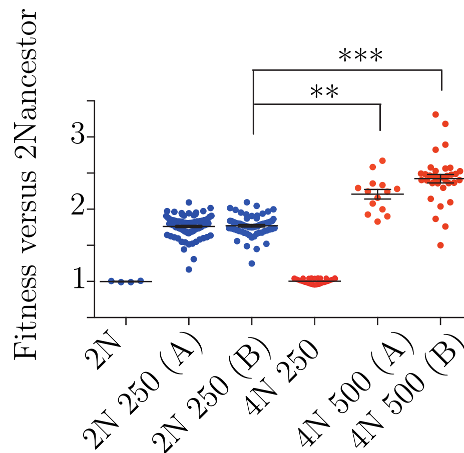


Figure A.4: Fitness comparison between diploid ancestor and a variety of evolved clones. First are the two replicate plates of diploid clones, evolved for 250 generations. Then a number of tetraploid clones that evolved for the same amount of generations are selected and evolved for 250 generations more.

I think the data so far presented fits pretty well the following scenario: 2N cells are much stabler than 4N cells so their fitness is much better in the new medium than 4N (shown in Figure A.3). 4N are much farther away from the optimum of the adaptive space than the 2N, so they can take bigger steps and the fitness change associated with each mutation is higher than the fitness change associated with mutation in 2N (Figure A.2)), for instance, by growing towards more stability or by preventing multivalent formation. 4N strains also adapt by improving

raffinose uptake, as shown in their SNP analysis. After 250 generations, some of the 4N clones (around 20%) reach back the fitness of the 2N ancestors, the status of every other 4N is currently unknown. One chromosome (XIII), is shown as having important role, maybe ploidy specific, in galactose pathway, which was already shown by a previous publication (Sirr et al., 2015).

To explore this theory, I downloaded the Supplementary Table 1 of Selmecki et al. (2015), that lists all SNP that were identified in the evolved strains. Of those, I extracted those occurring in the 4N and having non-synonymous mutations. I put this list of 50 genes on the *Saccharomyces* Genome Database (Cherry et al., 2012) to check what was the role of those genes. Just looking at the gene standard name, one can see that hexose-pathways genes are indeed present (SNF3, "sucrose Non Fermenting", which is discussed in their main text, RGT1 and RGT2, "restores glucose transport", PGU1, "polyGalactUronase") but I also found some interesting ones such as: CHD1 ("Chromatin organization modifier, Helicase, and DNA-binding domains"), MSC7 ("Meiotic Sister-Chromatid recombination"), SMC4 ("Structural Maintenance of Chromosomes") and other that might be very remotely linked (CDC31, "Cell Division Cycle", according to its overview, it is linked to the duplication of the Spindle Pole Body, involved in many aspects of yeast cell cycle, chromosome segregation among others). When one does the gene enrichment analysis for those 50 genes, no enrichment is found ever, using the proper correction, whether it be for hexose pathways or chromosome segregation/stability. The problem with this very raw analysis is that it discards the fact that some genes are found multiple times across the replication experiments.

To conclude on the paper, I think that obviously two things are going on, one is adaptation to raffinose processing and the second is related to the chromosomal stability of the 4N. This leads to the following observation: a proper control is lacking in the experimental set up. Indeed, competition experiments were carried between cells of the same ploidy level in the raffinose medium, but no competition experiment are made on glucose, the standard medium where yeast is grown. If this would have been done, we would have been able to discriminate between

the two effects for the 4N: stability problems on the glucose medium, stability problems plus hexose uptake in the raffinose medium. Moreover, it would have been particularly interesting to perform competition experiments between different levels of ploidy, to check if polyploids could outcompete normal diploid and haploid cells.

This paper is also highly relevant for the debate of the impacts of polyploidy on the evolutionary success of species. I think one of the pervasive question regarding polyploidy is whether or not polyploidy was fixed for adaptive reasons in some places of the tree of life or if neutral processes are enough to explain its appearance. In this experiment, we have seen that the 4N could have successfully established in a new environment only if it were devoid of 2N, otherwise they would have been outcompeted from the very start. Thus essentially, fixation of 4N would not have been dependent on their higher adaptability capacities shown by the study. Even though I'm doubtful that the data presented so far provides an example where adaptive process favors establishment of 4N populations, it could be that polyploids are better adapted to new environment. But apparently switching to raffinose for yeast is not one of such environment. Nevertheless, this pattern could theoritically correspond to a case of radiation time lag, where polyploid success can only occur with the dispersion in an environment where the diploid parent is not present.

Appendix B

Looking for other events of polyploidization in Teleostei

During my study, I endeavoured to identify in some teoleost groups polyploidization events to add data points for my aim of linking it to diversification. Here I describe the process and the work I have performed in that direction.

B.1 Identification of the clades of interest

I made use of a collection of chromosome counts of more than 3000 actinopterygians species, over a total of more than 25000, in order to try to identify potential groups of interest (Arai, 2011). I extracted the counts for each species and I grouped them using basic taxonomical information (Froese and Pauly, 2015). I built taxonomical equivalence database of fish by extracting the data for species synonymous names of FishBase and using taxonome (Kluyver and Osborne, 2013) to resolve uncertainties on sample names. I identified lowest taxonomical groups harboring at least two-fold chromosomal count variation and good enough sampling of chromosome counts suitable for trying to identify past polyploidy events. Chromosome number in actinopteryigians is generally well conserved and ranges between 40 and 50 at the diploid level (Arai, 2011), making polyploidy identification easier. Overall, I found seven potentially interesting groups, for which

polyploidy had been well documented before (Figure B.1).

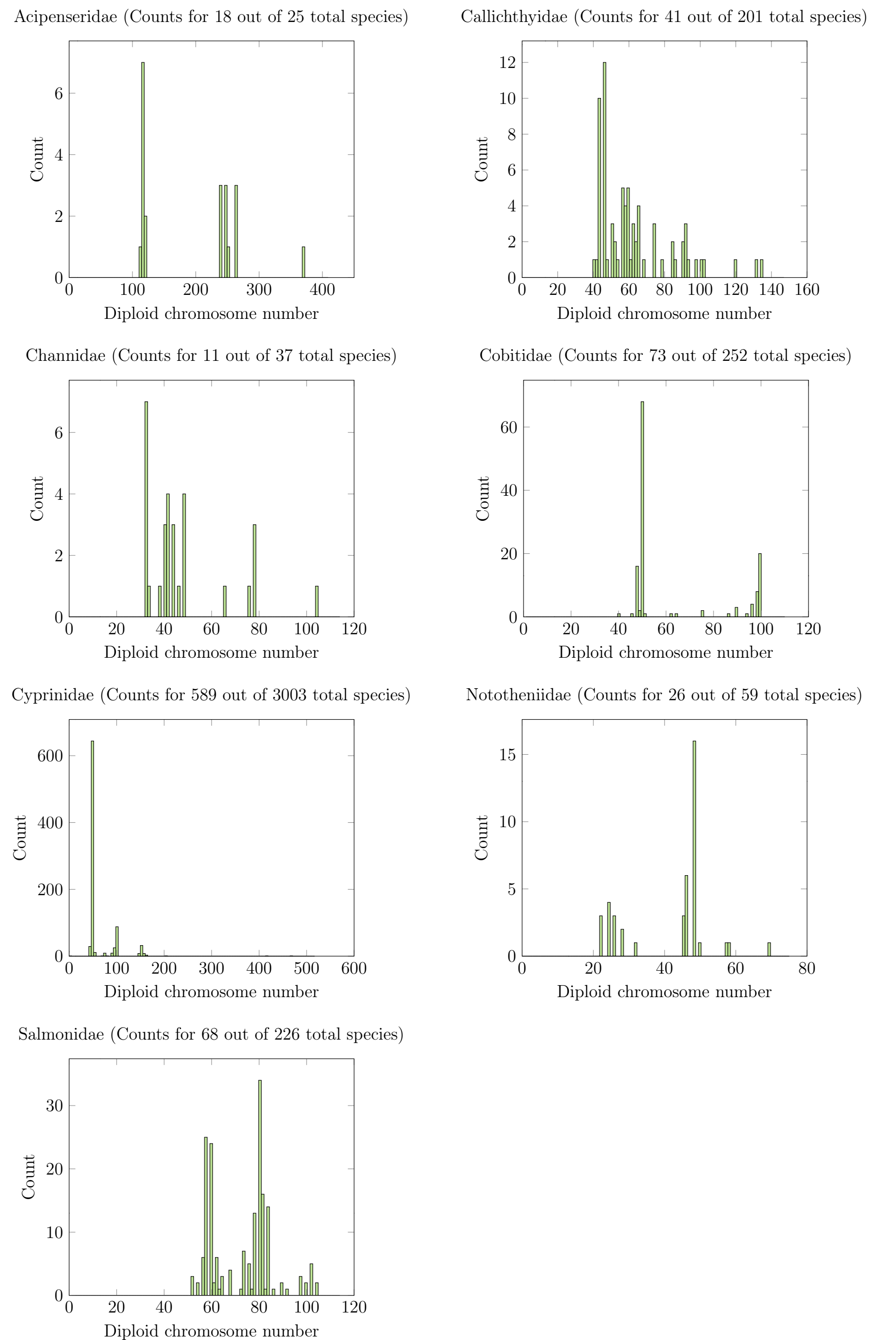
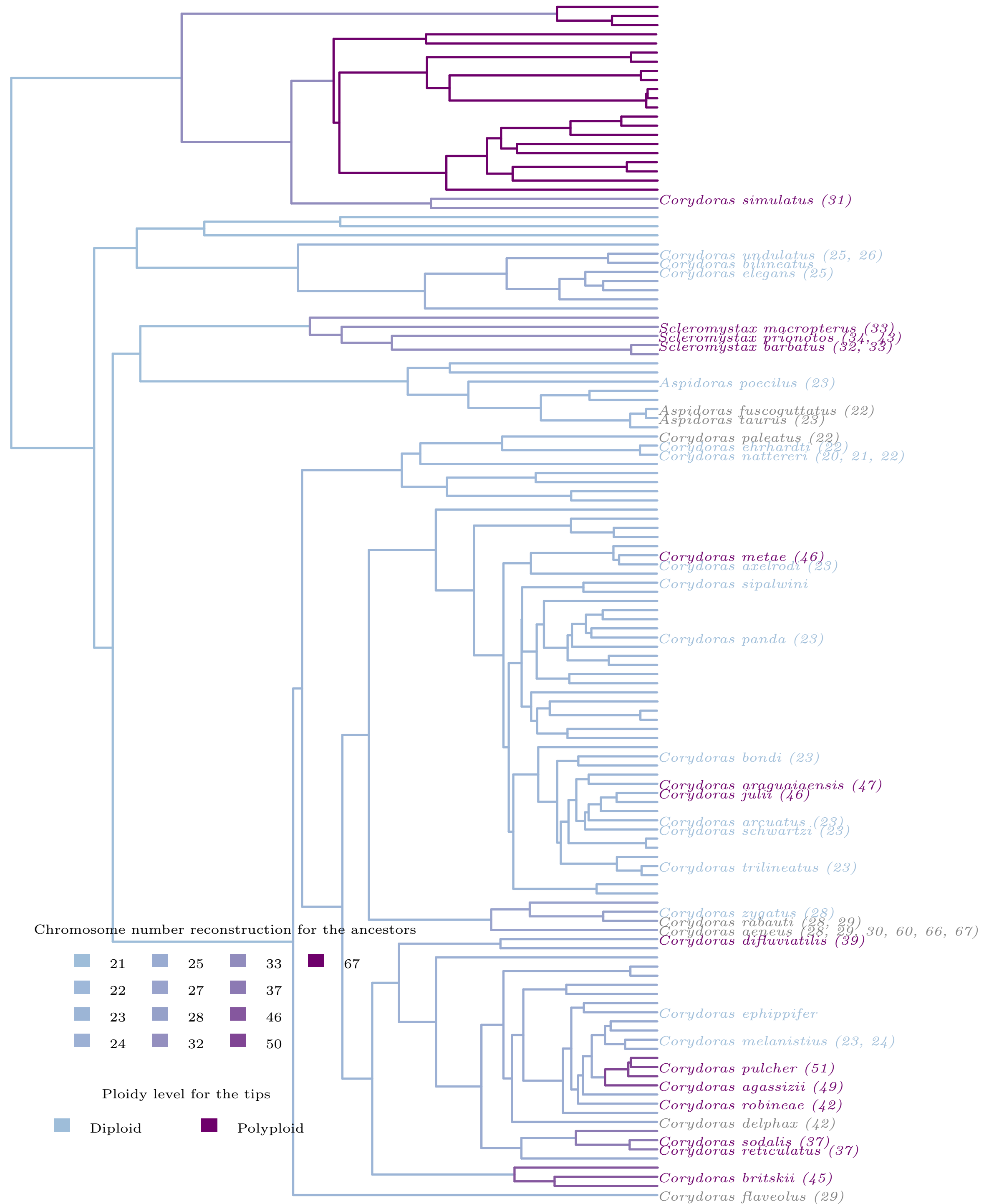


Figure B.1: Histogram of chromosome counts for selected actinopterygian groups

Following this, I chose to reconstruct the phylogenetic relationships for the Callichthyidae family. Members of this neotropical family are co-mimics and show important convergence of their color displays (Alexandrou et al., 2011), hence the species number ranging from 150 to 200 could be a large underestimate. Polyploidy has been frequently hinted at in this family (Leggatt and Iwama, 2003). I used the pipeline presented in section 4.3.1 to reconstruct the phylogeny for the 131 species that had molecular data available.

B.2 Polyploidy event determination in Callichthyidae

Chromosome counts in Callichthyidae, that includes the *Aspidoras, Corydoras* and *Schleromystax* genera, ranged from 20 to 67 at the gametophytic state. Fifty-six counts were available for 34 species out of the 131 that were included in our phylogeny (26% coverage). The four species of *Schleromystax* were recovered as monophyletic whereas *Aspidoras pauciradiatus* was found outside of the monophyletic group formed by 7 other *Aspidoras* species. Monophyly was supported for this genera, but only by phenotypic data (Britto, 2003). Some species had different chromosome counts but still appearing to be consistent with diploidy: *Corydoras undulatus* (25 or 26), *Corydoras melanistius* (23 or 24), and *Corydoras nattereri* (20, 21 or 22). Other species, scattered across the tree, showed intermediate chromosome counts, ranging from 28 to 33. An important portion of the species had more than 37 chromosomes, possibly indicating tetraploidy. Finally, 2 species had diverging chromosome counts that could indicate either different ploidy level in the same species, wrong species assignment of some counts, or lacking taxonomic information, which could be likely in a group were mimicry is prevalent (*Corydoras aeneus*, having counts of $28, 29, 30, 60, 66$ and 67, and *Scleromystax prionotos*, having as counts 34 and 43). I used chromEvol as described in section 4.3.2 to perform the ploidy assignment based on this data.



Corydoras simulatus (31)
Corydoras undulatus (25, 26)
Corydoras bilineatus
Corydoras elegans (25)
Scleromystax macropterus (33)
Scleromystax prionotos (34, 43)
Scleromystax barbatus (32, 33)
Aspidoras poecilus (23)
Aspidoras fuscoguttatus (22)
Aspidoras taurus (23)
Corydoras paleatus (22)
Corydoras ehrhardti (22)
Corydoras nattereri (20, 21, 22)
Corydoras metae (46)
Corydoras axelrodi (23)
Corydoras sipalwini
Corydoras panda (23)
Corydoras bondi (23)
Corydoras araguaiaensis (47)
Corydoras julii (46)
Corydoras arcuatus (23)
Corydoras schwartzi (23)
Corydoras trilineatus (23)
Corydoras zygatus (28)
Corydoras rabauti (28, 29)
Corydoras aeneus (28, 29, 30, 60, 66, 67)
Corydoras difluviatilis (39)
Corydoras ephippifer
Corydoras melanistius (23, 24)
Corydoras pulcher (51)
Corydoras agassizii (49)
Corydoras robineae (42)
Corydoras delphax (42)
Corydoras sodalis (37)
Corydoras reticulatus (37)
Corydoras britskii (45)
Corydoras flaveolus (29)
Chromosome number reconstruction for the ancestors
21
22
23
24
25
27
28
32
33
37
46
50
67
Ploidy level for the tips
Diploid
Polyploid

Figure B.2: (Previous page) Results for Callichthyidae for the chromosomal reconstruction, representing the most likely state of the ancestor with the tree colour, the calling of the species, and the number of chromosomes at the gametophytic state for the species, in parenthesis, as were recovered from our database of counts. Coloured blue or violet tips are species for which the confidence measure reached set threshold for significance and could be called either diploid or polyploid. Tips in gray are species for which chromosome counts were available but could not be confidently called according to chromEvol as either diploid or polyploid. The tree colour represent the most likely chromosome counts of the ancestors, but as significance was not reached for their children species, some tips can be left unknown although their ancestor was deemed has having more than twice the basal chromosome number.

Out of the 34 species that had chromosome counts, only 27 were categorized as diploid or polyploid: some because they had diverging counts (*Corydoras aenus*), others because they had intermediate chromosome counts (*Corydoras rabauti* and *Corydoras flaveolus*) and the rest for no obvious reason although they had chromosome compatible with a diploid (*Corydoras paleatus, Aspidoras taurus* and *Aspidoras fuscoguttatus*) or tetraploid (*Corydoras delphax*) status. Three species were deemed diploid although they had no associated counts (*Corydoras bilineatus, Corydoras sipalwini* and *Corydoras ephippifer*). Species that had more than 30 chromosome in their gametes were classified as polyploid (*Scleromystax prionotos, Corydoras simulatus* and *Scleromystax barbatus* whereas species with less as diploid (*Corydoras zygatus*). Count approaching 30 chromosome seem to indicate triploid organisms, considering that the basal chromosome number of the group is around 20, rather than tetraploids, but chromEvol only classifies species as diploids or polyploids as ouput.

Moreover, the ancestor of one clade (situated at the top of the tree on figure B.2, corresponding to lineage 1 in Alexandrou et al. (2011)) was deemed as having the maximum number of chromosome allowed during the simulation (67), whereas no extant children of this ancestor had chromosome data whatsoever, prompting possible inconsistency in the results of chromEvol. This seemingly wrong interferences coupled with the very low coverage of data prevented us to

confidently run a BiSSE analysis. Moreover, no old tetraploid clade was identified so that the time-dependent analysis presented in chapter 3 could be performed.

References

Abrouk, M., Murat, F., Pont, C., Messing, J., Jackson, S., Faraut, T., Tannier, E., Plomion, C., Cooke, R., Feuillet, C., and Salse, J. (2010). Palaeogenomics of plants: synteny-based modelling of extinct ancestors. *Trends in plant science*, 15(9):479–87. 6

Alexander, H. K., Lambert, A., and Stadler, T. (2015). Quantifying Age-dependent Extinction from Species Phylogenies. *Systematic biology*. 40

Alexandrou, M. a., Oliveira, C., Maillard, M., McGill, R. a. R., Newton, J., Creer, S., and Taylor, M. I. (2011). Competition and phylogeny determine community structure in Müllerian co-mimics. *Nature*, 469(7328):84–8. 97, 99

Alexandrou, M. a., Swartz, B. a., Matzke, N. J., and Oakley, T. H. (2013). Genome duplication and multiple evolutionary origins of complex migratory behavior in Salmonidae. *Molecular phylogenetics and evolution*, 69(3):514–23. 42, 55

Alfaro, M. E., Santini, F., Brock, C., Alamillo, H., Dornburg, A., Rabosky, D. L., Carnevale, G., and Harmon, L. J. (2009). Nine exceptional radiations plus high turnover explain species diversity in jawed vertebrates. *Proceedings of the National Academy of Sciences of the United States of America*, 106(32):13410–13414. 20, 24

Amborella Genome Project (2013). The Amborella genome and the evolution of flowering plants. *Science (New York, N.Y.)*, 342(6165):1241089. 8

Amores, A., Force, A. G., Yan, Y.-L., Joly, L., Amemiya, C. T., Fritz, A., Ho, R. K., Langeland, J., Prince, V., Wang, Y.-L., Westerfield, M., Ekker, M., and Postlethwait, J. H. (1998). Zebrafish hox clusters and vertebrate genome evolution. *Science*, 282(5394):1711–1714. 12

The Arabidopsis Genome Initiative (2000). Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. *Nature*, 408(6814):796–815. 6

Arai, R. (2011). *Fish Karyotypes: A Check List*. Springer. 95

Arnqvist, G., Sayadi, A., Immonen, E., Hotzy, C., Rankin, D., Tuda, M., Hjelmen, C. E., and Johnston, J. S. (2015). Genome size correlates with reproductive fitness in seed beetles. *Proceedings. Biological sciences / The Royal Society*, 282:11–14. 78

Ashman, T.-L., Kwok, a., and Husband, B. C. (2013). Revisiting the dioecy-polyploidy association: alternate pathways and research opportunities. *Cytogenetic and genome research*, 140(2-4):241–55. 81, 82

Baack, E. J. (2005). Ecological factors influencing tetraploid establishment in snow buttercups (Ranunculus adoneus, Ranunculaceae): Minority cytotype exclusion and barriers to triploid formation. *American Journal of Botany*, 92(11):1827–1835. 73

Baack, E. J., Melo, M. C., Rieseberg, L. H., and Ortiz-Barrientos, D. (2015). The origins of reproductive isolation in plants. *New Phytologist*, 207(4):968–984. 72

Barker, F. K., Burns, K. J., Klicka, J., Lanyon, S. M., and Lovette, I. J. (2013). Going to extremes: contrasting rates of diversification in a recent radiation of new world passerine birds. *Systematic biology*, 62(2):298–320. 21, 37

Barker, M. S., Arrigo, N., Baniaga, A. E., Li, Z., and Levin, D. A. (2015). On the relative abundance of autopolyploids and allopolyploids. *New Phytologist*. 73

Barker, M. S., Vogel, H., and Schranz, M. E. (2009). Paleopolyploidy in the Brassicales: analyses of the Cleome transcriptome elucidate the history of genome duplications in Arabidopsis and other Brassicales. *Genome biology and evolution*, 1:391–9. 6, 42

Beaulieu, J. M. and O'Meara, B. C. (2015). Extinction can be estimated from moderately sized molecular phylogenies. *Evolution; international journal of organic evolution*, 69(4):1036–43. 16

te Beest, M., Le Roux, J. J., Richardson, D. M., Brysting, A. K., Suda, J., Kubesová, M., and Pysek, P. (2012). The more the better? The role of polyploidy in facilitating plant invasions. *Annals of botany*, 109(1):19–45. 13

Benson, D. a., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., and Wheeler, D. L. (2005). GenBank. *Nucleic acids research*, 33(Database issue):D34—-8. 63

Benton, M. J. (1995). Diversification and extinction in the history of life. *Science*, 268(5207):52–58. 21, 40

Berthelot, C., Brunet, F., Chalopin, D., Juanchich, A., Bernard, M., Noël, B., Bento, P., Da Silva, C., Labadie, K., Alberti, A., Aury, J.-M., Louis, A., Dehais, P., Bardou, P., Montfort, J., Klopp, C., Cabau, C., Gaspin, C., Thorgaard, G. H., Boussaha, M., Quillet,

E., Guyomard, R., Galiana, D., Bobe, J., Volff, J.-N., Genêt, C., Wincker, P., Jaillon, O., Crollius, H. R., and Guiguen, Y. (2014). The rainbow trout genome provides novel insights into evolution after whole-genome duplication in vertebrates. *Nature Communications*, 5. 42, 43, 45, 55

Bertrand, Y. J. K., a. C. Scheen, Marcussen, T., Pfeil, B. E., de Sousa, F., and Oxelman, B. (2015). Assignment of Homoeologs to Parental Genomes in Allopolyploids for Species Tree Inference, with an Example from Fumaria (Papaveraceae). *Systematic Biology*, 64(3):448–471. 80

Betancur-R, R., Broughton, R. E., Wiley, E. O., Carpenter, K., López, J. A., Li, C., Holcroft, N. I., Arcila, D., Sanciangco, M., Ii, J. C. C., Zhang, F., Campbell, M. A., Ballesteros, J. A., Roa-varon, A., Willis, S., Borden, W. C., Hough, D. J., and Lu, G. (2013). The Tree of Life and a New Classification of Bony Fishes. *Plos current ToL*, 0732988. 45, 56

Bininda-Emonds, O. R. P., Cardillo, M., Jones, K. E., MacPhee, R. D. E., Beck, R. M. D., Grenyer, R., Price, S. a., Vos, R. a., Gittleman, J. L., and Purvis, A. (2007). The delayed rise of present-day mammals. *Nature*, 446(7135):507–512. 24, 29, 37

Birchler, J. a. and Veitia, R. a. (2010). The gene balance hypothesis: Implications for gene regulation, quantitative traits and evolution. *New Phytologist*, 186(1):54–62. 10

Blanc, G. and Wolfe, K. H. (2004). Widespread paleopolyploidy in model plant species inferred from age distributions of duplicate genes. *The Plant Cell*, 16(July):1667–1678. 6

Bogart, J. P. (1980). *Evolutionary Implications of Polyploidy in Amphibians and Reptiles*, pages 341–378. Springer US, Boston, MA. 81

Brawand, D., Wagner, C. E., Li, Y. I., Malinsky, M., Keller, I., Fan, S., Simakov, O., Ng, A. Y., Lim, Z. W., Bezault, E., Turner-Maier, J., Johnson, J., Alcazar, R., Noh, H. J., Russell, P., Aken, B., Alföldi, J., Amemiya, C. T., Azzouzi, N., Baroiller, J.-F., Barloy-Hubler, F., Berlin, A., Bloomquist, R., Carleton, K. L., Conte, M. a., D'Cotta, H., Eshel, O., Gaffney, L., Galibert, F., Gante, H. F., Gnerre, S., Greuter, L., Guyon, R., Haddad, N. S., Haerty, W., Harris, R. M., Hofmann, H. a., Hourlier, T., Hulata, G., Jaffe, D. B., Lara, M., Lee, A. P., MacCallum, I., Mwaiko, S., Nikaido, M., Nishihara, H., Ozouf-Costaz, C., Penman, D. J., Przybylski, D., Rakotomanga, M., Renn, S. C. P., Ribeiro, F. J., Ron, M., Salzburger, W., Sanchez-Pulido, L., Santos, M. E., Searle, S., Sharpe, T., Swofford, R., Tan, F. J., Williams, L., Young, S., Yin, S., Okada, N., Kocher, T. D., Miska, E. a., Lander, E. S., Venkatesh, B., Fernald, R. D., Meyer, A., Ponting, C. P., Streelman, J. T., Lindblad-Toh, K., Seehausen, O., and Di Palma, F. (2014). The genomic substrate for adaptive radiation in African cichlid fish. *Nature*. 12

Britto, M. R. (2003). Phylogeny of the subfamily Corydoradinae Hoedeman, 1952 (Siluriformes: Callichthyidae), with a definition of its genera. *Proceedings of the Academy of Natural Sciences of Philadelphia*, 153(1):119–154. 97

Bromham, L., Hua, X., Lanfear, R., and Cowman, P. F. (2015). Exploring the Relationships between Mutation Rates, Life History, Genome Size, Environment, and Species Richness in Flowering Plants. *The American Naturalist*, pages 000–000. 17, 78

Bronzati, M., Montefeltro, F. C., and Langer, M. C. (2015). Diversification events and the effects of mass extinctions on Crocodyliformes evolutionary history. *Royal Society Open Science*, 2. 40

Buggs, R. J. A. and Pannell, J. R. (2006). Rapid displacement of a monoecious plant lineage is due to pollen swamping by a dioecious relative. *Current biology : CB*, 16(10):996–1000. 81

Buggs, R. J. A., Wendel, J. F., Doyle, J. J., Soltis, D. E., Soltis, P. S., and Coate, J. E. (2014). The legacy of diploid progenitors in allopolyploid gene expression patterns. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 369. 73

Campbell, M. A., López, J. A., Sado, T., and Miya, M. (2013). Pike and salmon as sister taxa: detailed intraclade resolution and divergence time estimation of Esociformes + Salmoniformes based on whole mitochondrial genome sequences. *Gene*, 530(1):57–65. 55

Cañestro, C., Albalat, R., Irimia, M., and Garcia-Fernàndez, J. (2013). Impact of gene gains, losses and duplication modes on the origin and diversification of vertebrates. *Seminars in cell & developmental biology*, 24:83–94. 12

Cantalapiedra, J. L., Hernandez Fernandez, M., Azanza, B., and Morales, J. (2015). Congruent phylogenetic and fossil signatures of mammalian diversification dynamics driven by tertiary abiotic. *Evolution*, pages 1–33. 40

Chalhoub, B., Denoeud, F., Liu, S., Parkin, I. a. P., Tang, H., Wang, X., Chiquet, J., Belcram, H., Tong, C., Samans, B., Correa, M., Da Silva, C., Just, J., Falentin, C., Koh, C. S., Le Clainche, I., Bernard, M., Bento, P., Noel, B., Labadie, K., Alberti, a., Charles, M., Arnaud, D., Guo, H., Daviaud, C., Alamery, S., Jabbari, K., Zhao, M., Edger, P. P., Chelaifa, H., Tack, D., Lassalle, G., Mestiri, I., Schnel, N., Le Paslier, M.-C., Fan, G., Renault, V., Bayer, P. E., Golicz, a. a., Manoli, S., Lee, T.-H., Thi, V. H. D., Chalabi, S., Hu, Q., Fan, C., Tollenaere, R., Lu, Y., Battail, C., Shen, J., Sidebottom, C. H. D., Canaguier, a., Chauveau, A., Berard, a., Deniot, G., Guan, M., Liu, Z., Sun, F., Lim, Y. P., Lyons, E., Town, C. D., Bancroft, I., Meng, J., Ma, J., Pires, J. C., King, G. J., Brunel, D., Delourme, R., Renard, M., Aury, J.-M., Adams, K. L., Batley, J., Snowdon, R. J., Tost, J., Edwards, D., Zhou, Y., Hua, W., Sharpe, a. G., Paterson, A. H., Guan, C., and Wincker, P. (2014). Early allopolyploid

evolution in the post-Neolithic Brassica napus oilseed genome. *Science*, 345(6199):950–953. 8

Chen, Z. J. (2007). Genetic and epigenetic mechanisms for gene expression and phenotypic variation in plant polyploids. *Annu Rev Plant Biol*, 58:377–406. 72

Cherry, J. M., Hong, E. L., Amundsen, C., Balakrishnan, R., Binkley, G., Chan, E. T., Christie, K. R., Costanzo, M. C., Dwight, S. S., Engel, S. R., Fisk, D. G., Hirschman, J. E., Hitz, B. C., Karra, K., Krieger, C. J., Miyasato, S. R., Nash, R. S., Park, J., Skrzypek, M. S., Simison, M., Weng, S., and Wong, E. D. (2012). Saccharomyces Genome Database: the genomics resource of budding yeast. *Nucleic acids research*, 40(Database issue):D700–5. 92

Christoffels, A., Koh, E. G. L., Chia, J.-M., Brenner, S., Aparicio, S., and Venkatesh, B. (2004). Fugu genome analysis provides evidence for a whole-genome duplication early during the evolution of ray-finned fishes. *Molecular biology and evolution*, 21(6):1146–51. 56

Collares-Pereira, M. J., Matos, I., Morgado-Santos, M., and Coelho, M. M. (2013). Natural pathways towards polyploidy in animals: the Squalius alburnoides fish complex as a model system to study genome size and genome reorganization in polyploids. *Cytogenetic and genome research*, 140(2-4):97–116. 42

Comai, L. (2005). The advantages and disadvantages of being polyploid. *Nature reviews. Genetics*, 6(11):836–46. 41

Conant, G. C., Birchler, J. A., and Pires, J. C. (2014). Dosage, duplication, and diploidization: clarifying the interplay of multiple models for duplicate gene evolution over time. *Current Opinion in Plant Biology*, 19(June):91–98. 10

Conant, G. C. and Wolfe, K. H. (2008). Turning a hobby into a job: how duplicated genes find new functions. *Nature reviews. Genetics*, 9(12):938–950. 9, 10

Condamine, F. L., Clapham, M. E., and Kergoat, G. J. (2016). Global patterns of insect diversification: towards a reconciliation of fossil and molecular evidence? *Scientific Reports*, 6. 40

Condamine, F. L., Rolland, J., and Morlon, H. (2013). Macroevolutionary perspectives to environmental change. *Ecology letters*, 16 Suppl 1:72–85. 21

Cox, M. P., Dong, T., Shen, G., Dalvi, Y., Scott, D. B., and Ganley, A. R. D. (2014). An interspecific fungal hybrid reveals cross-kingdom rules for allopolyploid gene expression patterns. *PLoS genetics*, 10(3):e1004180. 72

Crisp, M. D. and Cook, L. G. (2009). Explosive radiation or cryptic mass extinction? Interpreting signatures in molecular phylogenies. *Evolution*, 63(9):2257–65. 35

Crow, K. D. and Wagner, G. P. (2006). What is the role of genome duplication in the evolution of complexity and diversity? *Molecular biology and evolution*, 23(5):887–92. 12, 41, 56

Darriba, D., Taboada, G. L., Doallo, R., and Posada, D. (2012). j{M}odel{T}est 2: more models, new heuristics and parallel computing. *Nat. Methods*, 9(8):772. 63

Davis, J. C. and Petrov, D. A. (2004). Preferential duplication of conserved proteins in eukaryotic genomes. *PLoS biology*, 2(3):E55. 10

Davis, M. P., Midford, P. E., and Maddison, W. P. (2013). Exploring power and parameter estimation of the BiSSE method for analyzing species diversification. *BMC evolutionary biology*, 13(38). 62, 75

Dehal, P. and Boore, J. L. (2005). Two rounds of whole genome duplication in the ancestral vertebrate. *PLoS Biol*, 3(10):e314. 12

Díaz-Pérez, a. J., Sharifi-Tehrani, M., Inda, L. a., and Catalán, P. (2014). Polyphyly, gene-duplication and extensive allopolyploidy framed the evolution of the ephemeral Vulpia grasses and other fine-leaved Loliinae (Poaceae). *Molecular phylogenetics and evolution*, 79:92–105. 62

Dodsworth, S., Chase, M. W., and Leitch, A. R. (2015). Is post-polyploidization diploidization the key to the evolutionary success of angiosperms ? *Botanical Journal of the Linnean Society*. 14, 41

Donoghue, P. C. J. and Purnell, M. (2005). Genome duplication, extinction and vertebrate evolution. *Trends in ecology & evolution*, 20(6):312–9. 41

Dos Reis, M., Inoue, J., Hasegawa, M., Asher, R. J., Donoghue, P. C. J., and Yang, Z. (2012). Phylogenomic datasets provide both precision and accuracy in estimating the timescale of placental mammal phylogeny. *Proceedings. Biological sciences / The Royal Society*, 279(1742):3491–500. 37

Drummond, A. J., Suchard, M. A., Xie, D., and Rambaut, A. (2012a). Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular biology and evolution*, 29(8):1969–73. 40

Drummond, A. J., Suchard, M. A., Xie, D., and Rambaut, A. (2012b). Bayesian Phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution*, 29(8):1969–1973. 63

Eble, G. J. (1999). On the dual nature of chance in evolutionary biology and paleobiology. *Paleobiology*, 25(1):75–87. 21

Endersby, J. (2013). Mutant utopias: evening primroses and imagined futures in early twentieth-century america. *Isis*, 104(3):471–503. 3

Erwin, D. H. (2006). *Extinction: how life on earth nearly ended 250 million years ago*. Princeton University Press, Princeton. 21

Escudero, M., Martín-Bravo, S., Mayrose, I., Fernández-Mazuecos, M., Fiz-Palacios, O., Hipp, A. L., Pimentel, M., Jiménez-Mejías, P., Valcárcel, V., Vargas, P., and Luceño, M. (2014). Karyotypic changes through dysploidy persist longer over evolutionary time than polyploid changes. *PloS one*, 9(1):e85266. 71

Estep, M. C., McKain, M. R., Vela Diaz, D., Zhong, J., Hodge, J. G., Hodkinson, T. R., Layton, D. J., Malcomber, S. T., Pasquet, R., and Kellogg, E. A. (2014). Allopolyploidy, diversification, and the Miocene grassland expansion. *Proceedings of the National Academy of Sciences*. 61

Etienne, R. S., Haegeman, B., Stadler, T., Aze, T., Pearson, P. N., Purvis, A., and Phillimore, A. B. (2012). Diversity-dependence brings molecular phylogenies closer to agreement with the fossil record. *Proceedings. Biological sciences / The Royal Society*, 279(1732):1300–9. 20, 40

Faller, B., Pardi, F., and Steel, M. (2008). Distribution of phylogenetic diversity under random extinction. *Journal of theoretical biology*, 251(2):286–96. 21

Faller, B. and Steel, M. (2012). Trait-dependent extinction leads to greater expected biodiversity loss. *SIAM Journal on Discrete Mathematics*, 26(2):472–481. 21, 37

Fawcett, J. a., Maere, S., and Van de Peer, Y. (2009). Plants with double genomes might have had a better chance to survive the Cretaceous-Tertiary extinction event. *Proceedings of the National Academy of Sciences of the United States of America*, 106(14):5737–42. 14, 60

Feduccia, A. (2014). Avian extinction at the end of the Cretaceous: Assessing the magnitude and subsequent explosive radiation. *Cretaceous Research*, 50:1–15. 40

Feldman, M., Levy, A., Chalhoub, B., and Kashkush, K. (2012). Genomic plasticity in polyploid wheat. In *Polyploidy and Genome Evolution*, pages 103–135. Springer-Verlag Berlin Heidelberg. 5

Fischer, S., Brunk, B. P., Chen, F., Gao, X., Harb, O. S., Iodice, J. B., Shanmugam, D., Roos, D. S., and Stoeckert, C. J. (2011). Using OrthoMCL to assign proteins to OrthoMCL-DB groups or to cluster proteomes into new ortholog groups. *Current protocols in bioinformatics*, 35. 63

FitzJohn, R. G., Maddison, W. P., and Otto, S. P. (2009). Estimating trait-dependent speciation and extinction rates from incompletely resolved phylogenies. *Systematic biology*, 58(6):595–611. 20

Food and Agriculture Organisation (2011). FAO Statistics. 5

Force, A. G., Lynch, M., Pickett, F. B., Amores, A., Yan, Y.-l., and Postlethwait, J. (1999). Preservation of duplicate genes by complementary , degenerative mutations. *Genetics*, 151:1531–1545. 9

Freeling, M., Scanlon, M. J., and Fowler, J. F. (2015). Fractionation and subfunctionalization following genome duplications : mechanisms that drive gene content and their consequences. *Current Opinion in Genetics & Development*, 35:110–118. 11

Freeling, M. and Thomas, B. C. (2006). Gene-balanced duplications, like tetraploidy, provide predictable drive to increase morphological complexity. *Genome research*, 16:805–814. 60

Freeling, M., Woodhouse, M. R., Subramaniam, S., Turco, G., Lisch, D. R., and Schnable, J. C. (2012). Fractionation mutagenesis and similar consequences of mechanisms removing dispensable or less-expressed DNA in plants. *Current Opinion in Plant Biology*, 15(2):131–139. 11

Froese, R. and Pauly, D. (2015). FishBase. 46, 95

Gaines, T. A., Zhang, W., Wang, D., Bukun, B., Chisholm, S. T., Shaner, D. L., Nissen, S. J., Patzoldt, W. L., Tranel, P. J., Culpepper, A. S., Grey, T. L., Webster, T. M., Vencill, W. K., Sammons, R. D., Jiang, J., Preston, C., Leach, J. E., and Westra, P. (2010). Gene amplification confers glyphosate resistance in Amaranthus palmeri. *Proceedings of the National Academy of Sciences of the United States of America*, 107(3):1029–34. 9

Gallardo, M. H., Bickham, J. W., Honeycutt, R. L., Ojeda, R. A., and Köhler, N. (1999). Discovery of tetraploidy in a mammal. *Nature*, 401(6751):341. 81

Gallardo, M. H., González, C. A., and Cebrián, I. (2006). Molecular cytogenetics and allotetraploidy in the red vizcacha rat, Tympanoctomys barrerae (Rodentia, Octodontidae). *Genomics*, 88(2):214–21. 81

Givnish, T. J., Spalink, D., Ames, M., Lyon, S. P., Hunter, S. J., Zuluaga, A., Iles, W. J. D., Clements, M. A., Arroyo, M. T. K., Leebens-mack, J., Endara, L., Kriebel, R., Neubig, K. M., Whitten, W. M., Williams, N. H., and Cameron, K. M. (2015). Orchid phylogenomics and multiple drivers of their extraordinary diversification. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 282. 40

Glennon, K. L., Ritchie, M. E., and Segraves, K. A. (2014). Evidence for shared broad-scale climatic niches of diploid and polyploid plants. *Ecology Letters*, 17(5):574–582. 72, 73

Glick, L. and Mayrose, I. (2014). ChromEvol: assessing the pattern of chromosome number evolution and the inference of polyploidy along a phylogeny. *Molecular biology and evolution*, 31(7):1914–22. 43, 64

Glover, N. M., Redestig, H., and Dessimoz, C. (2016). Homoeologs: What Are They and How Do We Infer Them? *Trends in Plant Science*, xx:1–13. 80

Gould, S. J. and Eldredge, N. (1977). Punctuated Equilibria : The Tempo and Mode of Evolution Reconsidered. *Paleobiology*, 3(2):115–151. 71

Gout, J.-F., Duret, L., and Kahn, D. (2009). Differential retention of metabolic genes following whole-genome duplication. *Molecular biology and evolution*, 26(5):1067–72. 11

Gout, J.-F., Kahn, D., and Duret, L. (2010). The Relationship among Gene Expression, the Evolution of Gene Dosage, and the Rate of Protein Evolution. *PLoS Genetics*, 6(5):e1000944. 11

Grandont, L., Jenczewski, E., and Lloyd, a. (2013). Meiosis and its deviations in polyploid plants. *Cytogenetic and genome research*, 140(2-4):171–84. 14, 61

Guignard, S., Nichols, R. A., Knell, R. J., Macdonald, A., Romila, C.-a., Trimmer, M., Leitch, I. J., and Leitch, A. R. (2016). Genome size and ploidy influence angiosperm species ' biomass under nitrogen and phosphorus limitation. *New Phytologist*. 81

Guinot, G. and Cavin, L. (2015). 'Fish' (Actinopterygii and Elasmobranchii) diversification patterns through deep time. *Biological Reviews*, 41:n/a–n/a. 52

Hagen, O., Hartmann, K., Steel, M., and Stadler, T. (2015). Age-dependent speciation can explain the shape of empirical phylogenies. *Systematic biology*, 0(0):1–9. 40

Harmon, L. J. and Harrison, S. (2015). Species Diversity Is Dynamic and Unbounded at Local and Continental Scales. *The American Naturalist*, pages 000–000. 40

Hartmann, K., Wong, D., and Stadler, T. (2010). Sampling trees from evolutionary models. *Systematic biology*, 59(4):465–76. 22, 36

Havelka, M., Kašpar, V., Hulák, M., and Flajšhans, M. (2011). Sturgeon genetics and cytogenetics : a review related to ploidy levels and interspecific hybridization. *Folia Zoologica*, 60(2):93–103. 42, 43

Hegarty, M., Coate, J. E., Sherman-Broyles, S., Abbott, R., Hiscock, S., and Doyle, J. J. (2013). Lessons from natural and artificial polyploids in higher plants. *Cytogenetic and genome research*, 140(2-4):204–25. 72

Heimberg, A. M., Sempere, L. F., Moy, V. N., Donoghue, P. C. J., and Peterson, K. J. (2008). MicroRNAs and the advent of vertebrate morphological complexity. *Proceedings of the National Academy of Sciences of the United States of America*, 105(8):2946–50. 78

Hijmans, R. J., Gavrilenko, T., Stephenson, S., Bamberg, J., Salas, A., and Spooner, D. M. (2007). Geographical and environmental range expansion through polyploidy in wild potatoes (Solanum section Petota). *Global Ecology and Biogeography*, 16(4):485–495. 72

Hinchliff, C. E., Smith, S. A., Allman, J. F., Burleigh, J. G., Chaudhary, R., Coghill, L. M., Crandall, K. A., Deng, J., Drew, B. T., Gazis, R., Gude, K., Hibbett, D. S., Katz, L. A., IV, H. D. L., McTavish, E. J., Midford, P. E., Owen, C. L., Reed, R. H., Rees, J. A., Soltis, D. E., Williams, T., and Cranston, K. A. (2015). Synthesis of phylogeny and taxonomy into a comprehensive tree of life. *Proceedings of the National Academy of Sciences*, 112(41). 79

Hoegg, S., Brinkmann, H., Taylor, J. S., and Meyer, A. (2004). Phylogenetic timing of the fish-specific genome duplication correlates with the diversification of teleost fish. *Journal of molecular evolution*, 59(2):190–203. 12, 42

Höhna, S. (2013). Fast simulation of reconstructed phylogenies under global time-dependent birth-death processes. *Bioinformatics*, 29(11):1367–74. 36

Hurles, M. (2004). Gene Duplication: The Genomic Trade in Spare Parts. *PLoS Biology*, 2(7):e206. 2

Hurley, I. A., Mueller, R. L., Dunn, K. A., Schmidt, E. J., Friedman, M., Ho, R. K., Prince, V. E., Yang, Z., Thomas, M. G., and Coates, M. I. (2007). A new time-scale for ray-finned fish evolution. *Proceedings of the Royal Society B: Biological Sciences*, 274(1609):489–498. 56

Husband, B. C. (2000). Constraints on polyploid evolution: a test of the minority cytotype exclusion principle. *Proceedings. Biological sciences / The Royal Society*, 267(1440):217–23. 14, 73

Husband, B. C. (2004). The role of triploid hybrids in the evolutionary dynamics of mixed-ploidy populations. *Biological Journal of the Linnean Society*, 82(4):537–546. 73

Innan, H. and Kondrashov, F. (2010). The evolution of gene duplications: classifying and distinguishing between models. *Nature Reviews Genetics*, 11(4):4. 9

The International Wheat Genome Sequencing Consortium (2014). A chromosome-based draft sequence of the hexaploid bread wheat (Triticum aestivum) genome. *Science (New York, N.Y.)*, 345(6194). 5, 11

Jetz, W., Thomas, G. H., Joy, J. B., Hartmann, K., and Mooers, a. O. (2012). The global diversity of birds in space and time. *Nature*, 491(7424):444–8. 21, 36, 37

Jiao, Y., Leebens-Mack, J., Ayyampalayam, S., Bowers, J. E., McKain, M. R., McNeal, J. R., Rolf, M., Ruzicka, D. R., Wafula, E., Wickett, N. J., Wu, X., Zhang, Y., Wang, J., Zhang, Y., Carpenter, E. J., Deyholos, M. K., Kutchan, T. M., Chanderbali, A. S., Soltis, P. S., Stevenson, D. W., McCombie, W. R., Pires, J. C., Wong, G. K.-S., Soltis, D. E., and Depamphilis, C. W. (2012). A genome triplication associated with early diversification of the core eudicots. *Genome biology*, 13(1):R3. 60

Jiao, Y., Wickett, N. J., Ayyampalayam, S., Chanderbali, A. S., Landherr, L., Ralph, P. E., Tomsho, L. P., Hu, Y., Liang, H., Soltis, P. S., Soltis, D. E., Clifton, S. W., Schlarbaum, S. E., Schuster, S. C., Ma, H., Leebens-Mack, J., and DePamphilis, C. W. (2011). Ancestral polyploidy in seed plants and angiosperms. *Nature*, 473(7345):97–100. 81

Jordan, I. K., Wolf, Y. I., and Koonin, E. V. (2004). Duplicated genes evolve slower than singletons despite the initial rate increase. *BMC evolutionary biology*, 4:22. 10

Kaessmann, H., Vinckenbosch, N., and Long, M. (2009). RNA-based gene duplication: mechanistic and evolutionary insights. *Nature Reviews Genetics*, 10(1):19–31. 2

Kagale, S., Robinson, S. J., Nixon, J., Xiao, R., Huebert, T., Condie, J., Kessler, D., Clarke, W. E., Edger, P. P., Links, M. G., Sharpe, a. G., and Parkin, I. a. P. (2014). Polyploid evolution of the brassicaceae during the cenozoic era. *The Plant Cell*, 26(July):2777–2791. 14

Katoh, K. and Toh, H. (2008). Recent developments in the MAFFT multiple sequence alignment program. *Briefings in Bioinformatics*, 9(4):286–298. 63

Kendall, D. G. (1948). On the generalized "birth-and-death" process. *Statistics*. 15

Kenny, N. J., Chan, K. W., Nong, W., Qu, Z., Maeso, I., Yip, H. Y., Chan, T. F., Kwan, H. S., Holland, P. W. H., Chu, K. H., and Hui, J. H. L. (2015). Ancestral whole-genome duplication in the marine chelicerate horseshoe crabs. *Heredity*, pages 1–10. 82

Kluyver, T. a. and Osborne, C. P. (2013). Taxonome: a software package for linking biological species data. *Ecology and evolution*, 3(5):1262–5. 95

Knight, C. A., Molinari, N. A., and Petrov, D. A. (2005). The Large Genome Constraint Hypothesis: Evolution, Ecology and Phenotype. *Annals of Botany*, 95(1):177–190. 17

Köhler, C., Mittelsten Scheid, O., and Erilova, A. (2010). The impact of the triploid block on the origin and evolution of polyploid plants. *Trends in genetics : TIG*, 26(3):142–8. 73

Kraaijeveld, K. (2010). Genome Size and Species Diversification. *Evolutionary biology*, 37(4):227–233. 17

Lambert, A. and Steel, M. (2013). Predicting the loss of phylogenetic diversity under non-stationary diversification models. *Journal of Theoretical Biology*, 337:111–24. 38

Larkin, K., Tucci, C., and Neiman, M. (2015). Effects of polyploidy and reproductive mode on life history trait expression. *Ecology and Evolution*, page 48. 81

Laurent, S., Robinson-Rechavi, M., and Salamin, N. (2015). Detecting patterns of species diversification in the presence of both rate shifts and mass extinctions. *BMC evolutionary biology*, 15:157. 77

Le Comber, S. C., Ainouche, M. L., Kovarik, A., and Leitch, A. R. (2010). Making a functional diploid : from polysomic to disomic inheritance. *New Phytologist*. 84

Leggatt, R. A. and Iwama, G. K. (2003). Occurrence of polyploidy in the fishes. *Review in Fish Biology and Fisheries*, 13:237–246. 97

Leitch, A. R. and Leitch, I. J. (2008). Genomic Plasticity and the Diversity of Polyploid Plants. *Science*, 320(4):702–711. 14

Leitch, I. J., Bennett, M. D., Gardens, R. B., and Tw, S. (2004). Genome downsizing in polyploid plants. *Biological Journal of the Linnean Society*, 82:651–663. 6

Levin, D. A. (1975). Minority Cytotype Exclusion in Local Plant. *Taxon*, 24(1):35–43. 61, 73

Levin, D. A. (1982). Polyploidy and novelty in flowering plants. *The American Naturalist*, 122(1):1–25. 14, 60

Li, Z., Baniaga, A. E., Sessa, E. B., Scascitelli, M., Graham, S. W., Rieseberg, L. H., and Barker, M. S. (2015). Early genome duplications in conifers and other seed plants. *Science Advences*, 1(November). 42, 82

Liu, B. (2003). Epigenetic phenomena and the evolution of plant allopolyploids. *Molecular Phylogenetics and Evolution*, 29(3):365–379. 72, 73

Lynch, M. (2007). Genomic expansion by gene duplication. In *The origins of genome architecture*, pages 193–237. Sinauer Associates. 2

Lynch, M. (2011). The lower bound to the evolution of mutation rates. *Genome Biology and Evolution*, 3(1):1107–1118. 83

Lynch, M. and Conery, J. S. (2000). The evolutionary fate and consequences of duplicate genes. *Science*, 290(5494):1151–1155. 10

Lynch, M., Force, A. G., Naturalist, T. A., and Dec, N. (2000). The origin of interspecific genomic incompatibility via gene duplication. *The American Naturalist*, 156(6):590–605. 12

Lysak, M. A., Cheung, K., Kitschke, M., and Bures, P. (2007). Ancestral chromosomal blocks are triplicated in Brassiceae species with varying chromosome number and genome size. *Plant physiology*, 145(2):402–10. 8

Mable, B. K. (2003). Breaking down taxonomic barriers in polyploidy research. *Trends in plant science*, 8(12):582–590. 82

Mable, B. K. (2004). 'Why polyploidy is rarer in animals than in plants': myths and mechanisms. *Biological Journal of the Linnean Society*, 82(4):453–466. 81

Mable, B. K., Alexandrou, M. a., and Taylor, M. I. (2011). Genome duplication in amphibians and fish: an extended synthesis. *Journal of Zoology*, 284(3):151–182. 42, 81

Maclean, C. J. and Greig, D. (2011). Reciprocal gene loss following experimental whole-genome duplication causes reproductive isolation in yeast. *Evolution; international journal of organic evolution*, 65(4):932–45. 41

Macqueen, D. J. and Johnston, I. A. (2014). A well-constrained estimate for the timing of the salmonid whole genome duplication reveals major decoupling from species diversification. *Proceedings. Biological sciences / The Royal Society*, 281. 42, 43, 55

Maddison, W. P., Midford, P. E., and Otto, S. P. (2007). Estimating a binary character's effect on speciation and extinction. *Systematic biology*, 56(5):701–10. 20, 44, 56, 62, 65

Madlung, A. (2013). Polyploidy and its effect on evolutionary success: old questions revisited with new tools. *Heredity*, 110(2):99–104. 13

Magnuson-Ford, K. and Otto, S. P. (2012). Linking the investigations of character evolution and species diversification. *The American naturalist*, 180(2):225–45. 65, 66, 74

Maia, R., Rubenstein, D. R., and Shawkey, M. D. (2013). Key ornamental innovations facilitate diversification in an avian radiation. *Proceedings of the National Academy of Sciences of the United States of America*, 110(26):10687–92. 40, 62

Majure, L. C., Puente, R., Griffith, M. P., Judd, W. S., Soltis, P. S., and Soltis, D. E. (2012). Phylogeny of Opuntia s.s. (Cactaceae): clade delineation, geographic origins, and reticulate evolution. *American journal of botany*, 99(5):847–64. 62

Marcet-Houben, M. and Gabaldón, T. (2015). Beyond the Whole-Genome Duplication: Phylogenetic Evidence for an Ancient Interspecies Hybridization in the Baker's Yeast Lineage. *PLoS biology*, 13(8):e1002220. 79

Mason, A. S. and Pires, J. C. (2015). Unreduced gametes: meiotic mishap or evolutionary mechanism? *Trends in genetics : TIG*, 31(1):5–10. 4, 83, 84

Mason-Gamer, R. J. (2013). Phylogeny of a genomically diverse group of elymus (poaceae) allopolyploids reveals multiple levels of reticulation. *PloS one*, 8(11):e78449. 62

May, M. R., Höhna, S., and Moore, B. R. (2015). A Bayesian Approach for Detecting Mass-Extinction Events When Rates of Lineage Diversification Vary. *bioRxiv*, pages 1–47. 77

May, M. R. and Moore, B. R. (2016). How Well Can We Detect Shifts in Rates of Lineage Diversification? A Simulation Study of Sequential AIC Methods. *Systematic biology*. 18, 32, 42

Mayr, E. (1992). Speciational Evolution or Punctuated Equilibria. In Peterson, A. S. and Steven, editors, *The dynamics of evolution*, pages 21–48. New York: Cornell University Press. 84

Mayrose, I., Zhan, S. H., Rothfels, C. J., Arrigo, N., Barker, M. S., Rieseberg, L. H., and Otto, S. P. (2014). Methods for studying polyploid diversification and the dead end hypothesis : a reply to Soltis et al . *New Phytologist*. 17, 62

Mayrose, I., Zhan, S. H., Rothfels, C. J., Magnuson-Ford, K., Barker, M. S., Rieseberg, L. H., and Otto, S. P. (2011). Recently formed polyploid plants diversify at lower rates. *Science*, 333(6047):1257. 17, 20, 62, 64, 67

McGrath, C. L., Gout, J.-F., Doak, T. G., Yanagi, A., and Lynch, M. (2014a). Insights into three whole-genome duplications gleaned from the Paramecium caudatum genome sequence. *Genetics*, 197(4):1417–28. 82

McGrath, C. L., Gout, J.-F., Johri, P., Doak, T. G., and Lynch, M. (2014b). Differential retention and divergent resolution of duplicate genes following whole-genome duplication. *Genome research*, 24:1665–1675. 10, 82

Meredith, R. W., Janečka, J. E., Gatesy, J., Ryder, O. a., Fisher, C. a., Teeling, E. C., Goodbla, A., Eizirik, E., Simão, T. L. L., Stadler, T., Rabosky, D. L., Honeycutt, R. L., Flynn, J. J., Ingram, C. M., Steiner, C., Williams, T. L., Robinson, T. J., Burk-Herrick, A., Westerman, M., Ayoub, N. a., Springer, M. S., and Murphy, W. J. (2011). Impacts of the Cretaceous Terrestrial Revolution and KPg extinction on mammal diversification. *Science*, 334(6055):521–4. 29, 36, 37

Merhej, J., Delaveau, T., Guitard, J., Palancade, B., Hennequin, C., Garcia, M., Lelandais, G., and Devaux, F. (2015). Yap7 is a transcriptional repressor of nitric oxide oxidase in yeasts, which arose from neofunctionalization after whole genome duplication. *Molecular Microbiology*, 96(March). 8

Moen, D. and Morlon, H. (2014). Why does diversification slow down? *Trends in Ecology & Evolution*, pages 1–8. 40

Moriyama, Y., Ito, F., Takeda, H., Yano, T., Okabe, M., Kuraku, S., Keeley, F. W., and Koshiba-Takeuchi, K. (2016). Evolution of the fish heart by sub/neofunctionalization of an elastin gene. *Nature Communications*, 7:10397. 57

Morlon, H. (2014). Phylogenetic approaches for studying diversification. *Ecology Letters*. 16, 40, 62

Morlon, H., Parsons, T. L., and Plotkin, J. B. (2011). Reconciling molecular phylogenies with the fossil record. *Proceedings of the National Academy of Sciences*, 108(39):16327–32. 37, 46

Morlon, H., Potts, M. D., and Plotkin, J. B. (2010). Inferring the dynamics of diversification: a coalescent approach. *PLoS biology*, 8(9). 36

Moyle, L. C., Olson, M. S., and Tiffin, P. (2004). Patterns of reproductive isolation in three angiosperm genera. *Evolution*, 58(6):1195–1208. 73

Muir, C. D. and Hahn, M. W. (2015). The limited contribution of reciprocal gene loss to increased speciation rates following whole-genome duplication. *The American Naturalist*, 185(1). 13, 41

Muller, H. J. (1925). Why polyploidy is rarer in animals than in plants. *The American Naturalist*, 59(663):346–353. 81

Müntzing, A. (1927). The evolutionary significance of autopolyploidy. *Hereditas*, 21. 4, 6

Murat, F., Louis, A., Maumus, F., Armero, A., Cooke, R., Quesneville, H., Crollius, H. R., and Salse, J. (2015). Understanding Brassicaceae evolution through ancestral genome reconstruction. *Genome Biology*, 16(1):262. 8

Near, T. J., Dornburg, A., Eytan, R. I., Keck, B. P., Smith, W. L., Kuhn, K. L., Moore, J. A., Price, S. A., Burbrink, F. T., Friedman, M., and Wainwright, P. C. (2013). Phylogeny and tempo of diversification in the superradiation of spiny-rayed fishes. *Proceedings of the National Academy of Sciences*, 110(31):12738–43. 36

Near, T. J., Eytan, R. I., Dornburg, A., Kuhn, K. L., Moore, J. a., Davis, M. P., Wainwright, P. C., Friedman, M., and Smith, W. L. (2012). Resolution of ray-finned fish phylogeny and timing of diversification. *Proceedings of the National Academy of Sciences of the United States of America*, 109(34):13698–703. 45, 52, 55, 56

Nee, S. (1997). Extinction and the Loss of Evolutionary History. *Science*, 278(5338):692–694. 21

Nee, S., Holmes, E. C., May, R., and Harvey, P. (1994a). Extinction rates can be estimated from molecular phylogenies. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 344(1307):77–82. 16, 20

Nee, S., May, R., and Harvey, P. (1994b). The reconstructed evolutionary process. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences.* 16, 61

Nei, M. (2005). Selectionism and neutralism in molecular evolution. *Molecular biology and evolution*, 22(12):2318–42. 1

Nei, M. and Nozawa, M. (2011). Roles of Mutation and Selection in Speciation: From Hugo de Vries to the Modern Genomic Era. *Genome Biology and Evolution*, 3(0):812–829. 3

Neiman, M., Kay, a. D., and Krist, a. C. (2013). Can resource costs of polyploidy provide an advantage to sex? *Heredity*, 110(2):152–9. 81

Ng, M. and Yanofsky, M. F. (2001). Function and evolution of the plant MADS-box gene family. *Nature reviews. Genetics*, 2(March):186–195. 9

Ogino, Y., Kuraku, S., Ishibashi, H., Miyakawa, H., Sumiya, E., Miyagawa, S., Matsubara, H., Yamada, G., Baker, M. E., and Iguchi, T. (2016). Neofunctionalization of Androgen Receptor by Gain-of-Function Mutations in Teleost Fish Lineage. *Molecular Biology and Evolution*, 33(1):228–244. 8

Ohno, S. (1970). *Evolution by gene duplication.* Springer Science & Business Media. 6

O'Leary, M. a., Bloch, J. I., Flynn, J. J., Gaudin, T. J., Giallombardo, A., Giannini, N. P., Goldberg, S. L., Kraatz, B. P., Luo, Z.-X., Meng, J., Ni, X., Novacek, M. J., Perini, F. a., Randall, Z. S., Rougier, G. W., Sargis, E. J., Silcox, M. T., Simmons, N. B., Spaulding, M., Velazco, P. M., Weksler, M., Wible, J. R., and Cirranello, A. L. (2013). The placental mammal ancestor and the post-K-Pg radiation of placentals. *Science*, 339(6120):662–7. 37, 78

Oliver, K. R. and Greene, W. K. (2009). Transposable elements: powerful facilitators of evolution. *BioEssays : news and reviews in molecular, cellular and developmental biology*, 31(7):703–14. 83

O'Meara, B. C. (2011). Evolutionary Inferences from Phylogenies: A Review of Methods. *Annual Review of Ecology, Evolution, and Systematics*, 43(1):120913143848009. 40

Orr, H. A. (1990). "Why polyploidy is rarer in animals than in plants" revisited. *The American Naturalist*, 136(6):759–770. 81

Orr, H. A. (1996). Dobzhanksy, Bateson, and the genetics of speciation. *Genetics*, 144:1331–1335. 13

Orr-Weaver, T. L. (2015). When bigger is better: the role of polyploidy in organogenesis. *Trends in genetics : TIG*, 31(6):307–315. 82

Oswald, B. P. and Nuismer, S. L. (2011). A unified model of autopolyploid establishment and evolution. *The American naturalist*, 178(6):687–700. 61

Otto, S. P. (2007). The evolutionary consequences of polyploidy. *Cell*, 131(3):452–62. 73

Otto, S. P. and Whitton, J. (2000). Polyploid incidence and evolution. *Annual review of genetics*, 34:401–437. 42, 81

Pandit, M. K., White, S. M., and Pocock, M. J. O. (2014). The contrasting effects of genome size , chromosome number and ploidy level on plant invasiveness : a global analysis. *New Phytologist*, 203:697–703. 17

Papp, B., Pál, C., and Hurst, L. D. (2003). Dosage sensitivity and the evolution of gene families in yeast. *Nature*, 424(6945):194–197. 10

Paradis, E. (2004). Can extinction rates be estimated without fossils? *Journal of theoretical biology*, 229(1):19–30. 16

Paradis, E. (2013). Molecular dating of phylogenies by likelihood methods: a comparison of models and a new information criterion. *Molecular phylogenetics and evolution*, 67(2):436–44. 40

Parisod, C. (2012). Polyploids integrate genomic changes and ecological shifts. 13th Congress of the European Society for Evolutionary Biology, Tuebingen, Germany, August 2011. *The New phytologist*, 193(2):297–300. 72

Parisod, C., Holderegger, R., and Brochmann, C. (2010). Evolutionary consequences of autopolyploidy. *The New phytologist*, 186(1):5–17. 14

Pennell, M. W., Harmon, L. J., and Uyeda, J. C. (2014). Is there room for punctuated equilibrium in macroevolution? *Trends in ecology & evolution*, 29(1):23–32. 71

Petit, C., Bretagnolle, F., and Felber, F. (1999). Evolutionary consequences of diploid–polyploid hybrid zones in wild species. *Trends in ecology & evolution*, 14(8):306–311. 73

Petit, M., Guidat, C., Daniel, J., Denis, E., Montoriol, E., Bui, Q. T., Lim, K. Y., Kovarik, A., Leitch, A. R., Grandbastien, M.-A., and Mhiri, C. (2010). Mobilization of retrotransposons in synthetic allotetraploid tobacco. *New Phytologist*, 186(1):135–147. 72

Pincheira-Donoso, D., Harvey, L. P., and Ruta, M. (2015). What defines an adaptive radiation? Macroevolutionary diversification dynamics of an exceptionally species-rich continental lizard radiation. *BMC Evolutionary Biology*, 15(1):153. 12

The Plant List (2013). The Plant List Version 1.1. 67

Puttick, M. N., Clark, J., and Donoghue, P. C. J. (2015). Size is not everything : rates of genome size evolution , not C -value , correlate with speciation in angiosperms. *Proceedings. Biological sciences / The Royal Society*, 282. 17

R Core Team (2013). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. 22

Rabosky, D. L. (2010). Extinction rates should not be estimated from molecular phylogenies. *Evolution; international journal of organic evolution*, 64(6):1816–24. 16

Rabosky, D. L. (2014). Automatic detection of key innovations, rate shifts, and diversity-dependence on phylogenetic trees. *PLoS ONE*, 9(2):e89543. 24, 25, 36

Rabosky, D. L., Donnellan, S. C., Talaba, A. L., and Lovette, I. J. (2007). Exceptional among-lineage variation in diversification rates during the radiation of Australia's most diverse vertebrate clade. *Proceedings. Biological sciences / The Royal Society*, 274(1628):2915–23. 20, 37

Rabosky, D. L. and Goldberg, E. E. (2015). Model inadequacy and mistaken inferences of trait-dependent speciation. *Systematic biology*, 64(2):340–355. 74

Rabosky, D. L. and Hurlbert, A. H. (2015). Species Richness at Continental Scales Is Dominated by Ecological Limits. *The American Naturalist*, pages 000–000. 40

Ramsey, J. and Ramsey, T. S. (2014). Ecological studies of polyploidy in the 100 years following its discovery. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 369:15–19. 4

Ramsey, J. and Schemske, D. W. (1998). Pathways, Mechanisms, and Rates of Polyploid Formation in Flowering Plants. *Annual Review of Ecology and Systematics*, 29(1):467–501. 12, 61

Raup, D. M. (1985). Mathematical models of cladogenesis. *Paleobiology*, 11(1):42–52. 40

Raup, D. M. (1992). *Extinction: bad genes or bad luck ?* W. W. Norton & Company, New York. 21

Raup, D. M. and Sepkoski, J. J. (1982). Mass extinction in the marine fossil record. *Science*, 215(4539):1501–1503. 21, 22

Renny-Byfield, S., Gong, L., Gallagher, J. P., and Wendel, J. F. (2015). Persistence of Subgenomes in Paleopolyploid Cotton after 60 My of Evolution. *Molecular biology and evolution*, pages 1–9. 6

Renny-Byfield, S., Kovarik, A., Kelly, L. J., Macas, J., Novak, P., Chase, M. W., Nichols, R. a., Pancholi, M. R., Grandbastien, M.-A., and Leitch, A. R. (2013). Diploidization and genome size change in allopolyploids is associated with differential dynamics of low- and high-copy sequences. *The Plant journal : for cell and molecular biology*, 74(5):829–39. 73

Rice, A., Glick, L., Abadi, S., Einhorn, M., Kopelman, N. M., Salman-Minkov, A., Mayzel, J., Chay, O., and Mayrose, I. (2014). The Chromosome Counts Database (CCDB) - a community resource of plant chromosome numbers. *The New phytologist*. 60, 64

Ricklefs, R. E. (2007). Estimating diversification rates from phylogenetic information. *Trends in ecology & evolution*, 22(11):601–10. 20

Rieseberg, L. H. and Willis, J. H. (2007). Plant speciation. *Science (New York, N.Y.)*, 317(5840):910–4. 60

Rolland, J., Condamine, F. L., Jiguet, F., and Morlon, H. (2014a). Faster speciation and reduced extinction in the tropics contribute to the mammalian latitudinal diversity gradient. *PLoS Biology*, 12(1):e1001775. 62

Rolland, J., Jiguet, F., Jønsson, K. A., Condamine, F. L., and Morlon, H. (2014b). Settling down of seasonal migrants promotes bird diversification. *Proceedings. Biological sciences / The Royal Society*, 281. 62

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D. L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M. A., and Huelsenbeck, J. P. (2012). MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic biology*, 61(3):539–42. 40

Roux, C. and Pannell, J. R. (2015). Inferring the mode of origin of polyploid species from next-generation sequence data. *Molecular ecology*, pages 1047–1059. 80

Rundell, R. J. and Price, T. D. (2009). Adaptive radiation, nonadaptive radiation, ecological speciation and nonecological speciation. *Trends in ecology & evolution*, 24(7):394–9. 12

Sanderson, M. J., Donoghue, M. J., Piel, W., and Eriksson, T. (1994). TreeBASE: a prototype database of phylogenetic analyses and an interactive tool for browsing the phylogeny of life. *American journal of botany*, 81(6):183. 79

Santini, F., Harmon, L. J., Carnevale, G., and Alfaro, M. E. (2009). Did genome duplication drive the origin of teleosts? A comparative study of diversification in ray-finned fishes. *BMC evolutionary biology*, 9:194. 12, 42

Scannell, D. R., Byrne, K. P., Gordon, J. L., Wong, S., and Wolfe, K. H. (2006). Multiple rounds of speciation associated with reciprocal gene loss in polyploid yeasts. *Nature*, 440(7082):341–5. 13, 41, 82

Scarpino, S. V., Levin, D. a., and Meyers, L. A. (2014). Polyploid formation shapes flowering plant diversity. *The American naturalist*, 184(4):456–65. 17, 61

Schranz, M. E., Mohammadin, S., and Edger, P. P. (2012). Ancient whole genome duplications, novelty and diversification: the WGD Radiation Lag-Time Model. *Current opinion in plant biology*, 15(2):147–53. 14, 41, 54, 83

Scott, A. D., Stenz, N., and Baum, D. A. (2015a). Whole genome duplication in coast redwood (Sequoia sempervirens) and its implications for explaining the rarity of polyploidy in conifers. *bioRxiv*. 15, 42

Scott, A. D., Stenz, N., and Baum, D. A. (2015b). Whole genome duplication in coast redwood (Sequoia sempervirens) and its implications for explaining the rarity of polyploidy in conifers. *New Phytologist*. 82

Selmecki, A. M., Maruvka, Y. E., Richmond, P. a., Guillet, M., Shoresh, N., Sorenson, A. L., De, S., Kishony, R., Michor, F., Dowell, R., and Pellman, D. (2015). Polyploidy can drive rapid adaptation in yeast. *Nature*. viii, 14, 87, 88, 89, 90, 92

Sémon, M. and Wolfe, K. H. (2006). Reciprocal gene loss between Tetraodon and zebrafish after whole genome duplication in their ancestor. *Trends in genetics : TIG*, 23(3):22. 13

Senerchia, N., Felber, F., and Parisod, C. (2014). Contrasting evolutionary trajectories of multiple retrotransposons following independent allopolyploidy in wild wheats. *New Phytologist*. 72

Shi, F.-X., Li, M.-R., Li, Y.-L., Jiang, P., Zhang, C., Pan, Y.-Z., Liu, B., Xiao, H.-X., and Li, L.-F. (2015). The impacts of polyploidy, geographic and ecological isolations on the diversification of Panax (Araliaceae). *BMC Plant Biology*, 15(1):297. 42

Silvestro, D., Schnitzler, J., Liow, L. H., Antonelli, A., and Salamin, N. (2014). Bayesian Estimation of Speciation and Extinction from Incomplete Fossil Occurrence Data. *Systematic biology*, 0(0):1–19. 40, 46, 66

Silvestro, D., Schnitzler, J., and Zizka, G. (2011). A Bayesian framework to estimate diversification rates and their variation through time and space. *BMC evolutionary biology*, 11(1):311. 20

Silvestro, D., Zizka, G., and Schulte, K. (2013). Disentangling the effects of key innovations on the diversification of Bromelioideae (bromeliaceae). *Evolution; international journal of organic evolution*, 68(1):163–75. 65

Sirr, A., Cromie, G. a., Jeffery, E. W., Gilbert, T. L., Ludlow, C. L., Scott, A. C., and Dudley, A. M. (2015). Allelic variation, aneuploidy, and nongenetic mechanisms suppress a monogenic trait in yeast. *Genetics*, 199(1):247–62. 89, 92

Slater, G. J. (2015). Iterative adaptive radiations of fossil canids show no evidence for diversity-dependent trait evolution. *Proceedings of the National Academy of Sciences*, 2014:201403666. 40

Slechtová, V., Bohlen, J., Freyhof, J., and Ráb, P. (2006). Molecular phylogeny of the Southeast Asian freshwater fish family Botiidae (Teleostei: Cobitoidea) and the origin of polyploidy in their evolution. *Molecular phylogenetics and evolution*, 39(2):529–41. 42, 43

Slotte, T., Huang, H., Lascoux, M., and Ceplitis, A. (2008). Polyploid speciation did not confer instant reproductive isolation in Capsella (Brassicaceae). *Molecular Biology and Evolution*, 25(7):1472–1481. 73

Smith, D. M. and Marcot, J. D. (2015). The fossil record and macroevolutionary history of the beetles. *Proceedings. Biological sciences / The Royal Society*, 282. 40

Smith, J. J. and Keinath, M. C. (2015). The sea lamprey meiotic map improves resolution of ancient vertebrate genome duplications. *Genome research*, 25:1081–1090. 12

Soltis, D. E., Albert, V. a., Leebens-Mack, J., Bell, C. D., Paterson, A. H., Zheng, C., Sankoff, D., Depamphilis, C. W., Wall, P. K., and Soltis, P. S. (2009). Polyploidy and angiosperm diversification. *American journal of botany*, 96(1):336–48. 8, 12, 17, 60, 62

Soltis, D. E., Buggs, R. J. A., Doyle, J. J., Soltis, P. S., Soltis, D. E., Buggs, R. J. A., Doyle, J. J., and Soltis, P. S. (2010). What we still don't know about polyploidy. *Taxon*, 59(5):1387–1403. 72

Soltis, D. E., Segovia-Salcedo, M. C., Jordon-Thaden, I., Majure, L. C., Miles, N. M., Mavrodiev, E. V., Mei, W., Cortez, M. B., Soltis, P. S., and Gitzendanner, M. A. (2014a). Are polyploids really evolutionary dead-ends (again)? A critical reappraisal of Mayrose et al . (2011). *New Phytologist*. 17, 62, 63

Soltis, D. E., Visger, C. J., and Soltis, P. S. (2014b). The polyploidy revolution then...and now: Stebbins revisited. *American journal of botany*, 101(7):1057–1078. 14

Soltis, P. S., Liu, X., Marchant, D. B., Visger, C. J., and Soltis, D. E. (2014c). Polyploidy and novelty: Gottlieb's legacy. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 369. 60

Soria, P. S., McGary, K. L., and Rokas, A. (2014). Functional divergence for every paralog. *Molecular biology and evolution*, 31(4):984–92. 9

Spriggs, E. L., Christin, P.-A., and Edwards, E. J. (2014). C4 photosynthesis promoted species diversification during the miocene grassland expansion. *PloS one*, 9(5):e97722. 62

Stadler, T. (2011a). Mammalian phylogeny reveals recent diversification rate shifts. *Proceedings of the National Academy of Sciences of the United States of America*, 108(15):6187–92. 20, 21, 24, 26, 37, 45

Stadler, T. (2011b). Simulating trees with a fixed number of extant species. *Systematic biology*, 60(5):676–84. 22

Stadler, T. and Bokma, F. (2013). Estimating speciation and extinction rates for phylogenies of higher taxa. *Systematic biology*, 62(2):220–30. 20

Stebbins, G. L. J. (1950). *Variation and evolution in plants.* Columbia University Press, New York and London. 13, 60

Stenberg, P. and Saura, a. (2013). Meiosis and its deviations in polyploid animals. *Cytogenetic and genome research*, 140(2-4):185–203. 14

Stoltzfus, A. (1999). On the possibility of constructive neutral evolution. *Journal of Molecular Evolution*, 49:169–181. 9

Suda, J. and Herben, T. (2013). Ploidy frequencies in plants with ploidy heterogeneity: fitting a general gametic model to empirical population data. *Proceedings. Biological sciences / The Royal Society*, 280(November 2012):20122387. 74

Svartman, M., Stone, G., and Stanyon, R. (2005). Molecular cytogenetics discards polyploidy in mammals. *Genomics*, 85(4):425–30. 81

Tank, D. C., Eastman, J. M., Pennell, M. W., Soltis, P. S., Soltis, D. E., Hinchliff, C. E., Brown, J. W., Sessa, E. B., and Harmon, L. J. (2015). Nested radiations and the pulse of angiosperm diversification : increased diversification rates often follow whole genome duplications. *New Phytologist*. 18, 41, 42, 62, 80

Taylor, J. S., Braasch, I., Frickey, T., Meyer, A., and Van de Peer, Y. (2003). Genome duplication, a trait shared by 22000 species of ray-finned fish. *Genome research*, 13(3):382–90. 12, 42

Taylor, J. S. and Raes, J. (2004). Duplication and divergence: the evolution of new genes and old ideas. *Annual review of genetics*, 38:615–43. 1, 8, 41

Teta, P., Pardiñas, U. F. J., Sauthier, D. E. U., and Gallardo, M. H. (2014). A new species of the tetraploid vizcacha rat Tympanoctomys (Caviomorpha, Octodontidae) from central Patagonia, Argentina. *Journal of Mammalogy*, 95(1):60–71. 81

Theodoridis, S., Randin, C., Broennimann, O., Patsiou, T., and Conti, E. (2013). Divergent and narrower climatic niches characterize polyploid species of European primroses in Primula sect. Aleuritia. *Journal of Biogeography*, 40(7):1278–1289. 72

Vanneste, K., Baele, G., Maere, S., and Van de Peer, Y. (2014). Analysis of 41 plant genomes supports a wave of successful genome duplications in association with the Cretaceous-Paleogene boundary. *Genome research*. 13, 14, 17, 60

Vanneste, K., Sterck, L., Myburg, Z., Van de Peer, Y., and Mizrachi, E. (2015). Horsetails are ancient polyploids: evidence from Equisetum giganteum. *The Plant cell*, pages 1–13. 82

Vanneste, K., Van De Peer, Y., and Maere, S. (2013). Inference of genome duplications from age distributions revisited. *Molecular Biology and Evolution*, 30(1):177–190. 6

Vaux, F., Trewick, S. A., and Morgan-Richards, M. (2015). Lineages, splits and divergence challenge whether the terms anagenesis and cladogenesis are necessary. *Biological Journal of the Linnean Society*, pages 165–176. 71

Volff, J.-N. (2005). Genome evolution and biodiversity in teleost fish. *Heredity*, 94(3):280–94. 56

de Vries, H. (1904). Species and varieties, their origin by mutation. In *Lectures Delivered at the University of California*. 3

de Vries, H. (1915). The coefficient of mutation in Oenothera biennis L. *The Botanical Gazette*, 59(3):169–196. 3

de Vries, H. and Boedijn, K. (1923). On the distribution of mutant characters among the chromosome of Oenothera lamarckiana. *Genetics*, 8(233). 3

Wagner, A. (2008). Neutralism and selectionism: a network-based reconciliation. *Nature Reviews Genetics*, 9. 1

Wagner, P. J. and Estabrook, G. F. (2014). Trait-based diversification shifts reflect differential extinction among fossil taxa. *Proceedings of the National Academy of Sciences of the United States of America*, 111(46). 40

Weber, M. G. and Agrawal, a. a. (2014). Defense mutualisms enhance plant diversification. *Proceedings of the National Academy of Sciences*, 111(46). 62

Wendel, J. F. (2015). The wondrous cycles of polyploidy in plants. *American journal of botany*, 102(11):1–4. 60

Wertheim, B., Beukeboom, L. W., and Van De Zande, L. (2013). Polyploidy in animals: Effects of gene expression on sex determination, evolution and ecology. *Cytogenetic and Genome Research*, 140(2-4):256–269. 82

Wiens, J. J. (2015). Faster diversification on land than sea helps explain global biodiversity patterns among habitats and animal phyla. *Ecology letters*. 40

Wilson, G. P. (2005). Mammalian faunal dynamics during the last 1.8 Million years of the Cretaceous in Garfield county, Montana. *Journal of Mammalian Evolution*, 12(1-2):53–76. 37

Wood, T. E., Takebayashi, N., Barker, M. S., Mayrose, I., Greenspoon, P. B., and Rieseberg, L. H. (2009). The frequency of polyploid speciation in vascular plants. *Proceedings of the National Academy of Sciences of the United States of America*, 106(33):13875–9. 60

Woodhouse, M. R., Cheng, F., Pires, J. C., Lisch, D., Freeling, M., and Wang, X. (2014). Origin, inheritance, and gene regulatory consequences of genome dominance in polyploids. *Proceedings of the National Academy of Sciences of the United States of America*, 111(14):5283–8. 11

Wright, K. M., Arnold, B. J., Xue, K., Šurinová, M., O'Connell, J., and Bomblies, K. (2014). Selection on Meiosis Genes in Diploid and Tetraploid Arabidopsis arenosa. *Molecular biology and evolution*, 32(4):944–955. 84

Xu, P., Zhang, X., Wang, X., Li, J., Liu, G., Kuang, Y., Xu, J., Zheng, X., Ren, L., Wang, G., Zhang, Y., Huo, L., Zhao, Z., Cao, D., Lu, C., Li, C., Zhou, Y., Liu, Z., Fan, Z., Shan, G., Li, X., Wu, S., Song, L., Hou, G., Jiang, Y., Jeney, Z., Yu, D., Wang, L., Shao, C., Song, L., Sun, J., Ji, P., Wang, J., Li, Q., Xu, L., Sun, F., Feng, J., Wang, C., Wang, S., Wang, B., Li, Y., Zhu, Y., Xue, W., Zhao, L., Wang, J., Gu, Y., Lv, W., Wu, K., Xiao, J., Wu, J., Zhang, Z., Yu, J., and Sun, X. (2014). Genome sequence and genetic diversity of the common carp, Cyprinus carpio. *Nature genetics*, 46(11):1212–9. 42

Yant, L., Hollister, J. D., Wright, K. M., Arnold, B. J., Higgins, J. D., Franklin, F. C. H., and Bomblies, K. (2013). Meiotic adaptation to genome duplication in Arabidopsis arenosa. *Current biology : CB*, 23(21):2151–6. 84

Yoo, M.-J., Liu, X., Pires, J. C., Soltis, P. S., and Soltis, D. E. (2014). Nonadditive gene expression in polyploids. *Annual review of genetics*, 48:485–517. 72

Yule, G. U. (1925). A mathematical theory of evolution based on the conclusions of Dr. J. C. Willis, F.R.S. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 213(402):21–87. 15

Zanne, A. E., Tank, D. C., Cornwell, W. K., Eastman, J. M., Smith, S. a., FitzJohn, R. G., McGlinn, D. J., O'Meara, B. C., Moles, A. T., Reich, P. B., Royer, D. L., Soltis, D. E., Stevens, P. F., Westoby, M., Wright, I. J., Aarssen, L., Bertin, R. I., Calaminus, A., Govaerts, R., Hemmings, F., Leishman, M. R., Oleksyn, J., Soltis, P. S., Swenson, N. G., Warman, L., and Beaulieu, J. M. (2014). Three keys to the radiation of angiosperms into freezing environments. *Nature*, 506(7486):89–92. 69

Zhan, S. H., Glick, L., Tsigenopoulos, C. S., Otto, S. P., and Mayrose, I. (2014). Comparative analysis reveals that polyploidy does not decelerate diversification in fish. *Journal of evolutionary biology*, pages 1–13. 42, 43, 44, 48, 50, 56

Zhang, J. (2003). Evolution by gene duplication: an update. *Trends in Ecology & Evolution*, 18(6):292–298. 2

Zhao, Z.-X., Cao, D.-C., Xu, J., Xu, R., Li, J.-T., Zhang, Y., Xu, P., and Sun, X.-W. (2015). Diversification of the duplicated Rab1a genes in a hypoxia-tolerant fish, common carp (Cyprinus carpio). *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology*, 188:54–62. 9

Zhong, Z., Yang, L., Zhang, Y. E., Xue, Y., and He, S. (2015). Correlated expression of retrocopies and parental genes in zebrafish. *Molecular Genetics and Genomics*. 2