https://doi.org/10.35662/unine-thesis-2733

UNIVERSITÉ DE NEUCHÂTEL

FONDS NATIONAL SUISSE DE LA RECHERCHE SCIENTIFIQUE

Comparative phylogeography of diploid and allopolyploid wild wheats *Aegilops* species

Thèse présentée à la Faculté des Sciences, Institut de Biologie
Université de Neuchâtel, Suisse

Par

Stella Huynh

Composition du jury :

PD Dr. Felber François, directeur de thèse
Pr. Dr. Koella Jacob, co-directeur
Pr. Dr. Parisod Christian, co-directeur et rapporteur
Dr. Marcussen Thomas, rapporteur
Dr. Naciri Yamama, rapporteur

2018

UNIVERSITÉ DE NEUCHÂTEL

Faculté des Sciences
Secrétariat-décanat de Faculté
Rue Emile-Argand 11
2000 Neuchâtel – Suisse
Tél : + 41 (0)32 718 21 00
E-mail : secretariat.sciences@unine.ch

IMPRIMATUR POUR THESE DE DOCTORAT

La Faculté des sciences de l'Université de Neuchâtel autorise l'impression de la présente thèse soutenue par

Madame Stella HUYNH

Titre:

"Comparative phylogeography of diploid and allopolyploid wild wheats *aegilops spp*"

sur le rapport des membres du jury composé comme suit:

- Dr François Felber, directeur de thèse, Université de Neuchâtel, Suisse
- Prof. Jacob Koella, co-directeur de thèse, Université de Neuchâtel, Suisse
- Prof. Christian Parisod, Université de Berne et Université de Neuchâtel, Suisse
- Dr Thomas Marcussen, Université d'Oslo, Norvège
- Dr Yamama Naciri, Conservatoire et Jardin botaniques de Genève

Neuchâtel, le 22 octobre 2018

Le Doyen, Prof. P. Felber

P. Fel

Table of contents

Abstract

Allopolyploidy is particularly widespread among flowering plants and have played an important role in speciation and diversification of species. Allopolyploid species generally show larger distribution than their progenitors. Allopolyploidy combines hybridization and whole-genome doubling, both of which lead to substantial genomic and epigenetic changes and are major activators of transposable elements (TEs). How allopolyploidization and the genomic changes following it affect the evolutionary history of species remains unclear.

The main objective of this thesis was to investigate the role of allopolyploidy in the evolution of species genome, distribution and ecology. The study was based on a model system composed of four allopolyploid wild wheat species, *Aegilops crassa* (DDMM), *Ae. cylindrica* (DDCC), *Ae. geniculata* (UUMM) and *Ae. triuncialis* (CCUU) that derived from the diploids *Ae. caudata* (CC), *Ae. comosa* (MM), *Ae. tauschii* (DD) and *Ae. umbellulata* (UU). Wheat species have a complex evolutionary history with recurrent homoploid and polyploid hybridization events as well as complex genomes filled with up to 90% of TEs.

The evolutionary origins and phylogeographic structure of diploid and allopolyploid *Aegilops* species were assessed across the species distribution using high-throughput amplicon sequencing of 30 nuclear markers and Sanger sequencing of 2 plastid markers. The allopolyploids likely appeared during glacial periods and have rapidly expanded their distribution after the Last Glacial Maximum about 10 000 years ago. Investigation of TE dynamics using fingerprinting assays for six TE families showed overall random activity of TEs across allopolyploid species ranges, however with located reactivation of TEs along the Balkan-Anatolian suture zone and at the species range margins for *Ae. cylindrica* and *Ae. triuncialis*. Finally, the ecological niche dynamics of the allopolyploids were inferred based on the modelled niches of their diploid progenitors using ordination technique. The ecological niches of the allopolyploids were mainly similar to the addition of their progenitors' niches, although with a slight niche expansion toward more temperate climate. These results, combined with the phylogeographic inferences, suggested that allopolyploidy conferred a broader ecological amplitude and higher genomic plasticity compared to diploids, enabling species to expand successfully without the need to shift their niche.

The allopolyploid *Aegilops* species offered a unique opportunity to address the long-term impact of allopolyploidy on species evolution, which reported genetic and ecological additivity of their diploid progenitors enabling species to rapidly expand their distribution.

Keywords: allopolyploidy, *Aegilops*, high-throughput amplicon sequencing, reticulate phylogeny, phylogeography, LTR retrotransposons dynamics, SSAP, ecological niche dynamic

Résumé

L'allopolyploidïe, notamment répandue chez les plantes à fleurs, a joué un rôle majeur dans la spéciation et la diversification des espèces. Les espèces allopolyploïdes ont une aire de distribution souvent plus large que leurs progéniteurs. L'allopolyploïdie combine hybridation et doublement du génome, activateurs majeurs des éléments transposables (ETs) et qui entraînent des changements génétiques et épigénétiques conséquents. Comment l'allopolyploïdization et les changements génomiques qu'elle induit affectent l'histoire évolutive des espèces restent obscures.

L'objectif principal de cette thèse est d'étudier le rôle de l'allopolyploïdie dans l'évolution du génome, de la distribution et de l'écologie des espèces. L'étude s'est basée sur quatre espèces de blé sauvage allopolyploïdes, *Aegilops crassa* (DDMM), *Ae. cylindrica* (DDCC), *Ae. geniculata* (UUMM) et *Ae. triuncialis* (CCUU) qui ont dérivé des diploïdes *Ae. caudata* (CC), *Ae. comosa* (MM), *Ae. tauschii* (DD) and *Ae. umbellulata* (UU). Les blés ont une histoire évolutive complexe avec des événements récurrents d'hybridation homoploïde et allopolyploïde ainsi que des génomes complexes composés jusqu'à 90% d'ETs.

Les origines évolutives et la structure phylogéographique des espèces d'*Aegilops* diploïdes et allopolyploïdes ont été étudiées à travers l'aire de distribution des espèces via le séquençage massif d'amplicons de 30 marqueurs nucléaires et le séquençage Sanger de 2 marqueurs chloroplastiques. Les allopolyploïdes seraient apparus durant les périodes glaciaires, ayant rapidement élargit distribution après le dernier maximum glaciaire, il y a environ 10 000 ans. L'étude de la dynamique de six familles d'ETs via la technique d'empreintes moléculaires a révélé une activité aléatoire des ETs à travers l'aire de distribution des espèces, avec des réactivations locales d'ETs autour de la zone de suture Balkans-Anatolie et en marge de distribution chez *Ae. cylindrica* et *Ae. triuncialis*. La dynamique de niche écologique chez les allopolyploïdes, inferrée sur la base des niches de leur progéniteurs diploïdes via la technique d'ordination, a révélé des niches globalement similaires à l'addition des niches parentales, avec une légère expansion de niche vers des climats plus tempérés. Ces résultats, combinés aux inférences phylogéographiques, suggèrent que l'allopolyploïdie confère une plus grande amplitude écologique et une plus grande plasticité génomique comparée à la diploïdie, permettant aux espèces d'étendre leur distribution sans avoir besoin de changer de niche.

Les espèces d'*Aegilops* allopolyploïdes ont offert l'opportunité unique d'addresser ici l'impact à long terme de l'allopolyploïdie sur l'évolution des espèces, reportant chez les

allopolyploïdes étudiés une additivité génétique et écologique de leur progéniteurs diploïdes permettant à ces espèces d'étendre efficacement leur aire de distribution.

Mots clés : allopolyploïdie, *Aegilops*, séquençage haut débit d'amplicons, phylogénie réticulée, phylogéographie, dynamique des rétrotransposons LTR, SSAP, dynamique de niche écologique

Abbreviations

AFLP	amplified fragment length polymorphism
LGM	last glacial maximum
LTR-RT	long terminal repeat retrotransposon
MCMC	Markov chain Monte Carlo
MSC	multispecies coalescence
PCA	principal component analysis
PCR	polymerase chain reaction
SSAP	sequence-specific amplified polymorphism
siRNA	small interferring ribonucleic acid
TE	transposable element
WGD	whole-genome duplication

Acknowledgment

First of all, I have to thank my supervisor Christian Parisod and my thesis director François Felber, for having given me the chance to carry this PhD project, for their support and wise advices during these four years.

I would also like to thank the committee members of my PhD defense, who granted me the doctoral degree and gave me precious advices on my manuscript and for my research career.

I would also like to thank all the people who have contributed to the achievement of this project and to the amazing time I spent in Switzerland:

- The staff members of the Botanical Garden of Neuchâtel, for their help in taking care of my plants and for their warm welcome when I was there for the experiments.
- Thomas Marcussen, my savior in the inextricable complexity of inferring polyploid and reticulate phylogenies.
- The staff of the Genetic Diversity Center of ETH Zurich, who helped me through all the molecular and bioinformatics steps of NGS sequencing and gave me precious advices and tips (thank you so much for that Silvia!).
- The CUSO and DPOB doctoral programs, for all the great courses I have attended.
- The best colleagues of UniNE: Guillaume, Julie, Magali, Olivier, Benjamin, Ursula, Nikhil, Giacomo, Alex, Mulot, Sandra, Max & Juan, Tom, Andrew, Michael, etc…
- All the (totally crazy!) guys who I shared so many wonderful moments with: Vinciane, Marine, Jonas, Daniel, Chris, Amandine, Coco, Dalinda, Quentin, Clément&Serena, Clément & Ricarda, Alfonso, Andrea, Monica, Gaylord, etc…
- Special thanks to Noémie, who was in my office for few months and quickly became a precious friend. I miss you and your so lovely Cobalt (the best dog ever!)!
- Very special thanks to Rimjhim, my office partner who started the PhD adventure at the same time as me. You have always been here for me. Thanks for having supported me until the end! You are not just a colleague but a very precious friend. I promise to come back to India so that you can show me more of its beauty!
- Last but not least, many thanks to my family for having supported me for 28 years (already!), and especially during the long writing of this PhD. **Now it's done !**

General introduction

Introduction

Role of polyploidy in species evolution: a long-term debate

Polyploidization, or whole genome duplication (WGD), has first been described about a century ago (Stebbins, 1940) and is since recognized as a powerful mechanism of speciation and diversification, especially among flowering plants (Levin, 2002; Wood *et al.*, 2009; Jiao *et al.*, 2011). Polyploidy indeed seems recurrent within the Angiosperms, which experienced at least two rounds of ancient polyploidization events (paleopolyploidization) predating their origin and radiation (Jiao *et al.*, 2011; Van de Peer *et al.*, 2017). About 15 % of the recent speciation events among Angiosperms are estimated to be associated with polyploidy (Wood *et al.*, 2009). Despite playing an important role in the evolution of plants, the question of polyploidy as an evolutionary advantage or as a hindrance still raises long-standing and animated debates (Stebbins, 1971; Mayrose *et al.*, 2011; Arrigo & Barker, 2012; Soltis *et al.*, 2014). The main argument for polyploidy as a hindrance came from Stebbins (1971) who postulated that even though polyploidy may initially confer advantages, it had limited long-term evolutionary potential and did not contribute to species diversification unlike diploids. Stebbins' assumption was later referred to as polyploidy being an 'evolutionary dead-end', a hypothesis supported by phylogenetic inferences in Mayrose *et al.* (2011) and Arrigo and Barker (2012). However, most (if not all) living organisms are thought to have undergone at least one polyploidization event including the most diversified plant phyla, the Angiosperms (Van de Peer *et al.*, 2017). Recurrent WGD may have produced particular genomic combinations that generated evolutionary novelties being advantageous for species evolution or adaptation, and thereby serving as a catalyst for species diversification (Soltis & Soltis, 2016).

The majority of polyploid species identified are allopolyploids. In contrast with autopolyploids that form within or among populations of the same species, allopolyploids involve hybridization between two distinct species followed by WGD. Even though many autopolyploids probably remain cryptic due to their morphological similarity with their progenitors, allopolyploidy is particularly recurrent and widespread among species (Stebbins, 1950; Grant, 1981).

Impact of allopolyploidy on nascent species establishment (revolutionary changes)

Allopolyploidy may be considered as a hindrance to the establishment of species as it induces epigenetic instability or changes in the dosage of maternal and paternal genomes (Comai,

2005). However, allopolyploidy also confers evolutionary advantages such as masking of deleterious alleles through gene redundancy and permanent fixation of individual's heterozygosity (Comai, 2005). The genetic divergence between allopolyploid's progenitors indeed favour the pairing of homologous chromosomes (i.e. diploid-like disomic inheritance) while preventing homeologous pairing (intergenomic recombination). This fixed heterozygosity increases the individual fitness, also called "hybrid vigor" or "heterosis", by maintaining the allelic variants from the allopolyploid's progenitors but also by masking deleterious alleles. Allopolyploidy is often associated to a selfing reproductive regime, which increases the effect of heterosis but also prevent the nascent allopolyploid individuals to backcross toward their progenitors, thus avoiding reproductive competition with their established progenitors (i.e. minority cytotype disadvantage; Levin, 1975).

Allopolyploidization induces a genome reorganization (Figure 1) due to genomic incompatibilities between the progenitors' genomes merged together, which can have both positive or deleterious effect on the allopolyploids (Comai, 2005).

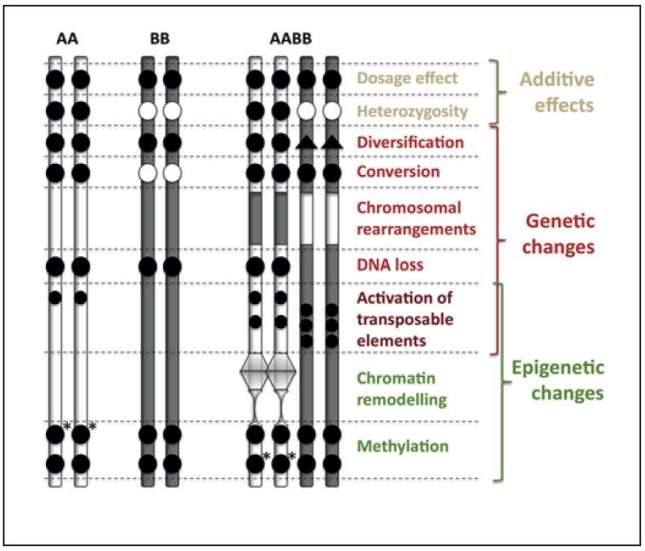


Figure 1. Main features of genome reorganization after allopolyploidization events. Figure taken from Tayalé and Parisod (2013).

These changes include structural changes, with elimination of DNA sequences and gene losses through chromosomal rearrangements, and functional changes such as methylation repatterning that regulates gene expression (Soltis & Soltis, 1999; Levin, 2002; Doyle *et al.*, 2008; Tayalé & Parisod, 2013). Such genome reorganization, commonly observed in the first generations after allopolyploid formation, highlight departure of allopolyploid genome from the expected addition of progenitors' genomes. Allopolyploids show a general trend toward

genome downsizing, i.e. decreased genome size as compared to the addition of its progenitors' genome sizes (Leitch & Bennett, 2004). Allopolyploid genomes are typically mainly filled with transposable elements (TEs), whose activation following allopolyploidization (Parisod & Senerchia, 2012) may lead to substantial structural changes contributing to the increase or decrease of allopolyploid genome sizes.

TEs are repetitive DNA fragments whose main function is to transpose themselves across the species ('host') genome. TEs transpose through either a "copy-and-paste" mechanism (i.e. retrotransponsons) or a "cut-and-paste" mechanism (i.e. transposons) (Wicker *et al.*, 2007). TE copies constitute the most dynamic part of the genome and contribute to genome size increase (Fedoroff, 2012), representing up to 90 % of allopolyploid genomes and especially of crops (Tenaillon *et al.*, 2010; Brenchley *et al.*, 2012). Bursts of TEs following allopolyploidization drastically affect the structure of allopolyploid genome, with a dynamic balance between increased genome size due to TE transposition and decreased genome size due to TE elimination by the host species. Allopolyploid species limit TE proliferation either through substantial structural mechanisms such as chromosomal rearrangements or through epigenetic mechanisms such as TEs silencing by the cytoplasmic small interfering RNAs (siRNAs).

The maternally-inherited siRNAs have evolved jointly with maternal TE copies but not with the paternal TE copies present in the allopolyploid individuals. The siRNAs might not specifically recognize paternal TE copies, as they come from a paternal species that is genetically divergent from the maternal species (qualitative differences), and may fail to silence these copies that can thus proliferate. Some studies indeed seem to highlight a higher TE proliferation of paternal copies as compared to the maternal copies (Blumenstiel & Hartl, 2005). In addition, the number of TEs copies may differ between paternal and maternal genomes (quantitative differences) also leading to imbalance between siRNA and TEs (the genome shock hypothesis; Parisod & Senerchia, 2012) The general trend toward genome downsizing following early generations in synthetic allopolyploids and natural populations suggests that despite TE proliferation, copies are immediately regulated through TE deletion during the formation of allopolyploids before genome stabilizes after few generations (Eilam *et al.*, 2008; Ha *et al.*, 2009). Hybridization appears to have a higher impact on TE activity than polyploidy *per se* (Ainouche *et al.*, 2009), which may be due to genomic and epigenetic incompatibilities between progenitor divergent genomes under the genome shock hypothesis (Senerchia *et al.*, 2015, 2016). To what extent the genetic divergence between progenitor

species can predict the intensity of the genomic conflicts and changes following allopolyploidization remains unclear.

Impact of allopolyploidy on long-term species evolution (evolutionary changes)

While the immediate effects of allopolyploidy have been extensively studied, mostly through synthetic neo-allopolyploids (Liu *et al.*, 1998; Gaeta *et al.*, 2007; Petit *et al.*, 2010) or natural neo-allopolyploids (Ainouche *et al.*, 2009), the long-term impact of allopolyploidy on species evolution have unfortunately been poorly investigated. Allopolyploid species may indeed evolve changes during their evolutionary history that are not due to allopolyploidy per se but instead facilitated by their allopolyploid state. Fixed heterozygosity and gene redundancy confer to allopolyploids the advantage of recruiting and modulating the expression of diversified gene duplicates in time (Flagel et al., 2008) or in different tissues (Adams et al., 2003; Adams & Wendel, 2005) through gene subfunctionalization, i.e. the partitioning of ancestral gene function between duplicates.

Gene subfunctionalization can imply transient epigenetic mechanisms silencing only one of the homeologous gene copies, whose function is complemented by the other copy, or affect both copies. Duplicated genes can lose their initial function (pseudogenization), which might affect the ecological speciation of the allopolyploid. Pseudogenization occur when genes are inactivated through genetic changes (substitutions, insertions or deletions) or epigenetic changes, or eliminated through structural rearrangements. On the opposite, duplicated genes can acquire new functions (neofunctionalization), thereby diversifying gene functions and subsequently foster diversification of species and of its ecological tolerance. New gene functions can be further recruited by natural selection to improve allopolyploid individual fitness. Several studies pointed out gene families that seemed specifically retained in allopolyploids and particularly sensitive to genome dosage (Sémon & Wolfe, 2007; Freeling *et al.*, 2012; Pires *et al.*, 2016) or genetic changes asymmetrically affecting genes copies either from paternal (Shaked *et al.*, 2001) or maternal subgenome (Ainouche *et al.*, 2009). However, the frequency and distribution of genetic novelties and their relative impact on the evolutionary success of allopolyploid species remains unclear. On the long-term, genome redundancy can potentially slow down species diversification and adaptation, as it dilutes the effects of new mutations including that of beneficial mutations, as compared to diploids (Otto, 2007). On the other hand, genome redundancy provide a high amount of genomic material to accumulate mutations that can further be recruited by the polyploid to rapidly adapt (e.g. to environmental stresses). Although deleterious mutations can also accumulate, allopolyploidy

reduces the incidence of homozygous recessive (Comai, 2005). The question of evolutionary success of allopolyploid species due to allopolyploidy *per se* or to the long-term evolution of allopolyploid genome remains open (Soltis *et al.*, 2016).

Impact of allopolyploidy on the ecology of species

Allopolyploid species generally show larger geographical distribution than their progenitors' distribution, which may result from higher dispersal abilities of allopolyploids. Allopolyploids typically exhibit higher genomic plasticity than their progenitors as a result of gene redundancy and fixed heterozygosity, as well as increased phenotypic plasticity, which may further facilitate species range expansion and colonization of new habitats (Levin, 2002; te Beest *et al.*, 2012). Large distribution of allopolyploids may indeed results from a wide range of ecological tolerance combining both progenitor's ecology (Otto & Whitton, 2000). Combination of high genomic plasticity and broad ecological tolerances potentially enabled allopolyploids to survive drastic environmental changes, as suggested by the high frequency of polyploid species whose ages were estimated around the Cretaceous-Tertiary extinction mass period, i.e. about 65 million years ago (Fawcett *et al.*, 2009; Soltis & Burleigh, 2009). By facilitating species expansion, wide ecological amplitude and phenotypic plasticity also confer to the allopolyploids a high potential of being invasive (Pandit *et al.*, 2011; Hahn *et al.*, 2012; te Beest *et al.*, 2012).

During their lifetime, allopolyploids may also evolve genetic novelties. Whether genetic novelties acquired during the evolution of allopolyploids underlie the ecological novelties remains to be investigated. Furthermore, allopolyploids often show recurrent and independent origins (Soltis & Soltis, 2009). Multiple origins not only increase the initial genetic diversity of nascent allopolyploids but it is also suggested to enhance ecological amplitude of allopolyploids (Meimberg *et al.*, 2009). Whether independent origins within an allopolyploid species give rise to lineages that are cryptic and/or locally adapted to different ecological niches remains an open question.

Allopolyploidy has mainly been investigated under the scope of genetic and epigenetic changes and their consequences on species evolution. However, studies focusing on the impact of allopolyploidy on species ecology are relatively rare (Ramsey & Ramsey, 2014). Allopolyploid species were initially identified by their intermediate morphology as compared to their progenitors' morphology, and further hypothesized to show combined or intermediate ecological requirements to that of their progenitors. Allopolyploids actually exhibit a mixture of their progenitors' morphological characteristics (Soltis *et al.*, 2004). However, polyploid

species generally exhibit bigger cell, stomata and pollen sizes than their progenitors (Ramsey & Schemske, 2002; Orr-Weaver, 2015). These morphological changes, which occur immediately following allopolyploidization, likely affect the individual fitness and adaptation to its local environment. In particular, allopolyploids have less numerous but bigger stomata, which reduces water deficiency and enable plants to be adapted to harsher environments and extreme climate compared to their progenitor species (Levin, 2002; te Beest *et al.*, 2012). Polyploidy was also suggested to increase in frequency with latitude (Stebbins, 1971). Allopolyploid species can possibly extend or shift their ecological requirements toward harsher conditions compared to their progenitor species. However, many polyploid plant species are also found in lower latitude and warmer climate, with distribution more or less overlapping that of their progenitors, such as the *Aegilops* species (van Slageren, 1994). To what extent the ecology of allopolyploids reflects the strict addition of their progenitors' ecology or results instead from a shift of their ecology from this addition is unclear.

The Aegilops species: a well-studied but complex diploid-allopolyploid system

The *Triticeae* tribe represents an ideal model system to investigate the consequences of allopolyploidy. Several species of this tribe have been sequenced (International Wheat Genome Sequencing Consortium *et al.*, 2013; Middleton *et al.*, 2014; The International Wheat Genome Sequencing Consortium (IWGSC) *et al.*, 2014) and studied for their agronomic value as they include major crop species. In particular, the wild wheats *Aegilops* species comprise many allopolyploid species, among which we focused on four species whose diploid progenitor species have been confidently identified (Badaeva *et al.*, 2002, 2004) and that showed multiple origins (Meimberg *et al.*, 2009).

The *Aegilops* species (Poaceae) are annual and predominantly selfing species comprising diploid, tetraploid and hexaploid species that are mainly distributed across the Mediterranean basin and the Iran-Turanian region (van Slageren, 1994). They are mainly found in open and disturbed habitats (Eig, 1929; van Slageren, 1994; Kilian *et al.*, 2011). Two allopolyploid species, *Ae. triuncialis* and *Ae. cylindrica*, have also been reported invasive or with a weedy behaviour in North America, i.e. outside of their native distribution (van Slageren, 1994; Lyons *et al.*, 2010). The genomes of *Aegilops* species were defined based on their meiotic behaviour in hybrid crossings (Kihara, 1954; Kimber & Tsunewaki, 1988), and can be classified under the following cytogenetic karyotypes: M, C, D, U, N, T or S genome (Table 1).

Table 1: Overview of *Aegilops* species and of their genome composition, modified from Kilian *et al.* (2011). Species sections are given according to van Slageren (1994). Genomic formulas of species are indicated into brackets, with the maternal genome underlined for polyploids, following Kimber and Tsunewaki (1988).

Aegilops infrageneric taxon	Diploid		Tetraploid		Hexaploid	
	Species		Species		Species	
Aegilops	*Ae. umbellulata* Zhuk.	(U)	*Ae. biuncialis* Vis.	(UM)		
			Ae. columnaris Zhuk.	(UM)		
			Ae. geniculata Roth	(UM)		
			Ae. kotschyi Boiss.	(US)		
			Ae. neglecta subsp. *neglecta* Req. ex Bertol.	(UM)	*Ae. neglecta* subsp. *recta* Req. ex Bertol.	(UMN)
			Ae. peregrina (Hack. In J. Fraser) Maire et Weiller	(US^s)		
			Ae. triuncialis var. *triuncialis*	(UC)		
			Ae. triuncialis var.*persica* (Boiss.) Eig	(UC)		
Vertebrata Zhuk. emend. Kihara	*Ae. tauschii* Coss.	(D)	*Ae. crassa subsp. crassa* (4x) Boiss.	(DM)	*Ae. crassa* subsp. *crassa* (6x) Boiss.	(DDM)
			Ae. ventricosa Tausch	(DN)		
					Ae. vavilovii (Zhuk.) Chennav.	(DMS)
					Ae. juvelanis (Thell.) Eig	(DMU)
Cylindricum (Jaub. et Spach) Zhuk.	*Ae. caudata* L.	(C)	*Ae. cylindrica* Host	(DC)		
Comopyrum (Jaub. et Spach) Zhuk.	*Ae. comosa* Sm. in Sibth. et Sm.	(M)				
	Ae. Uniaristata Vis.	(N)				
Sitopsis (Jaub. et Spach) Zhuk.	*Ae. bicornis* (Forssk.) Jaub. et Sp.	(S^b)				
	Ae. longissima Schweinf. et Muschl.	(S^l)				
	Ae. sharonensis Eig	(S^{sh})				
	Ae. searsii Feldman et Kislev ex Hammer	(S^s)				
	Ae. speltoides Tausch	(S)				
Subgenus *Amblyopyrum*	*Ae. mutica* (Boiss.) Eigs	(T)				

The *Aegilops* species have evolved through recurrent homoploid hybridization and allopolyploidization events. Numerous studies have investigated the phylogenetic relationships among the *Aegilops* and *Triticum* species (Petersen *et al.*, 2006; Kawahara, 2009; Baum *et al.*, 2012; Marcussen *et al.*, 2014; Bernhardt *et al.*, 2017). However, these studies were mostly based on chloroplastic markers, which only reflect the evolution of the maternal lineages and thus does not hold genomic footprints of hybridization. Phylogenies inferred from nuclear data remains scarce, a few studies using one to two loci (e.g. Petersen *et al.*, 2006) or limited sampling of taxa (e.g. Marcussen *et al.*, 2014). Both chloroplastic and nuclear phylogenies inferred to date showed conflicting topologies, especially with the placement of *Ae. speltoides* or *Ae. mutica* (Petersen *et al.*, 2006; Baum *et al.*, 2012; Bernhardt *et al.*, 2017), thus calling for phylogenetic studies based on a sufficient number of taxa and of nuclear markers.

All allopolyploid species were formed from diploid species having either the U (i.e. *Ae. umbellulata*) or the D genome (i.e. *Ae. tauschii*), with recurrent homoploid hybridizations also occurring among allopolyploids sharing one of these two genomes. Such biased recurrent hybridizations among allopolyploid species thus left the common U or D genome mainly unaltered as they have been evolving through homologous recombination, hence called "pivotal genomes", whereas the other genomes found in the allopolyploids were genetically different and thus evolved through homeologous recombination that structurally and epigenetically altered them, hence called "differential genomes" (Zohary & Feldman, 1962; Feldman & Levy, 2005).

Genomic recombinations impact the dynamics of TEs by deleting copies whereas epigenetic changes alter silencing of TE, which can be de-repressed (Senerchia *et al.*, 2016). TEs compose almost 90 % of the *Aegilops* genome (Tenaillon *et al.*, 2010; Luo *et al.*, 2013) and are mainly represented by long-terminal retrotransposon (LTR-RT) families in this genus (Sabot & Schulman, 2009). Recurrent allopolyploidization and homoploid hybridization among allopolyploid *Aegilops* species show to massively affect the expression and activity of TE copies (Ozkan *et al.*, 2001; Kashkush *et al.*, 2002; Senerchia *et al.*, 2014, 2015, 2016). Investigation of several LTR-RT families among four allopolyploid *Aegilops* species revealed differential evolutionary dynamics of TE copies among the species (Senerchia *et al.*, 2014). Some LTR-RT families showed uniform trends of proliferation (i.e. *BARE1*, *Romani*) or quiescence (i.e. *Egug*, *Nusif*), but specific families revealed species-specific trends (i.e. *Sabine*, *Xalax*). TE dynamics also differed between the allopolyploid genomes, with the

differential subgenomes generally showing more structural and functional alterations of its TE copies than the pivotal subgenomes (Senerchia *et al.*, 2014). The correlation observed between genetic divergence of progenitors' TE copies and TE proliferation in the allopolyploid potentially suggests a strong impact of genomic incompatibilities in TE activity (genome shock hypothesis). To what extent the genetic divergence between the progenitor species and their evolutionary histories affect the incompatibilities between maternal and paternal TE copies as well as TE dynamics in the allopolyploids remains unclear.

Genetic divergence among TE copies can also enhance the genomic incompatibilities between the differential genomes and contribute to substantial genomic rearrangements of these genomes. Differential introgression of active TE families in natural hybrids suggest that TEs may play a role in keeping species boundaries within hybrid zones (Senerchia *et al.*, 2016). In addition, other studies have highlighted the role of certain TE families or specific TE copies in regulating the expression of nearby genes (Feschotte, 2008; González & Petrov, 2009). Being highly dynamic, TEs can rapidly increase or decrease the expression of genes involved in stress responses and hence to cope with rapid environmental changes (Casacuberta & González, 2013). However, TE copies can also disrupt genes and spread erratically throughout the allopolyploid genome, which can lead to decreased fitness or death of the individual. Regarding the contrasted TE dynamics among the species and between their genomes (pivotal *vs.* differential), TEs might play a role in the apparent evolutionary success of the allopolyploid *Aegilops* species. However, whether the observed TE dynamics are even across the species range, and to what extent TE copies have been inherited from the allopolyploid's progenitors or have evolved during the evolutionary lifetime of the allopolyploid, remain completely unkown. Though some recent studies intended to model the long-term evolution of TE copies (Le Rouzic *et al.*, 2007; Szitenberg *et al.*, 2016), empirical studies are still lacking.

The *Triticeae* tribe has shown to have a complex evolutionary history punctuated by recurrent homoploid and polyploid hybridization events making it particularly difficult to resolve despite extensive investigations (Zohary & Feldman, 1962; van Slageren, 1994; Marcussen *et al.*, 2014; Li *et al.*, 2015; Sandve *et al.*, 2015; Glémin *et al.*, 2018). The multiple origins of the allopolyploid *Aegilops* species may have occurred at different timescales or geographical locations, but also between different progenitors' lineages. Divergent lineages in allopolyploid species can have contrasted evolutionary histories that can in turn affect not only genome evolution and TE dynamics, but also local adaptation of the species.

Intraspecific allopolyploid lineages having different origins may not necessarily have the same fitness, which raises the question of whether there are genome combinations “better” than other and that may be the key success of allopolyploids. It thus appears important to extensively and carefully investigate the evolutionary history of allopolyploid species jointly with their progenitor species and identify the precise origins of allopolyploids across the whole distribution of the species. Such investigations will provide a better understanding on the genetic, ecological and historical processes acting on the evolution of allopolyploid species as well as the relative impact of TE dynamics on the species evolution.

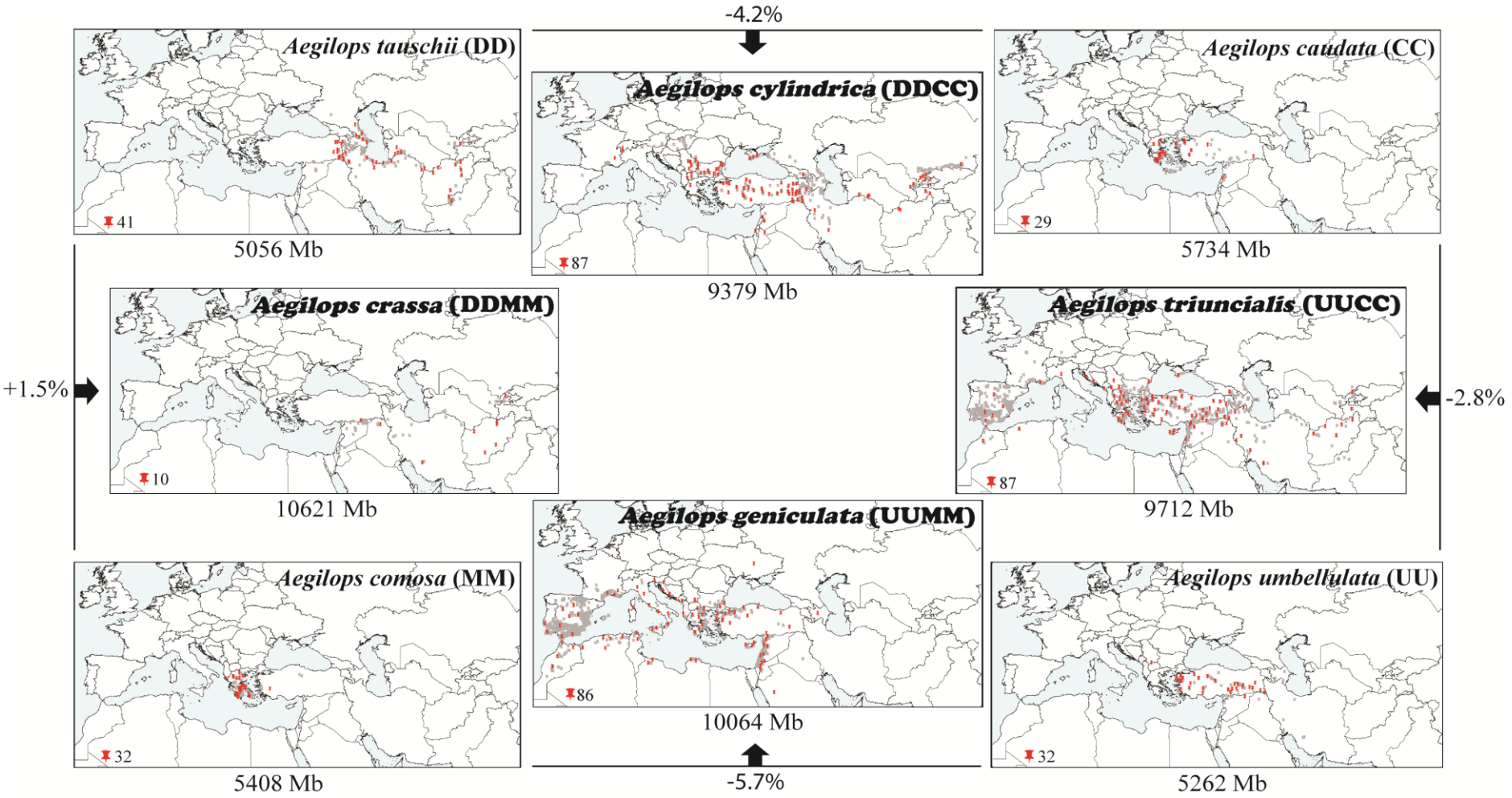


Figure 2: Evolutionary relationships and current distribution of four diploid and four of their derivated allopolyploid *Aegilops* species. Species distributions are represented with grey crosses for occurrences reported in GBIF database (https://www.gbif.org/) and red thumbtacks for genotyped individuals. Genomic formula are indicated into brackets following van Slageren (1994). Species genome sizes are reported according to Eilam *et al.*(2007, 2008), as well as the genome size reduction (%) in allopolyploids compared to the expected addition of progenitor genome sizes.

Chapter outline and context

The present PhD project intends to investigate the long-term impact of allopolyploidy on species genome (structural changes) and on species distribution and ecology (geographical expansion *vs.* new ecological requirements). In particular, the hypothesis of allopolyploid being the genetic and ecological additivity of its progenitors was tested throughout the chapter 2 (additivity of nuclear genome), chapter 3 (additivity of TE copies) and chapter 4 (additivity of ecological niches). The chapter 1 aimed at clarifying the phylogenetic relationships among the diploid *Aegilops* species prior to investigating the evolution of derived allopolyploids.

Chapter 1 intended to disentangle the evolutionary relationships among diploid *Aegilops-Triticum* species to better understand the genomic background of the diploid progenitors of the allopolyploid species. We used an innovative method to genotype 30 nuclear low-copy markers through simultaneous amplicon sequencing of 48 individuals using the Access Array IFC chip (Fluidigm) coupled with Illumina MiSeq of overlapping paired-end reads. Our method could detect and filter out paralogs while capturing all genetic variants. We detected more than one reticulation event at the origin of the D clade, which includes all diploid *Aegilops* species except the outcrossers *Ae. speltoides* (the B clade) and *Ae. mutica* (T).

Chapter 2 assessed the evolutionary history of the allopolyploid *Aegilops* species based on their progenitor origins, their relative ages, and their population genomic structure and diversity. We extended the NGS amplicon sequencing of Chapter 1 to genotype 402 individuals within four diploid species and their four derivative allopolyploid species across their distribution ranges. Our results showed that allopolyploids have likely appeared from different maternal and/or paternal progenitors during several glacial periods in the late Quaternary, and then progressively started to accumulate genetic mutations. *Ae. geniculata* appeared older than *Ae. triuncialis* and *Ae. cylindrica* (the youngest species). The current distribution of *Aegilops* species reflect the effect of range contraction in different refugia followed by postglacial expansion with secondary contact along the Balkans-Anatolia region. The wide range of allopolyploids thus suggests that these later have better colonization capacities than their diploid progenitors.

Chapter 3 investigated TEs dynamics across the allopolyploid ranges to assess the long-term and spatial evolution of TE copies. We tracked genome-wide TE insertions of six LTR-RT families, whose evolutionary trajectories were previously assessed (Senerchia *et al.*, 2014), using sequence-specific amplified polymorphism (SSAP) in the same 402 accessions of the

diploid and allopolyploid species genotyped in Chapter 2. Genome-wide restructuring of random sequences, assessed by similar technique (amplified fragment length polymorphism, AFLP), was removed from SSAP signal to highlight TE-specific dynamics. Comparisons of allopolyploids to their respective progenitors showed TE dynamics among species diverging from our expectations, as well as heterogeneous TE dynamics across species ranges suggesting that TE copies likely evolve through random processes. Some LTR-RT families, i.e. BARE, however reveal hotspots of recent TE transpositions in *Ae. triuncialis* and *Ae. cylindrica* close to contact zones identified in Chapter 2 or among marginal individuals, suggesting that specific evolutionary processes are acting on these TE copies.

Chapter 4 analyzed the impact of allopolyploidy on species ecology. A common ecological space of the eight *Aegilops* species was first modelled based on 20 bioclimatic variables randomly sampled across species ranges. The ecological niches of allopolyploid species were then modelled and compared within this ecological space through a PCA on previously genotyped accessions and 3265 geolocalized occurrences data. The ecological niches of allopolyploid species revealed mainly additive, although differentially overlapping that of their respective progenitor species. Most allopolyploid species also expanded their niches toward more temperate climate. Allopolyploid niche dynamics suggests that allopolyploidy *per se* did not induce an ecological niche shift but probably confer the adaptive potential for rapidly colonizing their habitats and potentially further expand to other habitats.

References

Adams KL, Cronn R, Percifield R, Wendel JF. **2003**. Genes duplicated by polyploidy show unequal contributions to the transcriptome and organ-specific reciprocal silencing. *Proceedings of the National Academy of Sciences* **100**: 4649–4654.

Adams K, Wendel J. **2005**. Novel patterns of gene expression in polyploid plants. *Trends in Genetics* **21**: 539–543.

Ainouche ML, Fortune PM, Salmon A, Parisod C, Grandbastien M-A, Fukunaga K, Ricou M, Misset M-T. **2009**. Hybridization, polyploidy and invasion: lessons from *Spartina* (Poaceae). *Biological Invasions* **11**: 1159–1173.

Arrigo N, Barker MS. **2012**. Rarely successful polyploids and their legacy in plant genomes. *Current Opinion in Plant Biology* **15**: 140–146.

Badaeva ED, Amosova AV, Muravenko OV, Samatadze TE, Chikida NN, Zelenin AV, Friebe B, Gill BS. **2002**. Genome differentiation in *Aegilops*. 3. Evolution of the D-genome cluster. *Plant Systematics and Evolution* **231**: 163–190.

Badaeva ED, Amosova AV, Samatadze TE, Zoshchuk SA, Shostak NG, Chikida NN, Zelenin AV, Raupp WJ, Friebe B, Gill BS. **2004**. Genome differentiation in *Aegilops*. 4. Evolution of the U-genome cluster. *Plant Systematics and Evolution* **246**: 45–76.

Baum BR, Edwards T, Mamuti M, Johnson DA. **2012**. Phylogenetic relationships among the polyploid and diploid *Aegilops* species inferred from the nuclear 5S rDNA sequences (Poaceae: Triticeae) (JP Gustafson, Ed.). *Genome* **55**: 177–193.

te Beest M, Le Roux JJ, Richardson DM, Brysting AK, Suda J, Kubesova M, Pysek P. **2012**. The more the better? The role of polyploidy in facilitating plant invasions. *Annals of Botany* **109**: 19–45.

Bernhardt N, Brassac J, Kilian B, Blattner FR. **2017**. Dated tribe-wide whole chloroplast genome phylogeny indicates recurrent hybridizations within Triticeae. *BMC Evolutionary Biology* **17**:141.

Blumenstiel JP, Hartl DL. **2005**. Evidence for maternally transmitted small interfering RNA in the repression of transposition in Drosophila virilis. *Proceedings of the National Academy of Sciences of the United States of America* **102**: 15965–15970.

Brenchley R, Spannagl M, Pfeifer M, Barker GLA, D'Amore R, Allen AM, McKenzie N, Kramer M, Kerhornou A, Bolser D, *et al.* 2012. Analysis of the bread wheat genome using whole-genome shotgun sequencing. *Nature* **491**: 705–710.

Casacuberta E, González J. **2013**. The impact of transposable elements in environmental adaptation. *Molecular Ecology* **22**: 1503–1517.

Comai L. **2005**. The advantages and disadvantages of being polyploid. *Nature Reviews Genetics* **6**: 836–846.

Doyle JJ, Flagel LE, Paterson AH, Rapp RA, Soltis DE, Soltis PS, Wendel JF. **2008**. Evolutionary genetics of genome merger and doubling in plants. *Annual Review of Genetics* **42**: 443–461.

Eig A. **1929**. *Monographisch-kritische uebersicht der gattung* Aegilops. Verlag des Repertoriums.

Eilam T, Anikster Y, Millet E, Manisterski J, Feldman M. **2008**. Nuclear DNA amount and genome downsizing in natural and synthetic allopolyploids of the genera *Aegilops* and *Triticum* (P Gustafson, Ed.). *Genome* **51**: 616–627.

Eilam T, Anikster Y, Millet E, Manisterski J, Sagi-Assif O, Feldman M. **2007**. Genome size and genome evolution in diploid Triticeae species (P Gustafson, Ed.). *Genome* **50**: 1029–1037.

Fawcett JA, Maere S, 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* **106**: 5737–5742.

Fedoroff NV. **2012**. Transposable elements, epigenetics, and genome evolution. *Science* **338**: 758–767.

Feldman M, Levy AA. **2005**. Allopolyploidy – a shaping force in the evolution of wheat genomes. *Cytogenetic and Genome Research* **109**: 250–258.

Feschotte C. **2008**. Transposable elements and the evolution of regulatory networks. *Nature Reviews Genetics* **9**: 397–405.

Flagel L, Udall J, Nettleton D, Wendel J. **2008**. Duplicate gene expression in allopolyploid *Gossypium* reveals two temporally distinct phases of expression evolution. *BMC Biology* **6**: 16.

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

Gaeta RT, Pires JC, Iniguez-Luy F, Leon E, Osborn TC. **2007**. Genomic changes in resynthesized *Brassica napus* and their effect on gene expression and phenotype. *The Plant Cell* **19**: 3403–3417.

Glémin S, Scornavacca C, Dainat J, Burgarella C, Viader V, Ardisson M, Sarah G, Santoni S, David J, Ranwez V. **2018**. Pervasive hybridizations in the history of wheat relatives. *bioRxiv*.

González J, Petrov DA. **2009**. The adaptive role of transposable elements in the *Drosophila* genome. *Gene* **448**: 124–133.

Grant V. **1981**. *Plant Speciation*. New York: Columbia University Press.

Ha M, Lu J, Tian L, Ramachandran V, Kasschau KD, Chapman EJ, Carrington JC, Chen X, Wang X-J, Chen ZJ. **2009**. Small RNAs serve as a genetic buffer against genomic shock in Arabidopsis interspecific hybrids and allopolyploids. *Proceedings of the National Academy of Sciences* **106**: 17835–17840.

Hahn MA, van Kleunen M, Müller-Schärer H. **2012**. Increased Phenotypic Plasticity to Climate May Have Boosted the Invasion Success of Polyploid Centaurea stoebe (S Allodi, Ed.). *PLoS ONE* **7**: e50284.

International Wheat Genome Sequencing Consortium, Jia J, Zhao S, Kong X, Li Y, Zhao G, He W, Appels R, Pfeifer M, Tao Y, *et al.* 2013. *Aegilops tauschii* draft genome sequence reveals a gene repertoire for wheat adaptation. *Nature* **496**: 91–95.

Jiao Y, Wickett NJ, Ayyampalayam S, Chanderbali AS, Landherr L, Ralph PE, Tomsho LP, Hu Y, Liang H, Soltis PS, *et al.* 2011. Ancestral polyploidy in seed plants and angiosperms. *Nature* **473**: 97–100.

Kashkush K, Feldman M, Levy AA. **2002**. Gene loss, silencing and activation in a newly synthesized wheat allotetraploid. *Genetics* **160**: 1651–1659.

Kawahara T. **2009**. Molecular phylogeny among *Triticum-Aegilops* species and of the tribe Triticeae. *Breeding Science* **59**: 499–504.

Kihara H. **1954**. Considerations on the evolution and distribution of *Aegilops* species based on the analyser-method. *CYTOLOGIA* **19**: 336–357.

Kilian B, Mammen K, Millet E, Sharma R, Graner A, Salamini F, Hammer K, Özkan H. **2011**. *Aegilops*. In: Kole C, ed. Wild Crop Relatives: Genomic and Breeding Resources. Berlin, Heidelberg: Springer Berlin Heidelberg, 1–76.

Kimber G, Tsunewaki K. **1988**. Genome symbols and plasma types in the wheat group. *Proceedings of the 7th International Wheat Genetics Symposium* **12**:1209-1210.

Le Rouzic A, Boutin TS, Capy P. **2007**. Long-term evolution of transposable elements. *Proceedings of the National Academy of Sciences* **104**: 19375–19380.

Leitch IJ, Bennett MD. **2004**. Genome downsizing in polyploid plants. *Biological Journal of the Linnean Society* **82**: 651–663.

Levin DA. **1975**. Minority Cytotype Exclusion in Local Plant Populations. *Taxon* **24**: 35–43.

Levin DA. **2002**. *The role of chromosomal change in plant evolution*. Oxford ; New York: Oxford University Press.

Li L-F, Liu B, Olsen KM, Wendel JF. **2015**. A re-evaluation of the homoploid hybrid origin of *Aegilops tauschii* , the donor of the wheat D-subgenome. *New Phytologist* **208**: 4–8.

Liu B, Vega JM, Segal G, Abbo S, Rodova M, Feldman M. **1998**. Rapid genomic changes in newly synthesized amphiploids of *Triticum* and *Aegilops*. I. Changes in low-copy noncoding DNA sequences. *Genome* **41**: 272–277.

Luo M-C, Gu YQ, You FM, Deal KR, Ma Y, Hu Y, Huo N, Wang Y, Wang J, Chen S, *et al.* 2013. A 4-gigabase physical map unlocks the structure and evolution of the complex genome of *Aegilops tauschii*, the wheat D-genome progenitor. *Proceedings of the National Academy of Sciences* **110**: 7940–7945.

Lyons KG, Shapiro AM, Schwartz MW. **2010**. Distribution and ecotypic variation of the invasive annual barb goatgrass (*Aegilops triuncialis*) on serpentine soil. *Invasive Plant Science and Management* **3**: 376–389.

Marcussen T, Sandve SR, Heier L, Spannagl M, Pfeifer M, The International Wheat Genome Sequencing Consortium, Jakobsen KS, Wulff BBH, Steuernagel B, Mayer KFX, *et al.* 2014. Ancient hybridizations among the ancestral genomes of bread wheat. *Science* **345**: 1250092–1250092.

Mayrose I, Zhan SH, Rothfels CJ, Magnuson-Ford K, Barker MS, Rieseberg LH, Otto SP. **2011**. Recently Formed Polyploid Plants Diversify at Lower Rates. *Science* **333**: 1257–1257.

Meimberg H, Rice KJ, Milan NF, Njoku CC, McKay JK. **2009**. Multiple origins promote the ecological amplitude of allopolyploid *Aegilops* (Poaceae). *American Journal of Botany* **96**: 1262–1273.

Middleton CP, Senerchia N, Stein N, Akhunov ED, Keller B, Wicker T, Kilian B. **2014**. Sequencing of Chloroplast Genomes from Wheat, Barley, Rye and Their Relatives Provides a Detailed Insight into the Evolution of the Triticeae Tribe (MW Davey, Ed.). *PLoS ONE* **9**: e85761.

Orr-Weaver TL. **2015**. When Bigger Is Better: The Role of Polyploidy in Organogenesis. *Trends in genetics : TIG* **31**: 307–315.

Otto SP. **2007**. The evolutionary consequences of polyploidy. *Cell* **131**: 452–462.

Otto SP, Whitton J. **2000**. Polyploid Incidence and Evolution. *Annual Review of Genetics* **34**: 401–437.

Ozkan H, Levy AA, Feldman M. **2001**. Allopolyploidy-induced rapid genome evolution in the wheat (*Aegilops–Triticum*) group. *The Plant Cell* **13**: 1735–1748.

Pandit MK, Pocock MJO, Kunin WE. **2011**. Ploidy influences rarity and invasiveness in plants. *Journal of Ecology* **99**: 1108–1115.

Parisod C, Senerchia N. **2012**. Responses of transposable elements to polyploidy. In: Grandbastien M-A, Casacuberta JM, eds. Plant Transposable Elements. Berlin, Heidelberg: Springer Berlin Heidelberg, 147–168.

Petersen G, Seberg O, Yde M, Berthelsen K. **2006**. Phylogenetic relationships of *Triticum* and *Aegilops* and evidence for the origin of the A, B, and D genomes of common wheat (*Triticum aestivum*). *Molecular Phylogenetics and Evolution* **39**: 70–82.

Petit M, Guidat C, Daniel J, Denis E, Montoriol E, Bui QT, Lim KY, Kovarik A, Leitch AR, Grandbastien M-A, *et al.* 2010. Mobilization of retrotransposons in synthetic allotetraploid tobacco. *New Phytologist* **186**: 135–147.

Pires ND, Bemer M, Müller LM, Baroux C, Spillane C, Grossniklaus U. **2016**. Quantitative Genetics Identifies Cryptic Genetic Variation Involved in the Paternal Regulation of Seed Development (G Gibson, Ed.). *PLOS Genetics* **12**: e1005806.

Ramsey J, Ramsey TS. **2014**. Ecological studies of polyploidy in the 100 years following its discovery. *Philosophical Transactions of the Royal Society B: Biological Sciences* **369**: 20130352–20130352.

Ramsey J, Schemske DW. **2002**. Neopolyploidy in Flowering Plants. *Annual Review of Ecology and Systematics* **33**: 589–639.

Sabot F, Schulman AH. **2009**. Genomics of transposable elements in the Triticeae. In: Muehlbauer GJ, Feuillet C, eds. Genetics and Genomics of the Triticeae. New York, NY: Springer US, 387–405.

Sandve SR, Marcussen T, Mayer K, Jakobsen KS, Heier L, Steuernagel B, Wulff BBH, Olsen OA. **2015**. Chloroplast phylogeny of *Triticum/Aegilops* species is not incongruent with an ancient homoploid hybrid origin of the ancestor of the bread wheat D-genome. *New Phytologist* **208**: 9–10.

Sémon M, Wolfe KH. **2007**. Consequences of genome duplication. *Current Opinion in Genetics & Development* **17**: 505–512.

Senerchia N, Felber F, North B, Sarr A, Guadagnuolo R, Parisod C. **2016**. Differential introgression and reorganization of retrotransposons in hybrid zones between wild wheats. *Molecular Ecology* **25**: 2518–2528.

Senerchia N, Felber F, Parisod C. **2014**. Contrasting evolutionary trajectories of multiple retrotransposons following independent allopolyploidy in wild wheats. *New Phytologist* **202**: 975–985.

Senerchia N, Felber F, Parisod C. **2015**. Genome reorganization in F1 hybrids uncovers the role of retrotransposons in reproductive isolation. *Proceedings of the Royal Society B: Biological Sciences* **282**: 20142874–20142874.

Shaked H, Kashkush K, Ozkan H, Feldman M, Levy AA. **2001**. Sequence elimination and cytosine methylation are rapid and reproducible responses of the genome to wide hybridization and allopolyploidy in wheat. *The Plant Cell* **13**: 1749–1760.

van Slageren MWSJM. **1994**. *Wild wheats: a monograph of* Aegilops *L. and* Amblyopyrum *(Jaub. & Spach) Eig (Poaceae)*. Wageningen, The Netherlands : Aleppo, Syria: Wageningen Agricultural University : International Center for Agricultural Research in the Dry Areas.

Soltis DE, Burleigh JG. **2009**. Surviving the K-T mass extinction: New perspectives of polyploidization in angiosperms. *Proceedings of the National Academy of Sciences of the United States of America* **106**: 5455–5456.

Soltis DE, Soltis PS. **1999**. Polyploidy: recurrent formation and genome evolution. *Trends in Ecology & Evolution* **14**: 348–352.

Soltis PS, Soltis DE. **2009**. The role of hybridization in plant speciation. *Annual Review of Plant Biology* **60**: 561–588.

Soltis PS, Soltis DE. **2016**. Ancient WGD events as drivers of key innovations in angiosperms. *Current Opinion in Plant Biology* **30**: 159–165.

Soltis DE, Soltis PS, Tate JA. **2004**. Advances in the study of polyploidy since plant speciation. *New Phytologist* **161**: 173–191.

Soltis DE, Visger CJ, Marchant DB, Soltis PS. **2016**. Polyploidy: Pitfalls and paths to a paradigm. *American Journal of Botany* **103**: 1146–1166.

Soltis DE, Visger CJ, Soltis PS. **2014**. The polyploidy revolution then...and now: Stebbins revisited. *American Journal of Botany* **101**: 1057–1078.

Stebbins GL. **1940**. The Significance of Polyploidy in Plant Evolution. *The American Naturalist* **74**: 54–66.

Stebbins GL. **1950**. *Variation and evolution in plants*. New York, NY.

Stebbins GL. **1971**. Chromosomal evolution in higher plants. *Chromosomal evolution in higher plants.*

Szitenberg A, Cha S, Opperman CH, Bird DM, Blaxter ML, Lunt DH. **2016**. Genetic drift, not life history or RNAi, determine long-term evolution of transposable elements. *Genome Biology and Evolution* **8**: 2964–2978.

Tayalé A, Parisod C. **2013**. Natural pathways to polyploidy in plants and consequences for genome reorganization. *Cytogenetic and Genome Research* **140**: 79–96.

Tenaillon MI, Hollister JD, Gaut BS. **2010**. A triptych of the evolution of plant transposable elements. *Trends in Plant Science* **15**: 471–478.

The International Wheat Genome Sequencing Consortium (IWGSC), Mayer KFX, Rogers J, Dole el J, Pozniak C, Eversole K, Feuillet C, Gill B, Friebe B, Lukaszewski AJ, *et al.* 2014. A chromosome-based draft sequence of the hexaploid bread wheat (Triticum aestivum) genome. *Science* **345**: 1251788–1251788.

Van de Peer Y, Mizrachi E, Marchal K. **2017**. The evolutionary significance of polyploidy. *Nature Reviews Genetics* **18**: 411–424.

Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, Flavell A, Leroy P, Morgante M, Panaud O, *et al.* 2007. A unified classification system for eukaryotic transposable elements. *Nature Reviews Genetics* **8**: 973–982.

Wood TE, Takebayashi N, Barker MS, Mayrose I, Greenspoon PB, Rieseberg LH. **2009**. The frequency of polyploid speciation in vascular plants. *Proceedings of the National Academy of Sciences* **106**: 13875–13879.

Zohary D, Feldman M. **1962**. Hybridization between amphidiploids and the evolution of polyploids in the wheat (*Aegilops-Triticum*) group. *Evolution* **16**: 44.

Chapter 1

Reticulate evolution of wheat relatives (*Aegilops-Triticum* group)

Stella Huynh[1], Thomas Marcussen[2], François Felber[1,3] and Christian Parisod[4]

[1]Institute of Biology, University of Neuchâtel, Switzerland

[2]Center for Ecological and Evolutionary Synthesis, University of Oslo, Norway

[3]Musée et Jardins botaniques cantonaux de Lausanne et Pont-de-Nant, Switzerland

[4]Institute of Plant Sciences, University of Bern, Switzerland

List of figures and tables

Abstract

The evolutionary relationships among wheats and relatives (*Aegilops* and *Triticum* species) have been extensively investigated, but phylogenetic studies have reported ambiguous insights and the diploid origin of wheats' B subgenome, in particular, remains elusive. Recent studies have pointed to a complex history of reticulated evolution in the Triticeae tribe that must be resolved to understand the evolutionary history of this plant group. Inference of species phylogenies from nuclear loci is challenging in the presence of reticulation, but we propose here to combine high-throughput genotyping of 30 nuclear low-copy loci on a representative sampling of 13 species to reconstruct the evolution of diploids in the *Aegilops-Triticum* group. Distance-based splits network and reticulate networks under the multi-species coalescent model clearly highlighted a hybridization event at the origin of a clade which included not only *Ae. tauschii* (the donor of the D subgenome of bread wheat), but also several selfing species with divergent genomes. Decomposition of phylogenetic signals of each of the A, B and D clades confirmed such a reticulated scenario with homoploid hybridization between a progenitor close to the *Triticum* donor of the A subgenome of wheats and a lineage related to *Ae. speltoides* at the origin of the majority of diploid relatives of wheats.

Keywords: Triticeae, reticulate species network, amplicon sequencing

Introduction

The grass tribe Triticeae comprises some of the world's most important cereal crops and especially the wheat group. The wild relatives of cultivated wheats include *Aegilops* and *Triticum* species that are annual grasses naturally occuring across the Mediterranean region up to Central Asia (Eig, 1929; van Slageren, 1994; Kilian *et al.*, 2011) and presenting multiple ploidy levels from diploids (2x = 14) to hexaploids (6x = 42). The phylogenetic relationships among wild wheats have long been studied (Eig, 1929; Kihara, 1975), with a special emphasis on the diploid progenitors of the hexaploid bread wheat *T. aestivum* (genome composition: BBAADD). Although the diploids *Ae. tauschii* (DD) and *T. monococcum* (AA) are commonly considered as progenitors of cultived wheat, having contributed to the D and A subgenomes respectively, the origin of the B genome remains debated. The diploid *Ae. speltoides* (SS), the only outcrossing species of the group with *Ae. mutica* (TT), has been suggested to be closely related to the donor of the B genome of wheat (Daud & Gustafson, 1996; Petersen *et al.*, 2006). Despite numerous phylogenetic studies attempting at shedding light on the evolution of wheats and relatives (e.g. Petersen *et al.*, 2006; Marcussen *et al.*, 2014; Baidouri *et al.*, 2017; Bernhardt *et al.*, 2017), no consensus has yet been reached (Li *et al.*, 2015b,a; Sandve *et al.*, 2015).

The evolutionary history of the *Aegilops* and *Triticum* group revealed indeed complex pattern of divergence, punctuated by recurrent homoploid and polyploid hybridization events (Zohary & Feldman, 1962; Marcussen *et al.*, 2014; Li *et al.*, 2015b) and introgression (Schoenenberger *et al.*, 2005; Parisod *et al.*, 2013; Senerchia *et al.*, 2016). Despite such a highly reticulated evolutionary history, numerous studies focused on phylogenies of maternal lineages using plastid markers and thus overlooked hybridization (Yamane & Kawahara, 2005; Dizkirici *et al.*, 2013; Gornicki *et al.*, 2014; Bernhardt *et al.*, 2017). Phylogenies based on nuclear DNA remains scarce in the wheat group and often used only few loci (e.g. Petersen *et al.*, 2006) or a limited sampling of taxa. In particular, Marcussen *et al.* (2014) used hundreds of nuclear loci, but specifically focused on the origin of *T. aestivum* and only included three candidate progenitors. Although this study unambiguously highlighted a hybrid origin of *Ae. tauschii*, sampling restricted to a few species may have distorted the complex evolutionary history of wheats. Li *et al.* (2015a) indeed suggested that *Aegilops* and *Triticum* species may have undergone multiple rounds of ancient and recent hybridization. However, a comprehensive nuclear-based investigation of reticulate evolution across the *Aegilops-Triticum* group supporting this hypothesis is still lacking.

Here, we developped an approach using high-throughput genotyping of multiple low-copy nuclear loci combined with specific statistical framework to infer the relationships among diploid representatives of the *Aegilops-Triticum* group while considering reticulate evolution. Using 30 low-copy nuclear loci genotyped on 37 accessions from 13 diploid *Aegilops-Triticum* species as well as three outgroups, a species network was inferred under the multispecies coalescent (MSC) model and highlighted reticulated nodes coherent with hybridization at the origin of the largest clade of *Aegilops* species (including the bread wheat progenitor *Ae. tauschii*). As processes such as incomplete lineage sorting may produce similar splits, the hybrid origin of this diploid clade was confirmed by iteratively removing one clade when inferring the species phylogenies under the MSC model (i.e. species tree inferred between A-B, A-D or B-D clades).

Material and Methods

Plant material

One to six accessions were selected for each of the 11 diploid *Aegilops* species and 2 diploid *Triticum* species as well as the Triticeae outgroups used here (i.e. Taeniatherum *caput-medusae*, *Agropyron mongolicum* and *Secale cereale*), for a total of 37 accessions (Table S1). The accessions were obtained from ARS-GRIN or ICARDA germplasms and from personal collections. The seeds were sown and grown outdoor at the Botanical garden of Neuchâtel until flowering and deposited as herbarium sheets at the Botanical Museum and Garden of Lausanne and Pont-de-Nant, Switzerland. Leaves from young seedlings were sampled and dried in silica gel for DNA extraction using the DNAeasy kit from Qiagen®, following the manufacturer's protocol. DNA concentration was then measured by Nanodrop and normalized at 50ng/µL for subsequent genotyping of nuclear markers. The samples were randomly distributed in a 96-well plate with one negative control.

Sequencing of biparentally-inherited nuclear markers

A total of 48 candidate low-copy nuclear loci were selected (see Table S2, Supplemental Material) based on available evidence from the literature (Sang, 2002; Duarte *et al.*, 2010). For each locus, sequences from *Aegilops tauschii* (DD) and *Triticum aestivum* (BBAADD) were retrieved from assembled genomes available on EnsemblPlants (Kersey *et al.*, 2014) through either their annotation or the best match following blastN using the corresponding locus from *Triticum aestivum*. Available sequences for each locus were aligned using MAFFT (Katoh, 2002) and the resulting alignments were used to design primers in exons encompassing 400 bp of intronic regions (i.e. one long or two to three short introns) using Primer3 (Koressaar & Remm, 2007; Untergasser *et al.*, 2012). All primer pairs were designed to achieve PCR amplification under similar conditions with a Tm around 60°C and tagged with a 22-bp sequence designed to be used with the Access Array barcode library for Illumina sequencers provided by Fluidigm.

Simultaneous PCR amplifications of all 48 loci were performed on the 37 samples in the 48.48 Access Array Integrative Fluidic Circuit (IFC) chip following Fluidigm instructions. Two replicates (i.e. 5.1 % of the whole dataset) and one negative were included, for a total of 40 samples processed. The IFC chip mixed the 40 samples and 48 primer pairs in all 1920 possible combinations for PCR amplification into independent nanochambers before pooling resulting amplicons for each sample. Each sample pool was barcoded with a specific 10 bp-

sequence attached at one side of the amplicons through short PCR (7 cycles). Barcoded sample pools were cleaned from remaining primers using AMPure beads (Agilent) at x0.7 ratio and their quality checked using High-Sensitive chips on the BioAnalyzer 2100 (Agilent). All sample pools were quantified twice through real-time PCRs using the LightCycler System in 384-well plates with four standards and their absolute concentration estimated based on the shift of their melting curve from the closest standard (here, the 0.2 nM standard). All samples were finally pooled at equimolar concentrations by individual pipetting, and this final pool was cleaned with AMPure beads and checked for quality with BioAnalyzer 2100. The concentration of this final pool was estimated through qPCR with four replicates and normalized at 2 nmol before loading into the Illumina MiSeq.

Sequencing of amplicons was performed using the MiSeq Reagent Kit v3 for paired-ends reads sequencing (2 x 300 bp). As recommended by Fluidigm, two specific sequencing primers containing custom Locked Nucleic Acid (LNA^{TM}) oligonucleotides from Exiqon were used at 10 µM to enhance the specificity and sensitivity of the sequencing. Sequencing quality was additionally ensured by adding PhiX Control v3 sequencing library from Illumina.

Genotyping of nuclear markers

Paired-end reads were merged using the BBmerge script in BBTools (http://jgi.doe.gov/data-and-tools/bbtools/) to correct for dropping quality towards the end of reads as targeted introns were 400 bp length and paired reads thus overlapped over 200 bp to accurately infer haplotypes. Merged reads were demultiplexed per sample based on barcode indexes without mismatch as well as per locus based on the sequence of primers allowing for 1 mismatch. After filtering by their mean Quality Control with a threshold of 20, reads were locally aligned by sample and locus using Muscle (Edgar, 2004).

Resulting alignments were cleaned with a custom R script based on the frequency of nucleotides, gaps and unknown base at each position. Variants present in less than 10% of the reads were removed. Remaining allelic variants characterized reads into haplotypes, among which only those present in at least 10% of the haplotypes were kept for further analyses. Individual haplotypes were then grouped per locus and locally aligned using MAFFT (Katoh, 2002). These alignments were further manually cleaned for chimeras, null alleles and repetitive regions by visual inspection. Loci that could not be confidently cleaned due to poor coverage were discarded and alignments of remaining loci were iteratively refined using SATé (Liu *et al.*, 2009) with the following settings for external tools: MAFFT as Aligner, Muscle as Merger, RAxML (Stamatakis, 2014) as Tree Estimator, GTRGAMMAI as

nucleotide substitution model, and for SATé-specific settings: SATé-II-ML, Centroid decomposition and "best alignment after last improvement". For each locus, diploid accessions for which only one haplotype was amplified were coded as homozygous (AA) whereas those having two haplotypes identified were coded as heterozygous (AB) for that locus.

To assess the origin of the bread wheat genome, corresponding sequences were extracted for each locus from the reference genome of *T. aestivum* (IWGSC1 assembly; The International Wheat Genome Sequencing Consortium (IWGSC) *et al.*, 2014) available on EnsemblPlants (ftp://ftp.ensemblgenomes.org/pub/plants/pre/fasta/triticum_aestivum/dna/). Additional outgroup sequences from *Brachypodium distachyon* were similarly extracted from the genome reference available (v1.0 INSDC assembly; The International Brachypodium Initiative, 2010) on EnsemblPlants website (ftp://ftp.ensemblgenomes.org/pub/plants/release-39/fasta/brachypodium_distachyon/dna/).

Phylogenetic analyses

The evolutionary relationships among taxa of the *Aegilops-Triticum* group were first assessed using the NeighborNet method (Bryant & Moulton, 2002) based on the number of pairwise SNP differences (i.e. uncorrected P) between concatenated sequences of the whole dataset and visualized in SplitsTree4 (Huson, 1998). When *T. aestivum* was included in the analyses, each of its subgenomes was considered as separate diploid 'individual' and the split network was similarly inferred.

To characterize the reticulation signal, a nuclear species phylogeny was inferred from a single accession per species using the MSC model in PhyloNet (Than *et al.*, 2008; Wen *et al.*, 2016). Inference of the number of reticulation events was computationally intensive, even with one accession per species, and only approached convergence.

Reticulated phylogeny was thus supported through an alternative strategy that decomposed the phylogenetic signal of each clade A, B and D. Accordingly, phylogenetic inferences first included all three clades and then iteratively removed one clade at a time (i.e. on dataset including either A & B clades, A & D clades or B & D clades). These four species phylogenies were all assessed under the MSC model using STACEY (Jones *et al.* 2014), a Bayesian method that infers species tree based on the gene trees, and implemented in BEAST2 v2.4.7 (Bouckaert *et al.* 2014). The species trees were inferred by considering each individual as a species and its haplotypes as the sampled species. A Jukes-Cantor (JC) model

of nucleotide substitution and a strict molecular clock for all loci was used to run the analysis for 500 million Markov chain Monte Carlo (MCMC) generations, with additional 1 million generations as burnin and tree sampling every 200 000 generations (i.e. 2500 MCMC trees). All phylogenetic analyses were checked for convergence, i.e. good mixing of MCMC search is ensured for parameters with an effective sample size (ESS) value higher than 200, on Tracer v1.7 (Rambaut *et al.*, 2018). Maximum Credibility Clade (MCC) trees were summarized by TreeAnnotator available in BEAST2 with a burnin percentage of 60% and viewed in FigTree v1. (http://tree.bio.ed.ac.uk/software/figtree/).

Node ages were compared between the four phylogenies by fixing the second most basal node (i.e. split from *S. cereale*) at 1 to ensure that different taxon samplings do not affect estimations of substitution rates and branch lengths. Relative node ages were further converted to million years ago (Mya) based on the relative time scale of the divergence between the *Aegilops-Triticum* species and *S. cereale*, provided that *S. cereale* diverged about 4.5 Mya (Bernhardt *et al.*, 2017). Such decomposition of the phylogenetic signal among the three clades enabled to compare the relative influence of each clade to the species tree topology and node ages. *Ae. mutica* was excluded from these inferences due to conflicting topologies leading to low node supports when it was included. Similar decomposition of the phylogenetic signal indeed showed *Ae. mutica* outside of the A, B and D clades, with a most recent divergence from clade A (Supporting results S1).

Results

Genotyping of nuclear markers

Amplicon sequencing of the 48 nuclear loci among the 37 accessions yielded about 3.9 millions of Illumina MiSeq reads that were stringently filtered and trimmed to 1.7 million reads. This dataset is nested within an extended one including additional accessions and polyploid species, which was used to filter out supposedly spurious loci based on low coverage, unspecific amplifications and/or detection of paralogs (details in Chapter 2). Provided that paralogs amplified despite specific primers are thoroughly filtered out, simultaneous PCRs of the 48.48 Access Array technology proved efficient and yielded 30 robust loci that were confidently used for subsequent analyses.

Species phylogeny inferences

The splits network presenting relationships among 37 samples of the *Aegilops-Triticum* group is coherent with the complex reticulate evolution of the group (Figure 1). The two *Triticum* species clustered together and formed the clade A, whereas *Ae. speltoides* (genome S) formed the clade B and was more closely related to *Ae. mutica* (genome T). All the other *Aegilops* species with genomes D, U, M, C, N and S* presented a common, reticulated node and were therefore included in an inclusive D clade containing *Aegilops tauschii* (genome D). The splits network including sequences from *T. aestivum* showed each of the three subgenomes A, B and D consistently branching with *T.urartu* in clade A, *Ae. speltoides* in clade B and *Ae. tauschii* in clade D respectively (Supplemental Material Figure S1).

The MSC phylogeny inferred with PhyloNet barely converged toward a network that was coherent with a homoploid hybrid origin of clade D, and in particular of *Ae. tauschii* (Figure 2). Although the clade A unambiguously contributed to the hybrid origin of clade D, the other progenitor was indirectly connected to clade B and appeared related to a more ancient common ancestor in the Triticeae tribe.

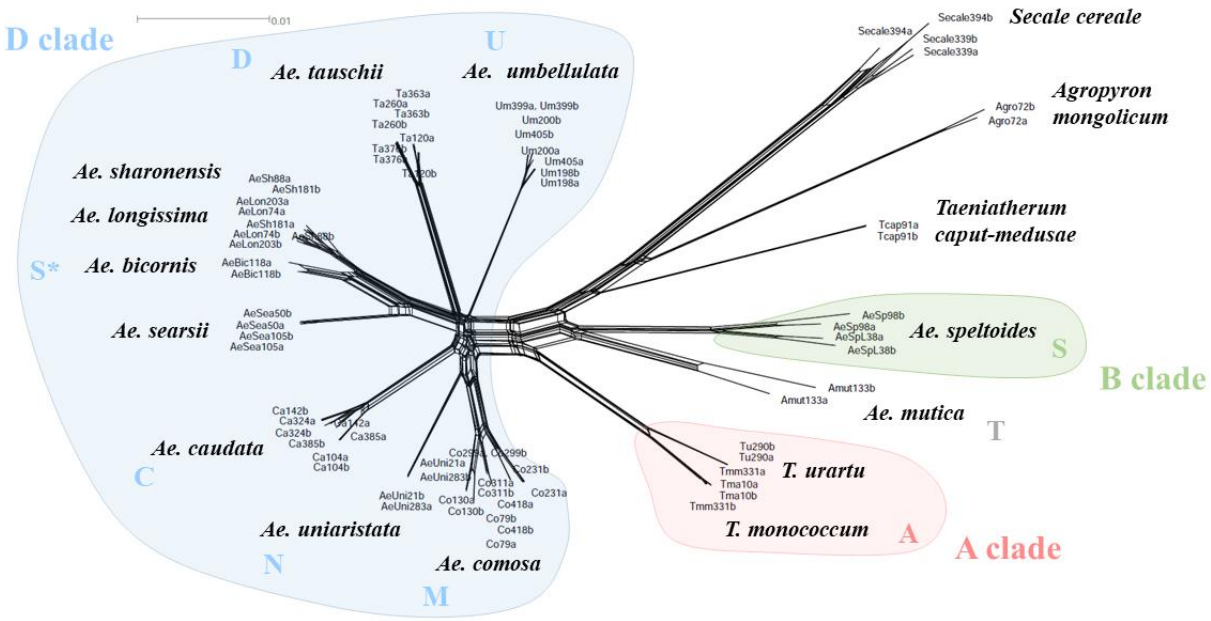


Figure 1: Reticulate network of diploid *Aegilops-Triticum* species. The network was based on 30 nuclear loci concatenated without phasing. Each diploid individual displayed two sequences accounting for individual heterozygosity. Species names are indicated next to corresponding accessions along with their genome composition, following van Slageren (1994). S* refer to all variants of the S genome. Species belonging to the A, B or D clade are differentially colored.

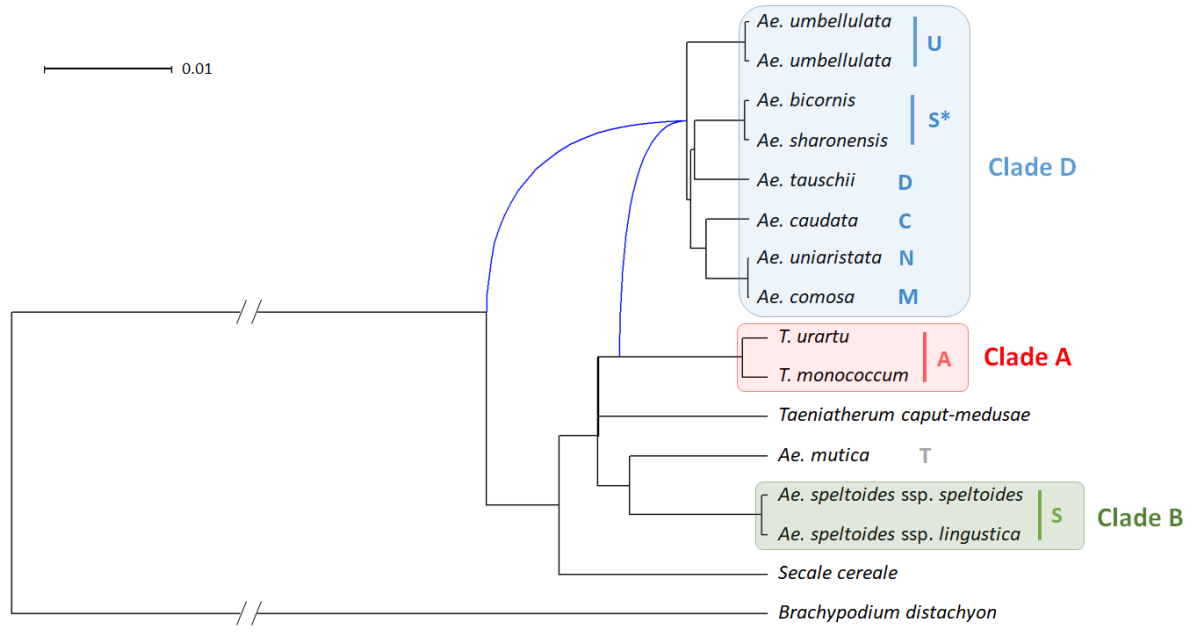


Figure 2: Reticulate species network of diploid of the *Aegilops-Triticum* clade. The species network was inferred from 30 nuclear low-copy loci on one accession per species using PhyloNet. *Brachypodium distachyon* was included as outgroup. The reticulation is shown by the blue branches. The genome composition of the species is indicated following van Slageren (1994), with S* referring to all variants of the S genome. Species belonging to the A, B or D clade are differentially colored.

Species phylogenies under the MSC model using STACEY converged rapidly and showed a polytomy at the base of all three clades (Figure 3). This pattern, coherent with reticulation events between these clades, was also apparent when including sequences from wheat subgenomes (Supplemental Material, Figure S2). The decomposition of phylogenetic signals from specific clades assessed that clades A and B diverged more deeply when the D clade was excluded (Figure 3b). These clades further diverged earlier than clades B and D (Figure 3c) or clades A and D (Figure 3d). The mean age of the A/B divergence indeed was dated at 0.92 relative to the divergence with *S. cereale* (i.e. about 4.14 Mya), whereas the age of divergence was estimated for B/D around 1.49 Mya and for A/D around 2.49 Mya (Table 1). The divergence age between clades B and D was the closest to the age of the polytomy between the three clades (1.76 Mya), supporting a homoploid hybrid origin of the D clade involving a representative of the B clade as genome progenitor.

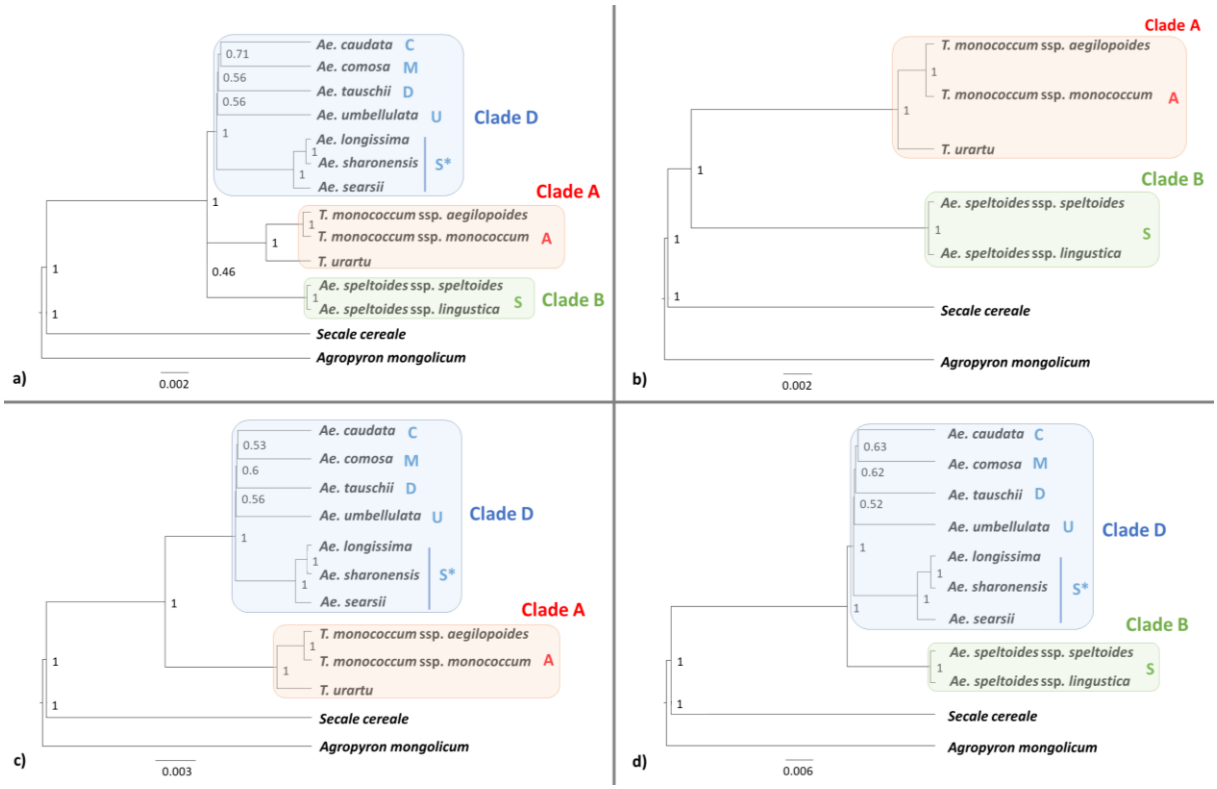


Figure 3: Decomposition of phylogenetic signals from A, B and D clades among *Aegilops* – *Triticum* species. Species phylogenies were inferred (a) on the three clades together, (b) on A and B clades only, (c) on B and D clades only, and (d) on A and D clades only. All species trees were inferred based on 30 low-copy nuclear loci with one accession per species using STACEY. Node supports and relative timescale are indicated. The genome composition of the species is indicated following van Slageren (1994), with S* referring to all variants of the S genome. Species belonging to the A, B or D clades are highlighted within light red, green or blue rectangles respectively.

Table 1: Relative ages of the most recent common ancestor of the A, B and D clades. Node ages are given relative to the divergence with *Secale cereale* and in million years ago (Mya), provided that *S. cereale* diverged about 4.5 Mya according to the plastid phylogeny of Bernhardt *et al.* (2017). The means and 95% credibility intervals (CI) are assessed based on a sampling of 1000 of the inferred Bayesian trees for the divergence of A/B/D, A/B, A/D and B/D clades.

		Divergence nodes			
		ABD	**AB**	**AD**	**BD**
Relative ages	Mean	0.39	0.92	0.55	0.33
	CI	0.33-0.49	0.75-0.99	0.33-0.68	0.28-0.40
Absolute ages (Mya)	Mean	1.76	4.14	2.48	1.49
	CI	1.49-2.21	3.38-4.46	1.49-3.06	1.26-1.80

Discussion

Decomposing the species phylogeny to detect reticulation signal

The splits network presented here based on a representative sampling of 30 nuclear loci sequenced on 37 accessions across the *Aegilops-Triticum* confirmed a complex reticulate evolution of the kind hypothesized by Li *et al.* (2015b) for this group. All *Aegilops* species with genomes M, C, D, U, N and S* grouped at the intersection of splits connecting clades A and B, matching the pattern expected under the hypothesis of a homoploid hybrid origin of such an inclusive clade D. This study thus expends the conclusions of Marcussen *et al* (2014) that *Ae. tauschii* is of hybrid origin to an entire clade comprising the majority of *Aegilops* species. The network is further coherent with clades A and B being progenitors of this D clade, followed by radiation into the current species with their differentially organized genomes. Such a postulated scenario was further substantiated by the decomposition of phylogenetic signals underlying polytomies commonly reported in the overall species trees (Figure 3a). Clades A and B indeed diverged first about 4.14 Mya, followed by the recent divergence of clade D from both clades A and B likely occurring around 1.49 Mya.

Those results also confirmed prior inferences (Petersen *et al.*, 2006; Wang *et al.*, 2013b; Marcussen *et al.*, 2014) that subgenomes of *T. aestivum* have been contributed by *T. urartu* (A subgenome) and *Ae. tauschii* (D subgenome), and a relative of *Ae. speltoides* (B subgenome). In particular, as previously postulated (Dvořák, 2009; Mizuno *et al.*, 2010), specific accessions of *Ae. tauschii* subsp. *strangulata* located along the coast of the Caspian Sea appear closely related to cultivated bread wheat.

The role of *Ae. speltoides* in the origin of the B subgenome of *T. aestivum* and more generally on the origin of the hybrid clade D remains controversial. All species phylogenies indeed indicated *Aegilops* as a highly polyphyletic group when this early-diverging species (Bernhardt *et al.* 2017) was included. Provided our large sampling of species included in the analyses, the still elusive identification of the progenitor of this clade as a relative of *Ae. speltoides* suggests this diploid taxon either remains to be discovered or went extinct, as also previously suspected (Provan *et al.*, 2004). Here, the two subspecies of *Ae. speltoides* appeared to have diverged particularly deeply as compared to taxa within other subspecies, consistent with a faster accumulation of mutations in this lineage and thus favoring extinction of the progenitor of the B subgenome of wheat as a likely scenario. As opposed to the mainly selfing *Aegilops* species, *Ae. speltoides* and the other taxon yielding conflicting topologies

(*Ae. mutica*; Supporting result S1) are mainly outcrossers, suggesting that differences in mating systems might add noise to inferences of species phylogenies by affecting the effective population sizes (Ne) of species. The impact of variable mating system on our ability to infer species trees under scenarios of reticulate evolution remains largely to be explored and, in the case of the wheat group, might explain the enduring ambiguities regarding the origin of the B subgenome of bread wheat. Appropriate methodologies allowing Ne to change along branches in response to changes in mating systems or hybridization events is currently developing, with the aim to improve phylogenetic inferences in clades with porous species boundaries such as wheats and their relatives.

References

Baidouri ME, Murat F, Veyssiere M, Molinier M, Flores R, Burlot L, Alaux M, Quesneville H, Pont C, Salse J. **2017**. Reconciling the evolutionary origin of bread wheat (*Triticum aestivum*). *New Phytologist* **213**: 1477–1486.

Bernhardt N, Brassac J, Kilian B, Blattner FR. **2017**. Dated tribe-wide whole chloroplast genome phylogeny indicates recurrent hybridizations within Triticeae. *BMC Evolutionary Biology* **17**:141.

Bryant D, Moulton V. **2002**. NeighborNet: An Agglomerative Method for the Construction of Planar Phylogenetic Networks. In: Guigó R, Gusfield D, eds. Algorithms in Bioinformatics. Berlin, Heidelberg: Springer Berlin Heidelberg, 375–391.

Daud HM, Gustafson JP. **1996**. Molecular evidence for Triticum speltoides as a B-genome progenitor of wheat (Triticum aestivum). *Genome* **39**: 543–548.

Dizkirici A, Kansu C, Onde S, Birsin M, Özgen M, Kaya Z. **2013**. Phylogenetic relationships among *Triticum* L. and *Aegilops* L. species as genome progenitors of bread wheat based on sequence diversity in trnT-F region of chloroplast DNA. *Genetic Resources and Crop Evolution* **60**: 2227–2240.

Duarte JM, Wall PK, Edger PP, Landherr LL, Ma H, Pires JC, Leebens-Mack J, dePamphilis CW. **2010**. Identification of shared single copy nuclear genes in *Arabidopsis*, *Populus*, *Vitis* and *Oryza* and their phylogenetic utility across various taxonomic levels. *BMC Evolutionary Biology* **10**: 61.

Dvořák J. **2009**. Triticeae Genome Structure and Evolution. In: Muehlbauer GJ, Feuillet C, eds. Genetics and Genomics of the Triticeae. New York, NY: Springer US, 685–711.

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

Eig A. **1929**. *Monographisch-kritische Uebersicht der Gattung Aegilops*. Verlag des Repertoriums.

Gornicki P, Zhu H, Wang J, Challa GS, Zhang Z, Gill BS, Li W. **2014**. The chloroplast view of the evolution of polyploid wheat. *New Phytologist* **204**: 704–714.

Huson DH. **1998**. SplitsTree: analyzing and visualizing evolutionary data. *Bioinformatics* **14**: 68–73.

Katoh K. **2002**. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research* **30**: 3059–3066.

Kersey PJ, Allen JE, Christensen M, Davis P, Falin LJ, Grabmueller C, Hughes DST, Humphrey J, Kerhornou A, Khobova J, *et al.* 2014. Ensembl Genomes 2013: scaling up access to genome-wide data. *Nucleic Acids Research* **42**: D546–D552.

Kihara H. **1975**. Origin of cultivated plants with special reference to wheat. *Seiken Ziho* **25**: 1–24.

Kilian B, Mammen K, Millet E, Sharma R, Graner A, Salamini F, Hammer K, Özkan H. **2011**. Aegilops. In: Kole C, ed. Wild Crop Relatives: Genomic and Breeding Resources. Berlin, Heidelberg: Springer Berlin Heidelberg, 1–76.

Koressaar T, Remm M. **2007**. Enhancements and modifications of primer design program Primer3. *Bioinformatics* **23**: 1289–1291.

Li L-F, Liu B, Olsen KM, Wendel JF. **2015a**. Multiple rounds of ancient and recent hybridizations have occurred within the Aegilops–Triticum complex. *New Phytologist* **208**: 11–12.

Li L-F, Liu B, Olsen KM, Wendel JF. **2015b**. A re-evaluation of the homoploid hybrid origin of *Aegilops tauschii* , the donor of the wheat D-subgenome. *New Phytologist* **208**: 4–8.

Liu K, Raghavan S, Nelesen S, Linder CR, Warnow T. **2009**. Rapid and Accurate Large-Scale Coestimation of Sequence Alignments and Phylogenetic Trees. *Science* **324**: 1561–1564.

Marcussen T, Sandve SR, Heier L, Spannagl M, Pfeifer M, The International Wheat Genome Sequencing Consortium, Jakobsen KS, Wulff BBH, Steuernagel B, Mayer KFX, *et al.* 2014. Ancient hybridizations among the ancestral genomes of bread wheat. *Science* **345**: 1250092–1250092.

Mizuno N, Hosogi N, Park P, Takumi S. **2010**. Hypersensitive Response-Like Reaction Is Associated with Hybrid Necrosis in Interspecific Crosses between Tetraploid Wheat and *Aegilops tauschii* Coss (E Newbigin, Ed.). *PLoS ONE* **5**: e11326.

Parisod C, Definod C, Sarr A, Arrigo N, Felber F. **2013**. Genome-specific introgression between wheat and its wild relative *Aegilops triuncialis*. *Journal of Evolutionary Biology* **26**: 223–228.

Petersen G, Seberg O, Yde M, Berthelsen K. **2006**. Phylogenetic relationships of *Triticum* and *Aegilops* and evidence for the origin of the A, B, and D genomes of common wheat (*Triticum aestivum*). *Molecular Phylogenetics and Evolution* **39**: 70–82.

Provan J, Wolters P, Caldwell KH, Powell W. **2004**. High-resolution organellar genome analysis of *Triticum* and *Aegilops* sheds new light on cytoplasm evolution in wheat. *TAG Theoretical and Applied Genetics* **108**: 1182–1190.

Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. **2018**. Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7 (E Susko, Ed.). *Systematic Biology*.

Sandve SR, Marcussen T, Mayer K, Jakobsen KS, Heier L, Steuernagel B, Wulff BBH, Olsen OA. **2015**. Chloroplast phylogeny of *Triticum/Aegilops* species is not incongruent with an ancient homoploid hybrid origin of the ancestor of the bread wheat D-genome. *New Phytologist* **208**: 9–10.

Sang T. **2002**. Utility of Low-Copy Nuclear Gene Sequences in Plant Phylogenetics. *Critical Reviews in Biochemistry and Molecular Biology* **37**: 121–147.

Schoenenberger N, Felber F, Savova-Bianchi D, Guadagnuolo R. **2005**. Introgression of wheat DNA markers from A, B and D genomes in early generation progeny of *Aegilops cylindrica* Host × *Triticum aestivum* L. hybrids. *Theoretical and Applied Genetics* **111**: 1338–1346.

Senerchia N, Felber F, North B, Sarr A, Guadagnuolo R, Parisod C. **2016**. Differential introgression and reorganization of retrotransposons in hybrid zones between wild wheats. *Molecular Ecology* **25**: 2518–2528.

van Slageren MWSJM. **1994**. *Wild wheats: a monograph of* Aegilops *L. and* Amblyopyrum *(Jaub. & Spach) Eig (Poaceae): a revision of all taxa closely related to wheat, excluding wild* Triticum *species, with notes on other genera in the tribe Triticeae, especially* Triticum. Wageningen, The Netherlands : Aleppo, Syria: Wageningen Agricultural University : International Center for Agricultural Research in the Dry Areas.

Stamatakis A. **2014**. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**: 1312–1313.

Than C, Ruths D, Nakhleh L. **2008**. PhyloNet: a software package for analyzing and reconstructing reticulate evolutionary relationships. *BMC Bioinformatics* **9**: 322.

The International Brachypodium Initiative. **2010**. Genome sequencing and analysis of the model grass *Brachypodium distachyon*. *Nature* **463**: 763–768.

The International Wheat Genome Sequencing Consortium (IWGSC), Mayer KFX, Rogers J, Dole el J, Pozniak C, Eversole K, Feuillet C, Gill B, Friebe B, Lukaszewski AJ, *et al.* 2014. A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome. *Science* **345**: 1251788.

Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, Rozen SG. **2012**. Primer3—new capabilities and interfaces. *Nucleic Acids Research* **40**: e115–e115.

Wang J, Luo M-C, Chen Z, You FM, Wei Y, Zheng Y, Dvorak J. **2013**. *Aegilops tauschii* single nucleotide polymorphisms shed light on the origins of wheat D-genome genetic diversity and pinpoint the geographic origin of hexaploid wheat. *New Phytologist* **198**: 925–937.

Wen D, Yu Y, Nakhleh L. **2016**. Bayesian Inference of Reticulate Phylogenies under the Multispecies Network Coalescent (S Edwards, Ed.). *PLOS Genetics* **12**: e1006006.

Yamane K, Kawahara T. **2005**. Intra- and interspecific phylogenetic relationships among diploid *Triticum-Aegilops* species (Poaceae) based on base-pair substitutions, indels, and microsatellites in chloroplast noncoding sequences. *American Journal of Botany* **92**: 1887–1898.

Zohary D, Feldman M. **1962**. Hybridization Between Amphidiploids and the Evolution of Polyploids in the Wheat (*Aegilops-Triticum*) Group. *Evolution* **16**: 44.

Supplementary Material

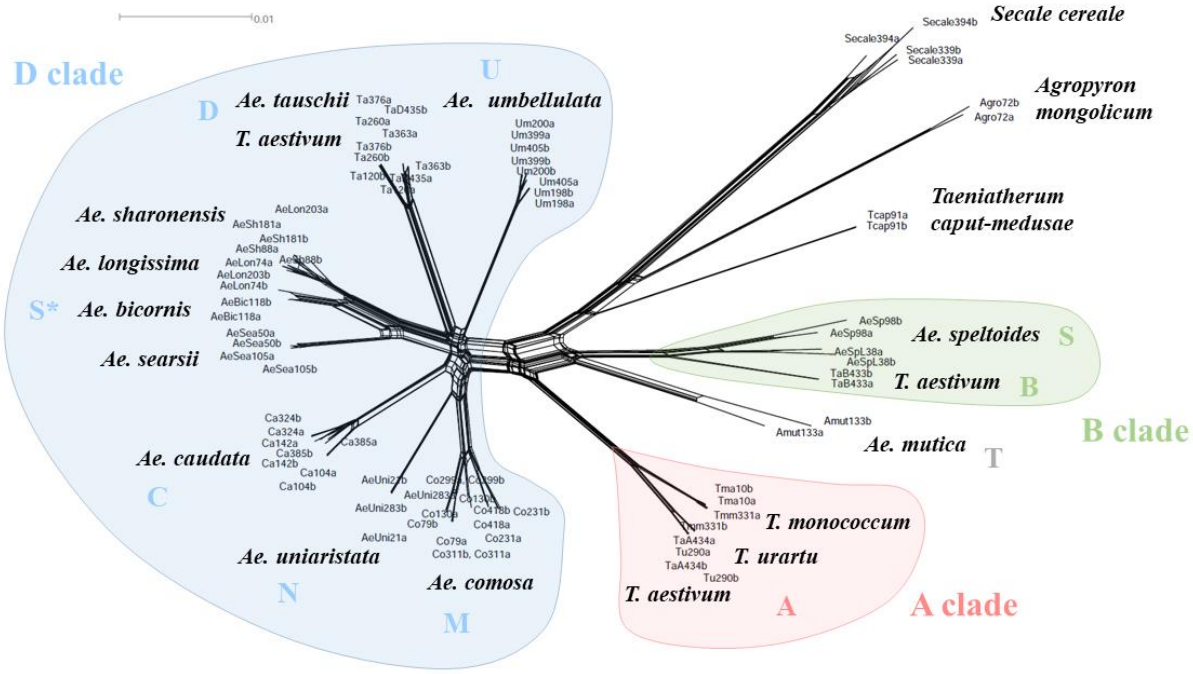


Figure S1: Reticulate network of diploid *Aegilops-Triticum* species and hexaploid bread wheat. The network was based on 30 nuclear loci concatenated without phasing. Each diploid individual displayed two concatenated sequences accounting for individual heterozygosity. The hexaploid bread wheat *T. aestivum* (BBAADD) is included with each of its subgenome treated separately. The genome composition of the species is indicated following van Slageren (1994), with S* referring to all variants of the S genome. Species belonging to the A, B or D clade are grouped by color and indicated on the figure.

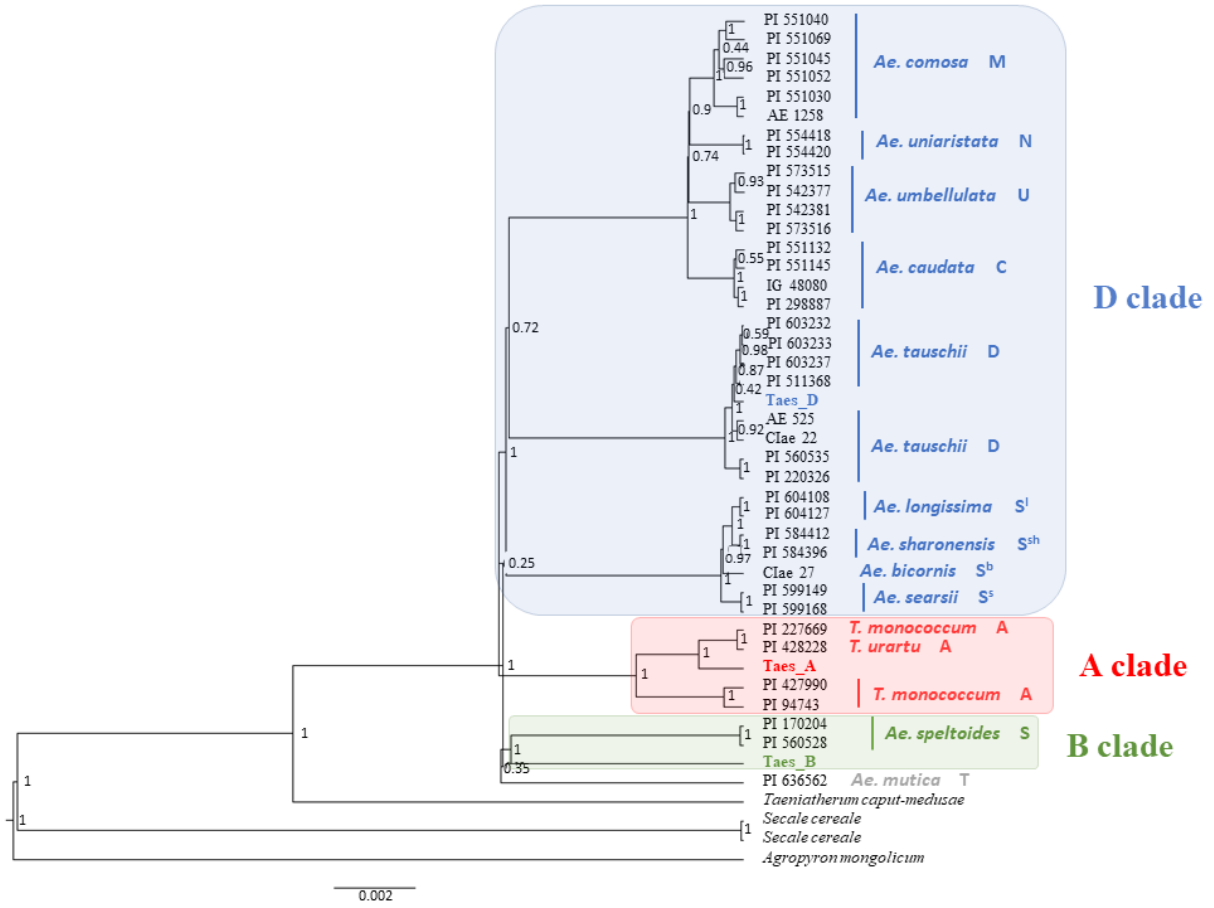


Figure S2: Nuclear phylogeny of diploid *Aegilops* – *Triticum* species. The phylogenetic tree was inferred under the Bayesian model, with JC substitutions and strict molecular clock. In this analysis, each accession was considered as a separate taxon and its genotyped haplotypes as specimens. The summarized tree following 1 billion generations, with a burnin of 100'000 and trees sampled every 200'000 generations, is represented here. Posterior node supports are shown on the tree. The labels report the repository ID of each *Aegilops-Triticum* accession and their genome composition indicated following van Slageren (1994). Each clade A, B or D is highlighted with colored rectangle. The genome reference of *T. aestivum* (The International Wheat Genome Sequencing Consortium (IWGSC) *et al.* 2014) used in the analysis is reported as "Taes" followed by the corresponding subgenome A, B or D and highlighted accordingly.

Supporting results S1: The specific case of Aegilops mutica = Amblyopyrum muticum

The evolutionary history of *Aegilops mutica* and its influence on *Aegilops-Triticum* species phylogenies revealed surprising. This species has been reported to be maternally close to *Ae. umbellulata* (Chapter 2; Bernhardt *et al.*, 2017) whereas the MSC species trees inferred from the nuclear markers genotyped here indicate *Ae. mutica* closer to the clade B (Figure A), as also suggested by previous studies (Petersen *et al* 2009, Bernhardt *et al* 2017). *Ae. mutica* is indeed located between the clades A and B when these two clades are included together in phylogenetic inferences while it has a more basal position when the clades A and B are included separately (Figure B). One hypothesis would be that *Ae. mutica* also has a hybrid origin at the same time as the D clade, which maybe explains why our attempts to infer the reticulation events using PhyloNet resulted in inconsistent topologies. The taxonomic and phylogenetic status of *Ae. mutica* (also named *Amblyopyrum muticum*), considered as sister-species of the *Aegilops* species and the only representative of the T genome, is still debated (van Slageren 1994, Kilian *et al* 2011, Petersen *et al* 2009, Bernhardt *et al* 2017). *Ae. mutica* should thus be further investigated towards resolution of the evolutionary history of the *Aegilops-Triticum* group.

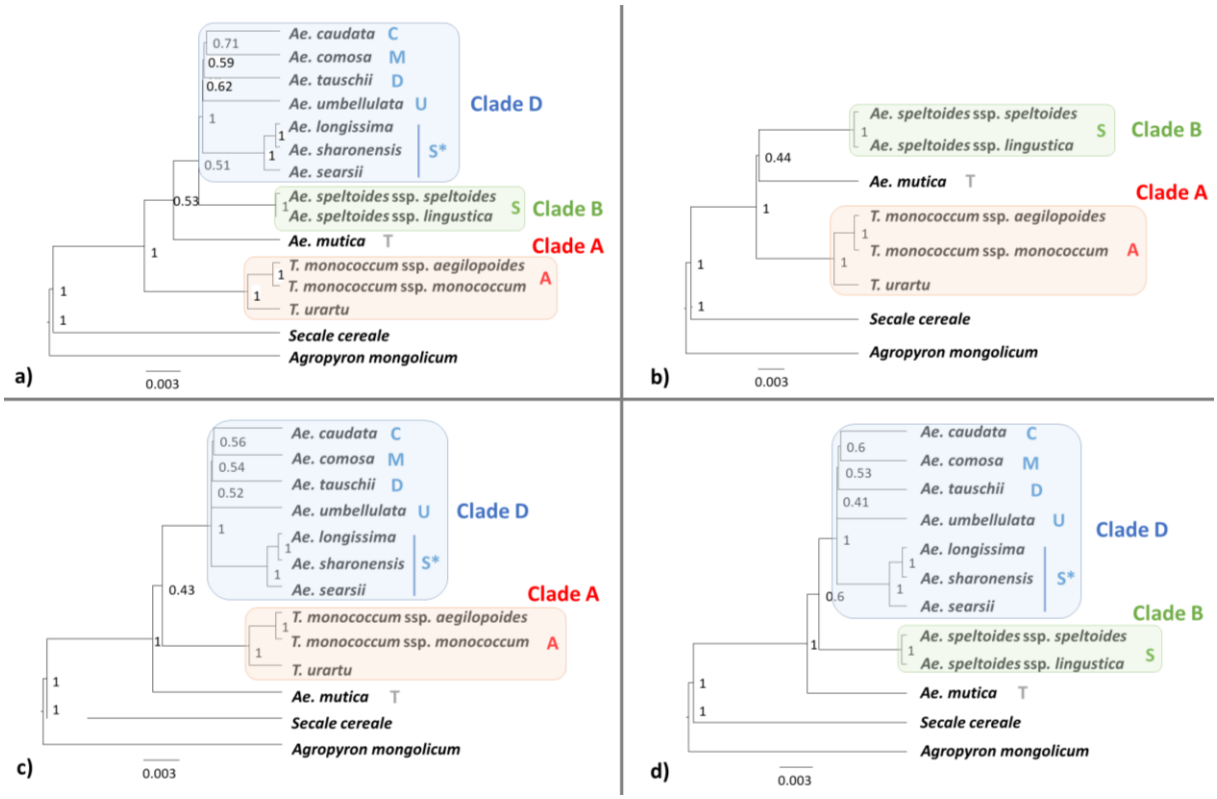


Figure A: Species phylogeny of diploid *Aegilops – Triticum* species based on 30 low-copy nuclear loci including *Ae. mutica*. The species trees were inferred from 30 low-copy nuclear loci under the multispecies coalescent model using JC model and strict clock in STACEY. The genome composition of the species is indicated following van Slageren (1994), with S* referring to all variants of the S genome. The species trees included the three clades A, B and D, or only two clades at a time (A/B, A/D and B/D)

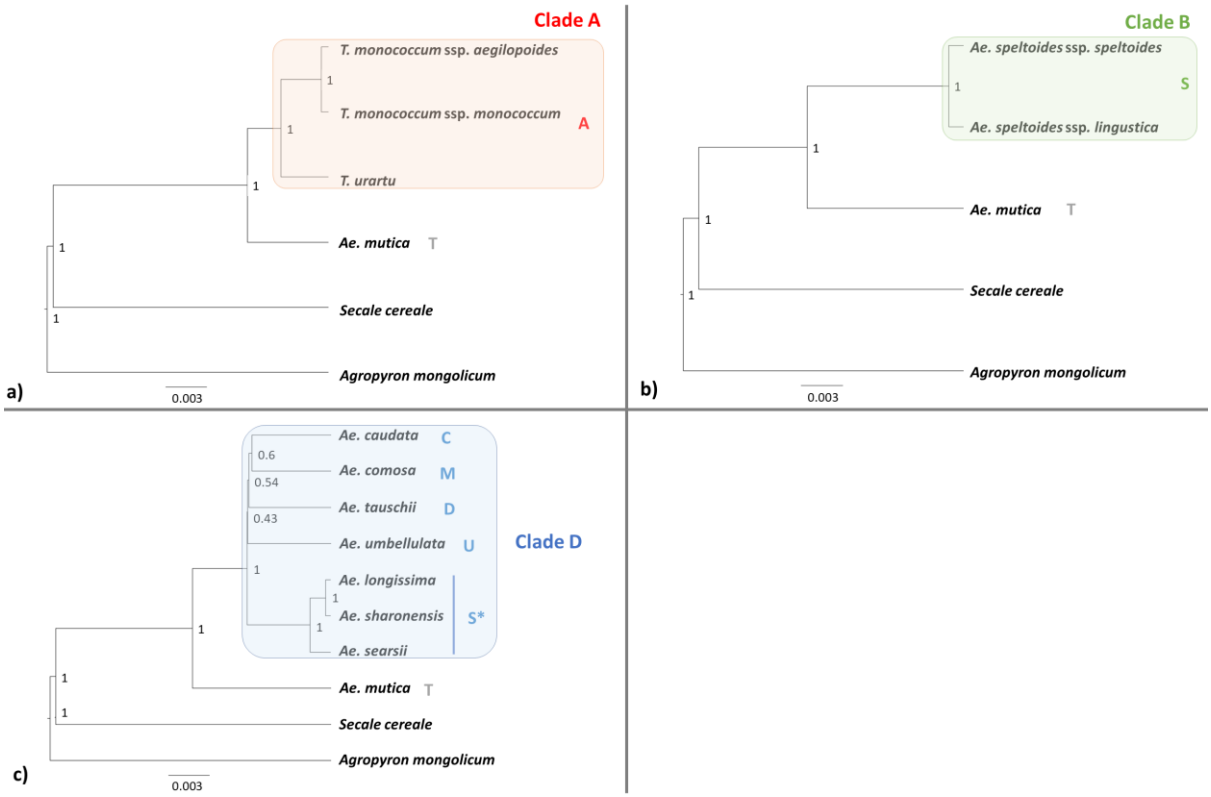


Figure B: Species phylogeny of diploid *Aegilops – Triticum* species based on 30 low-copy nuclear loci. The species trees were inferred under the multispecies coalescent model using JC model and strict clock in STACEY. One specimen was selected per species. The genome composition of the species is indicated following van Slageren (1994), with S* referring to all variants of the S genome. The species trees included only one clade (A, B or D) at time.

Table S1. List of the 37 accessions of the diploid *Aegilops*-*Triticum* species and outgroups used in the study. The genome composition of the species were reported following van Slageren (1994). The accessions were provided by USDA ARS-GRIN (PI; CIae) or IPK-Gatersleben (AE) genebanks.

Species	Genome	Accessions
Ae. caudata	CC	PI 551132, PI 203431, PI 551125, PI 551131
Ae. comosa	MM	AE 1258, PI 551030, PI 551021, PI 551019, PI 551045, PI 551069
Ae. tauschii	DD	AE 525, PI 203326, PI 431600, PI 662070
Ae. umbellulata	UU	PI 486261, PI 542381, PI 542377, PI 573516
Ae. uniaristata	NN	PI 554418, PI 554420
Ae. bicornis	S^bS^b	CIae 47
Ae. sharonensis	$S^{sh}S^{sh}$	PI 584396, PI 584412
Ae. searsii	S^sS^s	PI 599168, PI 599149
Ae. speltoides var. speltoides	SS	PI 170204
Ae. speltoides var. lingustica	SS	PI 560528
Ae. longissima	S^lS^l	PI 604127, PI 604108
Ae. mutica	TT	PI 636562
T. monococcum spp. *monococcum*	A^mA^m	PI 94743
T. monococcum spp. *aegilopoides*	A^mA^m	PI 427990
T. urartu	A^uA^u	PI 428228
Agropyron mongolicum		PI 499391
Taeniatherum caput-medusae		PI 577709
Secale cereale		PI 234656, PI 561793

Table S2: List of the 48 genotyped low-copy loci (Electronic material). The information reported in the table are as follows: name of the loci as annotated in *T. aestivum* reference genome, the description of the loci, the corresponding name of the loci found in *Brachypodium distachyon* and *Ae. tauschii*, the primer and reverse forward designed for amplicon sequencing with corresponding melting temperatures in degrees, the expected length of targeted sequences in number of base pairs and the results of the post-processing quality filtering steps. The table is an electronic supplemental material is provided separately in an Excel file format named "StellaHuynh_PhD_Chapter1_ SuppMat_TableS2–ListOf48LowCopyLoci.xlsx".

Chapter 2

Origin and postglacial colonization of diploid and allopolyploid Aegilops species following the Last Glacial Maximum

Stella Huynh[1], François Felber[1,2] and Christian Parisod[3]

[1]Institute of Biology, University of Neuchâtel, Switzerland

[2]Musée et Jardins botaniques cantonaux de Lausanne et Pont-de-Nant, Switzerland

[3]Institute of Plant Sciences, University of Bern, Switzerland

List of figures and tables

Abstract

The last glacial maximum (LGM) has drastically affected the current distribution and genetic structure of the species, with range contraction into refugia during the LGM followed by postglaciation recolonization about 10 000 years ago. Allopolyploidy combines hybridization between genetically divergent species and whole genome duplication (WGD), providing a high and fixed genetic diversity that may have conferred evolutionary advantages enabling them to rapidly colonize suitable habitats following glaciation. However, the role and evolution of allopolyploidy during the LGM have been under-investigated in the Mediterranean region. The present study thus investigated the evolutionary history of four allopolyploid and their four diploid progenitors *Aegilops* species across their whole distribution. Using 2 plastid and 30 nuclear markers, we identified several progenitor lineages involved in the multiple origins of the allopolyploids and assessed the genetic diversity and structure of the eight species. The *Aegilops* species showed several refugia mainly in the Balkans and in Anatolia, from where the intraspecific genetic lineages have expanded. The genetic lineages that were previously isolated have then further met through a secondary contact along the suture zone separating the Balkans and Anatolia. As expected, genetic novelties accumulate with the age of the allopolyploid species, as for *Ae. geniculata*, but their contribution to the allopolyploid success remains to be further investigated.

Keywords: *Aegilops*, phylogeography, Mediterranean region, refugia within refugia, extra-Mediterranean refugia, suture zone.

Introduction

Quaternary climatic oscillations have drastically influenced the spatial distribution of species (Hewitt, 2003), driving cycles of range contractions - expansions during glacial and interglacial periods, and thus shaped the genetic structure of the species (Hewitt, 2004). The current distribution of genetic diversity within and among species has been mostly influenced by the Last Glacial Maximum (LGM) (Taberlet *et al.*, 1998; Hewitt, 2004), which occurred between 10 000 and 50 000 years ago approximately (Clark *et al.*, 2009).

Polyploidy, or whole-genome doubling (WGD), has long been hypothesized as an evolutionary strategy in rapidly colonizing de-glaciated habitats (Stebbins, 1950; Brochmann *et al.*, 2004). The Arctic, which was covered by the ice sheet during the LGM, is one of the Earth most polyploid-rich regions (Brochmann *et al.*, 2004). Polyploidization events seem to occur more likely following stressful conditions (Ramsey & Schemske, 1998) such as extreme temperatures, which tend to increase the formation of polyploid species through unreduced gametes (Tayalé & Parisod, 2013). In particular, allopolyploid species combine hybridization between divergent progenitor species and WGD, and are generally reported to be more adapted to extreme climatic conditions than their related diploid species. The fixed heterozygosity inherent to allopolyploidy and genome doubling may indeed confer a high genetic plasticity enabling allopolyploid species to adapt to a broader range of environmental conditions and hence rapidly re-colonize their habitats or colonize new habitats after glaciation compared to diploid relatives (Levin, 2002; Brochmann *et al.*, 2004). Fixed heterozygosity also limits the loss of genetic diversity when species range contracts during glacial periods, further contributing to postglacial colonization abilities of allopolyploid species. Whether glacial periods promote the formation of allopolyploid species and to what extent allopolyploidy is evolutionarily advantageous in colonizing deglaciated habitats remains unclear.

The impact of the LGM on polyploid evolution have been mostly studied in the Arctic (Stebbins, 1984; Adamowicz *et al.*, 2002; Brochmann *et al.*, 2004), in North America (Hunter *et al.*, 2000; Novikova *et al.*, 2018) or in the Alps (Bavec *et al.*, 2004; Parisod & Besnard, 2007; Gao *et al.*, 2015), where the degree of glaciation was the highest, however it has been overlooked in other regions. In particular, the Mediterranean region is hypothesized to have served as a global refugium during the LGM, where the temperatures were globally cooler (the ‘Southern refugia’ hypothesis; Bennett *et al.*, 1991; Taberlet & Cheddadi, 2002). The Mediterranean region, known as a biodiversity hotspot (Médail & Quézel, 1999), is

characterized by heterogeneous habitats (Thompson, 2005; Blondel *et al.*, 2010), hence potentially providing suitable habitats for numerous species. Marques *et al* (2018) have estimated that about 36% of the Mediterranean plants are polyploids. In fact, the Mediterranean region contains several small refugia (the 'refugia-within-refugia' model; Gómez & Lunt, 2007), especially in the Balkans and in the Italian and Iberian peninsula, which may have played as major sources for postglacial recolonization (Taberlet & Cheddadi, 2002; Carrión *et al.*, 2003; Médail & Diadema, 2009; Feliner, 2011). Different scenarios of postglacial recolonization routes arose from phylogeography studies on Mediterranean species, mostly diploids, involving the Balkans and the Iberian and Italian peninsulas as the sources (i.e. served as refugia during the LGM) as first suggested by Hewitt (1999). Many studies have documented extra-Mediterranean refugia located in the Eastern Mediterranean (Varga, 2010; Schmitt & Varga, 2012), in cryptic Northern regions such as the Carpathians (Provan & Bennett, 2008; Wielstra *et al.*, 2017) or in the mountains (Kühne *et al.*, 2017), nevertheless essentially focused on animals. Whether Mediterranean plants, especially polyploid species, may have survived in these extra-Mediterranean refugia remains unclear despite existing phylogeographic studies on polyploid Mediterranean species (Mansion *et al.*, 2005; Pimentel & Sahuquillo, 2007).

The *Aegilops* species (Poaceae: Pooideae: tribe Triticeae) are Mediterranean wild grasses comprising ten diploid species (2x = 14 chromosomes) that differentially combined into fourteen allopolyploid species. All allopolyploid *Aegilops* species were formed from a progenitor species having either the U or D genome (called pivotal genome), which combined with another genome (C, M or S; called differential genome) from a related progenitor species. Those species mainly occur across the Mediterranean Basin, in Anatolia and Western Asia, with allopolyploid species usually exhibiting a broader range than the diploid species. The diploid *Aegilops* species belonging to clade D, which includes all diploids except *Ae. speltoides* and *Ae. mutica*, are estimated to have diverged about 1.49 million years ago (Mya; see Chapter 1). The diploid species have thus went through interglacial-glacial periods with rounds of range expansion-contraction. However, whether the allopolyploid species have formed during glacial or interglacial period and have experienced similar rounds of range expansion-contraction as their diploid relatives remains unknown. The allopolyploid *Aegilops* species are known to originate multiple times (Meimberg *et al.*, 2009; Bernhardt *et al.*, 2017), and such multiple origins hence increased their initial genetic diversity that was further maintained through fixed heterozygosity. Whether the allopolyploid species have originated

independently (at different geographical locations) and/or from different progenitor lineages is however unclear. Inferring the age and origins of the allopolyploid species across their whole distribution is crucial to evaluate the factors having contributed to their current geographical distribution and determine to what extent allopolyploidy would confer evolutionary advantages promoting their expansion.

Here, plastid and nuclear markers were used to comprehensively characterize the phylogeography of four diploid species and four of their derivated allopolyploid species composing the MCDU cluster, named based on the species genome composition (Figure 1): the diploids *Ae. caudata* (CC), *Ae. comosa* (MM), *Ae. tauschii* (DD) and *Ae. umbellulata* (UU), and the allopolyploids *Ae. crassa* (DDMM), *Ae. cylindrica* (DDCC), *Ae. geniculata* (UUMM) and *Ae. triuncialis* (UUCC). The relative age of the allopolyploid *Aegilops* species was estimated based on the divergence time of the diploid species. The distribution of genetic diversity within species, combined with spatial analyses of isolation-by-distance (IBD) pattern, further defined the evolution of intraspecific lineages and the postglacial colonization history of the diploid *vs.* allopolyploid *Aegilops* species. Such insights on the relative importance of polyploidy *per se* (i.e. the immediate advantages conferred by WGD and fixed heterozygosity) and of the long-term species evolution (i.e. genetic novelties evolved by the allopolyploid species) offers a better understanding of evolutionary processes shaping allopolyploid species.

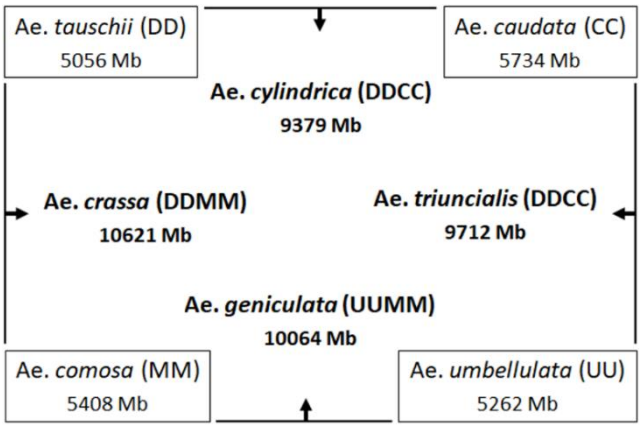


Figure 1: **Evolutionary relationships of four diploid and four derivative tetraploid *Aegilops* species of the MCDU cluster.** Genomic formula are indicated into brackets following van Slageren (1994) and genomes sizes reported according to Eilam *et al.* (2007, 2008).

Material and Methods

Plant material

A total of 420 geo-localized accessions of the eight *Aegilops* species, about 30 diploid and 80 polyploid accessions of each species, were selected to represent genetic variation across their distribution ranges (General Supplemental Material, Table S1). For *Ae. crassa*, whose distribution and seed availability were limited, only 10 accessions were obtained. One to three additional accessions of ten other diploid *Aegilops* and *Triticum* species, and of eight other species (*Agropyron mongolicum*, *Taeniatherum caput-medusae*, *Secale cereale*, *Thynopyrum junceum*, *Dasypyrum villosum*, *Hordeum vulgare*, *Bromus hordeaceus* and *Brachypodium dystachion*), for a total of 26 accessions, were included as outgroups for phylogenetic inferences. The 446 accessions were obtained from ARS-GRIN or ICARDA germplasms, and from personal collections. Seeds of all accessions were sown and grown outdoor at the Botanical garden of Neuchâtel until flowering and deposited as herbarium sheets at the Botanical Museum and Garden of Lausanne and Pont-de-Nant, Switzerland. Leaves from young seedlings were dried in silica gel for DNA extraction.

DNA was extracted using DNAeasy kit from Qiagen® following the manufacturer protocol. All DNA samples were distributed and randomized on five 96-well plates, including 26 accessions replicated once and one accession of *Ae. geniculata* (PI 487229) replicated on each plate (i.e. 6.04 % of the whole dataset) to ensure genotyping accuracy. One negative control was also added on each plate. DNA concentration was measured by Nanodrop and normalized at 10ng/µL and 50ng/µL for amplification of plastid and nuclear markers respectively.

Sequencing of maternally-inherited plastid markers

Two intergenic regions of the plastid genomes, *trnT-trnF* and *ndhF-rpl32,* were selected for genotyping based on the polymorphism reported among *Aegilops* species by Meimberg *et al.* (2009). Alignments of corresponding regions from *Ae. cylindrica* and *Ae. geniculata* (Middleton *et al.*, 2014) were used to design primers with Primer3 2.3.4 (Untergasser *et al.*, 2012) into conserved regions encompassing intergenic sequences for *trnT-trnF*, (5'-GTG CAG AGA CTC AAT GGA AGC-3' and 5'-CCT GAC CTT TTC TTG TGC ATC A-3') and for *ndhF-rpl32* (5'-AAT CAT TAT TGA CGG TCC AAG ACC -3' and 3'-GGG GGA TAA AAA GAA GGG AAA ATG A-5').

PCR amplifications have been performed on 10ng of genomic DNA in 25 µL (1X buffer, 1.5 mM $MgCl_2$, 0.5 nM dNTP, 5nM of each primers, 0.2 U GoTaq polymerase from Promega)

with the following cycles: 240s at 94°C, 36x (45s at 94°C, 45s at 58°C, 45s at 72°C), 300s at 72°C. Resulting PCR products were checked by electrophoresis on 1% Agarose gels and sequenced using BigDye v.3.1 (service provided by Macrogen Inc., Amsterdam, Netherlands). Sequences were manually cleaned using raw chromatograms and trimmed at both ends before being concatenated. Stretches of more than seven single nucleotide repeats were removed from the alignment following Meimberg *et al.* (2009), resulting in a total of 1760 bp. The whole alignment was inferred iteratively using ten to fifteen cycles of SATé (Liu *et al.*, 2009) with the following settings for external tools: MAFFT as Aligner, Muscle as Merger, RAxML (Stamatakis, 2014) as Tree Estimator, GTRGAMMAI as nucleotide substitution model, and for SATé-specific settings: SATé-II-ML, Centroid decomposition and "best alignment after last improvement".

The plastid markers were used to infer the maternal relationships among all 446 accessions of the *Aegilops* and outgroup species through a haplotype network of concatenated sequences based on the Median-Joining method performed following Bandelt *et al.* (1999) in SplitsTree4 (Huson & Bryant, 2006). Gaps and uninformative sites were excluded from the inference.

Sequencing and genotyping of biparentally inherited nuclear markers

From the 446 accessions genotyped at plastid markers, 425 accessions (including 24 replicates, i.e. 5.65 % of this dataset) were also genotyped at 48 low-copy nuclear loci using amplicon sequencing as in Chapter 1 (General Supplemental Material Table S1). The 425 accessions were simultaneously PCR-amplified with four negative controls using eight 48.48 Access Array Integrative Fluidic Circuit (IFC) chips following Fluidigm instructions. Samples were barcoded with a specific 10 bp-sequence, pooled into a single library and sequenced using the Illumina MiSeq Reagent Kit v3. Paired-end reads (2 x 300 bp) were merged using BBmerge (http://jgi.doe.gov/data-and-tools/bbtools/), filtered for a mean Quality Control above 20, demultiplexed and locally aligned by sample and locus using Muscle (Edgar, 2004). Variants present in less than 10 % of the reads were removed and remaining alleles were characterized into haplotypes for each locus, among which only those present in at least 10% of the haplotypes were kept for further analyses After alignement using SATé (Liu *et al.*, 2009), chimeras, null alleles and repetitive regions were detected by visual inspection and filtered out.

The presence of paralogs was assessed at each locus through a combination of the Robinson-Fould metric (RF distance, which measures the distance between an unrooted gene tree and the species tree; Robinson & Foulds, 1981) and the global node support of the gene tree. The

species tree was inferred from all loci on a subset of diploid individuals using the STACEY model (Jones, 2014) implemented in BEAST2 (Bouckaert *et al.*, 2014), considering each individual as a distinct population and assuming a Jukes-Cantor (JC) model for the nucleotide substitution rate and a strict molecular clock. Gene trees were inferred using BEAST2 on the same subset of individuals and under the same model. High RF values and high node supports indicated gene tree with strongly supported topologies, but appearing highly different from the species tree topology, and thus supported the presence of paralogs. Gene trees were further inspected for three types of topological features indicative of paralogy: (i) ingroup species (i.e. the eight studied *Aegilops* species) coalescing deeply, (ii) alleles within an individual coalescing deeply and (iii) outgroups coalescing shallowly. For the loci showing one allele of an outgroup being closely related to the ingroups, only the spurious allele was removed.

Allele dosage inferred by sequencing coverage of each haplotype was considered misleading due to possible biases introduced by PCR amplification. In particular, allele dosage estimated at $\frac{1}{3}/\frac{2}{3}$ were commonly reported in polyploid individuals, precluding $\frac{1}{4}/\frac{3}{4}$ to be distinguished from $\frac{1}{2}/\frac{1}{2}$ with confidence. Accordingly, allele dosage was conservatively assessed for diploid individuals presenting one (A) or two haplotypes (AB) as homozygous (AA) or heterozygous (AB) respectively. Similarly, tetraploid individuals presenting one (A), two (AB), three (ABC) or four (ABCD) haplotypes were assumed as having (AAAA), (AABB), (ABC-) and (ABCD) genotype respectively, with "-" corresponding to a missing sequence.

Assignment of homeologous sequences to allopolyploid subgenomes

For each nuclear locus in each polyploid species, haplotypes being identical to those segregating in diploid progenitor species were assigned to the corresponding subgenome or considered as shared between subgenomes when shared by diploids, otherwise they were considered as derived haplotypes and assigned to a specific subgenome using phylogenetic distances based on gene trees (Supplemental Material Figure S2). Such derived haplotypes were related to a specific diploid progenitor species and assigned to the corresponding subgenome when showing shorter branch lengths than with other diploids. Derived haplotypes showing unclear phylogenetic relationships with haplotypes from progenitor species were considered as of unknown origin. When a subgenome was assigned only one haplotype at a given locus, corresponding haplotype was considered as homozygous and

coded accordingly. When one of the subgenomes was not assigned any haplotype at a given locus, corresponding haplotypes were coded as missing data and considered as being lost.

Relative contribution of diploid progenitors to the genome of their derivated polyploid

We measured the relative contribution of diploid progenitors to their allopolyploid derivatives based on the assignment of allopolyploid nuclear haplotypes to their corresponding subgenomes. Haplotypes identical to those segregating in either diploid progenitor(s) were considered as "conserved haplotypes" whereas haplotypes specific to the allopolyploid were considered as "derived haplotypes". Haplotypes that could not be assigned or that were not amplified at a given subgenome and locus were categorized as "lost haplotypes". For each allopolyploid species, the proportions of conserved, derived and lost haplotypes were calculated as the number of corresponding haplotypes segregating within the species divided by the total number of haplotypes, i.e. four haplotypes (corresponding to the ploidy level of the species), and multiplied by the number of individuals within the species. The relative proportions of haplotypes from each subgenome segregating within an allopolyploid individual are expected to be equal (50% and 50% for a tetraploid individual) under the genetic additivity of progenitor species.

Phylogenetic inferences of bi-parental origins of the allopolyploid species

Allopolyploid multi-labeled species trees were inferred from sequences of nuclear loci using the AlloppNET model (Jones, 2014; Oxelman *et al.*, 2017) incorporating hybridization events, as implemented in BEAST v1.8.4 (Drummond *et al.*, 2012). The species trees were assessed on a data subset constituted of two individuals per genetic clusters inferred by STRUCTURE for each species (see below). AlloppNET inferred the origin of one tetraploid species at a time, by assigning each pair homologous sequences (i.e. its two subgenomes) to one of the premised progenitor species. In the case of *Ae. crassa*, we included one hexaploid individual by splitting its three subgenomes into two "virtual" tetraploid individuals, a first one containing the two subgenomes common with the actual tetraploid individuals and another one containing the third subgenome and an empty subgenome (i.e. constituted of only empty sequences). The use of an empty subgenome enabled us to identify the progenitor species of the third subgenome of the hexaploid individual without affecting the inferences (Jones, 2014). The analyses were run for 500 million Markov chain Monte Carlo (MCMC) generations, including 500 000 generations as burnin, and the trees were sampled every 100 000 generations. The phylogenetic analyses were checked for convergence, i.e. good mixing of MCMC search was ensured for parameters with an effective sample size (ESS)

value higher than 200, on Tracer v1.7 (Rambaut *et al.*, 2018). Maximum Credibility Clade (MCC) trees were summarized by TreeAnnotator available in BEAST2 with a burnin percentage of 60% and viewed in FigTree v1.4.3 (http://tree.bio.ed.ac.uk/software/figtree/). The node ages of the allopolyploid species trees were estimated as the relative branch length separating the allopolyploids from the split node between diploid *Aegilops* progenitors, whose age was inferred in Chapter 1 (i.e. age of clade D), also termed as "secondary calibrations". The relative ages of the allopolyploid species, estimated in million years ago (Mya), corresponded to the relative time scale between the divergence between diploid progenitors and the divergence of the allopolyploids from their diploid progenitors.

Analyses of genetic variation among species

For both plastid and nuclear loci, allelic richness (AR) within species was estimated as the expected number of alleles among *k* gene copies, *k* being defined as the minimum sample size of 10 and 40 gene copies reported in *Ae. crassa* for plastid and nuclear markers respectively. Gene diversity within and among species was estimated after correction for sample size following (Nei, 1978) from unordered alleles (*h*; i.e. all alleles considered equidistant) *vs* ordered alleles (*v*; i.e. considering the phylogenetic distances among alleles as their proportion of nucleotide substitutions) using SPAGeDi (Hardy & Vekemans, 2002). For clarity, the genetic diversity *v* will be referred here as "phylogenetic diversity" to distinguish it from Nei's genetic diversity *h*. Genetic diversity based on *h* and *v* was then partitioned within and among species to estimate genetic differentiation as G_{ST} and N_{ST} respectively. N_{ST} is equivalent to G_{ST} but accounting for the genetic distance between alleles. Comparisons showing significantly higher N_{ST} than G_{ST} thus indicated that closely related alleles occured within species rather than among species (Pons & Petit, 1996). Differences between G_{ST} and N_{ST} were tested through 10000 alleles permutations and significance considered from one-tailed test in SPAGeDi.

For nuclear loci, the observed heterozygosity (Ho) was estimated by SPAGeDi for each species as the average proportion of heterozygous individuals within the species. Additionally, the individual Ho (Ho_i) was manually calculated for each individual as the proportion of loci for which the individual was heterozygous and mapped accordingly to their geographical locations. Furthermore, patterns of genetic differentiation between the eight *Aegilops* species were assessed through a Principal Component Analysis (PCA) on a frequency matrix based on the presence or absence of each haplotype found for each locus

according to the ploidy level in each individual using the R-package *adegenet* (Jombart, 2008).

Analyses of genetic variation within species

Genetic structure within each species was assessed from nuclear loci using the Bayesian clustering method implemented in STRUCTURE 2.3.4 (Pritchard *et al.*, 2000). The optimal number of genetic subgroups (K) was selected through STRUCTURE HARVESTER (Earl & vonHoldt, 2012) following the *ad hoc* deltaK procedure, which estimated the rate of change in the log probability of data between successive K values (Evanno *et al.*, 2005), from analyses based on the admixture model and correlated alleles for K between one and eight or ten for diploid and polyploid species respectively. For each K, 500'000 MCMC generations with 100'000 as burnin were run and repeated ten times, and the optimal K value was selected as the highest peak showed by the corresponding deltaK. The population structure under the optimal K was further analyzed in STRUCTURE with 1'000'000 MCMC generations, 100'000 as burnin and 20 iterations. Individual assignment probabilities were averaged over all these iterations using the FullSearch algorithm in CLUMPP 1.1.2 (Jakobsson & Rosenberg, 2007). Stacked bar plot representations were generated using R package 'strataG' v0.9.4 (Archer *et al.*, 2017).

Patterns of isolation by distance (IBD)

The pairwise geographic distances among individuals were estimated using the R package *geosphere* (Hijmans, 2017), which computed the Euclidean distances between geographical locations by considering the Earth ellipsoid. The pairwise genetic distances among individuals were estimated in SPAGeDi (Hardy & Vekemans, 2002) as the Rousset's *a* statistic based on the probability of identity of genes within and among individuals, which is thus analogous to $F_{ST}/(1-F_{ST})$ (Rousset, 2000). For each species, signal of isolation by distance was tested by comparing the geographic and genetic pairwise distance matrices through mantel tests using the R-package *vegan*, with significance tested through 20000 permutations. Mantel tests were similarly performed within each genetic clusters inferred by STRUCTURE, with admixed individuals assigned to one cluster using values from the Q-matrix with a threshold at 0.5, to detect lineage-specific evolutionary history.

Results

Genotyping and haplotype network of plastid markers

The alignment of concatenated plastid sequences from the 446 individuals at the *ndhF-rpl32* and *trnT-trnF* loci was 1760 bp in length including gaps and presented 69 parsimony-informative sites. The species relationships showed by the haplotype network were consistent with previous studies and offered evidence that several polyploid species originated multiple times (Figure 2). In particular, the polyploid *Ae. cylindrica* (DDCC) shared two distinct haplotypes that appeared identical with those segregating within the diploid *Ae. tauschii* (DD), indicating recurrent origins of the polyploid involving different maternal lineages of the same progenitor species. Similarly, the polyploid *Ae. triuncialis* (UUCC) shared multiple haplotypes that appeared identical with plastid sequences from both the diploids *Ae. caudata* (CC) and *Ae. umbellulata* (UU), supporting at least two recurrent origins involving the two diploid species as maternal progenitors. The polyploid *Ae. geniculata* (UUMM) presented highly diversified haplotypes closely related to the diploid *Ae. umbellulata* but that appear specific to this species, supporting divergence of the polyploid from its maternal progenitor species. The polyploid *Ae. crassa* (DDMM) showed a unique haplotype that also appeared distinct from any of the surmised diploid progenitor species. Although generally consistent with multiple origins of polyploid species, plastid markers do not unambiguously resolve their precise maternal origin(s).

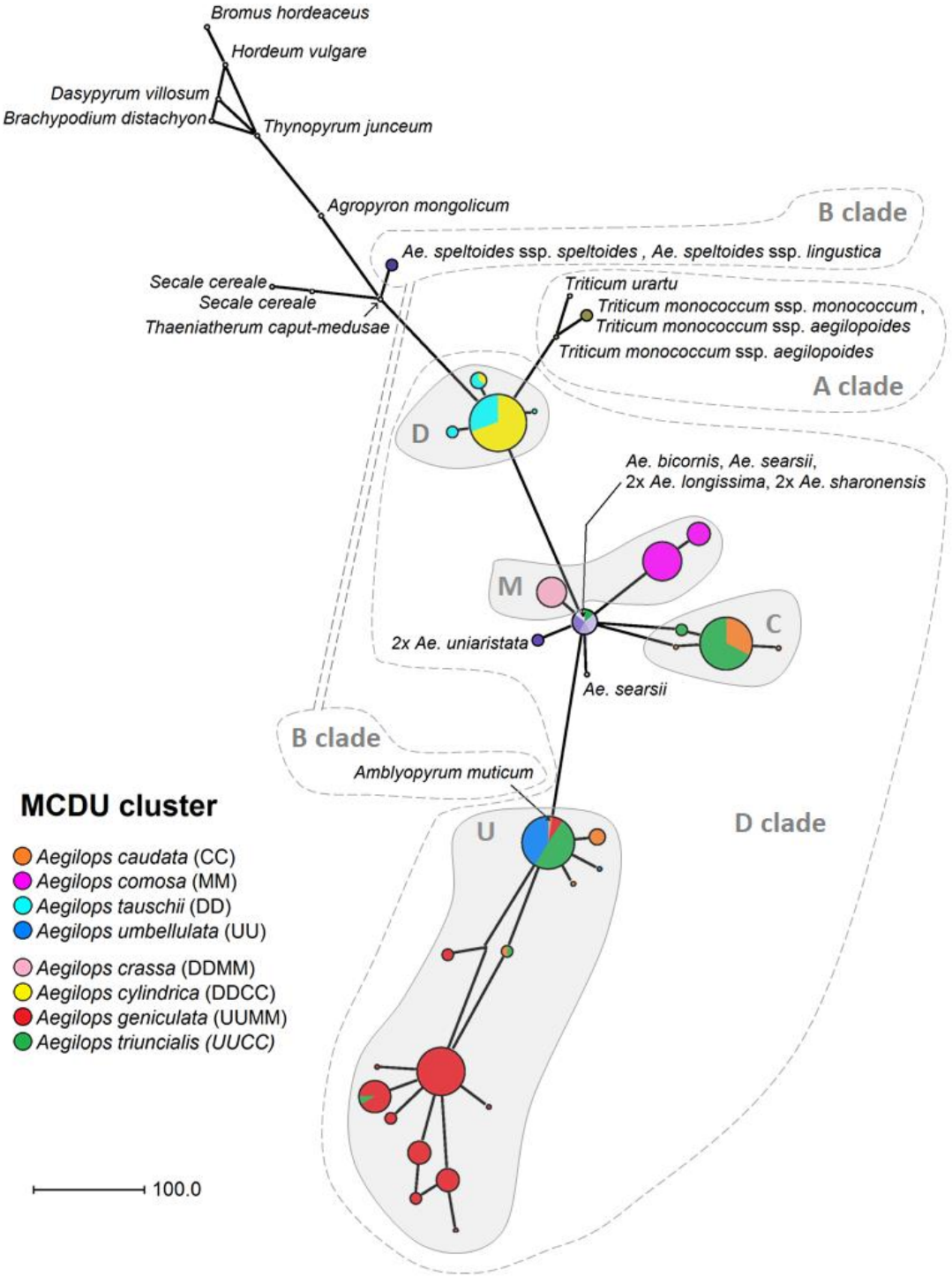


Figure 2: Maternal origins of the allopolyploid *Aegilops* species. The plastid haplotype network was based on concatenated intergenic regions *ndhF-rpl32* and *trnT-trnF* for a total of 1760 bp length. Nodes are scaled by the number of samples sharing the same haplotypes. Eight species of the MCDU cluster were colored following the legend. The clade A, B or D to which belongs the *Aegilops* species according to Macurssen *et al.* (2014) are indicated by dotted areas.

Genotyping of nuclear markers

The amplicon sequencing of the 48 nuclear loci among the 446 individuals (i.e. including previously genotyped individuals from Chapter 1) yielded about 43 millions of Illumina MiSeq reads that were stringently filtered and trimmed to 18 million reads representative of haplotypes segregating in the surveyed species. The same seven loci as in Chapter 1 (*Cmyc*, *CyC*, *Glycin*, *MCMC3*, *PCNA_6L*, *PHY_A* and *S1rna*) were poorly or unspecifically amplified and hence removed from further analyses.

The combination of RF distances and global node support of gene tree for each of the 41 remaining loci further detected patent paralogs. Four loci (*RPB2*, *RPA2A*, *CBS* and *2Fe2S*) showed both high RF distances and high node support (Supplemental Material Figure S1), and were removed. Seven additional loci (*Adh*, *Atrans*, *GTPase*, *mtCarrier*, *pyroP*, *SUI1* and *tRNAsyn*) presented suspicious gene tree topologies (Supplemental Material Figure S2) and were conservatively excluded. One haplotype of the *zinc* locus showed identical haplotypes in an outgroup and an ingroup species, and only this supposedly spurious haplotype was removed. This filtering step resulted in a stringent removal of 11 additional loci, leaving 30 robust loci that were used to analyze the evolutionary history of the species.

Population structure within Aegilops species

All *Aegilops* species revealed genetic clusters that were spatially structured (Figure 3 and Supplemental Material Figure S3). The optimal number of genetic clusters was K=3 for the diploid *Ae. comosa*, as well as the polyploids *Ae. geniculata* and *Ae. triuncialis*, while it was K=2 for the other species (Supplemental Material Figure S4). In most species, a specific cluster appeared geographically more widespread than the other cluster(s) and contact regions between such clusters presented a particularly high proportion of admixed individuals.

The diploid *Ae. caudata* and *Ae. comosa* presented all their genetic clusters in the Balkans and one cluster (Ca_k3) extending across Anatolia (Figure 3). Similarly, *Ae. umbellulata* showed multiple genetic clusters in the Western part of Anatolia, whereas a specific one extended across Anatolia. In contrast, *Ae. tauschii* showed a less intermingled distribution with a particularly divergent genetic cluster (Supplemental Material, Figure S5) being specifically located in the current center of the species range, along the Caspian Sea (Figure 3).

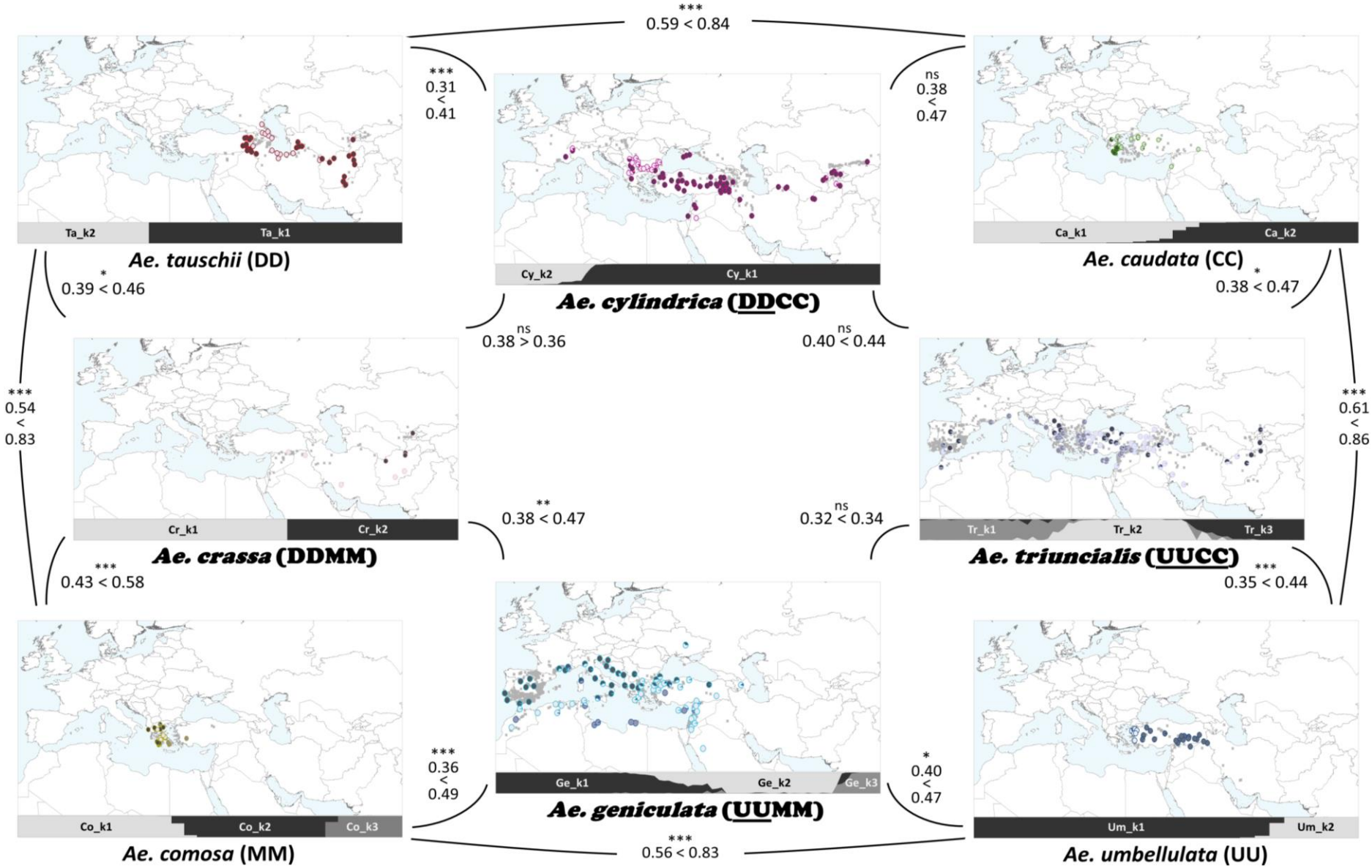

***
0.59 < 0.84
***
0.31
<
0.41
ns
0.38
<
0.47
Ta_k2
Ta_k1
Ae. tauschii (DD)
Cy_k2
Cy_k1
Ae. cylindrica (DDCC)
Ca_k1
Ca_k2
Ae. caudata (CC)
*
0.39 < 0.46
*
0.38 < 0.47
ns
0.38 > 0.36
ns
0.40 < 0.44
***
0.54
<
0.83
***
0.61
<
0.86
Cr_k1
Cr_k2
Ae. crassa (DDMM)
Tr_k1
Tr_k2
Tr_k3
Ae. triuncialis (UUCC)
**
0.38 < 0.47
ns
0.32 < 0.34
***
0.43 < 0.58
***
0.35 < 0.44
Co_k1
Co_k2
Co_k3
Ae. comosa (MM)
Ge_k1
Ge_k2
Ge_k3
Ae. geniculata (UUMM)
Um_k1
Um_k2
Ae. umbellulata (UU)
***
0.36
<
0.49
*
0.40
<
0.47
***
0.56 < 0.83

Figure 3: Genetic diversity and population structure of *Aegilops* species. Maternal origin of polyploids as evidenced by plastid markers is underlined (except for *Ae. crassa*). Genetic differentiation calculated as G_{ST} and N_{ST} following Pons and Petit (1996) is indicated between each coherent pairs of species. Comparisons between G_{ST} and N_{ST} were tested for significance -and associated p-value indicated as followed: (***) < 0.001, (**) < 0.01, (*) < 0.05 and (^{ns}) not significant. Species distributions are represented with grey crosses for occurrences reported in GBIF database and circles for genotyped individuals. Population structures were inferred for each species using STRUCTURE. The best number of genetic clusters K is indicated on the maps and clusters represented by pie charts and stacked histograms using a gradient of greys. The name of each cluster is indicated in the stacked bars. The phylogenetic relationships between all the genetic clusters found in the eight species is shown by four colors representing the M (yellow), C (green), D (red) and U (blue) genomes as assessed through Neighbor-Joining method.

Allopolyploid multilabeled species trees inferences

The multi-labeled species trees of the three polyploid species - *Ae. cylindrica*, *Ae. geniculata* and *Ae triuncialis* - showed unambiguous assignment of the allopolyploid subgenomes to the corresponding progenitor species (Figure 4). However, allopolyploid haplotypes usually formed mostly monophyletic clades sharing a common ancestor with their corresponding progenitor species rather than being indicative of specific progenitor. Accordingly, no specific diploid lineage was apparent at the origin of *Ae. geniculata* (Figure 4a) or *Ae. triuncialis* (Figure 4b). One lineage of *Ae. geniculata* (Ge_k3) however presented divergent haplotypes related to distinct lineages from both diploid progenitors (Figure 4a), consistent with an independent origin that would have been overlooked by plastid loci (Figure 2). Noticeably, the origin of *Ae. cylindrica* seemingly involved the same lineage of the diploid *Ae. tauschii* (Ta_k1) that was also involved as third subgenome of hexaploid *Ae. crassa* (Cr_k2; Figure 4c). *Ae. crassa* showed the deepest divergence from its progenitor species and exact diploid progenitor were hence hardly identified.

Assignment of homeologous haplotypes could be accurately assessed for all polyploids except *Ae. crassa*. Most haplotypes of *Ae. crassa* were species-specific and coalesced before one or both its surmised diploid progenitors. Half of the individuals genotyped were suspected to be hexaploids while the other half seemed tetraploids, which was further confirmed by chromosome counting and flow cytometry (data not shown). The inferred multilabeled species tree of *Ae. crassa* similarly showed ambiguous genome assignment of the selected individuals (Figure 4d). The hexaploid individual included in the species network inference however showed its two subgenomes similar to the subgenomes the tetraploid individuals whereas its third subgenome revealed clearly grouped with a specific lineage of its diploid progenitor *Ae. tauschii*.

Ae. crassa was seemingly the oldest allopolyploid species, with an estimated age between 1.75 and 1.37 Mya. *Ae. geniculata* appeared as the second oldest allopolyploid species with an estimated age of 1.28 Mya, followed by the seemingly youngest species *Ae. triuncialis* and *Ae. cylindrica* showing the shallowest divergence with 0.65 Mya and 0.77 Mya respectively.

Based on the population structure analyses and the paternal origins of the allopolyploids *Aegilops*, all species revealed distinct internal genetic structures that were consistent with the phylogenetic analyses (Figure 3 and 4d), the two genetic clusters inferred within *Ae. crassa* distinguished the tetraploid individuals (Cr_k1) in the West part of its distribution from the

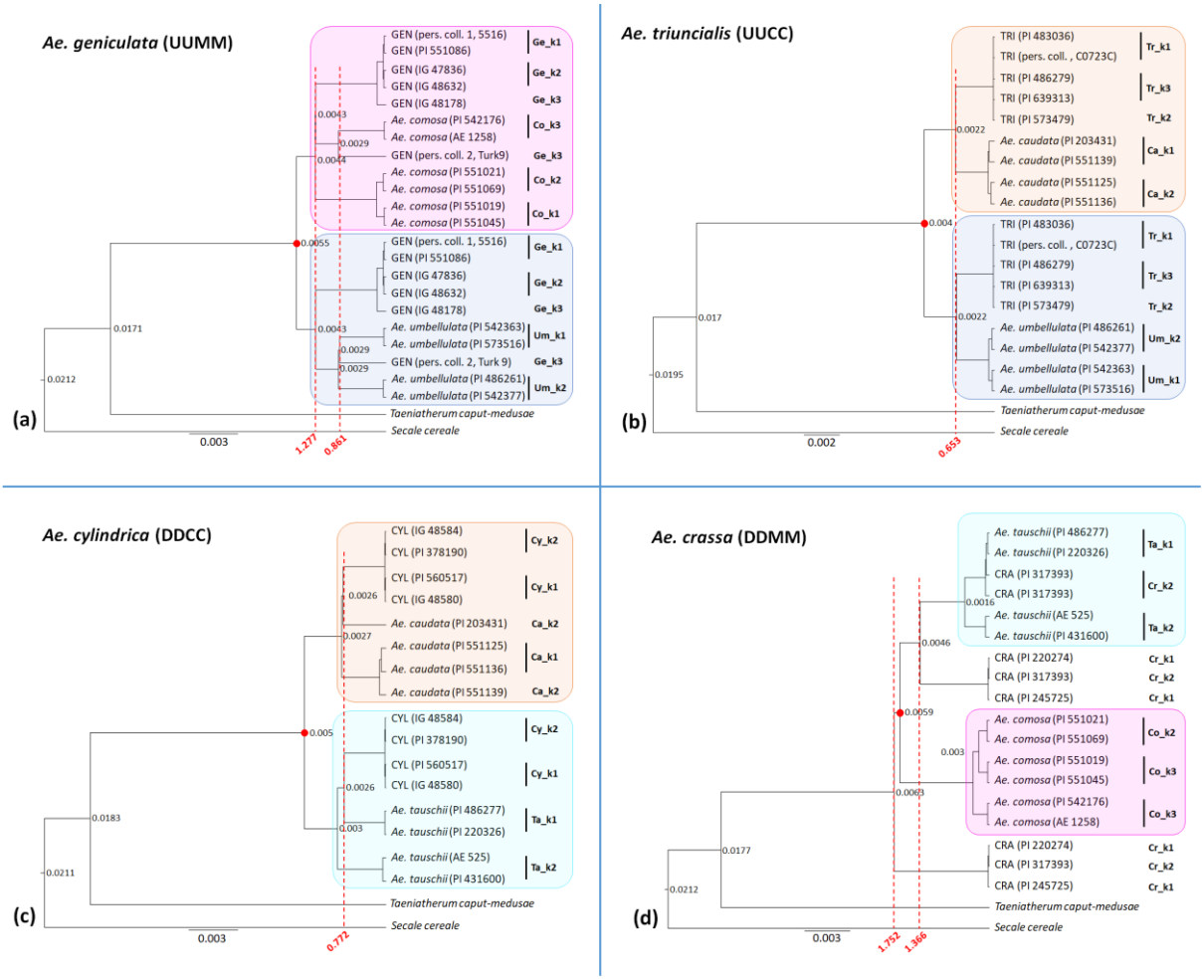


Figure 4: Parental origins of the allopolyploid *Aegilops* species. The origins of allopolyploids were assessed using multi-labeled species trees for each allopolyploid species and based on representative individuals of each diploid and allopolyploid species. Accessions ID of each individual and their genetic clusters as assessed by STRUCTURE are specified. Each allopolyploid individual has two branches corresponding to their homeologous pairs (subgenomes). The diploid progenitors groups are highlighted by colored rectangles. The allopolyploid individuals are abbreviated on branch tips as follow: GEN (*Ae. geniculata*), TRI (*Ae. triuncialis*), CYL (*Ae. cylindrica*) and CRA (*Ae. crassa*). The split of the allopolyploid species from their diploid progenitor species is indicated by the red dashed line, with the relative age (in million years ago; Mya) shown below and calculated based on the divergence age of the diploid species (indicated by the red circle) estimated in Chapter 1 at around 1.49 Mya.

hexaploid ones (Cr_k2) in the Northeast part of its distribution. *Ae. cylindrica* showed a genetic cluster mainly restricted to the Balkans (Cy_k1), whereas the other extended across Anatolia and Central Asia (Cy_k2). The two main genetic clusters of *Ae. geniculata* were mostly restricted to either the Northwest coast (Ge_k1) or the East and Southwest coast (Ge_k2) of the Mediterranean Sea whereas a third genetic cluster including as few as seven individuals (Ge_k3) presented a scattered distribution in Anatolia and North Africa. The three genetic clusters inferred within *Ae. triuncialis* showed a relatively specific distribution along the North and East coast of the Mediterranean Sea (Tr_k1), throughout Anatolia (Tr_k2) or scattered across the whole species distribution (Tr_k3).

Diploid progenitors' contribution to their polyploid genetic diversity

Subgenomes of each allopolyploid *Aegilops* species showed equal proportions of conserved, derived and lost haplotypes from their diploid progenitor species (Table 1 and Supplemental Material Table S1). However, allopolyploid species presented contrasted proportions of each haplotype category, with *Ae. cylindrica* showing the most conserved genome with 94.3% of conserved haplotypes (Supplemental Material Table S1) followed by *Ae. triuncialis* with 77.6% of conserved haplotypes and *Ae. geniculata* showing equal proportions of conserved (51.3%) and derived haplotypes (41.9%). In contrast to those polyploids species presenting relatively stable subgenomes with low proportions of lost haplotypes, *Ae. crassa* was chiefly composed of derived (58.4%) and lost (18.6%) haplotypes.

Genetic diversity within Aegilops species

As expected, genetic diversity within *Aegilops* species differs between nuclear and plastid markers, as well as between diploids and polyploids (Table 2). The plastid markers show relatively high allelic richness (AR = 5.29 - 9.02) and genetic diversity (h = 0.79 - 0.98) in all species, except for the diploid *Ae. umbellulata* (AR = 3.47, h = 0.52) and the polyploids *Ae. cylindrica* (AR = 3.17, h = 0.38) and *Ae. crassa* (AR=3, h=0.35). *Ae. comosa* showed the highest AR (8.24) and h (0.95) among diploids while *Ae. geniculata* showed the highest AR (9.02) and h (0.98) among polyploids. The phylogenetic diversity showed a contrasted pattern, with the diploid *Ae. caudata* (v = 0.0015) and the polyploid *Ae. triuncialis* (v = 0.0013) showing higher v than the other species (v = 0-0.0007).

Table 1: Relative proportion of "conserved haplotypes" *vs.* "new haplotypes" among polyploid species. Conserved haplotypes designates polyploid haplotypes being identical to those segregating in diploid progenitor species. New haplotypes designates haplotypes specifically segregating within polyploid species. Proportions of loss correspond to the proportion of subgenomes for which any haplotype could be assigned at a given locus and individual or that could not be assigned. The pivotal genomes are underlined according to the "pivotal-differential genome" hypothesis (Zohary & Feldman, 1962).

Species	Genome	Conserved haplotypes				Derived haplotypes				Lost haplotypes			
		D	**U**	**M**	**C**	**D**	**U**	**M**	**C**	**D**	**U**	**M**	**C**
Ae. cylindrica	DDCC	0.495			0.448	0.001			0.035	0.002			0.019
Ae. crassa	DDMM	0.227		0.003		0.317		0.267		0.053		0.133	
Ae. geniculata	UUMM		0.268	0.245			0.207	0.212			0.016	0.052	
Ae. triuncialis	UUCC		0.400		0.376		0.087		0.112		0.019		0.006

Table 2: Genetic diversity within *Aegilops* species of the MCDU cluster. AR is the allelic richness calculated as the expected number of alleles among the minimum number gene copies found in a species, i.e. *Ae. crassa*, which was 40 and 10 gene copies for nuclear markers and plastid markers respectively. Ho is the observed heterozygosity. *h* is the genetic diversity corrected for sample size as defined by Nei (1978) and *v* is the genetic diversity with ordered alleles, i.e. with phylogenetic distances between alleles taken into account.

Species	Genome	Number of occurrences	Number of genotyped	Plastid markers			Nuclear markers			
				AR	*h*	*v*	AR	*h*	*v*	Ho
Ae. caudata	CC	102	29	7.41	0.92	0.0015	3.50	0.41	0.005	0.12
Ae. comosa	MM	67	30	8.24	0.95	0.0004	4.77	0.50	0.007	0.06
Ae. tauschii	DD	197	41	5.94	0.80	0.0003	2.84	0.40	0.003	0.00
Ae. umbellulata	UU	76	26	3.47	0.52	0.0001	2.78	0.33	0.002	0.15
Ae. cylindrica	DDCC	328	82	2.80	0.35	0.0002	2.30	0.50	0.013	0.64
Ae. geniculata	UUMM	1366	81	9.02	0.98	0.0007	5.95	0.61	0.013	0.56
Ae. triuncialis	UUCC	1076	76	5.29	0.79	0.0013	3.81	0.56	0.013	0.58
Ae. crassa	DDMM	53	10	3.00	0.38	0.0000	3.30	0.63	0.017	0.66

The nuclear diversity showed similar trends to plastid markers, with AR in *Ae. comosa* and *Ae. geniculata* highlighting genetically highly diverse species, whereas *Ae. cylindrica* appeared genetically homogenous. The phylogenetic diversity *v* was twice lower in diploids (*v* = 0.002-0.007) as compared to polyploids (*v* = 0.013-0.017). The observed heterozygosity Ho was relatively low for diploids (maximum 0.15), but consistently close to 0.5 in polyploids, reflecting both the selfing regime of *Aegilops* species and the fixed heterozygosity expected in allopolyploid species.

Inspecting the geographical distribution of individual observed heterozygosity (Ho_i) among species, three main hotspots presenting high proportions of individuals with high Ho_i were located in a region encompassing the south of Balkans and the western part of Anatolia, in the eastern part of Anatolia and in Central Asia (Supplemental Material Figure S6). The diploids *Ae. caudata* and *Ae. comosa* showed particularly high Ho_i along the South and East coast of Greece, but Ho_i presented no specific pattern for allopolyploids, although *Ae. triuncialis* showed particularly higher Ho_i among individuals from Anatolia (Supplemental Material Figure S6).

Genetic differentiation among Aegilops species

Genetic differentiation at nuclear loci was even and relatively high among diploid species (Figure 4) with G_{ST} (0.54-0.61) always significantly lower than N_{ST} (0.83-0.86). It was slightly lower (G_{ST} = 0.32-0.40) and generally not significantly different from N_{ST} for allopolyploids.

The comparison of allopolyploids to their respective diploid progenitors showed relatively similar N_{ST} but significantly higher N_{ST} than G_{ST}, indicating that a considerable proportion of the diploid phylogenetic diversity has been transmitted to the derived species during polyploid speciation. Although relatively similar, the phylogenetic diversity seemed slightly contrasted genetic differentiation from each of their progenitor, with *Ae. cylindrica* and *Ae. crassa* being less differentiated from *Ae. tauschii* than from their differential progenitors, whereas *Ae. geniculata* and *Ae. triuncialis*, similarly showed lower differentiation with *Ae. umbellulata* as compared to their specific progenitor.

Patterns of isolation by distance (IBD)

Mantel tests yielded contrasted results among diploid and allopolyploid species (Table 3), with significant IBD in all diploid species except *Ae. tauschii*, and non-significant IBD in all allopolyploids except *Ae. cylindrica*. Significant IBD was also consistently observed within

the main of the genetic cluster of each species, except *Ae. geniculata* and *Ae. triuncialis* that showed no IBD at all.

Table 3: Mantel tests of isolation by distance within *Aegilops* species. Pairwise genetic distance within species were estimated as 'Rousset *a*' indice. Mantel tests correlation values were reported in the table with the associated p-values indicated as followed: (***) < 0.001, (**) < 0.01, (*) < 0.05 and (^{ns}) not significant.

(a)

Ae. caudata **(CC)**		***Ae. comosa*** **(MM)**		***Ae. tauschii*** **(DD)**		***Ae. umbellulata*** **(UU)**	
Total	0.455 ***	**Total**	0.226 ***	**Total**	-0.004 ns	**Total**	0.483 ***
Ca_k1	0.577 ***	**Co_k1**	-0.241 ns	**Ta_k1**	0.367 ***	**Um_k1**	0.181 ns
Ca_k2	0.221 ns	**Co_k2**	0.301 *	**Ta_k2**	-0.125 ns	**Um_k2**	0.852 *
		Co_k3	-0.177 ns				

(b)

Ae. crassa **(DDMM)**		***Ae. cylindrica*** **(DDCC)**		***Ae. geniculata*** **(UUMM)**		***Ae. triuncialis*** **(UUCC)**	
Total	0.172 ns	**Total**	0.145 *	**Total**	0.042 ns	**Total**	0.055 ns
Cr_k1	0.855 *	**Cy_k1**	0.184 *	**Ge_k1**	0.112 ns	**Tr_k1**	-0.025 ns
Cr_k2	0.631 *	**Cy_k2**	0.169 ns	**Ge_k2**	0.124 ns	**Tr_k2**	-0.025 ns
				Ge_k3	0.304 ns	**Tr_k3**	-0.070 ns

Discussion

Multiple and independent origins of the allopolyploid species

Maternal and phylogenetic relationships among *Aegilops* species (Figure 2 & 4) are mostly consistent with previous studies (Petersen *et al.*, 2006; Meimberg *et al.*, 2009; Bernhardt *et al.*, 2017) and here further support multiple origins of allopolyploid species such as *Ae. cylindrica*, *Ae. geniculata* and *Ae. triuncialis*. The origins of *Ae. triuncialis* from both diploid species as maternal progenitor was particularly apparent, whereas data for *Ae. cylindrica* and *Ae. geniculata* were rather coherent with origin(s) from a single maternal progenitor. Although in slight constrast with Meimberg *et al.* (2009) who reported bi-maternal origins of *Ae. cylindrica*, maternal relationships presented here also confirmed previous observations of accessions from *Ae. mutica* (T) and *Ae. caudata* (CC) sharing similar plastid haplotypes with *Ae. umbellulata* (UU) indicative of chloroplast capture (Badaeva *et al.*, 2004; Bernhardt *et al.*, 2017).

Multiple nuclear loci further enabled inference of the paternal origin(s) of polyploids (Figure 4). Although coherent with a single paternal origin of *Ae. triuncialis*, nuclear loci clearly identified multiple paternal origins involving different lineages of diploid progenitors in *Ae. cylindrica* as well as *Ae. geniculata.* Noticeably, the genetic lineage of *Ae. tauschii* involved here in the evolution of *Ae. cylindrica* appeared different from the lineage identified along the Caspian Sea that also contributed the D genome of the hexaploid wheat *T. aestivum* (Dvorak *et al.*, 1998; Petersen *et al.*, 2006; Wang *et al.*, 2013b).

Progenitors of the allopolyploid *Ae. crassa* (DDMM) seem to have been misidentified in prior works (Badaeva *et al.*, 1998). This species indeed appeared older than its postulated progenitor *Ae. comosa* (MM) in the multilabeled tree, suggesting the likely involvement of a relatively ancient common ancestor and also explaining the lack of conserved M haplotypes reported here. Hexaploid individuals of *Ae. crassa* seem to occur in the northeast part of the species distribution corresponding to Central Asia. The genome of the hexaploids showed clearly similar to the genome of tetraploids *Ae. crassa* and have an additional genome identical to the widespread lineage *Ae. tauschii* (Ta_k1), supporting the origin of hexaploids through the merging of tetraploid *Ae. crassa* and *Ae. tauschii* (Badaeva *et al.*, 1998, 2002). Given the unclear origin of *Ae. crassa*, results regarding the genetic diversity and population structure inferred for this species should thus be carefully interpreted. Thorough phylogenetic

analyses involving all *Aegilops* and *Triticum* species as well as diverse related species is necessary to untangle the evolutionary history of *Ae. crassa*.

Impact of LGM on species evolution and distribution range

The inferred age of the allopolyploid species is unambiguously coherent with their origin before the LGM. *Ae. crassa* seem likely to have formed an allopolyploid species first, followed by *Ae. geniculata*. In contrast, *Ae. triuncialis* and *Ae. cylindrica* appeared younger. Rough age estimates based on the divergence among the four diploid species (around 1.49 Mya; Chapter 1) can hardly offer precise absolute dating, but *Ae. crassa* and *Ae. geniculata* seem to have originated during climatic oscillations of the Donau period around 1.75-1.28 Mya, whereas *Ae. triuncialis* and the recent lineage of *Ae. geniculata* (Ge_k3) may have evolved during the Gunz period some 0.86-0.77 Mya, and *Ae. cylindrica* seemingly appeared at the beginning of the Mindel period 0.65 Mya. Although their origin during a glacial or interglacial remains elusive, allopolyploid *Aegilops* species have likely endured multiple glacial-interglacial periods resulting in cycles of range contraction-expansion that have considerably impacted the current distribution of their genetic diversity.

Despite such complex spatio-temporal dynamics, *Aegilops* species all showed distinct genetic clusters being spatially coherent and generally separated by admixed individuals. Such a pattern appears coherent with the allopatric evolution of independent genetic clusters in distinct refugia during range contraction of the LGM, followed by secondary contact during postglacial expansion (Clark *et al.*, 2009). Significant IBD within widespread genetic lineages of each species likely reflects their expansion along migration routes from their last refugia. The higher individual heterozygosity (Ho_i) observed along secondary contact zones supports this hypothesis. Interestingly, patterns of individual heterozygosity appeared similar for multiple species, highlighting zones of secondary contact of general importance in the Balkans, in the West and in the East of Anatolia. Such geographical clustering of contact zones looks characteristic of suture zones first defined by Remington (1968) and here appears consistent with the phylogeographic break already reported between the Balkans and the Anatolian region (Bilgin, 2011). Phylogeographic break between West and East of Anatolia has also been observed in several studies (listed in Bilgin, 2011), but the phylogeographic pattern of the *Aegilops* species. does not correlate with the intra-Anatolia suture zone. Nevertheless, the Anatolian region has likely served as refugia for most *Aegilops* species, hence constituting an extra-Mediterranean refugium that may have played a role in surviving the LGM as important as the Southern refugia (Schmitt & Varga, 2012).

The restricted distribution of diploid *Aegilops* species and of their genetic clusters in the Balkans-West Anatolia for *Ae. comosa* and *Ae. caudata*, in Anatolia for *Ae. umbellulata* and around the Caspian Sea for *Ae. tauschii* is coherent with limited dispersal from their last refugia. On the contrary, allopolyploid species showed large distribution ranges and widespread genetic clusters indicative of a greater dispersal ability than diploids. The exact location of refugia having hosted widespread allopolyploids such as *Ae. cylindrica* and *Ae. geniculata* remains elusive due to the genetic uniformity of their main lineages, but likely include Anatolia and the Southern refugia (i.e. the Balkans and/or the Iberian and Italian peninsulas). Coherent with its relatively ancient origin, *Ae. geniculata* showed a strong phylogeographic structure with clusters mostly restricted to the North-Western and the South-Eastern parts of the Mediterranean basin (Figure 3). A third genetic lineage (Ge_k3) showed a disjoint distribution around Anatolia and in the North of Africa. Although human-mediated migration may have shaped the current distribution of genetic variation in *Ae. geniculata* (Arrigo *et al.*, 2010), a scenario involving the persistence of a relictual lineage across multiple of contraction-expansion cycles and the swamping by other post-glacial recolonisers may likely explain such a disjunction. Similarly, the allopolyploid *Ae. triuncialis* showed phylogeographic breaks located within Anatolia and likely survived the LGM in multiple refugia before presenting a rapid recolonization leading to a patchy distribution of genetic clusters. Given that all *Aegilops* species investigated here are selfers and present unspecialized dispersal units similar to species with either a D or a U genome, such a contrasted pattern is likely due to specificities of being allopolyploid. Multiple origins of the allopolyploid species indeed certainly increased their initial genetic diversity, which was mostly maintained through fixed heterozygosity. Such an allopolyploid genetic system thus most probably have limited the consequences of inbreeding and genetic drift inherent to cycles of range contraction-expansion, and thereby facilitated the expansion of polyploid species as compared to their diploid progenitors.

Relative contribution of parental genomes and genetic novelties to the colonization success of allopolyploid species

The comparison of phylogeographies from *Aegilops* allopolyploids offers further insights on the impact of their relative ages or specific genome combinations on their wide distributions following the LGM. Surveyed allopolyploid species presented here a majority of conserved haplotypes from their diploid progenitor species, with the young allopolyploid *Ae. cylindrica* presenting up to 95% of the genetic background of its progenitors whereas it dropped to about

50% in the older species *Ae. geniculata*. Haplotypes from both progenitors were conserved in derived allopolyploid species and presented relatively similar N_{ST} indicative of similar contribution and supporting balanced reorganization of subgenomes. This pattern contrasts with expectations of the pivotal-differential genome hypothesis, whereby allopolyploid species sharing a common pivotal genome recurrently hybridize and see this subgenome conserved under homologous recombination, whereas the other (differential) subgenome goes through considerable changes following homeologous recombination (Zohary & Feldman, 1962; Feldman & Levy, 2005). Such observations based on low-copy loci support previous insights from repetitive sequences to conclude on limited consequences of introgression among species (Senerchia *et al.*, 2014), but studies including functional genes underlying phenotypic differentiation are necessary (Morjan & Rieseberg, 2004; Feldman *et al.*, 2012). Available data rather support random accumulation of genetic mutations as the time passes, but to what extent allopolyploids evolved genetic novelties supporting range expansion through time (e.g. adaptation to a wider range of ecological niches) or, in contrast, accumulated deleterious changes remains elusive. Here, not only relatively young allopolyploids such as *Ae. cylindrica* and *Ae. triuncialis* but also older ones such as *Ae. geniculata* showed higher genetic diversity going along with greater spatial extension than their respective diploid progenitors, supporting the hypothesis that polyploidy is advantageous. Noticeably, another ancient allopolyploid *Ae. crassa* showed a highly contrasted pattern, with a scarce distribution likely reflecting a once widespread range and evidence of additional rounds of allopolyploidization. To what extent this pattern is indicative of an extinction meltdown and represents more than an exception remains unknown.

References

Adamowicz SJ, Gregory TR, Marinone MC, Hebert PDN. **2002**. New insights into the distribution of polyploid Daphnia: the Holarctic revisited and Argentina explored. *Molecular Ecology* **11**: 1209–1217.

Archer FI, Adams PE, Schneiders BB. **2017**. strataG: An R package for manipulating, summarizing and analysing population genetic data. *Molecular Ecology Resources* **17**: 5–11.

Arrigo N, Felber F, Parisod C, Buerki S, Alvarez N, David J, Guadagnuolo R. **2010**. Origin and expansion of the allotetraploid *Aegilops geniculata*, a wild relative of wheat. *New Phytologist* **187**: 1170–1180.

Badaeva ED, Amosova AV, Muravenko OV, Samatadze TE, Chikida NN, Zelenin AV, Friebe B, Gill BS. **2002**. Genome differentiation in *Aegilops*. 3. Evolution of the D-genome cluster. *Plant Systematics and Evolution* **231**: 163–190.

Badaeva ED, Amosova AV, Samatadze TE, Zoshchuk SA, Shostak NG, Chikida NN, Zelenin AV, Raupp WJ, Friebe B, Gill BS. **2004**. Genome differentiation in *Aegilops*. 4. Evolution of the U-genome cluster. *Plant Systematics and Evolution* **246**: 45–76.

Badaeva ED, Friebe B, Zoshchuk SA, Zelenin AV, Gill BS. **1998**. Molecular cytogenetic analysis of tetraploid and hexaploid *Aegilops crassa*. *Chromosome Research 1998* **6**: 629–637.

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

Bavec M, Tulaczyk SM, Mahan SA, Stock GM. **2004**. Late Quaternary glaciation of the Upper Soča River Region (Southern Julian Alps, NW Slovenia). *Sedimentary Geology* **165**: 265–283.

Bennett KD, Tzedakis PC, Willis KJ. **1991**. Quaternary Refugia of North European Trees. *Journal of Biogeography* **18**: 103.

Bernhardt N, Brassac J, Kilian B, Blattner FR. **2017**. Dated tribe-wide whole chloroplast genome phylogeny indicates recurrent hybridizations within Triticeae. *BMC Evolutionary Biology* **17**:141.

Bilgin R. **2011**. Back to the Suture: The Distribution of Intraspecific Genetic Diversity in and Around Anatolia. *International Journal of Molecular Sciences* **12**: 4080–4103.

Blondel J, Aronson J, Bodiou J-Y, Boeuf G. **2010**. *The Mediterranean Region: Biological Diversity in Space and Time*. Oxford, New York: Oxford University Press.

Bouckaert R, Heled J, Kühnert D, Vaughan T, Wu C-H, Xie D, Suchard MA, Rambaut A, Drummond AJ. **2014**. BEAST 2: A Software Platform for Bayesian Evolutionary Analysis (A Prlic, Ed.). *PLoS Computational Biology* **10**: e1003537.

Brochmann C, Brysting AK, Alsos IG, Borgen L, Grundt HH, Scheen A-C, Elven R. **2004**. Polyploidy in arctic plants. *Biological Journal of the Linnean Society* **82**: 521–536.

Carrión JS, Yll EI, Walker MJ, Legaz AJ, Chaín C, López A. **2003**. Glacial refugia of temperate, Mediterranean and Ibero- North African flora in south-eastern Spain: new evidence from cave pollen at two Neanderthal man sites. *Global Ecology and Biogeography* **12**: 119–129.

Clark PU, Dyke AS, Shakun JD, Carlson AE, Clark J, Wohlfarth B, Mitrovica JX, Hostetler SW, McCabe AM. **2009**. The Last Glacial Maximum. *Science* **325**: 710–714.

Drummond AJ, Suchard MA, Xie D, Rambaut A. **2012**. Bayesian Phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution* **29**: 1969–1973.

Dvorak J, Luo M-C, Yang Z-L, Zhang H-B. **1998**. The structure of the *Aegilops tauschii* genepool and the evolution of hexaploid wheat. *TAG Theoretical and Applied Genetics* **97**: 657–670.

Earl DA, vonHoldt BM. **2012**. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources* **4**: 359–361.

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

Eilam T, Anikster Y, Millet E, Manisterski J, Feldman M. **2008**. Nuclear DNA amount and genome downsizing in natural and synthetic allopolyploids of the genera *Aegilops* and *Triticum* (P Gustafson, Ed.). *Genome* **51**: 616–627.

Eilam T, Anikster Y, Millet E, Manisterski J, Sagi-Assif O, Feldman M. **2007**. Genome size and genome evolution in diploid Triticeae species (P Gustafson, Ed.). *Genome* **50**: 1029–1037.

Evanno G, Regnaut S, Goudet J. **2005**. Detecting the number of clusters of individuals using the software structure: a simulation study. *Molecular Ecology* **14**: 2611–2620.

Feldman M, Levy AA. **2005**. Allopolyploidy – a shaping force in the evolution of wheat genomes. *Cytogenetic and Genome Research* **109**: 250–258.

Feldman M, Levy AA, Fahima T, Korol A. **2012**. Genomic asymmetry in allopolyploid plants: wheat as a model. *Journal of Experimental Botany* **63**: 5045–5059.

Feliner GN. **2011**. Southern European glacial refugia: A tale of tales. *Taxon* **60**: 365–372.

Gao Y-D, Zhang Y, Gao X-F, Zhu Z-M. **2015**. Pleistocene glaciations, demographic expansion and subsequent isolation promoted morphological heterogeneity: A phylogeographic study of the alpine *Rosa sericea* complex (Rosaceae). *Scientific Reports* **5**:11698.

Gómez A, Lunt DH. **2007**. Refugia within refugia: Patterns of phylogeographic concordance in the Iberian Peninsula. In: Weiss S, Ferrand N, eds. Phylogeography of Southern European Refugia. Dordrecht: Springer Netherlands, 155–188.

Hardy OJ, Vekemans X. **2002**. spagedi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Molecular Ecology Notes* **2**: 618–620.

Hewitt GM. **1999**. Post-glacial re-colonization of European biota. *Biological Journal of the Linnean Society* **68**: 87–112.

Hewitt GM. **2003**. Ice Ages: their impact on species distributions and evolution. In: Rothschild LJ, Lister AM, eds. Evolution of Planet Earth. Academic Press.

Hewitt GM. **2004**. Genetic consequences of climatic oscillations in the Quaternary. *Philosophical Transactions of the Royal Society B: Biological Sciences* **359**: 183–195.

Hijmans RJ. **2017**. geosphere: Spherical trigonometry. R package version 1. 5-5. https://CRAN.R-project.org/package=geosphere.

Hunter KL, Betancourt JL, Riddle BR, Van Devender TR, Cole KL, Geoffrey SW. **2000**. Ploidy race distributions since the Last Glacial Maximum in the North American desert shrub, *Larrea tridentata*. *Geologia Sudetica* **33**: 521–533.

Huson DH, Bryant D. **2006**. Application of Phylogenetic Networks in Evolutionary Studies. *Molecular Biology and Evolution* **23**: 254–267.

Jakobsson M, Rosenberg NA. **2007**. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* **23**: 1801–1806.

Jombart T. **2008**. adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* **24**: 1403–1405.

Jones GR. **2014**. Species delimitation and phylogeny estimation under the multispecies coalescent. *bioRxiv*.

Kühne G, Kosuch J, Hochkirch A, Schmitt T. **2017**. Extra-Mediterranean glacial refugia in a Mediterranean faunal element: the phylogeography of the chalk-hill blue *Polyommatus coridon* (Lepidoptera, Lycaenidae). *Scientific Reports* **7**: 43533.

Levin DA. **2002**. *The role of chromosomal change in plant evolution*. Oxford ; New York: Oxford University Press.

Liu K, Raghavan S, Nelesen S, Linder CR, Warnow T. **2009**. Rapid and Accurate Large-Scale Coestimation of Sequence Alignments and Phylogenetic Trees. *Science* **324**: 1561–1564.

Mansion G, Zeltner L, Bretagnolle F. **2005**. Phylogenetic Patterns and Polyploid Evolution within the Mediterranean Genus Centaurium (Gentianaceae - Chironieae). *Taxon* **54**: 931–950.

Marcussen T, Sandve SR, Heier L, Spannagl M, Pfeifer M, The International Wheat Genome Sequencing Consortium, Jakobsen KS, Wulff BBH, Steuernagel B, Mayer KFX, *et al.* 2014. Ancient hybridizations among the ancestral genomes of bread wheat. *Science* **345**: 1250092–1250092.

Marques I, Loureiro J, Draper D, Castro M, Castro S. **2018**. How much do we know about the frequency of hybridisation and polyploidy in the Mediterranean region? (J Arroyo, Ed.). *Plant Biology* **20**: 21–37.

Médail F, Diadema K. **2009**. Glacial refugia influence plant diversity patterns in the Mediterranean Basin. *Journal of Biogeography* **36**: 1333–1345.

Médail F, Quézel P. **1999**. Biodiversity Hotspots in the Mediterranean Basin: Setting Global Conservation Priorities. *Conservation Biology* **13**: 1510–1513.

Meimberg H, Rice KJ, Milan NF, Njoku CC, McKay JK. **2009**. Multiple origins promote the ecological amplitude of allopolyploid *Aegilops* (Poaceae). *American Journal of Botany* **96**: 1262–1273.

Middleton CP, Senerchia N, Stein N, Akhunov ED, Keller B, Wicker T, Kilian B. **2014**. Sequencing of Chloroplast Genomes from Wheat, Barley, Rye and Their Relatives Provides a Detailed Insight into the Evolution of the Triticeae Tribe (MW Davey, Ed.). *PLoS ONE* **9**: e85761.

Morjan CL, Rieseberg LH. **2004**. How species evolve collectively: implications of gene flow and selection for the spread of advantageous alleles. *Molecular Ecology* **13**: 1341–1356.

Nei M. **1978**. Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* **89**: 583–590.

Novikova PY, Hohmann N, Van de Peer Y. **2018**. Polyploid *Arabidopsis* species originated around recent glaciation maxima. *Current Opinion in Plant Biology* **42**: 8–15.

Oxelman B, Brysting AK, Jones GR, Marcussen T, Oberprieler C, Pfeil BE. **2017**. Phylogenetics of Allopolyploids. *Annual Review of Ecology, Evolution, and Systematics* **48**: 543–557.

Parisod C, Besnard G. **2007**. Glacial in situ survival in the Western Alps and polytopic autopolyploidy in *Biscutella laevigata* L. (Brassicaceae). *Molecular Ecology* **16**: 2755–2767.

Petersen G, Seberg O, Yde M, Berthelsen K. **2006**. Phylogenetic relationships of *Triticum* and *Aegilops* and evidence for the origin of the A, B, and D genomes of common wheat (*Triticum aestivum*). *Molecular Phylogenetics and Evolution* **39**: 70–82.

Pimentel M, Sahuquillo E. **2007**. Infraspecific variation and phylogeography of the high-polyploid Iberian endemic *Anthoxanthum amarum* Brot. (Poaceae; Pooideae) assessed by random amplified polymorphic DNA markers (RAPDs) and morphology. *Botanical Journal of the Linnean Society* **155**: 179–192.

Pons O, Petit RJ. **1996**. Measuring and testing genetic differentiation with irdered *versus* unordered alleles. *Genetics* **144**: 1237–1245.

Pritchard JK, Stephens M, Donnelly P. **2000**. Inference of population structure using multilocus genotype data. *Genetics* **155**: 945–959.

Provan J, Bennett K. **2008**. Phylogeographic insights into cryptic glacial refugia. *Trends in Ecology & Evolution* **23**: 564–571.

Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. **2018**. Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7 (E Susko, Ed.). *Systematic Biology*.

Ramsey J, Schemske DW. **1998**. Pathways, mechanisms, and rates of polyploid formation in flowering plants. *Annual Review of Ecology and Systematics* **29**: 467–501.

Remington CL. **1968**. Suture-Zones of Hybrid Interaction Between Recently Joined Biotas. In: Evolutionary Biology. Springer, Boston, MA, 321–428.

Robinson DF, Foulds LR. **1981**. Comparison of phylogenetic trees. *Mathematical Biosciences* **53**: 131–147.

Rousset. **2000**. Genetic differentiation between individuals. *Journal of Evolutionary Biology* **13**: 58–62.

Schmitt T, Varga Z. **2012**. Extra-Mediterranean refugia: The rule and not the exception? *Frontiers in Zoology* **9**: 22.

Senerchia N, Felber F, Parisod C. **2014**. Contrasting evolutionary trajectories of multiple retrotransposons following independent allopolyploidy in wild wheats. *New Phytologist* **202**: 975–985.

van Slageren MWSJM. **1994**. *Wild wheats: a monograph of* Aegilops *L. and* Amblyopyrum *(Jaub. & Spach) Eig (Poaceae): a revision of all taxa closely related to wheat, excluding wild* Triticum *species, with notes on other genera in the tribe Triticeae, especially* Triticum. Wageningen, The Netherlands : Aleppo, Syria: Wageningen Agricultural University : International Center for Agricultural Research in the Dry Areas.

Stamatakis A. **2014**. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**: 1312–1313.

Stebbins GL. **1950**. *Variation and evolution in plants*. New York, NY.

Stebbins GL. **1984**. Polyploidy and the distribution of the arctic-alpine flora : new evidence and a new approach. *Botanica Helvetica* **94**: 1–13.

Taberlet P, Cheddadi R. **2002**. Quaternary refugia and persistence of biodiversity. *Science* **297**: 2009–2010.

Taberlet P, Fumagalli L, Wust-Saucy A-G, Cosson J-F. **1998**. Comparative phylogeography and postglacial colonization routes in Europe. *Molecular Ecology* **7**: 453–464.

Tayalé A, Parisod C. **2013**. Natural Pathways to Polyploidy in Plants and Consequences for Genome Reorganization. *Cytogenetic and Genome Research* **140**: 79–96.

Thompson JD. **2005**. *Plant Evolution in the Mediterranean*. Oxford, New York: Oxford University Press.

Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, Rozen SG. **2012**. Primer3—new capabilities and interfaces. *Nucleic Acids Research* **40**: e115–e115.

Varga Z. **2010**. Extra-Mediterranean Refugia, Post-Glacial Vegetation History and Area Dynamics in Eastern Central Europe. In: Habel JC, Assmann T, eds. Relict Species. Berlin, Heidelberg: Springer Berlin Heidelberg, 57–87.

Wang J, Luo M-C, Chen Z, You FM, Wei Y, Zheng Y, Dvorak J. **2013**. *Aegilops tauschii* single nucleotide polymorphisms shed light on the origins of wheat D-genome genetic diversity and pinpoint the geographic origin of hexaploid wheat. *New Phytologist* **198**: 925–937.

Wielstra B, Zieliński P, Wiesław B. **2017**. The Carpathians hosted extra-Mediterranean refugia-within-refugia during the Pleistocene Ice Age: genomic evidence from two newt genera. *Biological Journal of the Linnean Society* **122**: 605–613.

Zohary D, Feldman M. **1962**. Hybridization Between Amphidiploids and the Evolution of Polyploids in the Wheat (*Aegilops-Triticum*) Group. *Evolution* **16**: 44-61.

Supplementary Material

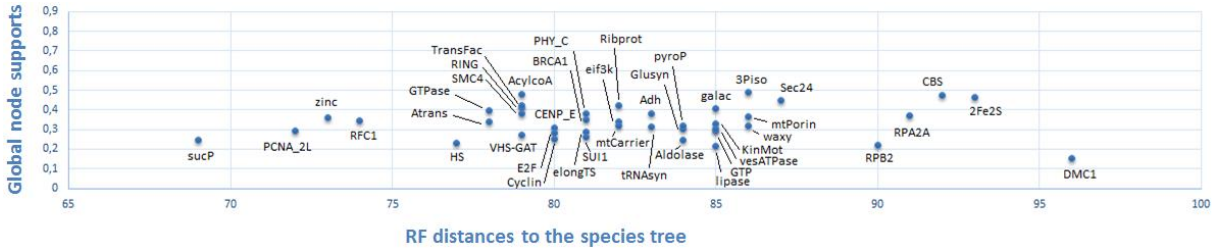
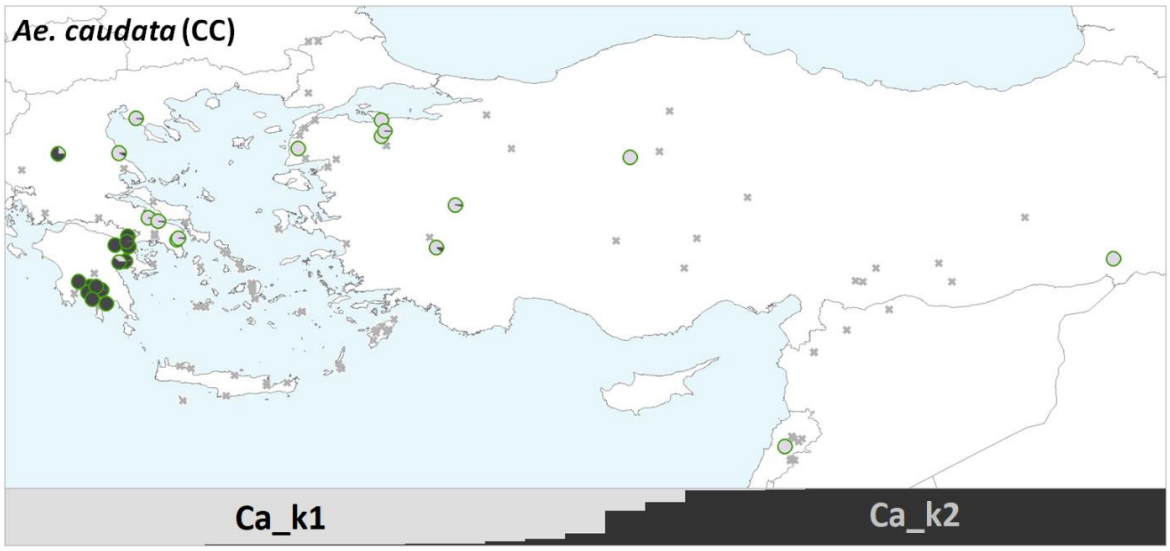


Figure S1: Distribution of gene trees for each nuclear locus based on their global node support and Robinson-Fould distance values. The full name and details of each locus are indicated in the electronic supplemental material Table S2 from Chapter 1: "StellaHuynh_PhD_Chapter1_ SuppMat_TableS2–ListOf48LowCopyLoci.xlsx".

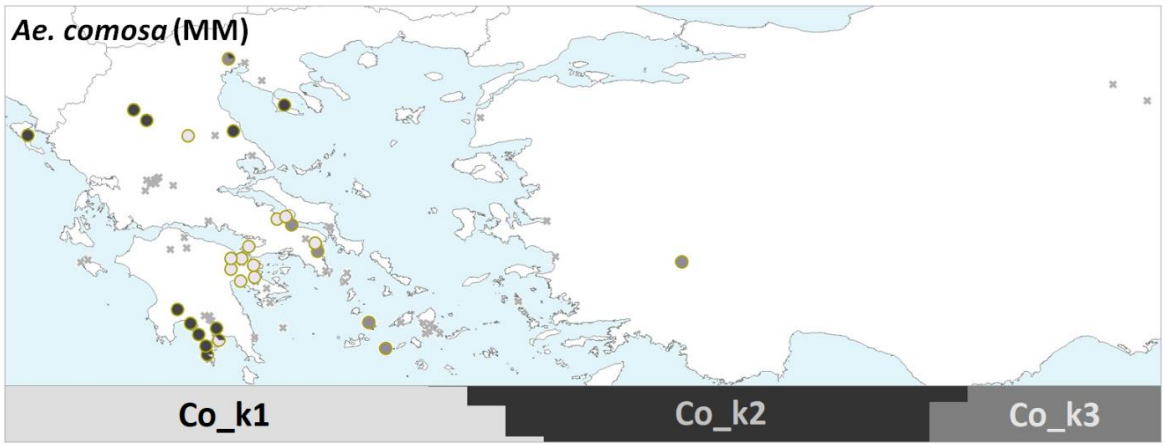
Figure S2: Gene trees of the 43 low-copy loci (Electronic Material). The electronic material can be found under the following PowerPoint document: "StellaHuynh_PhD_Chapter2_SuppMat_FigureS2-GeneTreesOf43Loci.pptx" For each loci, the gene tree was inferred based on all unique haplotypes sequenced among the studied species using BEAST2 for 500 million MCMC generations and 10% burnin. Trees were summarized with TreeAnnotator with 80 % burnin and then viewed in FigTree. Tips of summarized trees were manually labelled with species or individual sharing the corresponding haplotype(s) followed by a number (corresponding to the number of single haplotypes sharing the tip) or by "x" (corresponding to more than 4 single haplotypes sharing the tip) into brackets. Tips only represented by a single MCDU *Aegilops* species were colored accordingly to the species as follows: orange (*Ae. caudata*), pink-purple (*Ae. comosa*), light pink (*Ae. crassa*), yellow (*Ae. cylindrica*), red (*Ae. geniculata*), turquoise (*Ae. tauschii*), green (*Ae. triuncialis*) and dark blue (*Ae. umbellulata*).

Figure S3: Detailed population structure of the eight MCDU *Aegilops* species. Each genotyped accession is represented by a pie chart with gradients of grey corresponding to the genetic clusters inferred by STRUCTURE under admixture model. The outer color of pie charts corresponds to the genome associated to the genetic cluster as follows: (M) as dark yellow, (C) as green, (D) as gradients of red and (U) as gradients of blue. Occurrences of each species as recorded in GBIF database are represented by grey crosses.

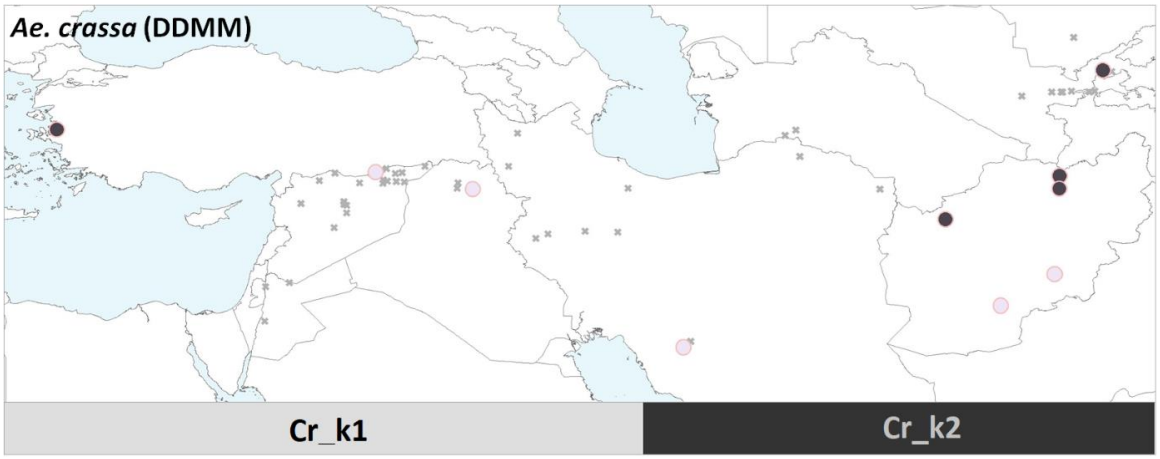
(a)



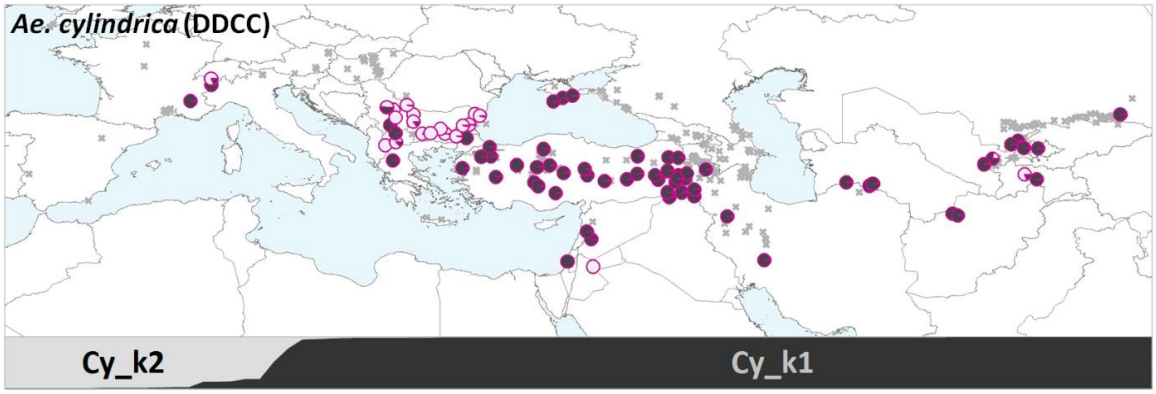
(b)



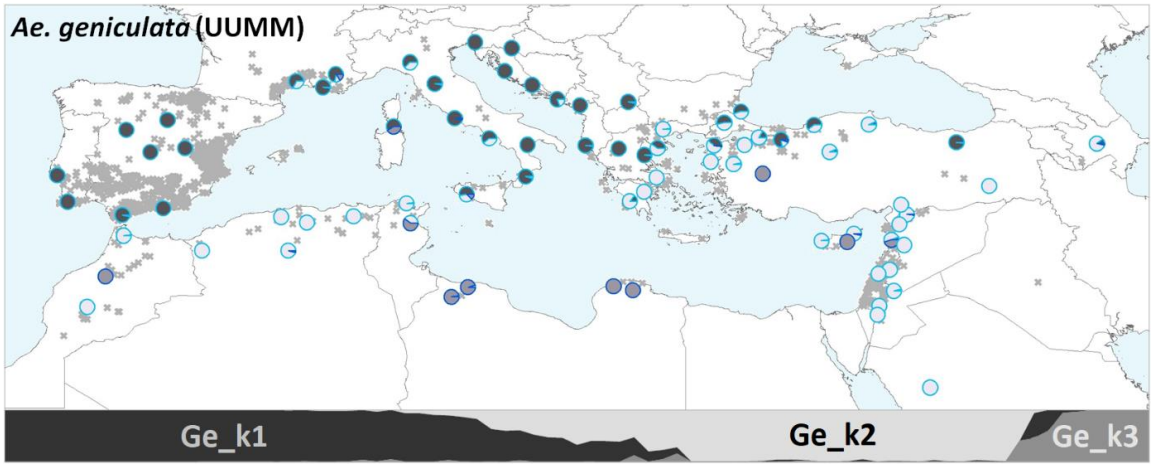
(c)



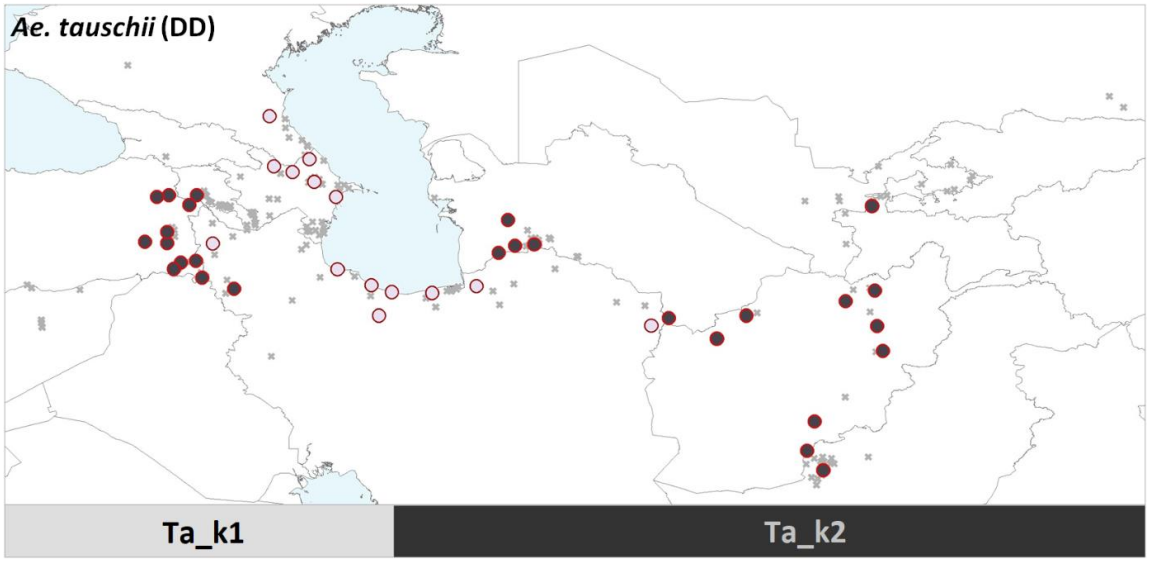
(d)


Ae. cylindrica (DDCC)
Cy_k2
Cy_k1

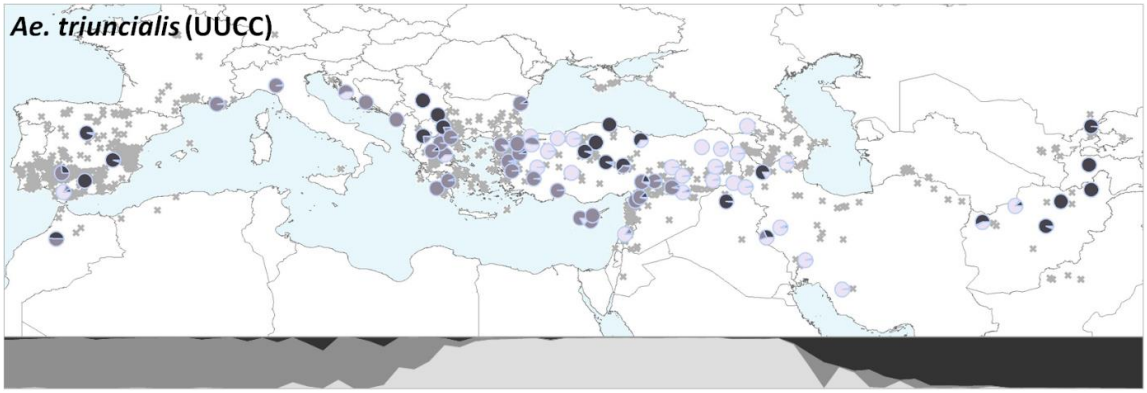
(e)


Ae. geniculata (UUMM)
Ge_k1
Ge_k2
Ge_k3

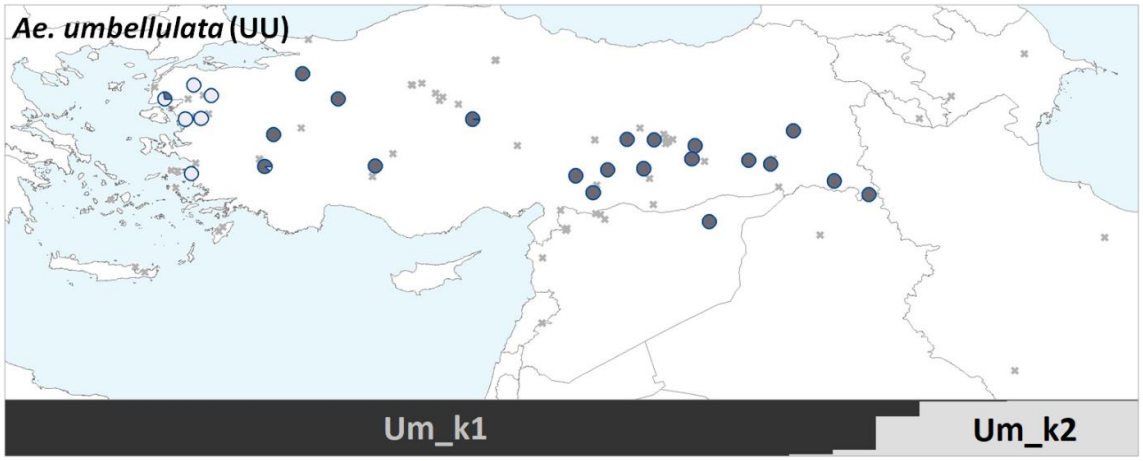
(f)


Ae. tauschii (DD)
Ta_k1
Ta_k2

(g)



(h)



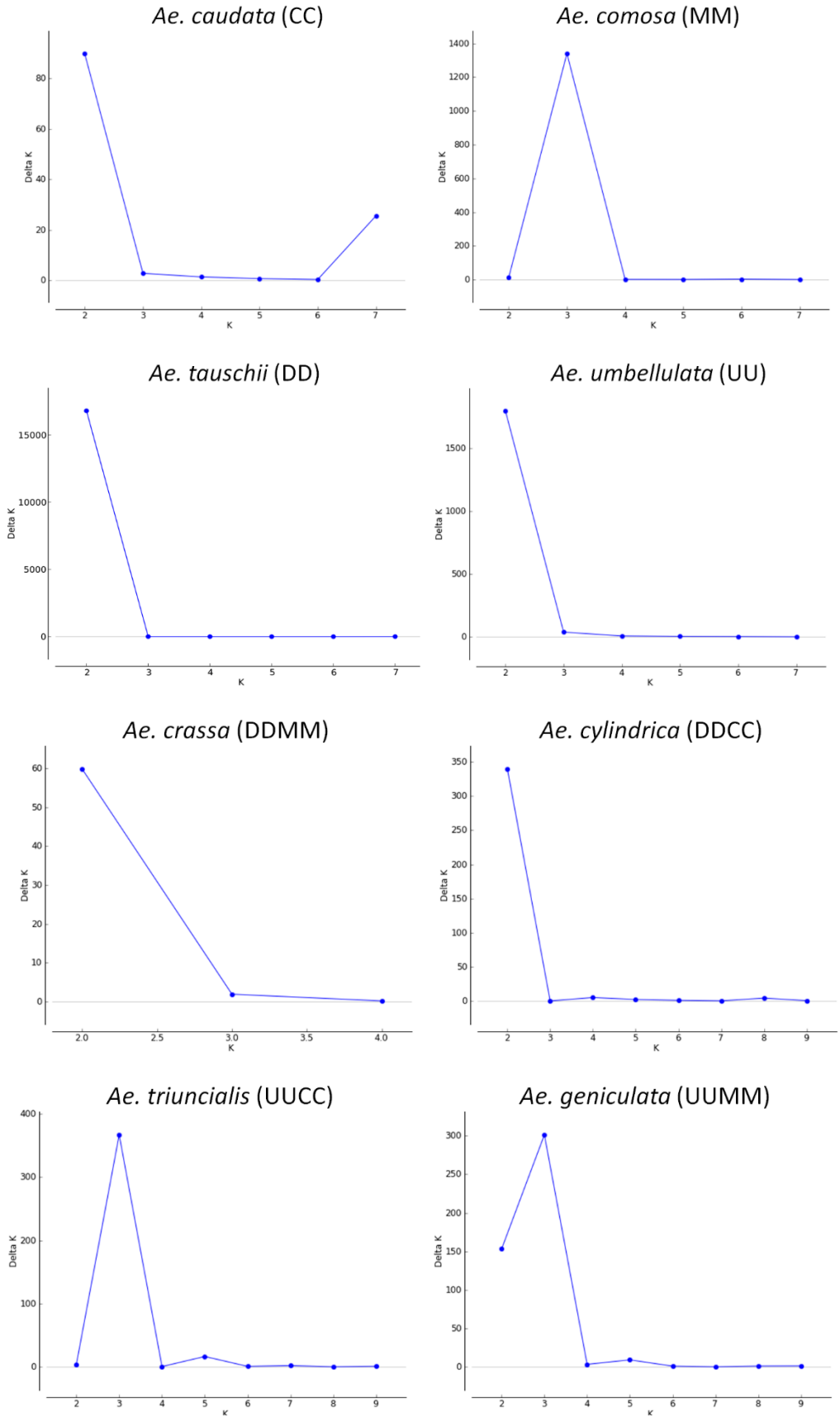


Figure S4: Optimal number of genetic clusters of the eight *Aegilops* species. The calculation of optimal number of genetic clusters was based on deltaK following Evanno *et al.* (2005).

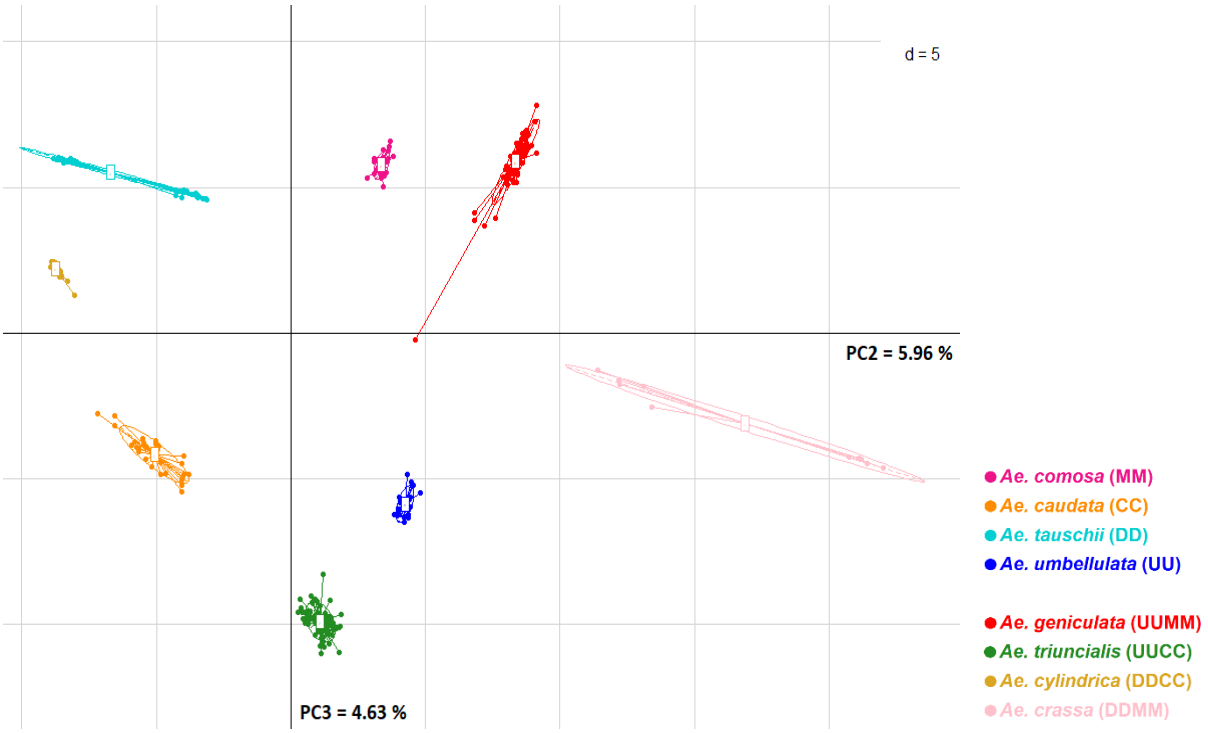


Figure S5: Genetic differentiation among the eight MCDU *Aegilops* species assessed by Principal Component Analysis (PCA). The PCA is based on haplotypes frequencies for 30 nuclear loci. The first PC, which explains 6.58% of the total genetic variance, is only driven by *Ae. crassa*. The second and third PCs, summarizing 5.96% and 4.63% of the variance respectively, however better discriminates the eight *Aegilops* species. Colors corresponding to each species is shown in the above legend.

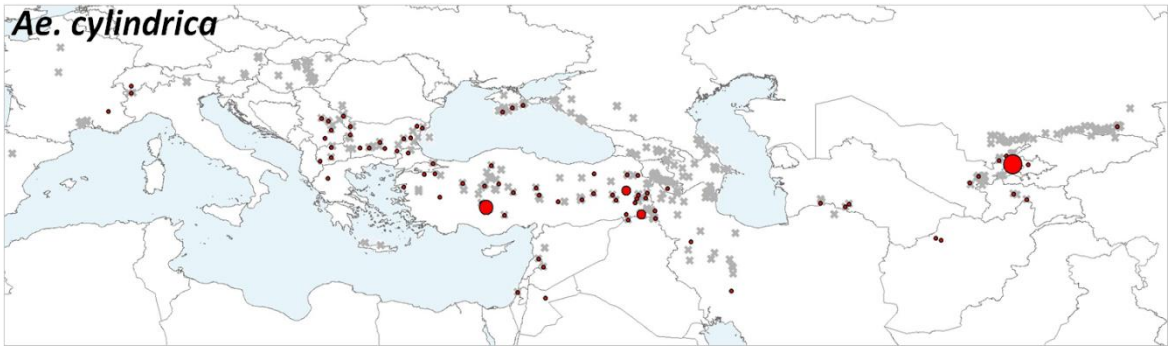
Figure S6: Individual observed heterozygosity (Ho) within *Aegilops* species. Individual Ho was calculated as the proportion of heterozygous loci among the 30 nuclear loci for each genotyped individual. For each species (a to h) and all eight *Aegilops* species (i), individuals are represented by a circle proportional to their Ho value divided into 2-8 classes with break values defined following the quantile method implemented in ArcGIS.

(a)

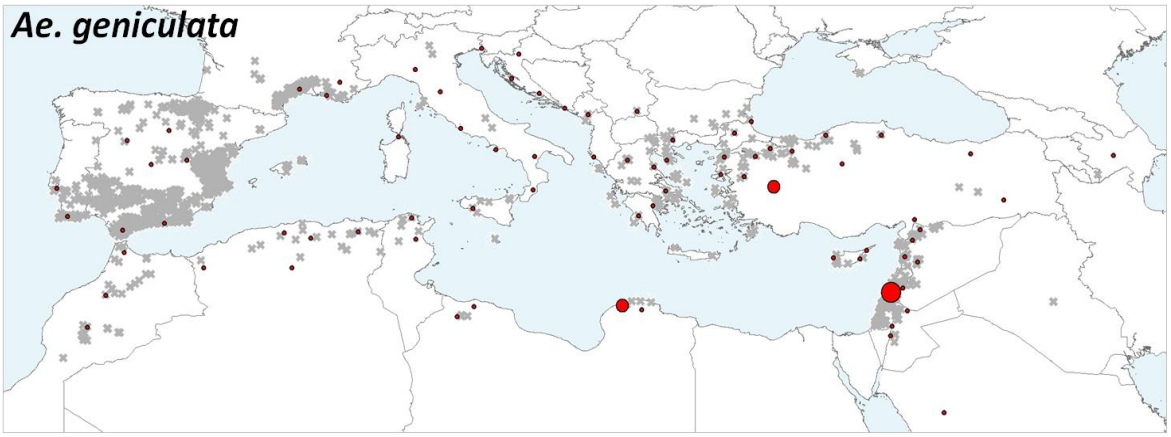
(b)

(c)

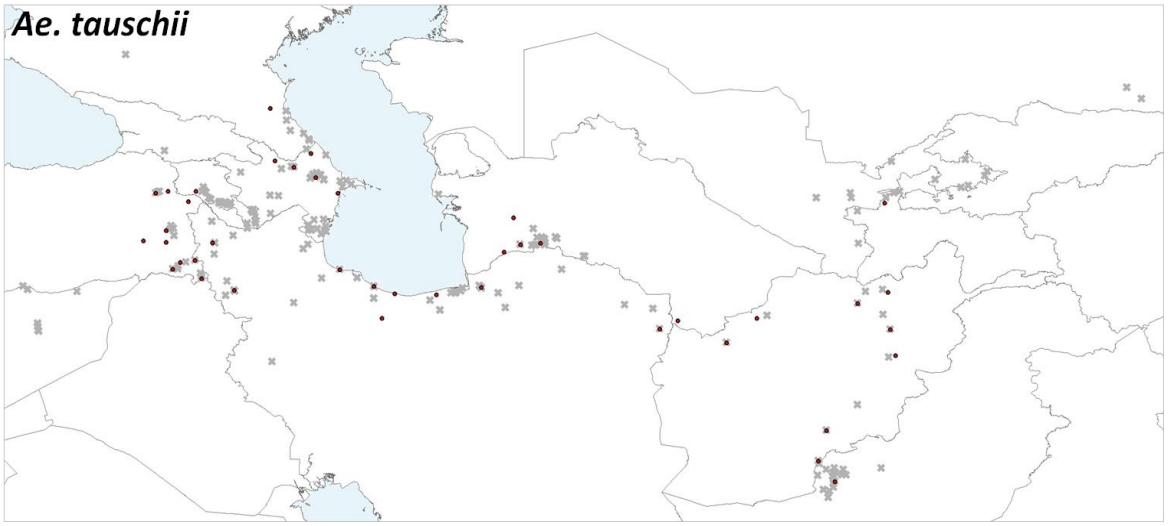
(d)



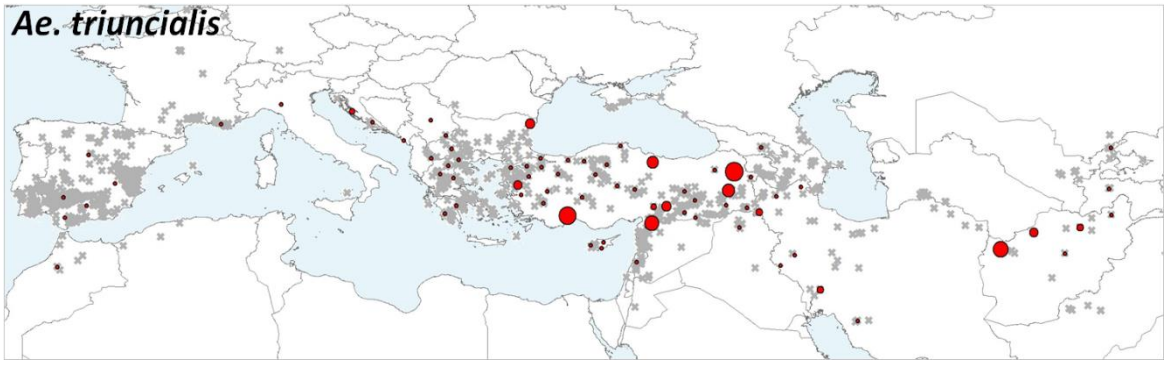
(e)



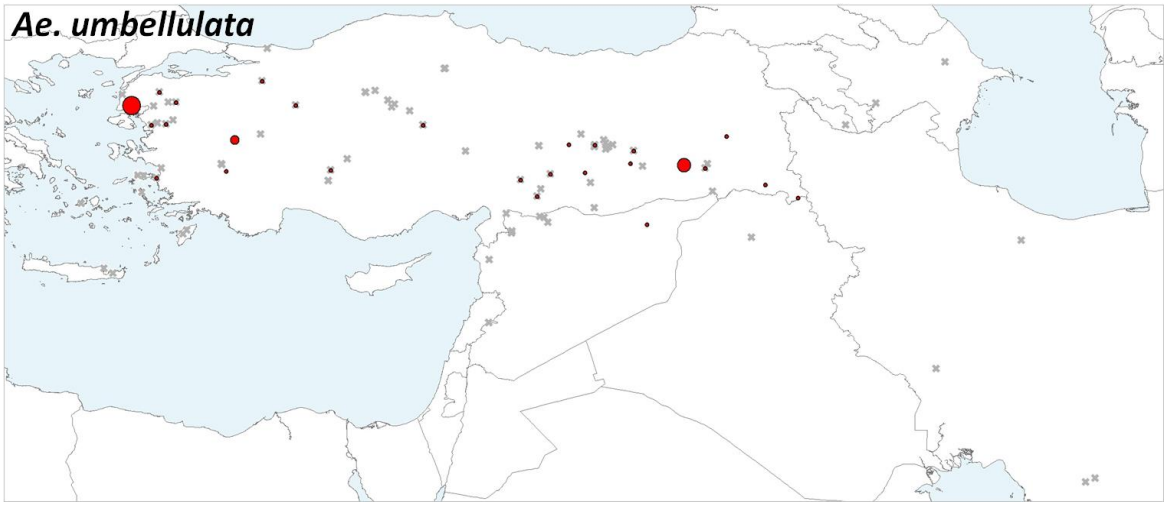
(f)



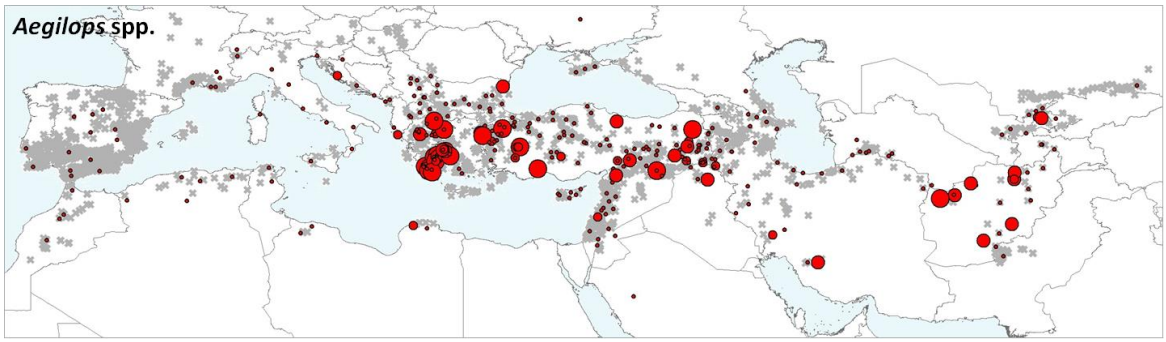
(g)



(h)



(i)



Chapter 3

Dynamics of retrotransposons across the distribution ranges of allopolyploid wild wheats

Stella Huynh[1], François Felber[1,2] and Christian Parisod[3]

[1]Institute of Biology, University of Neuchâtel, Switzerland

[2]Musée et Jardins botaniques cantonaux de Lausanne et Pont-de-Nant, Switzerland

[3]Institute of Plant Sciences, University of Bern, Switzerland

List of figures and tables

Abstract

Transposable elements (TEs) play an important role in plant evolution and diversification. In particular, long-terminal repeat retrotransposons (LRT-RT) represent the major fraction of crop genomes and are rapidly activated following allopolyploidization. However, the long-term evolution of LTR-RT copies within their allopolyploid hosts remains largely unknown. We thus investigated TE dynamics in four allopolyploid *Aegilops* species across their natural distribution ranges and tested the non-additivity of progenitors' loci to distinguish new loci from lost loci within allopolyploids. Six retrotransposons families were amplified using SSAP fingerprinting and compared to genome-wide AFLP to distinguish recent TE proliferation or deletion (SSAP) from random genetic changes (AFLP) in 402 individuals from the four allopolyploid species and their four diploid progenitor species. Our study highlighted long-term evolution of TE copies that is species-specific and likely driven by random processes. It was also coherent with punctual re-activation of TE copies from specific LTR-RT families in accessions locared in the Balkans-Anatolia suture zone and at the margins of species distribution. Losses of TE copies were mainly observed in the differential genomes, hypothesized as evolving through homeologous recombinations, than in the pivotal genomes, also called 'unaltered' as evolving through homologous recombinations. Although plausible, the hypothesis of uncontrolled paternal TE copies by the maternal siRNA machinery cannot fully explain the reactivation or asymmetrical losses of TEs observed in the allopolyploids, especially in *Ae. triuncialis* considering its bi-maternal origin. The differential TE losses between allopolyploid genomes however support the pivotal-differential genome hypothesis. Nevertheless, other molecular mechanisms or extrinsic processes may also underlie the evolution of pivotal and differential genomes within these allopolyploid species.

Keywords: transposable elements, LTR retrotransposons, allopolyploid, *Aegilops*, individual TE dynamics, SSAP.

Introduction

Transposable elements (TEs), also called "selfish genes", are repetitive DNA fragments found in all organisms and whose primary function stands in self-replicating throughout the genome. TEs thus represent the most dynamic part of the genome and have played an important role in species evolution and diversification (Feldman & Levy, 2005; Lisch, 2013; Oliver *et al.*, 2013; Belyayev, 2014; Senerchia *et al.*, 2015). TEs replicate through two processes and are classified accordingly (Wicker *et al.*, 2007): through a "copy-and-paste" mechanism (Class I), which necessitate an RNA intermediate, or through a "cut-and-paste" mechanism (Class II). TEs from Class I are particularly predominant among plants and are mainly represented by long terminal repeat retrotransposons (LTR-RT; Kumar & Bennetzen, 1999). Retrotransposons constitute up to 90 % of the genome among crops (Tenaillon *et al.*, 2010; Luo *et al.*, 2013) and might thus have played an important role in their evolutionary history. Hybridization and polyploidization events, which is known recurrent throughout the evolution of crops (Feldman & Levy, 2015), are indeed the major activators of TE transposition (i.e. genome shock; Parisod & Senerchia, 2012).

Allopolyploidy, i.e. the merging of genetically divergent genomes into a single organism by a combination of hybridization and whole genome doubling, provides an ideal substrate for investigating TE dynamics across large plant populations. Allopolyploids show a general tendency to genome downsizing compared to the expected additivity of their progenitor genome (Song *et al.*, 1995; Ozkan, 2003; Leitch *et al.*, 2008), indicative of substantial genome restructuring and epigenetic changes (Doyle *et al.*, 2008). Allopolyploid genomes indeed regulate the number of TE copies through either stable mechanisms, i.e. genomic rearrangements suppressing large portions of DNA containing TE copies (Montgomery *et al.*, 1991; Devos, 2002), or labile mechanisms, i.e. epigenetic silencing via maternally-inherited cytoplasmic small interference RNA (siRNA) repressing nuclear TE activity (Lisch & Slotkin, 2011; Fedoroff, 2012). Such changes may occur immediately following allopolyploidization (i.e. 'revolutionary changes') but can also take place during the species lifetime (i.e. 'evolutionary changes') (Feldman & Levy, 2012). While revolutionary changes contribute to the establishment of nascent allopolyploid species, the evolutionary changes contribute to the buildup of the genetic diversity within the species and therefore is related to the overall fitness and colonizing capacities of the allopolyploid species (Feldman & Levy, 2012). However, the immediate consequences of allopolyploidy on the evolutionary dynamics of retrotransposon copies has been extensively investigated (Kashkush *et al.*, 2003; Feldman

& Levy, 2009; Parisod & Senerchia, 2012) whereas the long-term evolution of TE copies across allopolyploid species has been poorly addressed. Some studies suggested that the long-term evolution of TEs is mainly driven by genetic drift (Le Rouzic *et al.*, 2007; Szitenberg *et al.*, 2016), but whether similar evolutionary processes underlie TEs dynamics across allopolyploid species ranges is unknown.

The *Aegilops* species (Poaceae) are annual, predominantly selfing, wild wheats, with a native distribution including the Mediterranean Sea, Anatolia and Central Asia. *Aegilops* comprises diploid, allotetraploid and allohexaploid species that share a complex evolutionary history involving recurrent homoploid and polyploid hybridization events among the diploids and allopolyploid species (Chapter 1 & 2; van Slageren, 1994; Meimberg *et al.*, 2009). In particular, two diploid species, *Ae. tauschii* (DD) and *Ae. umbellulata* (UU), have recurrently combined with other diploid species to form the *Aegilops* allopolyploid species, among which those sharing a common U or D genome also recurrently hybridize (Zohary & Feldman, 1962; Badaeva *et al.*, 2002, 2004). According to Zohary and Feldman (1962), the common genomes U and D (also called 'pivotal genomes') thus mainly evolved through homologous recombination and thereby remain unaltered, whereas the other recombining genomes (also called 'differential genomes') evolved through homeologous recombinations leading to substantial restructuring of these genomes (the 'pivotal-differential genome' hypothesis). Hybridizations among allopolyploid species sharing common pivotal genomes have indeed been reported and are still recurrently occurring (Feldman, 1965; Pazy & Zohary, 1965; Senerchia *et al.*, 2016). The pivotal-differential genome hypothesis thus results in genetic asymmetry, formerly referring to structural asymmetry (Zohary & Feldman, 1962), but also includes functional asymmetry (Feldman & Levy, 2012), i.e. the evolutionary changes distinctly affecting pivotal and differential genomes through subfunctionalization (division of the initial functions of a gene among its duplicated copies) or neofunctionalization (alteration of the initial functions of a gene). Recurrent interspecific hybridizations, combined with intraspecific hybridizations in the suture zones (see Chapter 2), likely enhance pivotal-differential asymmetry as well as TEs bursts.

Retrotransposons make up to 90 % of the Aegilops genomes according to Luo et al.(2013), among which 50 % are represented by only eight LTR-RT families including Nusif or BARE1 (Choulet et al., 2010). Higher restructuring of the differential genome than of the pivotal genome was mainly supported by Senerchia et al. (2014) through the survey of the evolutionary dynamics of the main LTR-RT families among four allopolyploid Aegilops

species Ae. cylindrica (DDCC), Ae. crassa (DDMM), Ae. geniculata (UUMM) and Ae. triuncialis (UUCC). In contrast, comparison of nuclear low-copy loci among the same four allotetraploid species and their diploid progenitor species – Ae. comosa (MM), Ae. caudata (CC), Ae. tauschii (DD) and Ae. umbellulata (UU) – showed no asymmetrical gene losses toward one of the progenitor genome (see Chapter 2), suggesting that restructuring of differential genomes might differently affect nuclear low-copy loci and TE copies (Shaked et al., 2001).

Whether the pivotal-differential genome restructuring observed by Senerchia *et al.* (2014) is consistent across species ranges or may differ among species lineages remains unknown, as the authors investigated TE dynamics on a limited sampling of three accessions per species. The present study aimed at understanding the evolutionary processes and molecular mechanisms acting on the long-term evolution of LTR-RT families across the distribution range of the four allopolyploid *Aegilops* species. The evolutionary dynamics of TEs, defined here as the deviation of allopolyploid TE copies from the expected addition of both its diploid progenitor copies (i.e. TE proliferation or TE loss), was investigated on six abundant LTR-RT families with contrasted evolutionary dynamics according to Senerchia *et al.* (2014): two quiescent (*Egug*, *Nusif*), two active (*BARE1*, *Romani*) and two with species-specific dynamics (*Sabine* and *Xalax*). Their evolutionary dynamics was assessed through 402 accessions covering the whole distribution of the four allopolyploids and their diploid progenitors (Figure 1). Retrotransposons isolated through sequence-specific amplified polymorphism (SSAP) were compared to genome-wide random sequences assessed by amplified fragment length polymorphism (AFLP) to highlight the dynamics of TEs and their possible geographical components in the light of the evolutionary history of the allopolyploid *Aegilops* species inferred in Chapter 2.

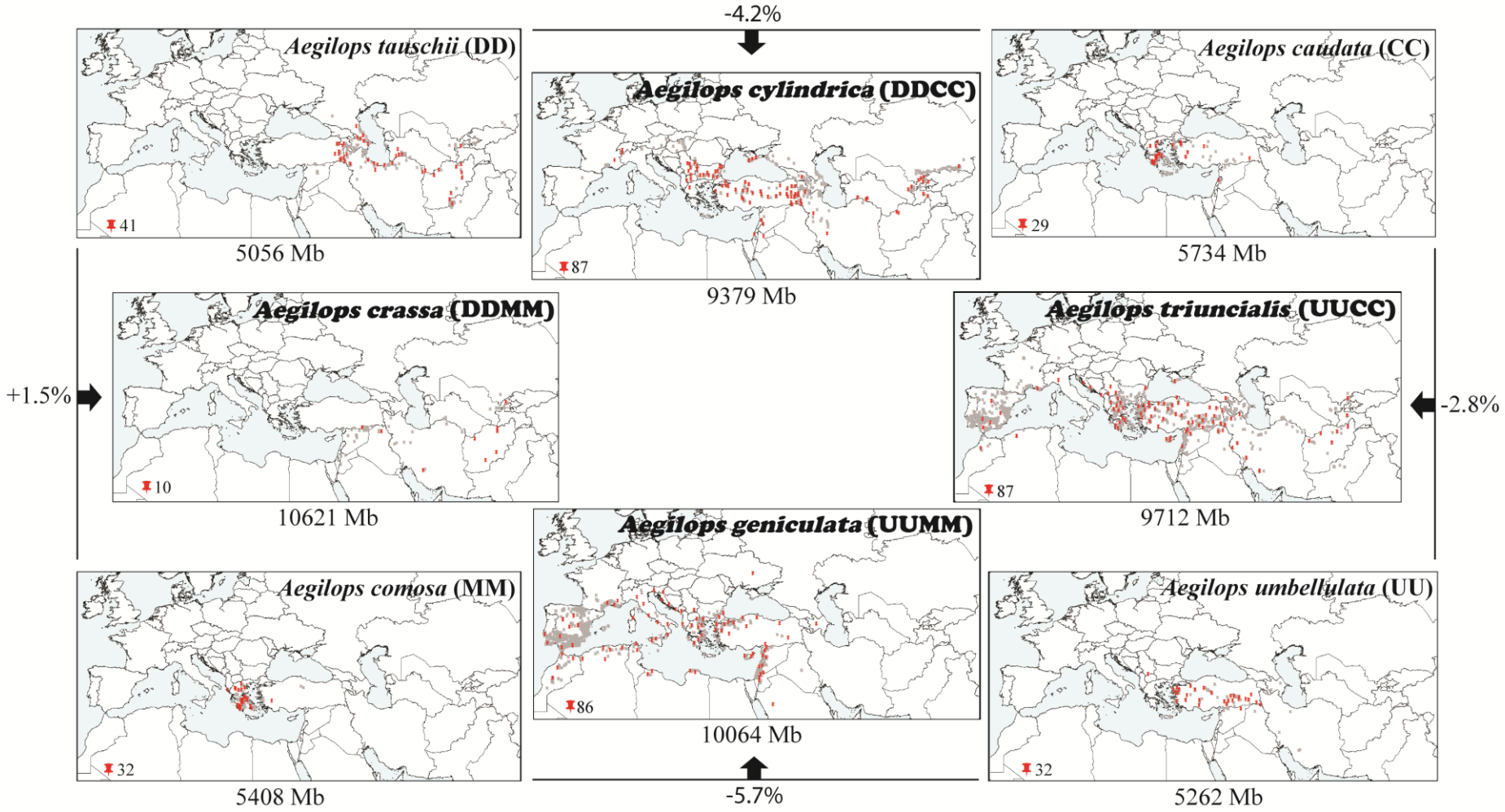


Figure 1: Evolutionary relationships and current distribution of four diploid and their four derivated allopolyploid *Aegilops* species. Species distributions are represented with grey crosses for occurrences reported in GBIF database (https://www.gbif.org/) and red thumbtacks for genotyped individuals. Genomic formula are indicated into brackets following van Slageren (1994). Species genome sizes are reported according to Eilam *et al.* (2007, 2008), as well as the genome size difference (%) in allopolyploids compared to the expected addition of progenitor genome sizes.

Material and Methods

Plant material

A total of 402 accessions of the eight *Aegilops* species were selected for genotyping to represent variation across the species distribution ranges. All accessions were geo-localized except four from Senerchia *et al.* (2014), with an average precision below 5 km. Seeds of about 30 accessions for each diploid and 80 accessions for each allopolyploid species were obtained from ARS-GRIN, ICARDA and IPK Gaterseleben (General Supplemental Material, Table S1). A few additional accessions were obtained from ICARDA germplasm and from field collections (Arrigo *et al.*, 2010; Senerchia *et al.*, 2014). For *Ae. crassa*, whose distribution and seed availability were limited, only 10 accessions were obtained. Seeds were sown and grown outdoor at the Botanical garden of Neuchâtel until flowering and deposited as herbarium sheets at the Botanical Museum and Garden of Lausanne and Pont-de-Nant, Switzerland. Leaves from young seedlings were sampled and dried in silica gel for DNA extraction using DNAeasy kit from Qiagen® following the manufacturer protocol. The quality of each individual DNA was checked on 1% agarose gel, its concentration measured by Nanodrop and then normalized at 10ng/µL. Individuals were randomly positioned in 96-well plates with 12 accessions replicated twice, one accession (*Ae. geniculata*, PI 487229) and one negative control replicated in each plate, for a total of 428 individuals. Our dataset also included 7 out of the 12 allopolyploid accessions genotyped and analyzed by Senerchia *et al.* (2014) for cross-validation of the molecular fingerprinting techniques (Supplementary Material Table S1).

Molecular fingerprinting techniques

Amplified fragment length polymorphism (AFLP) and sequence-specific amplified polymorphism (SSAP) were carried out following the protocol of Parisod and Christin (2008). Briefly, digestion of genomic DNA with *Eco*RI and *Mse*I (New England Biolabs Inc.) and then ligation of double-stranded adaptors with T4 ligase (Promega) were performed at 37°C. After inactivation of enzymes at 65°C for 20 min, preselective amplification was carried out using primers corresponding to the adaptors with GoTaq®DNA polymerase (Promega). Twenty-fold diluted PCR products were selectively amplified with a touchdown PCR using primers with three additional nucleotides. Unlabelled *Mse*I primers were used together with fluorescently labelled *Eco*RI primers for AFLP and with fluorescently labelled LTR-RT family-specific primers designed by Senerchia *et al* (2014) for SSAP (Supplemental Material Table S2). SSAP thus mostly amplified the region encompassing the termini of

retrotransposon insertions and their flanking genomic sequences. To minimize experimental inconsistencies, digestion and ligation steps were performed simultaneously and PCRs were carried out on thermocyclers with fixed ramp rate. Genotyping error rates were estimated based on the 26 replicates (i.e. ~ 6.2 % of the dataset) following Bonin *et al.* (2004). The 402 *Aegilops* accessions were characterized on six *Eco*RI-*Mse*I combinations of AFLP selective primers and on three combinations of TE-specific and *Mse*I selective primers for each LTR-RT family based on Senerchia *et al* (2014) (Supplemental Material Table S2). Selective PCR products amplified with FAM™, VIC® or NED™ fluorescent dye were pooled together with GeneScan™ 500 LIZ™ Size Standard and separated with an ABI 3730XL DNA Analyzer. Resulting electropherograms were visualized and scored with GeneMarker® (SoftGenetics®) using manual bin sets and default peak detection parameters for AFLP. The scoring was manually checked and loci recorded as present (1) or absent (0) in binary matrices.

Binary matrices were further cleaned by removing loci under the following criteria: (i) loci below 60 bp length to prevent size homoplasy (Vekemans *et al.* 2002), (ii) loci that were found differently scored in more than 15 % of the replicates, (iii) loci that were present in less than 5% in the whole dataset, (iv) loci that were present in less than 5% of the individuals within a species.

Genetic diversity, differentiation and structure of Aegilops species assessed by AFLP and SSAP loci

The genetic diversity of each *Aegilops* species was estimated as Nei's diversity (*Ndiv*; Nei, 1978) for AFLP and each LTR-RT family using the Bayesian method with non-uniform prior distribution of allele frequencies (assuming $F_{IS} = 1$) as implemented in AFLPsurv (Vekemans *et al.*, 2002). Pairwise genetic distances among species were calculated in AFLPsurv as Nei's distances (D) modified by Lynch and Milligan (1994) to correct Nei's D for variance in allelic sampling among species. Correlations between matrices of pairwise Nei's genetic distances among species for AFLP with each of the six LTR-RT families were tested by Mantel tests in R with the Pearson method and 100 000 permutations. Significance was assessed with a sequential Bonferroni correction, here α starting at 8.33×10^{-3} for a significance equivalent to 0.05, to account for multiple comparisons (Rice, 1989).

The genetic structure within each *Agilops* species was inferred based on genome-wide AFLP loci using the c-fuzzy means clustering method, a non-model based algorithm enabling investigation of selfing species, for which Hardy-Weinberg equilibrium is not expected. The c-fuzzy means clustering was performed through the *cmeans* function from the package *e107*

on R cran, with a fuzziness parameter of 1.07, and tested for K = 1 to K = 8 through 1000 repetitions for each K, K being the number of genetic clusters. For each species, the best K was chosen based on the intergroup variance that was best maximized, a proxy of clustering accuracy analogous to the strategy proposed by Evanno *et al.* (2005) and adapted by Arrigo *et al.* (2010).

Analyses of allopolyploid band profiles

The intraspecific genetic variation captured by the large sampling within each *Aegilops* species allowed to accurately infer the observed band profiles in diploid and allopolyploid, and thereby to better assess the expected allopolyploid band profiles through combinational probabilities following Senerchia *et al.*(2014). For each locus *j*, the observed band profile was estimated as the frequencies of band presence (1) and band absence (0) present at the given locus among accessions of the diploid progenitors ($D_1(1)_j$, $D_1(0)_j$ and $D_2(1)_j$, $D_2(0)_j$, respectively) and for each accession *i* of the derived allopolyploid ($A(1)_{ij}$, $A(0)_{ij}$, respectively). The allopolyploid band profile expected under the additivity of its progenitors band profiles was defined at each locus *j* as the probability of band presence $E(1)_j$ and band absence $E(0)_j$ based on the observed frequencies of band presence and band absence in the progenitors. The expected allopolyploid band profile at each locus *j* was calculated as $E(1)_j = D_1(1)_j \times D_2(1)_j + D_1(1)_j \times D_2(0)_j + D_1(0)_j \times D_2(1)_j$ and $E(0)_j = D_1(0)_j \times D_2(0)_j$ following Senerchia *et al.* (2014). The observed allopolyploid band profile of each accession *i* was then compared to the expected allopolyploid band profile at each locus *j*. Probabilities of additive bands (i.e. identity between expected and observed allopolyploid profiles) was assessed as $E(1)_j \times A(1)_j + E(0)_j \times A(0)_j$. Non-additive profiles were assessed as probabilities of new bands (i.e. presence of a band in the observed allopolyploid profile despite absence in the expected profile) as $E(0)_j \times A(1)_j$ or of lost bands (i.e. band absence in the observed allopolyploid profile despite predicted presence in the expected profile) as $E(1)_j \times A(0)_j$.

Individual probabilities of additive (A), new (N) and lost (L) bands calculated for each locus were summed across loci and divided by the total number of bands to obtain individual proportions of A, N and L bands for AFLP and each LTR-RT family. For each allopolyploid accession, the ratio (R) of lost bands divided by new bands was individually calculated and log-transformed so that R = log(L/N). The ratio R was interpreted as TE proliferation if log(L/N) < 0 or TE deletion if log(L/N) > 0.

The probabilities of A, N and L bands and the ratio R assessed for the seven accessions taken from Senerchia *et al.* (2014) were compared with the probabilities formerly inferred by these authors to evaluate the impact of large sampling across species ranges on molecular fingerprinting results.

Origin of lost bands in allopolyploid species

The origin of lost bands was assessed for each tetraploid species on loci presenting band presence in only one progenitor (i.e. either the pivotal or the differential genome in the allopolyploid). For each allopolyploid species, and then for each LTR-RT family, differences in proportions of lost bands among pivotal and differential genomes were assessed using one-way ANOVA test on R cran after validation of distribution normality (not shown).

Departure of TE dynamics (SSAP) from genome-wide activity (AFLP)

For each LTR-RT family, individual proportions of A, N and L bands and R were compared with corresponding individual proportions measured for AFLP to highlight TE-specific restructuring from 'background noise', i.e. technical noise due to PCR-based fingerprinting methods unlinked to TE-specific activity as assessed by AFLP. Comparisons were carried out for each allopolyploid species through one-way ANOVA followed by pairwise comparisons with Tukey's honest significant differences (HSD) *post hoc* tests to assess LTR-RT families significantly differing from AFLP at 5% levels, using the package *agricolae* on R cran.

To investigate the intraspecific TE dynamics, the departure of individual proportions of A, L and N bands and the ratio R measured for each LTR-RT family from the corresponding proportions measured for AFLP was assessed and referred to as "z-scores". For each allopolyploid accession and LTR-RT family, the individual z-score was calculated as the ratio of band proportion (A, L, N or R) from the LTR-RT family divided by the corresponding band proportion from AFLP, and the ratio was further subtracted by 1 to obtain positive or negative z-score. Standard deviations (sd) of individual z-scores (z) were assessed within species for each LTR-RT family and used as significance threshold for interpreting individual TE dynamics as follows: TEs were considered as significantly eliminated if $z < -0.5 \times sd$, proliferating if $z > 0.5 \times sd$, or quiescent if $-0.5 \times sd < z < 0.5 \times sd$. For each LTR-RT family and category of band proportions (A, L, N and R), the individual z-scores were mapped over the geographical location of each accession with corresponding TE activity (only figures corresponding to N and R bands were shown here). To detect geographically structured patterns of recent TE activity, spatial autocorrelation tests were performed across each allopolyploid species range and for each LTR-RT family and each category of band

proportions (A, L, N and R) using *EcoGenetics* on R cran (Roser *et al.*, 2017). Spatial autocorrelation was assessed using the Moran's I statistics and the k-nearest method (k) as the weighting mode for local analysis, with k=10 (i.e. using the 10 nearest neighbors to assess Moran's I) and significance assessed with 100'000 permutations.

Results

Comparison of individual genotyping from AFLP and SSAP techniques

The stringent scoring resulted in 522 bands for AFLP and from 146 bands (*Sabine*) to 287 bands (*Egug*) for SSAP loci (Supplemental Material Table S3), with an error rate ranging from 3.31 % for AFLP to 5.84 % for *Nusif* (Supplemental Material Table S2). The correlation coefficient *r* between fragment sizes and their relative frequencies ranged from *r* = -0.6827 for *Nusif* to *r* = -0.2960 for *Romani*, indicating relatively high frequency of short fragments in our dataset. Comparison of the seven accessions in common with Senerchia *et al.* (2014) revealed relatively consistent band proportions (Supplemental Material Table S1), with a majority of additive bands (A = 0.704 – 0.891) followed by lost bands (L = 0.085 – 0.213) and a minority of new bands (N = 0.025 – 0.112).

Genetic diversity, differentiation and structure in diploid and tetraploid Aegilops species

The Nei's genetic diversity (*Ndiv*) varied among species and LTR-RT families (Table 1). The genetic diversity revealed higher among diploids, with *Ndiv* ranging from 0.088 to 0.234 for *Xalax* and *Sabine* in *Ae. caudata*, than among allopolyploids, with *Ndiv* ranging from 0.059 for *BARE1* in *Ae. cylindrica* to 0.170 for *Nusif* in *Ae. geniculata*. In overall, the LTR-RT families *BARE1*, *Sabine* and *Nusif* showed higher *Ndiv* among species than AFLP. Comparisons of Nei's distances between each LTR-RT families and AFLP among pairwise species through Mantel tests showed all LTR-RT families significantly correlated with AFLP after Bonferroni correction (data not shown), indicating similar patterns of divergence among species for LTR-RT families and genome-wide AFLP.

Population structure analyses based on genome-wide AFLP loci showed all *Aegilops* species composed of two to three genetic clusters that were geographically structured (Supplemental Material Figure S1). Three genetic clusters were found for the diploids *Ae. comosa* (M) and *Ae. tauschii* (D) as well as the allopolyploids *Ae. geniculata* (UM) and *Ae. triuncialis* (UC) whereas two genetic clusters were detected for the other species. *Ae. geniculata* appear genetically uniform across its distribution except in Italy where the three genetic clusters co-occur. *Ae. cylindrica* was genetically mostly uniform with no clear geographic structure of its two genetic clusters. *Ae. tauschii* revealed a finer genetic structure of the accessions occurring along the coast of the Caspian Sea, with a small genetic cluster (Ta_k3) located on the South coast and spreading eastward. The genetic structure of *Ae. crassa* consistently distinguished

Table 1: Genetic diversity within *Aegilops* species. Nei's gene diversity within diploid and allopolyploid species is presented with standard-errors (±) for random sequences (assessed by AFLP) as well as the 6 TE families (assessed by SSAP).

	Ae. caudata **(genome C)**	***Ae. comosa*** **(genome M)**	***Ae. tauschii*** **(genome D)**	***Ae. umbellulata*** **(genome U)**	***Ae. crassa*** **(genome DM)**	***Ae. cylindrica*** **(genome DC)**	***Ae. geniculata*** **(genome UM)**	***Ae. triuncialis*** **(genome UC)**
AFLP	0.121 ± 0.007	0.144 ± 0.008	0.138 ± 0.008	0.101 ± 0.007	0.102 ± 0.008	0.063 ± 0.005	0.140 ± 0.007	0.085 ± 0.006
Egug	0.096 ± 0.009	0.161 ± 0.011	0.132 ± 0.010	0.107 ± 0.010	0.121 ± 0.012	0.060 ± 0.007	0.139 ± 0.010	0.087 ± 0.008
Nusif	0.163 ± 0.014	0.151 ± 0.013	0.133 ± 0.013	0.167 ± 0.014	0.160 ± 0.015	0.103 ± 0.011	0.170 ± 0.013	0.144 ± 0.012
BARE1	0.114 ± 0.011	0.175 ± 0.012	0.156 ± 0.012	0.131 ± 0.011	0.118 ± 0.012	0.059 ± 0.007	0.138 ± 0.010	0.127 ± 0.010
Romani	0.091 ± 0.012	0.111 ± 0.012	0.116 ± 0.013	0.063 ± 0.010	0.094 ± 0.014	0.068 ± 0.009	0.110 ± 0.011	0.061 ± 0.010
Sabine	0.234 ± 0.017	0.128 ± 0.015	0.117 ± 0.014	0.120 ± 0.015	0.103 ± 0.016	0.106 ± 0.012	0.109 ± 0.012	0.147 ± 0.014
Xalax	0.088 ± 0.010	0.125 ± 0.011	0.100 ± 0.010	0.081 ± 0.010	0.071 ± 0.010	0.061 ± 0.008	0.120 ± 0.010	0.086 ± 0.009

the tetraploid (Cr_k2) and hexaploid (Cr_k1) accessions, in the East and West part of the species distribution respectively, as previously assessed in Chapter 2.

Contrasted allopolyploid genome restructuring among LTR-RT families

Allopolyploid band profiles, i.e. proportions of additive (A), new (N) and lost (L) bands and the log-scaled ratio L/N (R), showed differential deviation from expected progenitor additivity among species and LTR-RT families (Figure 2). All four allopolyploid species showed mainly additive (Supplemental Material Table S4), with proportions of A bands around 0.80, and higher genetic losses (L bands ranging from 0.095 to 0.196) than genetic novelties (N bands ranging from 0.033 to 0.087).

All LTR-RT families showed contrasted evolutionary trajectories among species (Figure 2). The LTR-RT families *Nusif* and *Egug*, assessed as quiescent in Senerchia *et al.* (2014), showed here significantly different from random sequences (AFLP) with higher proportion of N and L bands among *Ae. geniculata* for *Egug* and higher proportion of L bands among *Ae. triuncialis* for *Nusif*. The expected active LTR-RT families *BARE1* and *Romani* showed significantly high proportions of N bands and a low R (i.e. TE proliferation) among *Ae. triuncialis* for *BARE1* and among *Ae. cylindrica* for *Romani*. The LTR-RT families *Sabine* and *Xalax* showed constrasted evolutionary trajectories but differing from the trajectories previously assessed by Senerchia *et al.* (2014). *Sabine* similarly showed significantly low proportions of N bands and high R (i.e. TE elimination) among *Ae. geniculata* but high proportions of L and N bands among *Ae. triuncialis* and *Ae. cylindrica* indicating a high TE dynamics with turnover between deletion and proliferation of TE copies. *Xalax* showed high proportions of N bands and low R exclusively among *Ae. cylindrica* (i.e. TE proliferation).

Origin of lost bands in the allopolyploid species

All allopolyploids except *Ae. geniculata* showed significantly higher proportion of lost bands from the differential genome than from the pivotal genome (Figure 3) with p-value $P < 0.01$. *Ae. geniculata* showed non-significant difference in the overall proportions of lost bands between the two genomes ($P = 0.63$), which is consistent with the differences observed at the LTR-RT family level (Supplemental Material Table S5). *Ae. geniculata* indeed showed significantly higher proportions of lost bands from differential genome for *Romani* ($P< 0.05$) whereas *BARE1* significantly showed opposite proportions towards pivotal genome ($P< 0.05$).

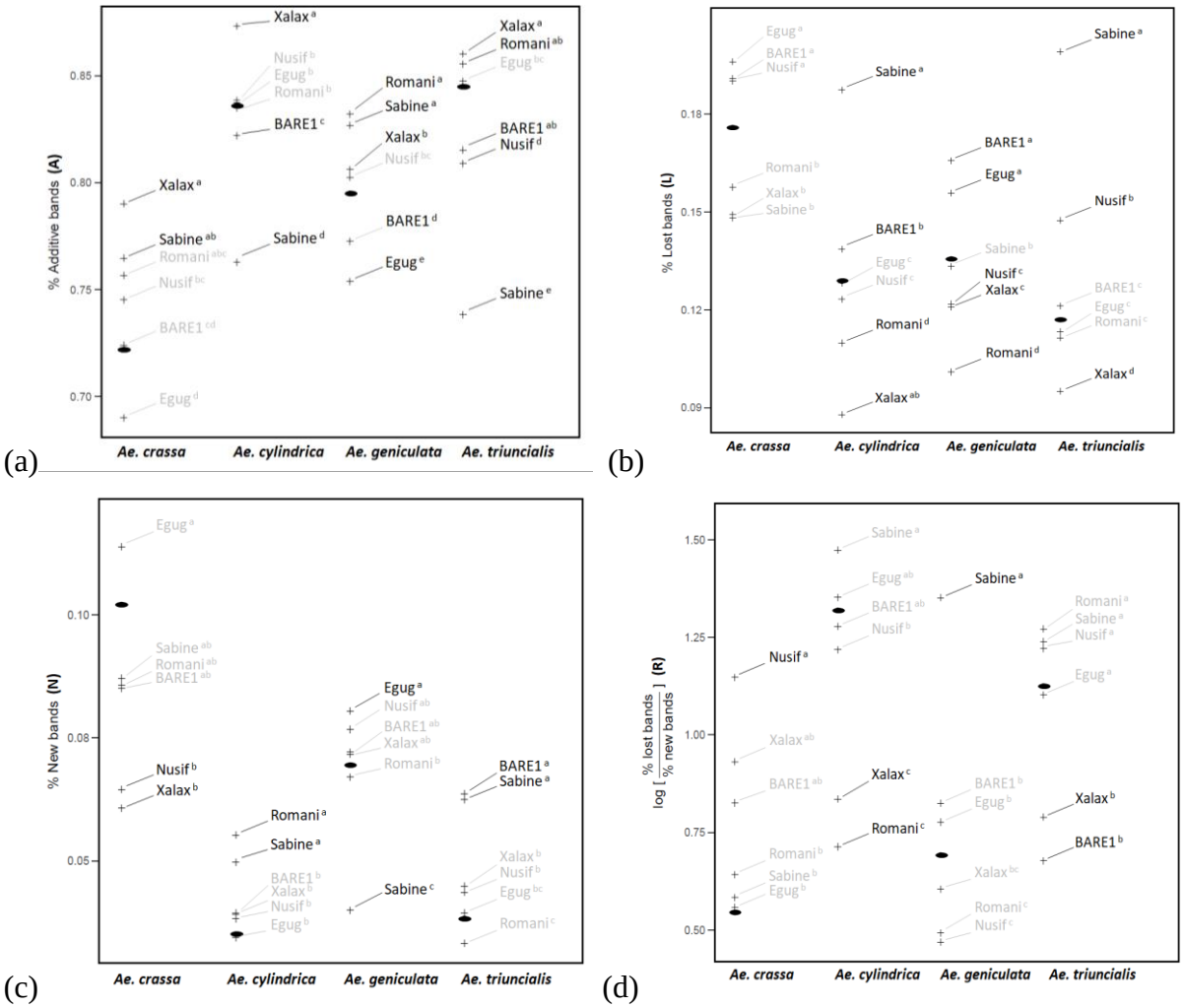


Figure 2: Deviation from the expected additivity of progenitors in *Aegilops* allopolyploid species. Proportions of deviating bands for six TE families investigated by SSAP (crosses) compared to genome-wide proportions of AFLPs (black ellipses). (a) Proportions of additive bands (A). (b) Proportions of lost bands (N). (c) Proportions of new bands (C). (d) Log-scaled ratio between proportions of lost and new bands (R). TE families having significantly different proportions do not share small letters and are presented in grey if not significantly different from AFLPs, as assessed by Tukey tests.

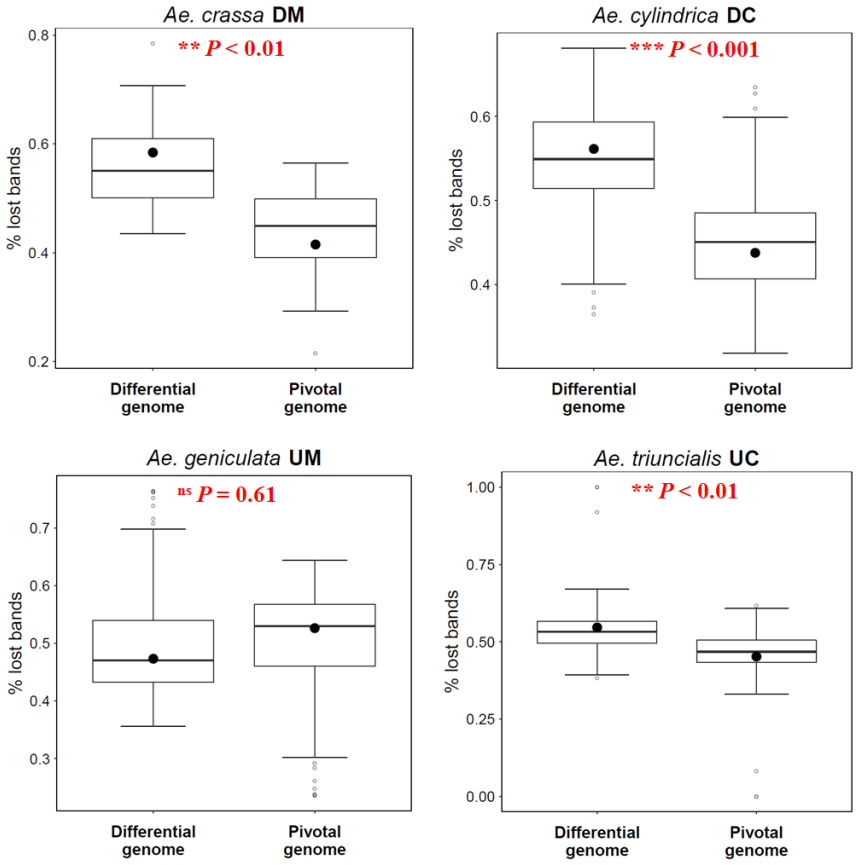


Figure 3: Proportion of lost bands originating from the differential or the pivotal genomes in four *Aegilops* allopolyploid species. Significant differences were assessed for all proportions of TEs families (assessed by SSAPs) within each *Aegilops* species through one-way ANOVA tests. Significance of the tests are reported in each plot as p-value (*P*). Medians and quartiles are shown as boxes and whiskers shows minimum and maximum values within 1.5 interquartile ranges. White circles represent outlier data whereas black circles represent proportions of lost random sequences as assessed by AFLPs.

Individual departure of LTR-RT family (SSAP) dynamics from AFLP genome-wide variation

LTR-RT families mostly showed high intraspecific variation of individual proportions of A, L and N bands and of R across species (data not shown). However, some LTR-RT families exhibited general trends in a given species, e.g. *Sabine* showed a clear trend toward more L bands among all *Ae. triuncialis* and *Ae. cylindrica* individuals compared to other species.

Overall, the geographic distribution of individual TE-specific z-scores showed no strong spatial structure across allopolyploid species ranges (Figure 4 and Supplemental Material Figures S2-S7). Among the 96 patterns of spatial autocorrelation tests assessed across individual z-scores, fourteen were assessed as significant, eight of which appeared in *Ae. geniculata* (Table 2).

Table 2: Spatial structure of individual TE dynamics within allopolyploid host species. Individual TE dynamics were measured as departures from AFLP scores for the additive bands (A), new bands (N), lost bands (L) and ratio of lost bands over new bands (R). The presence of spatial structure of individual TE dynamics was tested through spatial autocorrelation within four allopolyploid *Aegilops* species for six TE families (*BARE1*, *Egug*, *Nusif*, *Romani*, *Sabine*, *Xalax*). Significant correlations are noted '+' and non-significant correlations are noted '-'.

	***Ae. crassa* (DM)**				***Ae. cylindrica* (DC)**				***Ae. geniculata* (UM)**				***Ae. triuncialis* (UC)**			
	A	N	L	R	A	N	L	R	A	N	L	R	A	N	L	R
Egug	-	-	-	-	-	-	-	-	-	-	-	+	-	+	-	-
Nusif	-	-	-	-	-	-	-	+	-	-	+	-	-	-	-	-
BARE1	-	-	-	-	-	-	-	-	+	+	-	+	-	-	-	-
Romani	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Sabine	-	-	-	-	-	+	-	+	-	-	+	-	-	-	-	-
Xalax	-	-	-	-	-	-	-	-	+	+	-	-	-	+	-	+

For instance, *BARE1* showed particularly high individual z-scores for the proportion of N bands in *Ae. geniculata* accessions occurring in the Balkans, the Levant and North Africa, whereas those located on the Iberian or Italian peninsulas showed particularly low individual z-scores (Figure 4). In addition to these few spatial autocorrelation patterns, the region between the Balkans and Anatolia and the margins of species distribution seemed to concentrate recent TE activities. In particular, *Ae. cylindrica* showed especially high TE activity in the West and East regions of Anatolia while *Ae. triuncialis* showed high TE activity in the region between the Balkans and Anatolia (Figure 4 and Supplemental Material Figure S2-S7). In addition, high z-scores ratios (R bands), indicative of higher proportion of TE proliferation than of TE deletion, revealed particularly marked for *Xalax*, *Egug*, *Sabine* and *BARE1* families at the range margins of *Ae. triuncialis* (Figure 5).

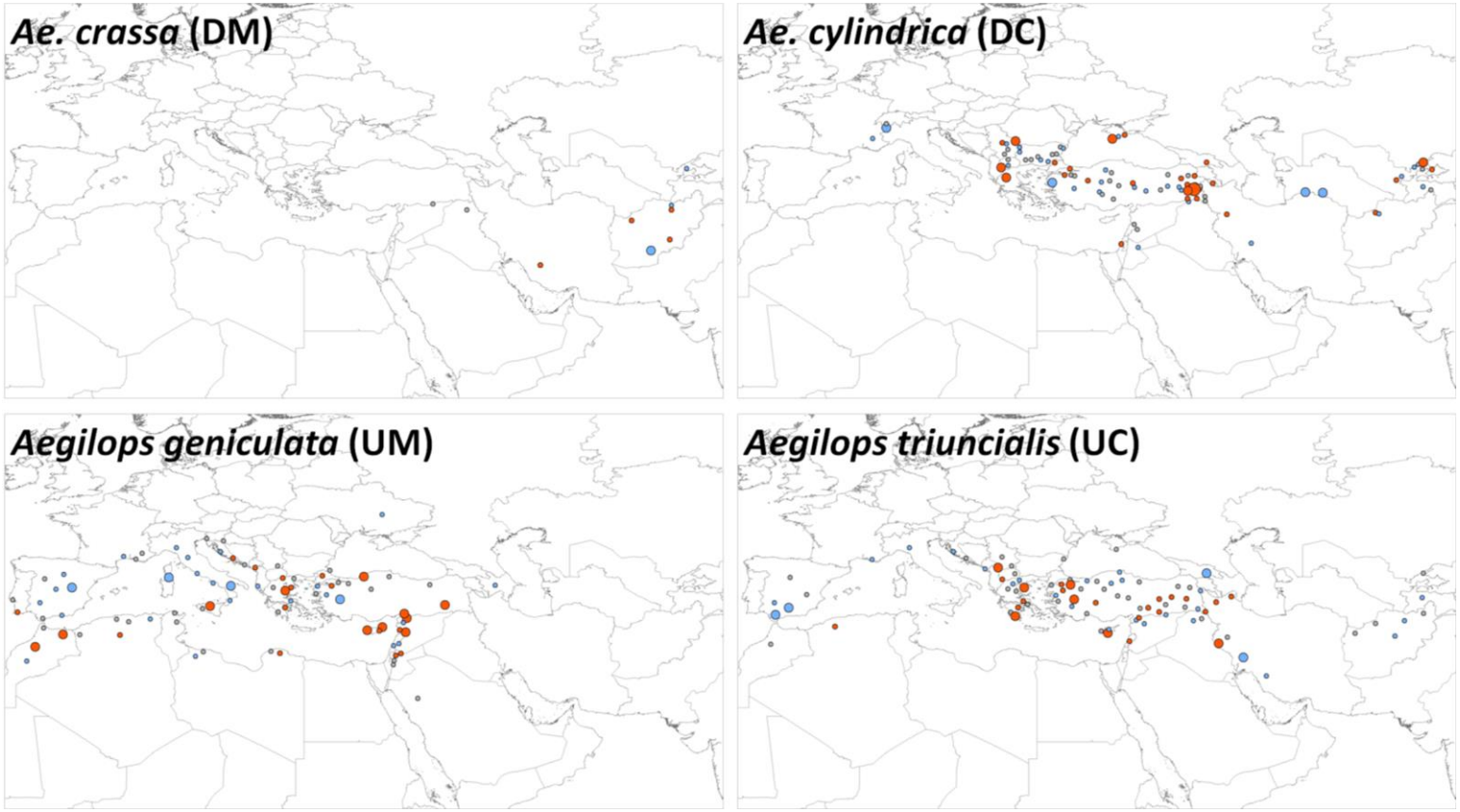


Figure 4: Levels of *BARE1* proliferation within genotyped accessions across allopolyploid *Aegilops* species range. Proliferation of *BARE1* copies is expressed here as the departure of proportions (noted as 'z-scores') of new bands of *BARE1* from genome-wide variation assessed by AFLP. Negative z-scores are represented by blue circles with increasing sizes for values lower than -0.5/-1.5/-2.5 standard deviation respectively, indicative of low TE proliferation. Positive z-scores are represented by red circles with increasing sizes for values higher than 0.5/1.5/2.5 standard deviation respectively, indicative of high TE proliferation. Values between -0.5 and 0.5 standard deviation, considered as not significantly different from AFLP (i.e. quiescent TE copies), are represented by small grey circles.

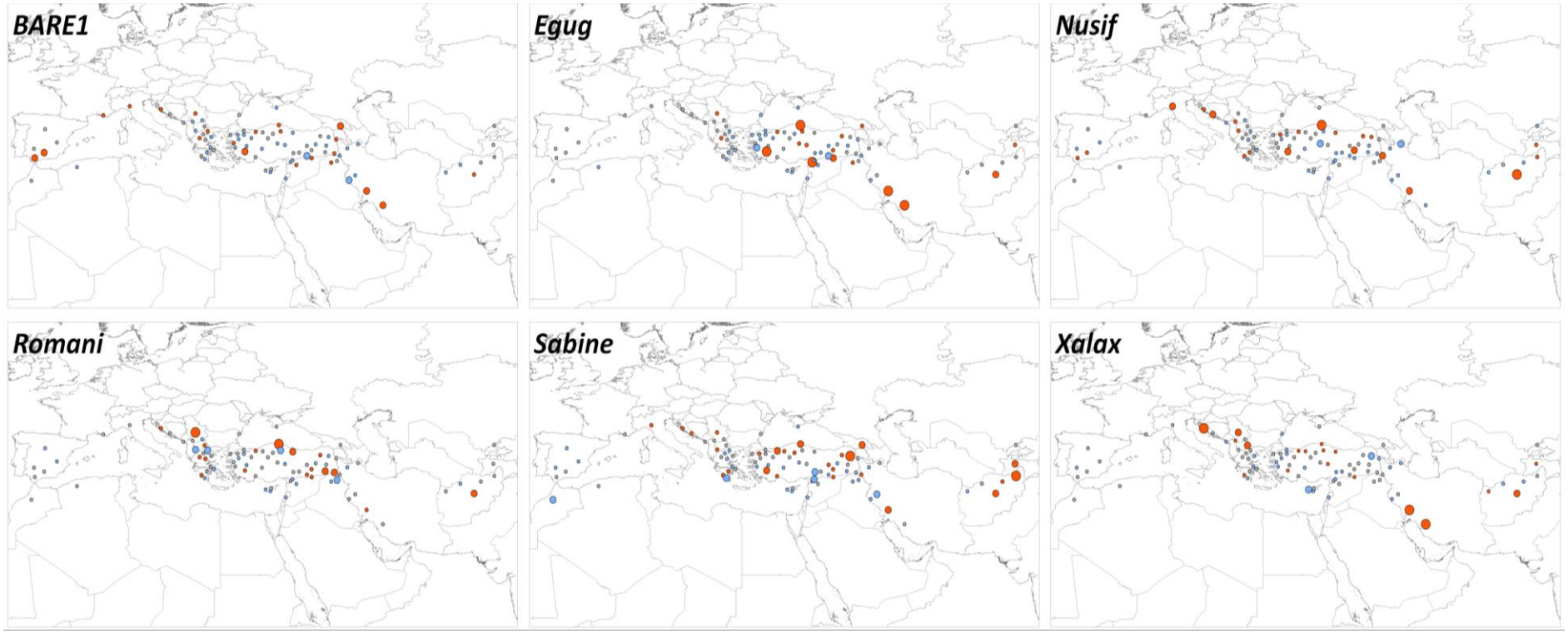


Figure 5: High individual z-scores of log-ratios of lost/new bands (R) at the range margins of the allopolyploid *Ae. triuncialis*. Z-score ratios are expressed as the departure from genome-wide variation as assessed by AFLP (noted as 'z-scores') of the ratios of lost bands over new bands (R). Negative z-score ratios are represented by blue circles with increasing sizes for values lower than -0.5/-1.5/-2.5 standard deviation respectively, indicative of higher proportions of lost bands than of new bands (i.e. TEs mainly deleted). Positive z-score ratios are represented by red circles with increasing sizes for values higher than 0.5/1.5/2.5 standard deviation respectively, indicative of higher proportions of new bands than of lost bands (i.e. TE mainly proliferating). Values between -0.5 and 0.5 standard deviation, represented by small grey circles, represent either quiescent TEs or are due to equal proportion of TE proliferation/deletion indicative of high TE turnover.

Discussion

Patterns of genetic variation among individuals

The proportions of A, L and N bands were mostly consistent with Senerchia *et al.* (2014). The proportions of A bands inferred in this study were overall about 10% higher than assessed by Senerchia *et al.* (2014), along with an overall decreased proportion of L bands of about 10% than formerly assessed. These slight differences are explained by our higher sampling sizes across species ranges, which affected not only the observed band proportions within diploid and allopolyploid species but also the expected probabilities of band presence and absence estimated for allopolyploids. Though the combination probabilities method developed by Senerchia *et al.* (2014) likely limited the impact of low intra-specific sampling size on the inferences of evolutionary trajectories of LTR retrotransposons in allopolyploids, high sampling size covering species distribution thus seem necessary for accurately comparing the genetic variants segregating in allopolyploids with those segregating in their progenitor species and hence for confidently assessing TE dynamics in allopolyploids.

The phylogeographic break between the Balkans and Anatolia previously highlighted in the genetic structure of most *Aegilops* species (see Chapter 2) was recovered only for *Ae. triuncialis* and *Ae. umbellulata.* The lower level of genetic structures inferred here in allopolyploids, compared to those inferred in Chapter 2, is attributable to the genetic information provided by the genome-wide screening from AFLP used in the present study whereas 30 low-copy nuclear loci were used in Chapter 2. Diploids showed higher genetic diversity (*Ndiv*) than allopolyploids, which is mostly congruent with the genetic diversity inferred among species in Chapter 2. The significant correlation between Nei's genetic distances among TE copies and among random sequences (genome-wide AFLP) across the eight *Aegilops* species suggests species-specific evolution of TE copies, despite recurrent hybridization occurring among and within allopolyploid species (Chapter 2).

Contrasted evolutionary trajectories of TEs among allopolyploids

TE dynamics highly differed among allopolyploid species and confirmed departure of allopolyploids from expected progenitors' additivity, but with noticeable differences in the evolutionary trajectories of each LTR-RT family with Senerchia *et al.* (2014). All six LTR-RT families surveyed indeed show contrasted evolutionary trajectories among the allopolyploid species, with the expected quiescent *Egug* and *Nusif* respectively showing a high turnover of TE copies within *Ae. geniculata* and significant loss of TE copies within *Ae.*

triuncialis. The specific elimination of *Sabine* copies in *Ae. geniculata* is consistent with Senerchia *et al.* (2014), but our study also highlight a high turnover of *Sabine* copies in *Ae. cylindrica* and *Ae. triuncialis*. Surprisingly *BARE1*, which is known to be active and proliferating among Triticeae (Vicient *et al.*, 1999; Senerchia *et al.*, 2013), shows here proliferating only in *Ae. triuncialis*. Investigating TE dynamics on a large sampling covering species ranges affects the inference of TE dynamics in allopolyploids, as suggested by the differences in band proportions inferred for the common accessions with Senerchia *et al.* (2014). Larger sampling size may thus explain the divergent trend observed for the evolutionary trajectories of *BARE1*, highlighting the importance of sampling size and range.

Contrasted evolutionary trajectories of TEs within allopolyploid species

Though the early generations of nascent allopolyploids are often characterized by bursts of TE proliferation (Petit *et al.*, 2010) subsequent to genomic stresses induced by whole genome doubling and hybridization, the long-term evolution of TE copies seem however driven by random processes. TE dynamics is indeed highly variable across the allopolyploid species ranges.

However, recent and located TE proliferation seemingly occur, especially at the margins of species distribution and along suture zones identified in Chapter 2 such as the Balkans-Anatolia gap or the east region of Anatolia. Secondary contact between diverging genetic clusters along suture zones may have generated genetic conflicts due to hybridization among divergent genomes that further re-activated specific LTR-RT families, e.g. *Egug* in *Ae. triuncialis* (Supplemental Material Figure S2a). The high TE proliferation observed in marginal allopolyploid accessions, such as *BARE1* and *Sabine* in *Ae. cylindrica* and *Xalax* and *Sabine* in *Ae. triuncialis* (Supplemental Material Figures S2-S7), may reflect extrinsic stresses specific to margin zones. As species expands its range through dispersal, gene flow between marginal and core (here, the center of species range) populations might progressively decrease, resulting in lower genetic diversity and higher impact of genetic drift at margin ranges (Vucetich & Waite; Lönn & Prentice, 2002). However, the evolution of margin populations depends on a combination of factors, including gene flow and dispersal abilities of the species as well as the habitat heterogeneity, and whose relative impact remains debated (Alleaume-Benharira *et al.*, 2006; Bridle & Vines, 2007; Eckert *et al.*, 2008).

Molecular mechanisms underlying pivotal vs. differential genome

TE copies confirmed non-randomly eliminated in pivotal vs. differential genome among allopolyploid species, with higher restructuring of the differential genome than of the pivotal genome for all allopolyploid species except Ae. geniculata, which is consistent with Senerchia et al.(2014).

The overall trend observed toward higher restructuring of differential genomes than of pivotal genomes in the allopolyploid species might results from nuclear-cytoplasmic interactions involving the repression of nuclear TE copies by the cytoplasmic siRNA machinery. Under the genome shock hypothesis, the merging of progenitor TE copies having divergent copy numbers or nucleotide composition results in quantitative and qualitative imbalance between cytoplasmic siRNAs and nuclear TE copies that can lead to de-repression of TEs (Parisod & Senerchia, 2012). This imbalance should mainly affect paternal TE copies, which are not recognized and silenced by the maternal siRNA machinery. The allopolyploid Ae. cylindrica (DC) maternally originated from its pivotal D genome donor Ae. tauschii only (see Chapter 2). The higher restructuring of the allopolyploid differential genome (C) may thus reflect specific proliferation of paternal TE copies that failed to be repressed by its maternal siRNA machinery.

The allopolyploid Ae. triuncialis (UC) has maternally originated from both its progenitor species (see Chapter 2) and is thus expected to have siRNAs regulating paternal as well as maternal TE copies, i.e. displaying symmetrical restructuring of its pivotal and differential genomes. However, this species also exhibits asymmetrical genetic losses from its differential genome (C), though consistently matching the asymmetrical structural changes expected under the pivotal-differential hypothesis from Zohary and Feldman (1962). The allopolyploid Ae. geniculata (UM) originated from the pivotal U genome as single maternal progenitor (see Chapter 2). The contrasted patterns of genome restructuring among LTR-RT families observed in this allopolyploid, in particular BARE1 and Romani that showed significant but opposite restructuring of pivotal vs. differential genomes, contradicts to some extent the expected asymmetry under both the pivotal-differential and the genome shock hypotheses. Other molecular mechanisms or extrinsic processes probably act on the evolution of pivotal-differential genomes in the allopolyploids Aegilops derived from the pivotal U genome that remain to be assessed.

Conclusions and perspectives

TE copies confirmed differentially restructured between the pivotal and differential genomes, confirming asymmetrical structural changes under the pivotal-differential genome hypothesis. Large individual sampling across species ranges of both allopolyploids and their diploid progenitors enabled to infer the long-term evolution of TE families seemingly driven by random processes. Nevertheless, recent re-activation of certain TEs along suture zones highlight the complex evolution processes underlying the evolution of TEs that needs to be clarified through additional population genomics focusing on TE dynamics in these particular zones.

References

Alleaume-Benharira M, Pen IR, Ronce O. **2006**. Geographical patterns of adaptation within a species' range: interactions between drift and gene flow. *Journal of Evolutionary Biology* **19**: 203–215.

Arrigo N, Felber F, Parisod C, Buerki S, Alvarez N, David J, Guadagnuolo R. **2010**. Origin and expansion of the allotetraploid *Aegilops geniculata*, a wild relative of wheat. *New Phytologist* **187**: 1170–1180.

Badaeva ED, Amosova AV, Muravenko OV, Samatadze TE, Chikida NN, Zelenin AV, Friebe B, Gill BS. **2002**. Genome differentiation in *Aegilops*. 3. Evolution of the D-genome cluster. *Plant Systematics and Evolution* **231**: 163–190.

Badaeva ED, Amosova AV, Samatadze TE, Zoshchuk SA, Shostak NG, Chikida NN, Zelenin AV, Raupp WJ, Friebe B, Gill BS. **2004**. Genome differentiation in *Aegilops*. 4. Evolution of the U-genome cluster. *Plant Systematics and Evolution* **246**: 45–76.

Belyayev A. **2014**. Bursts of transposable elements as an evolutionary driving force. *Journal of Evolutionary Biology* **27**: 2573–2584.

Bonin A, Bellemain E, Bronken Eidesen P, Pompanon F, Brochmann C, Taberlet P. **2004**. How to track and assess genotyping errors in population genetics studies. *Molecular Ecology* **13**: 3261–3273.

Bridle JR, Vines TH. **2007**. Limits to evolution at range margins: when and why does adaptation fail? *Trends in Ecology & Evolution* **22**: 140–147.

Choulet F, Wicker T, Rustenholz C, Paux E, Salse J, Leroy P, Schlub S, Le Paslier M-C, Magdelenat G, Gonthier C, *et al.* 2010. Megabase Level Sequencing Reveals Contrasted Organization and Evolution Patterns of the Wheat Gene and Transposable Element Spaces. *The Plant Cell* **22**: 1686–1701.

Devos KM. **2002**. Genome Size Reduction through Illegitimate Recombination Counteracts Genome Expansion in Arabidopsis. *Genome Research* **12**: 1075–1079.

Doyle JJ, Flagel LE, Paterson AH, Rapp RA, Soltis DE, Soltis PS, Wendel JF. **2008**. Evolutionary genetics of genome merger and doubling in plants. *Annual Review of Genetics* **42**: 443–461.

Eckert CG, Samis KE, Lougheed SC. **2008**. Genetic variation across species' geographical ranges: the central–marginal hypothesis and beyond. *Molecular Ecology* **17**: 1170–1188.

Evanno G, Regnaut S, Goudet J. **2005**. Detecting the number of clusters of individuals using the software structure: a simulation study. *Molecular Ecology* **14**: 2611–2620.

Fedoroff NV. **2012**. Transposable Elements, Epigenetics, and Genome Evolution. *Science* **338**: 758–767.

Feldman M. **1965**. Further evidence for natural hybridization between tetraploid species of *Aegilops* section *Pleioxathera*. *Evolution* **19**: 162–174.

Feldman M, Levy AA. **2005**. Allopolyploidy – a shaping force in the evolution of wheat genomes. *Cytogenetic and Genome Research* **109**: 250–258.

Feldman M, Levy AA. **2009**. Genome evolution in allopolyploid wheat—a revolutionary reprogramming followed by gradual changes. *Journal of Genetics and Genomics* **36**: 511–518.

Feldman M, Levy AA. **2012**. Genome evolution due to allopolyploidization in wheat. *Genetics* **192**: 763–774.

Feldman M, Levy AA. **2015**. Origin and evolution of wheat and related Triticeae species. In: Molnár-Láng M, Ceoloni C, Doležel J, eds. Alien Introgression in Wheat. Cham: Springer International Publishing, 21–76.

Kashkush K, Feldman M, Levy AA. **2003**. Transcriptional activation of retrotransposons alters the expression of adjacent genes in wheat. *Nature Genetics* **33**: 102–106.

Kumar A, Bennetzen JL. **1999**. Plant Retrotransposons. *Annual Review of Genetics* **33**: 479–532.

Le Rouzic A, Boutin TS, Capy P. **2007**. Long-term evolution of transposable elements. *Proceedings of the National Academy of Sciences* **104**: 19375–19380.

Leitch IJ, Hanson L, Lim KY, Kovarik A, Chase MW, Clarkson JJ, Leitch AR. **2008**. The Ups and Downs of Genome Size Evolution in Polyploid Species of *Nicotiana* (Solanaceae). *Annals of Botany* **101**: 805–814.

Lisch D. **2013**. How important are transposons for plant evolution? *Nature Reviews Genetics* **14**: 49–61.

Lisch D, Slotkin RK. **2011**. Strategies for Silencing and Escape. In: International Review of Cell and Molecular Biology. Elsevier, 119–152.

Lönn M, Prentice HC. **2002**. Gene diversity and demographic turnover in central and peripheral populations of the perennial herb Gypsophila fastigiata. *Oikos* **99**: 489–498.

Luo M-C, Gu YQ, You FM, Deal KR, Ma Y, Hu Y, Huo N, Wang Y, Wang J, Chen S, *et al.* 2013. A 4-gigabase physical map unlocks the structure and evolution of the

complex genome of *Aegilops tauschii*, the wheat D-genome progenitor. *Proceedings of the National Academy of Sciences* **110**: 7940–7945.

Meimberg H, Rice KJ, Milan NF, Njoku CC, McKay JK. **2009**. Multiple origins promote the ecological amplitude of allopolyploid *Aegilops* (Poaceae). *American Journal of Botany* **96**: 1262–1273.

Montgomery EA, Huang S-M, Langley CH, Judd BH. **1991**. Chromosome rearrangement by ectopic recombination in *Drosophila melanogaster*: Genome structure and evolution. *Genetics* **129**: 1085–1098.

Nei M. **1978**. Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* **89**: 583–590.

Oliver KR, McComb JA, Greene WK. **2013**. Transposable elements: Powerful contributors to angiosperm evolution and diversity. *Genome Biology and Evolution* **5**: 1886–1901.

Ozkan H. **2003**. Nonadditive changes in genome size during allopolyploidization in the wheat (*Aegilops-Triticum*) group. *Journal of Heredity* **94**: 260–264.

Parisod C, Christin P-A. **2008**. Genome-wide association to fine-scale ecological heterogeneity within a continuous population of *Biscutella laevigata* (Brassicaceae). *New Phytologist* **178**: 436–447.

Parisod C, Senerchia N. **2012**. Responses of transposable elements to polyploidy. In: Grandbastien M-A, Casacuberta JM, eds. Plant Transposable Elements. Berlin, Heidelberg: Springer Berlin Heidelberg, 147–168.

Pazy B, Zohary D. **1965**. The process of introgression between *Aegilops* polyploids: Natural hybridization between *A. variabilis*, *A. ovata*, and *A. biuncialis*. *Evolution* **19**: 385–394.

Petit M, Guidat C, Daniel J, Denis E, Montoriol E, Bui QT, Lim KY, Kovarik A, Leitch AR, Grandbastien M-A, *et al.* 2010. Mobilization of retrotransposons in synthetic allotetraploid tobacco. *New Phytologist* **186**: 135–147.

Rice WR. **1989**. Analyzing tables of statistical tests. *Evolution* **43**: 223–225.

Roser LG, Ferreyra LI, Saidman BO, Vilardi JC. **2017**. EcoGenetics: An R package for the management and exploratory analysis of spatial data in landscape genetics. *Molecular Ecology Resources* **17**: e241–e250.

Senerchia N, Felber F, North B, Sarr A, Guadagnuolo R, Parisod C. **2016**. Differential introgression and reorganization of retrotransposons in hybrid zones between wild wheats. *Molecular Ecology* **25**: 2518–2528.

Senerchia N, Felber F, Parisod C. **2014**. Contrasting evolutionary trajectories of multiple retrotransposons following independent allopolyploidy in wild wheats. *New Phytologist* **202**: 975–985.

Senerchia N, Felber F, Parisod C. **2015**. Genome reorganization in F1 hybrids uncovers the role of retrotransposons in reproductive isolation. *Proceedings of the Royal Society B: Biological Sciences* **282**: 20142874–20142874.

Senerchia N, Wicker T, Felber F, Parisod C. **2013**. Evolutionary dynamics of retrotransposons assessed by high-throughput sequencing in wild relatives of wheat. *Genome Biology and Evolution* **5**: 1010–1020.

Shaked H, Kashkush K, Ozkan H, Feldman M, Levy AA. **2001**. Sequence elimination and cytosine methylation are rapid and reproducible responses of the genome to wide hybridization and allopolyploidy in wheat. *The Plant Cell* **13**: 1749–1760.

van Slageren MWSJM. **1994**. *Wild wheats: a monograph of* Aegilops *L. and* Amblyopyrum *(Jaub. & Spach) Eig (Poaceae)*. Wageningen, The Netherlands : Aleppo, Syria: Wageningen Agricultural University : International Center for Agricultural Research in the Dry Areas.

Song K, Lu P, Tang K, Osborn TC. **1995**. Rapid genome change in synthetic polyploids of *Brassica* and its implications for polyploid evolution. *Proceedings of the National Academy of Sciences* **92**: 7719–7723.

Szitenberg A, Cha S, Opperman CH, Bird DM, Blaxter ML, Lunt DH. **2016**. Genetic drift, not life history or RNAi, determine long-term evolution of transposable elements. *Genome Biology and Evolution* **8**: 2964–2978.

Tenaillon MI, Hollister JD, Gaut BS. **2010**. A triptych of the evolution of plant transposable elements. *Trends in Plant Science* **15**: 471–478.

Vekemans X, Beauwens T, Lemaire M, Roldán-Ruiz I. **2002**. Data from amplified fragment length polymorphism (AFLP) markers show indication of size homoplasy and of a relationship between degree of homoplasy and fragment size. *Molecular Ecology* **11**: 139–151.

Vicient CM, Suoniemi A, Anamthawat-Jónsson K, Tanskanen J, Beharav A, Nevo E, Schulman AH. **1999**. Retrotransposon *BARE-1* and its role in genome evolution in the genus *Hordeum*. *Plant Cell* **11**: 1769–1784.

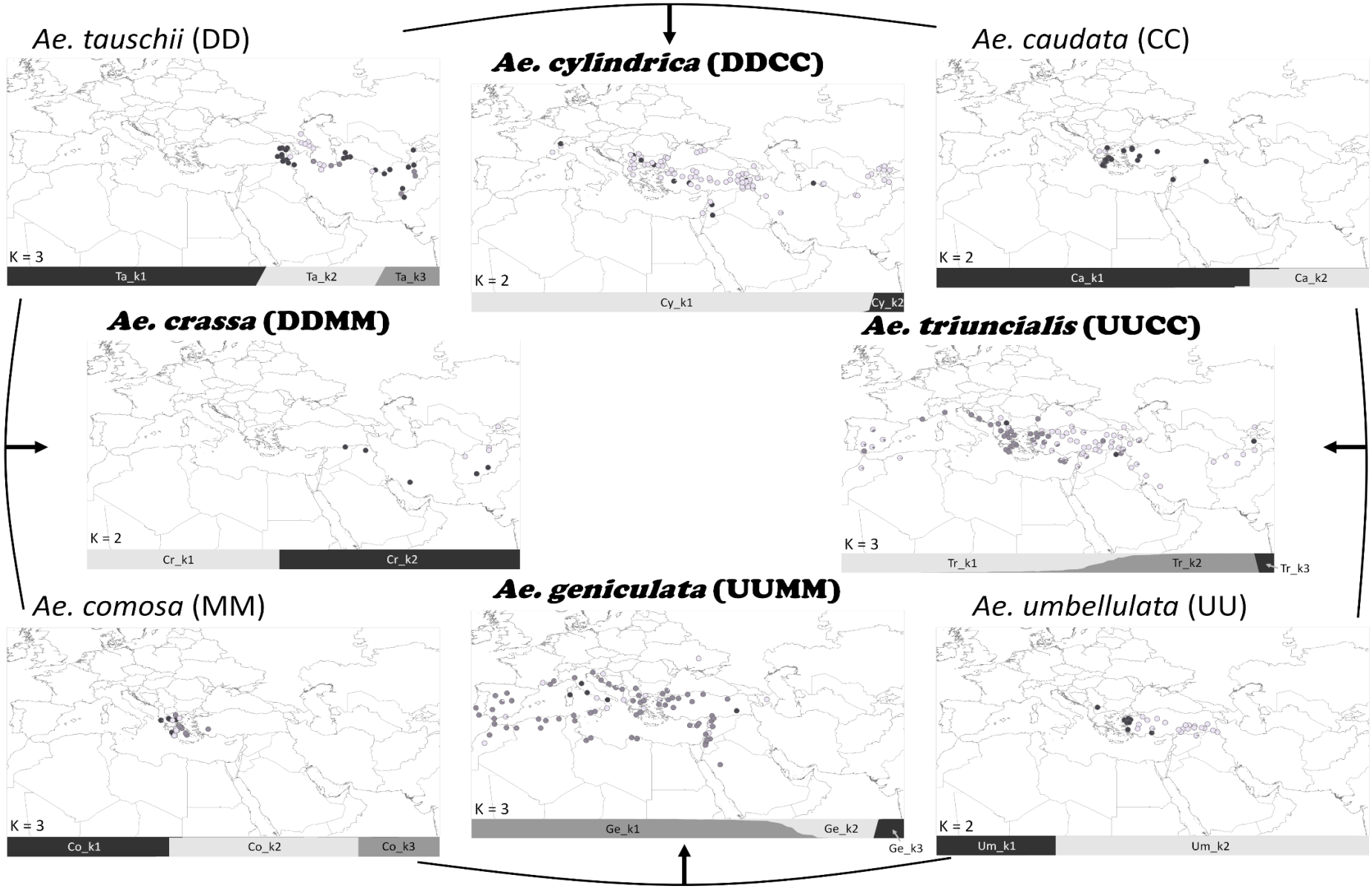
Vucetich JA, Waite TA. **2003.** Spatial patterns of demography and genetic processes across the species' range: Null hypotheses for landscape conservation genetics. *Conservation Genetics* **4**: 639-645.

Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, Flavell A, Leroy P, Morgante M, Panaud O, *et al.* 2007. A unified classification system for eukaryotic transposable elements. *Nature Reviews Genetics* **8**: 973–982.

Zohary D, Feldman M. **1962**. Hybridization between amphidiploids and the evolution of polyploids in the wheat (*Aegilops-Triticum*) group. *Evolution* **16**: 44-61

Supplementary Material

Figure S1: Population structure of *Aegilops* species assessed by AFLP markers (below). The accessions were genotyped on 6 AFLP markers. The best number of genetic clusters K is indicated on the distribution maps. The genetic clusters are represented by pie charts and stacked histrograms using a gradient of greys, and their names indicated in the stacked histograms. Evolutionary relationships among the *Aegilops* species are indicated by black arrows. The *Aegilops* genome compositions are reported into brackets following van Slageren (1994).


Ae. tauschii (DD)
K = 3
Ta_k1
Ta_k2
Ta_k3
Ae. cylindrica (DDCC)
K = 2
Cy_k1
Cy_k2
Ae. caudata (CC)
K = 2
Ca_k1
Ca_k2
Ae. crassa (DDMM)
K = 2
Cr_k1
Cr_k2
Ae. triuncialis (UUCC)
K = 3
Tr_k1
Tr_k2
Tr_k3
Ae. comosa (MM)
K = 3
Co_k1
Co_k2
Co_k3
Ae. geniculata (UUMM)
K = 3
Ge_k1
Ge_k2
Ge_k3
Ae. umbellulata (UU)
K = 2
Um_k1
Um_k2

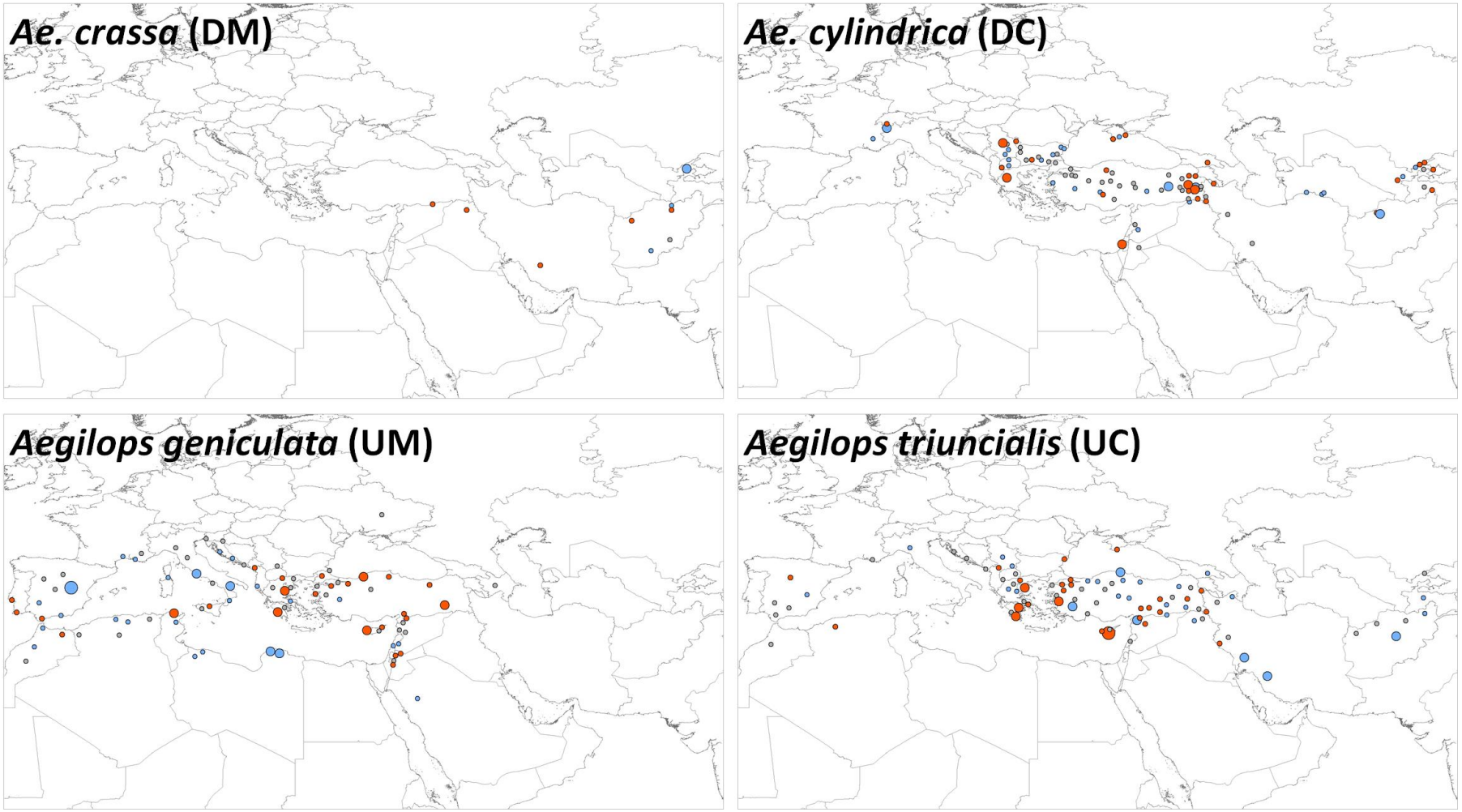


Figure S2a: Distribution **of individual *Egug* dynamics in terms of new bands (N) within allopolyploid *Aegilops* species.** *Egug* dynamics is expressed here as the departure of proportions (noted as 'z-scores') of new bands of *Egug* from genome-wide variation as assessed by AFLP. Blue circles represent negative z-scores with increasing sizes for values lower than -0.5/-1.5/-2.5 standard deviation respectively. Red circles represent positive z-scores with increasing sizes for values higher than 0.5/1.5/2.5 standard deviation respectively. Values between -0.5 and 0.5 standard deviation are represented by small grey circles.

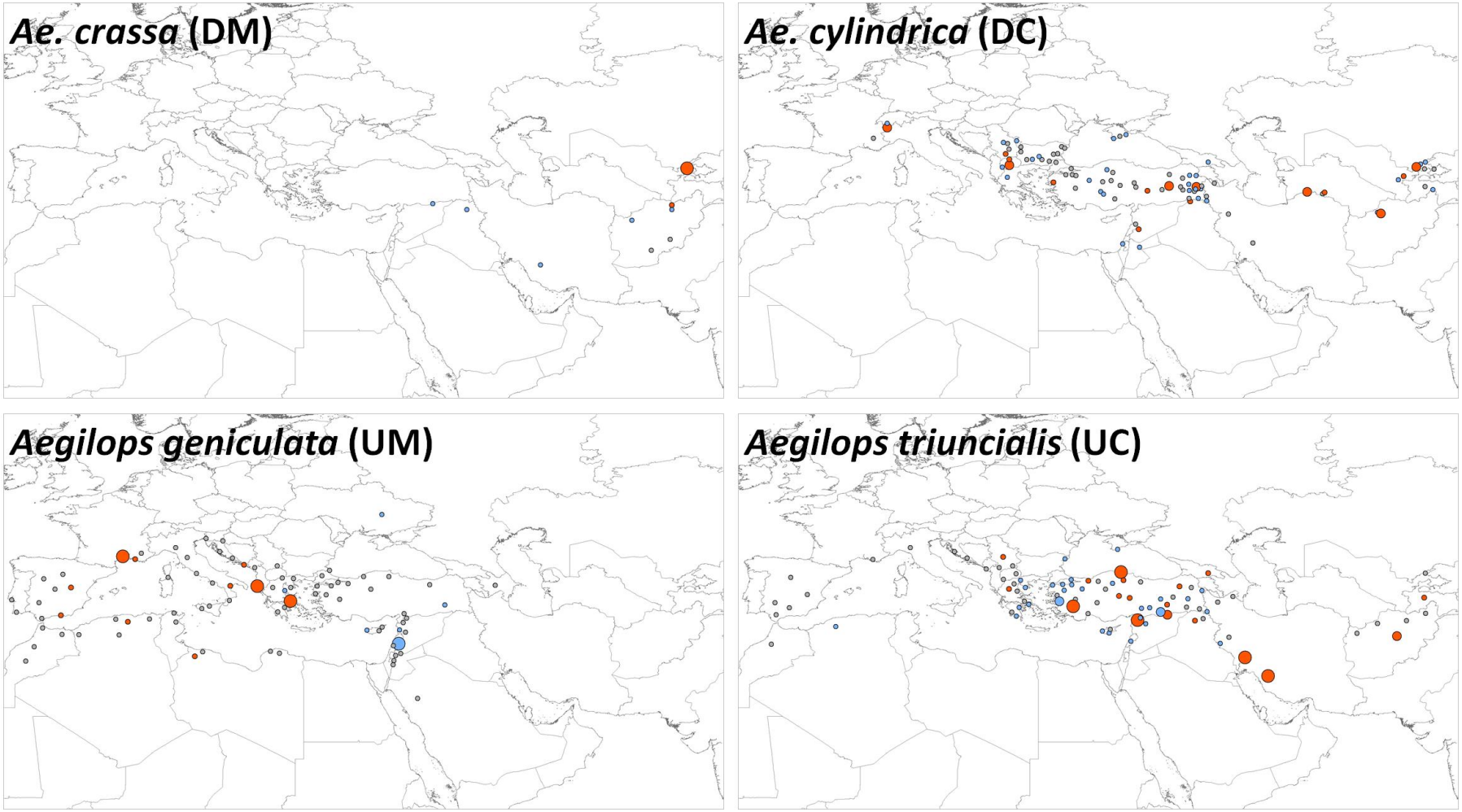


Figure S2b: Distribution of individual *Egug* dynamics in terms of balance of lost bands over new bands (R) within allopolyploid *Aegilops* species. *Egug* dynamics is expressed here as the departure (noted as 'z-scores') of of balance of lost bands over new bands of *Egug* from genome-wide variation as assessed by AFLP. Blue circles represent negative z-scores with increasing sizes for values lower than -0.5/-1.5/-2.5 standard deviation respectively. Red circles represent positive z-scores with increasing sizes for values higher than 0.5/1.5/2.5 standard deviation respectively. Values between -0.5 and 0.5 standard deviation are represented by small grey circles

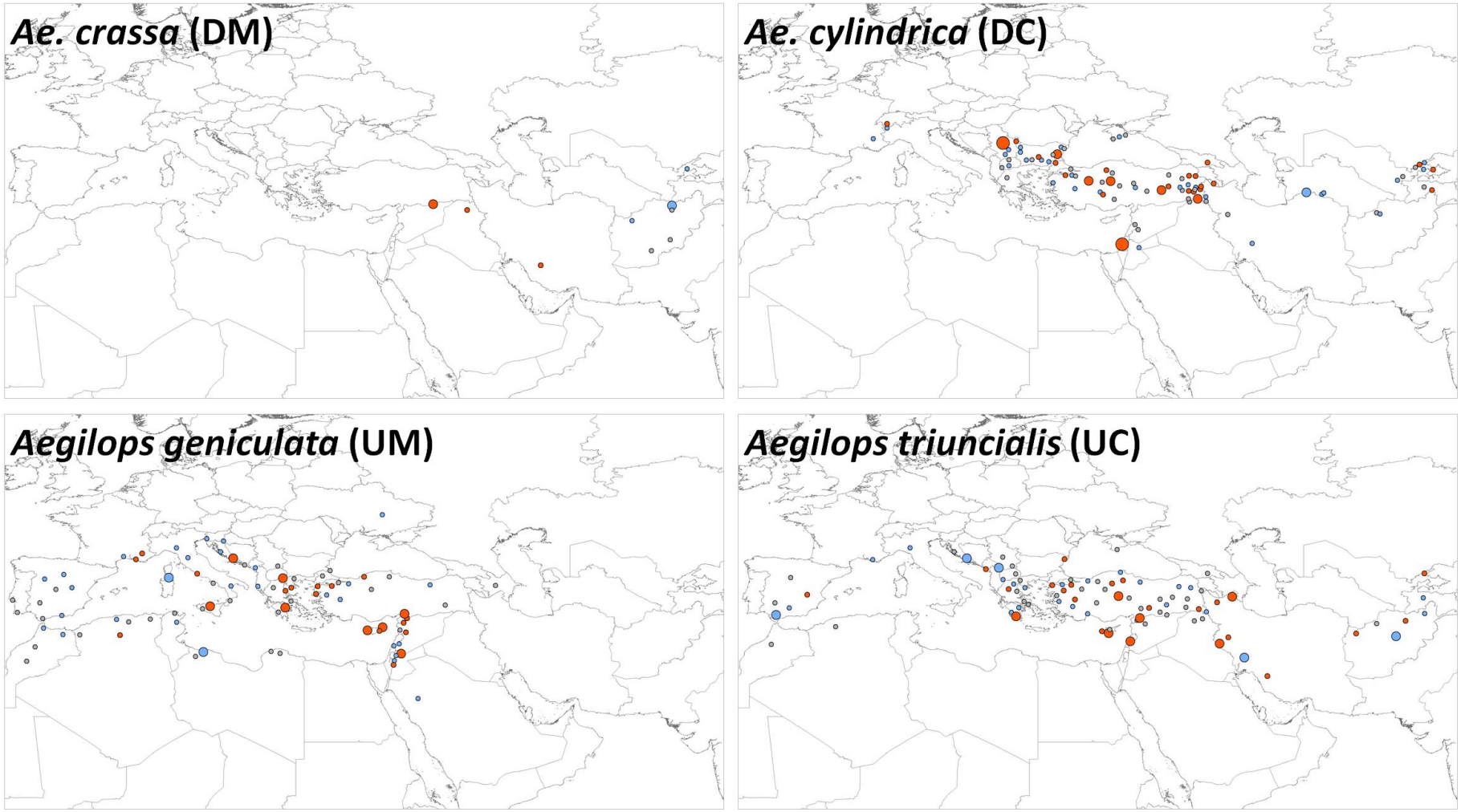


Figure S3a: Distribution of individual *Nusif* dynamics in terms of new bands (N) within allopolyploid *Aegilops* species. *Nusif* dynamics is expressed here as the departure of proportions (noted as 'z-scores') of new bands of *Nusif* from genome-wide variation as assessed by AFLP. Blue circles represent negative z-scores with increasing sizes for values lower than -0.5/-1.5/-2.5 standard deviation respectively. Red circles represent positive z-scores with increasing sizes for values higher than 0.5/1.5/2.5 standard deviation respectively. Values between -0.5 and 0.5 standard deviation are represented by small grey circles.

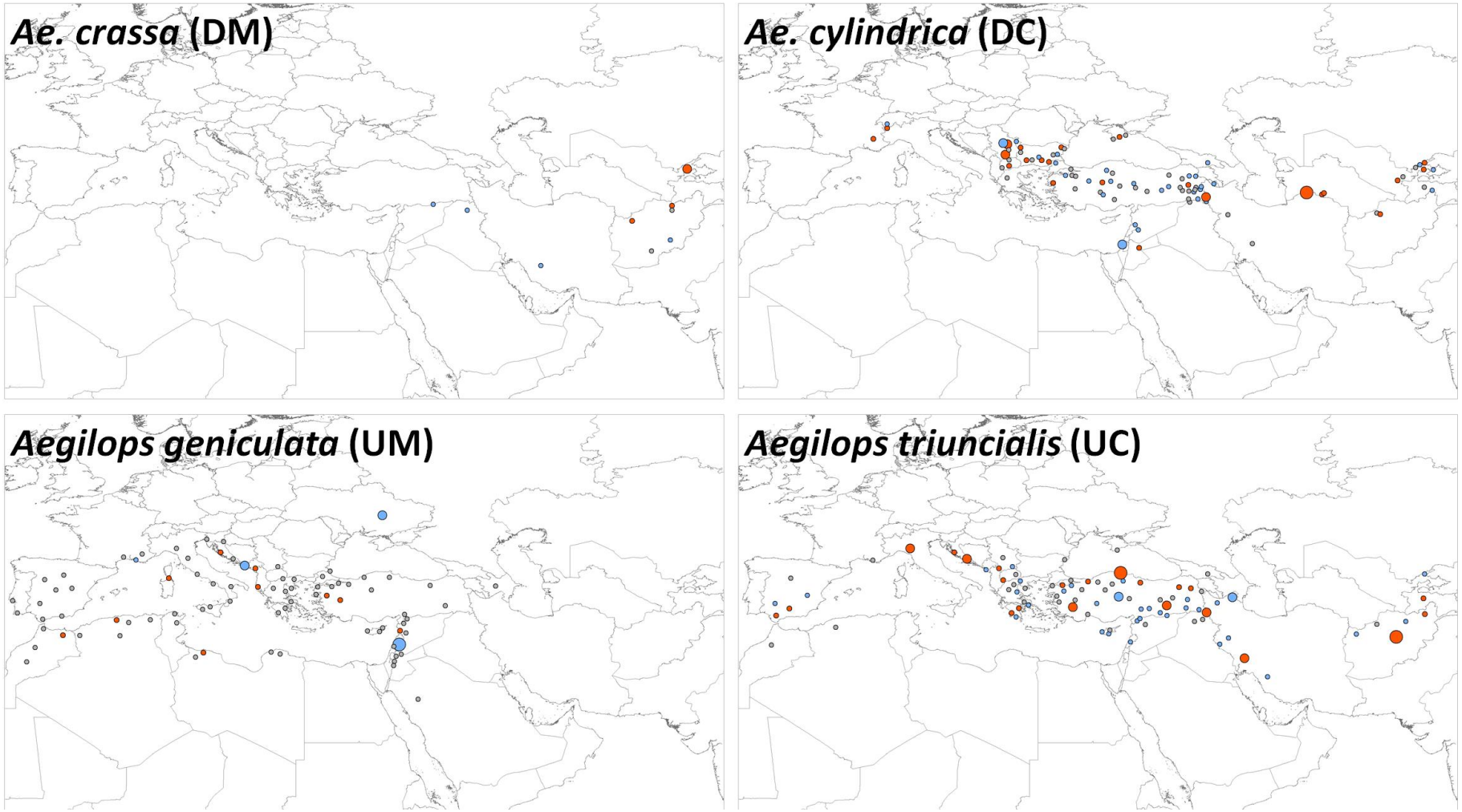


Figure S3b: Distribution of individual *Nusif* dynamics in terms of balance of lost bands over new bands (R) within allopolyploid *Aegilops* species. *Nusif* dynamics is expressed here as the departure (noted as 'z-scores') of balance of lost bands over new bands of *Nusif* from genome-wide variation as assessed by AFLP. Blue circles represent negative z-scores with increasing sizes for values lower than -0.5/-1.5/-2.5 standard deviation respectively. Red circles represent positive z-scores with increasing sizes for values higher than 0.5/1.5/2.5 standard deviation respectively. Values between -0.5 and 0.5 standard deviation are represented by small grey circles.

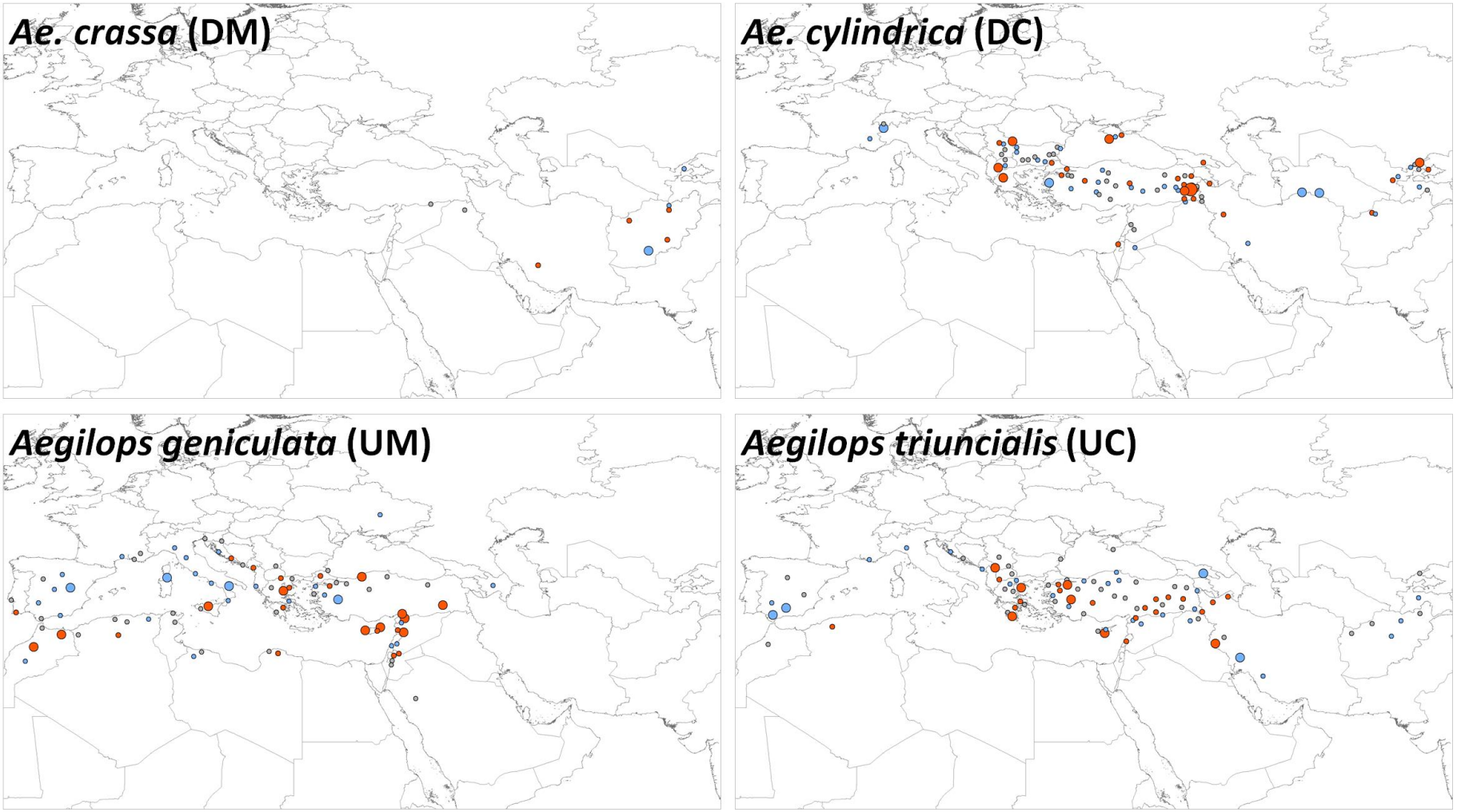


Figure S4a: Distribution of individual *BARE1* dynamics in terms of new bands (N) within allopolyploid *Aegilops* species. *BARE1* dynamics is expressed here as the departure of proportions (noted as 'z-scores') of new bands of *BARE1* from genome-wide variation as assessed by AFLP. Blue circles represent negative z-scores with increasing sizes for values lower than -0.5/-1.5/-2.5 standard deviation respectively. Red circles represent positive z-scores with increasing sizes for values higher than 0.5/1.5/2.5 standard deviation respectively. Values between -0.5 and 0.5 standard deviation are represented by small grey circles.

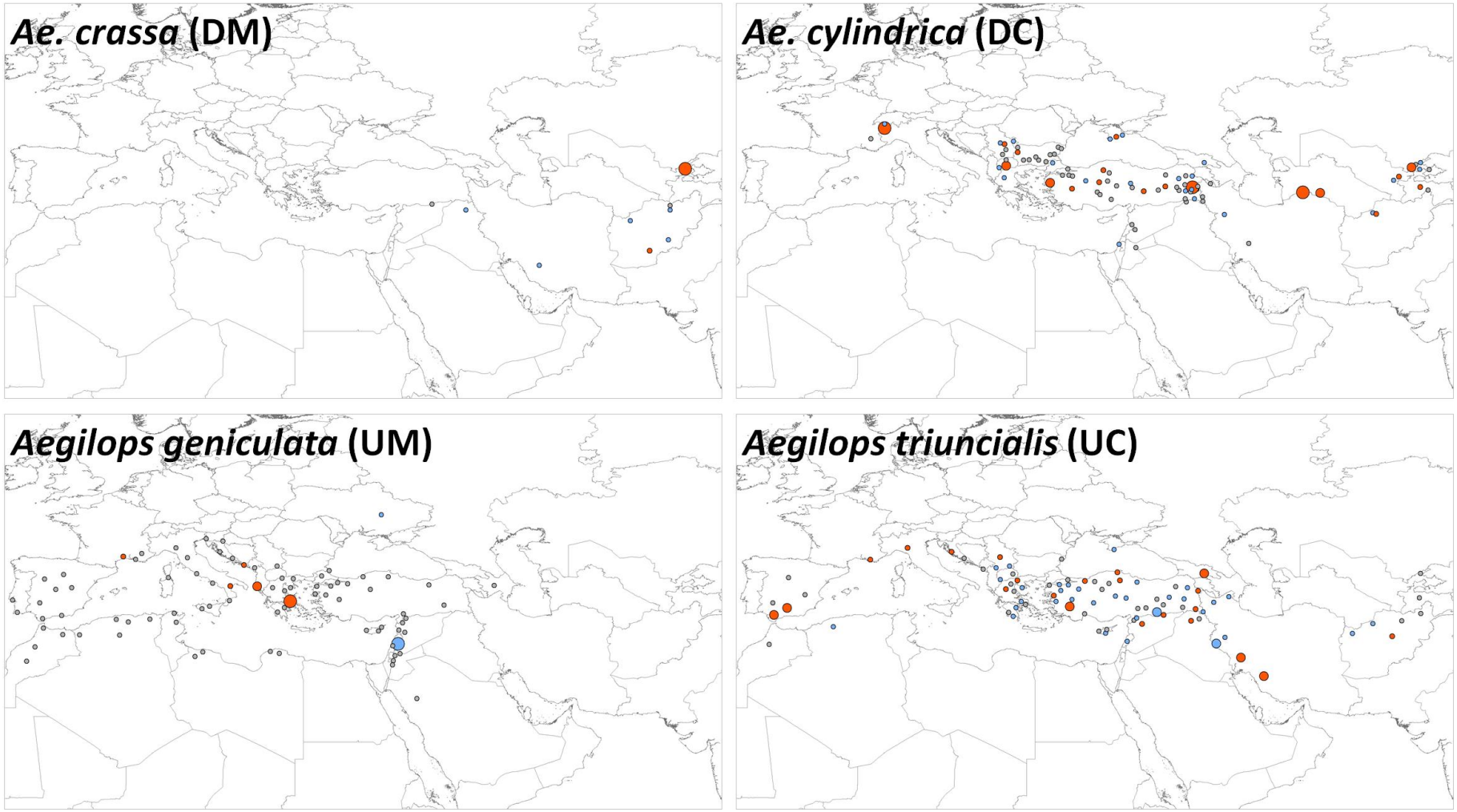


Figure S4b: Distribution of individual *BARE1* dynamics in terms of balance of lost bands over new bands (R) within allopolyploid *Aegilops* species. *BARE1* dynamics is expressed here as the departure (noted as 'z-scores') of balance of lost bands over new bands of *BARE1* from genome-wide variation as assessed by AFLP. Blue circles represent negative z-scores with increasing sizes for values lower than -0.5/-1.5/-2.5 standard deviation respectively. Red circles represent positive z-scores with increasing sizes for values higher than 0.5/1.5/2.5 standard deviation respectively. Values between -0.5 and 0.5 standard deviation are represented by small grey circles.

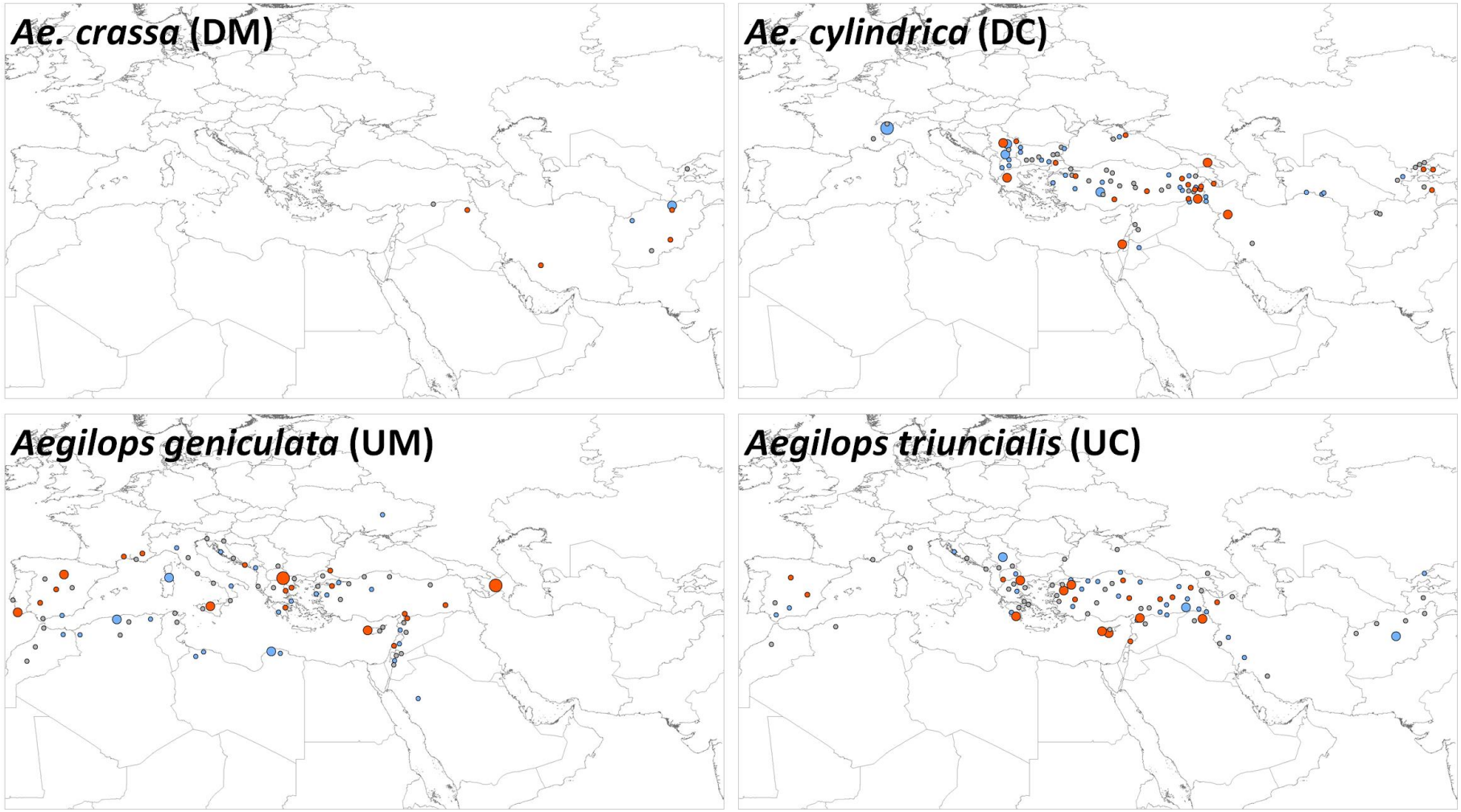


Figure S5a: Distribution of individual *Romani* dynamics in terms of new bands (N) within allopolyploid *Aegilops* species. *Romani* dynamics is expressed here as the departure of proportions (noted as 'z-scores') of new bands of *Romani* from genome-wide variation as assessed by AFLP. Blue circles represent negative z-scores with increasing sizes for values lower than -0.5/-1.5/-2.5 standard deviation respectively. Red circles represent positive z-scores with increasing sizes for values higher than 0.5/1.5/2.5 standard deviation respectively. Values between -0.5 and 0.5 standard deviation are represented by small grey circles.

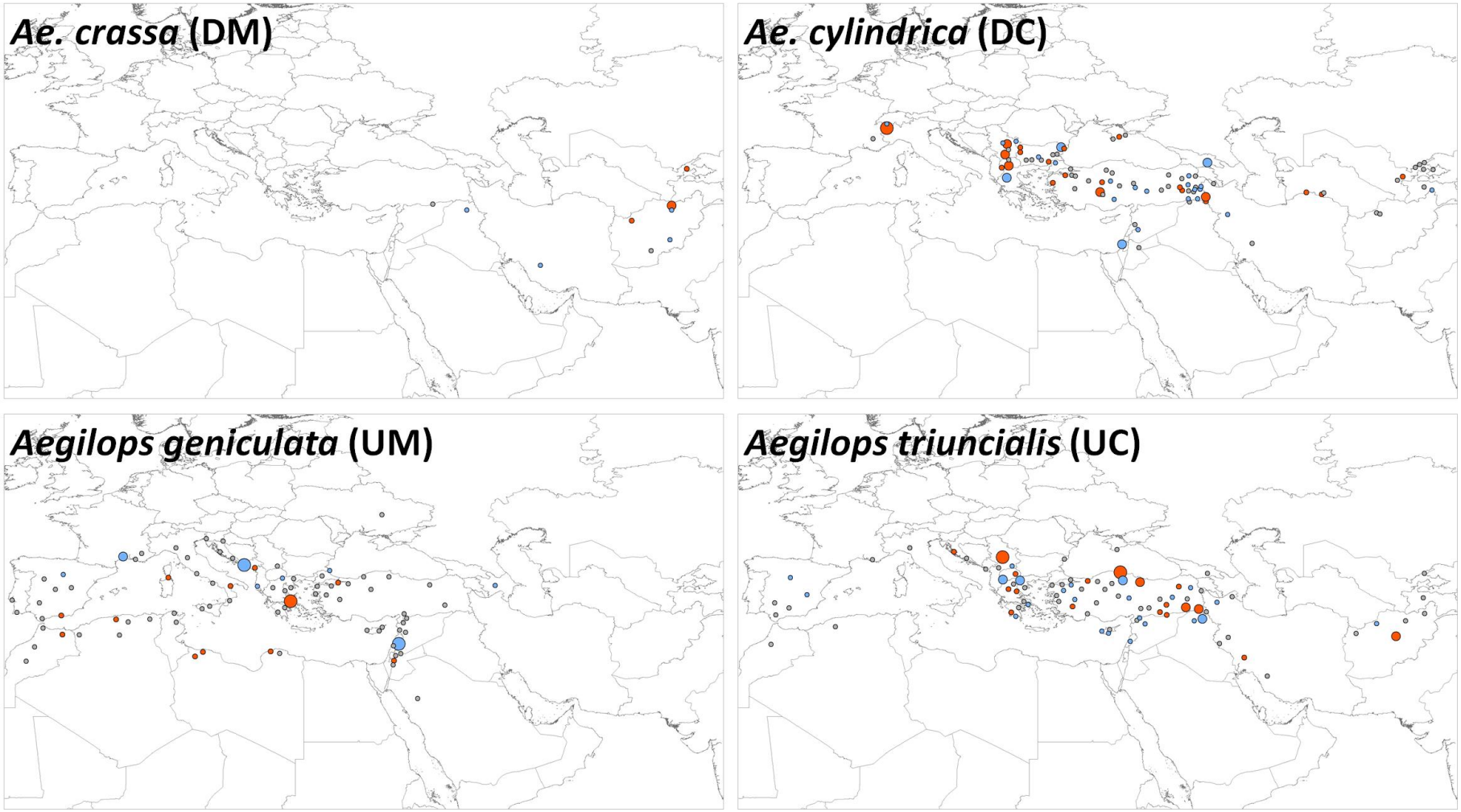


Figure S5b: Distribution of individual *Romani* dynamics in terms of balance of lost bands over new bands (R) within allopolyploid *Aegilops* species. *Romani* dynamics is expressed here as the departure (noted as 'z-scores') of balance of lost bands over new bands of *Romani* from genome-wide variation as assessed by AFLP. Blue circles represent negative z-scores with increasing sizes for values lower than -0.5/-1.5/-2.5 standard deviation respectively. Red circles represent positive z-scores with increasing sizes for values higher than 0.5/1.5/2.5 standard deviation respectively. Values between -0.5 and 0.5 standard deviation are represented by small grey circles.

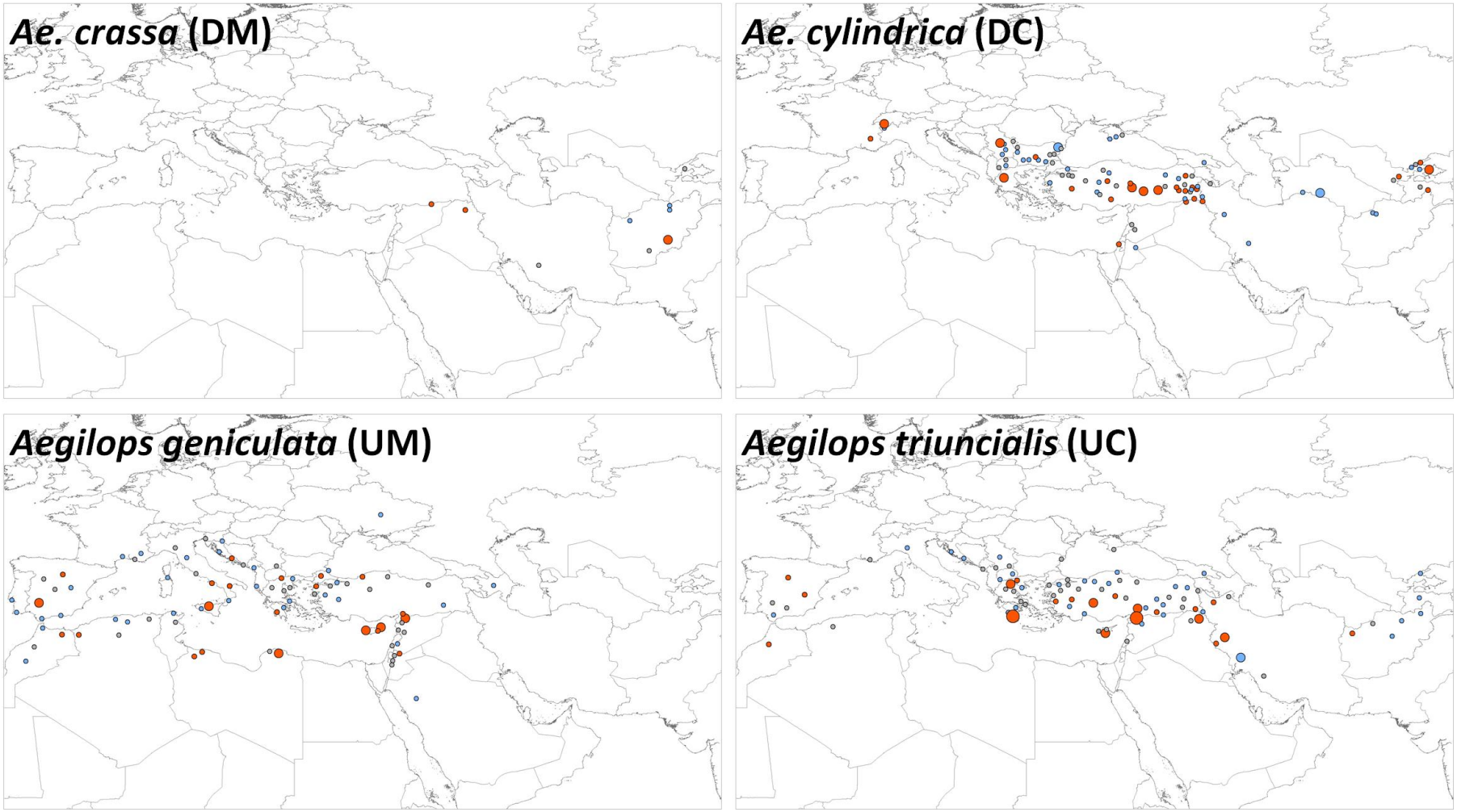


Figure S6a: Distribution of individual *Sabine* dynamics in terms of new bands (N) within allopolyploid *Aegilops* species. *Sabine* dynamics is expressed here as the departure of proportions (noted as 'z-scores') of new bands of *Sabine* from genome-wide variation as assessed by AFLP. Blue circles represent negative z-scores with increasing sizes for values lower than -0.5/-1.5/-2.5 standard deviation respectively. Red circles represent positive z-scores with increasing sizes for values higher than 0.5/1.5/2.5 standard deviation respectively. Values between -0.5 and 0.5 standard deviation are represented by small grey circles.

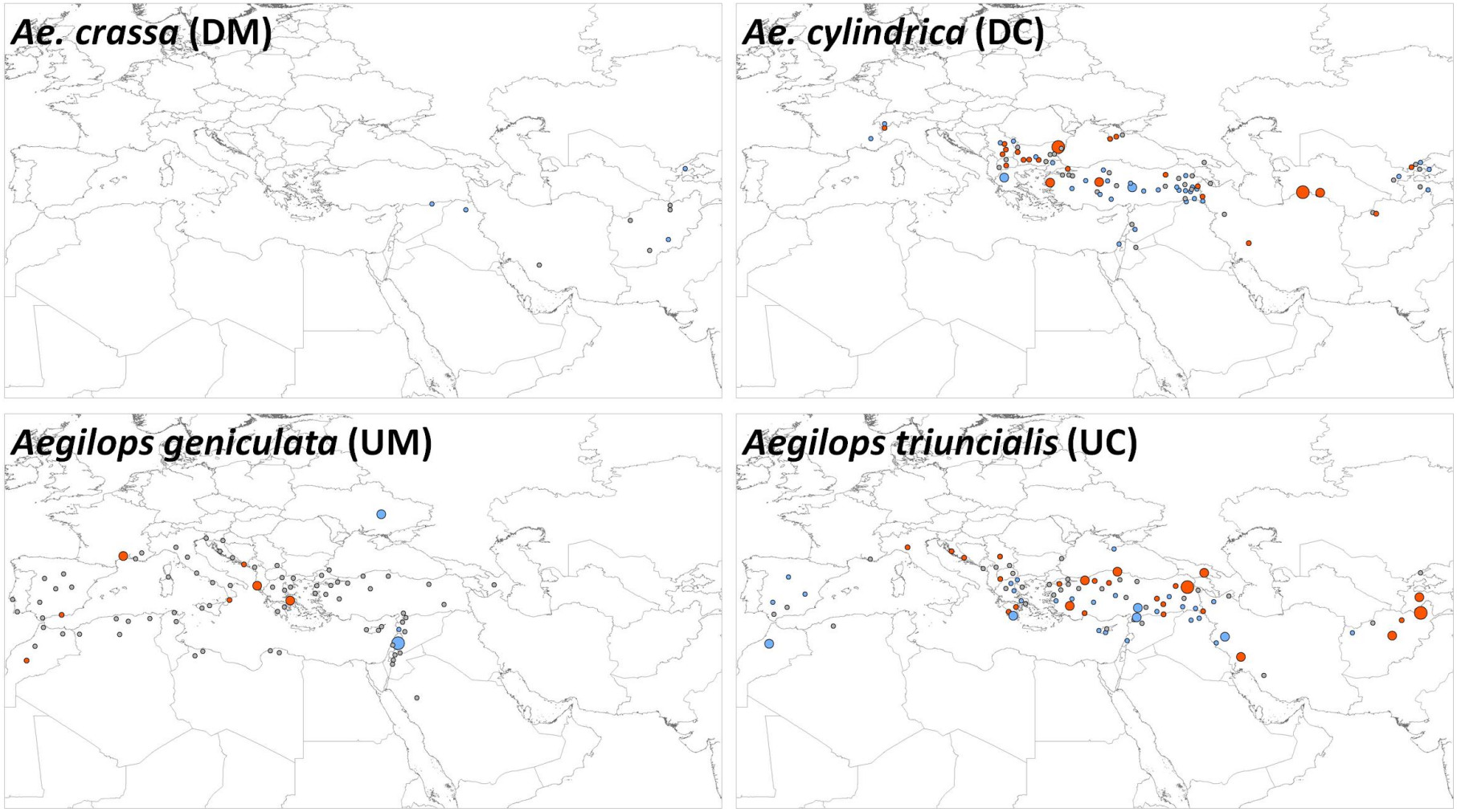


Figure S6b: Distribution of individual *Sabine* dynamics in terms of balance of lost bands over new bands (R) within allopolyploid *Aegilops* species. *Sabine* dynamics is expressed here as the departure (noted as 'z-scores') of balance of lost bands over new bands of *Sabine* from genome-wide variation as assessed by AFLP. Blue circles represent negative z-scores with increasing sizes for values lower than -0.5/-1.5/-2.5 standard deviation respectively. Red circles represent positive z-scores with increasing sizes for values higher than 0.5/1.5/2.5 standard deviation respectively. Values between -0.5 and 0.5 standard deviation are represented by small grey circles.

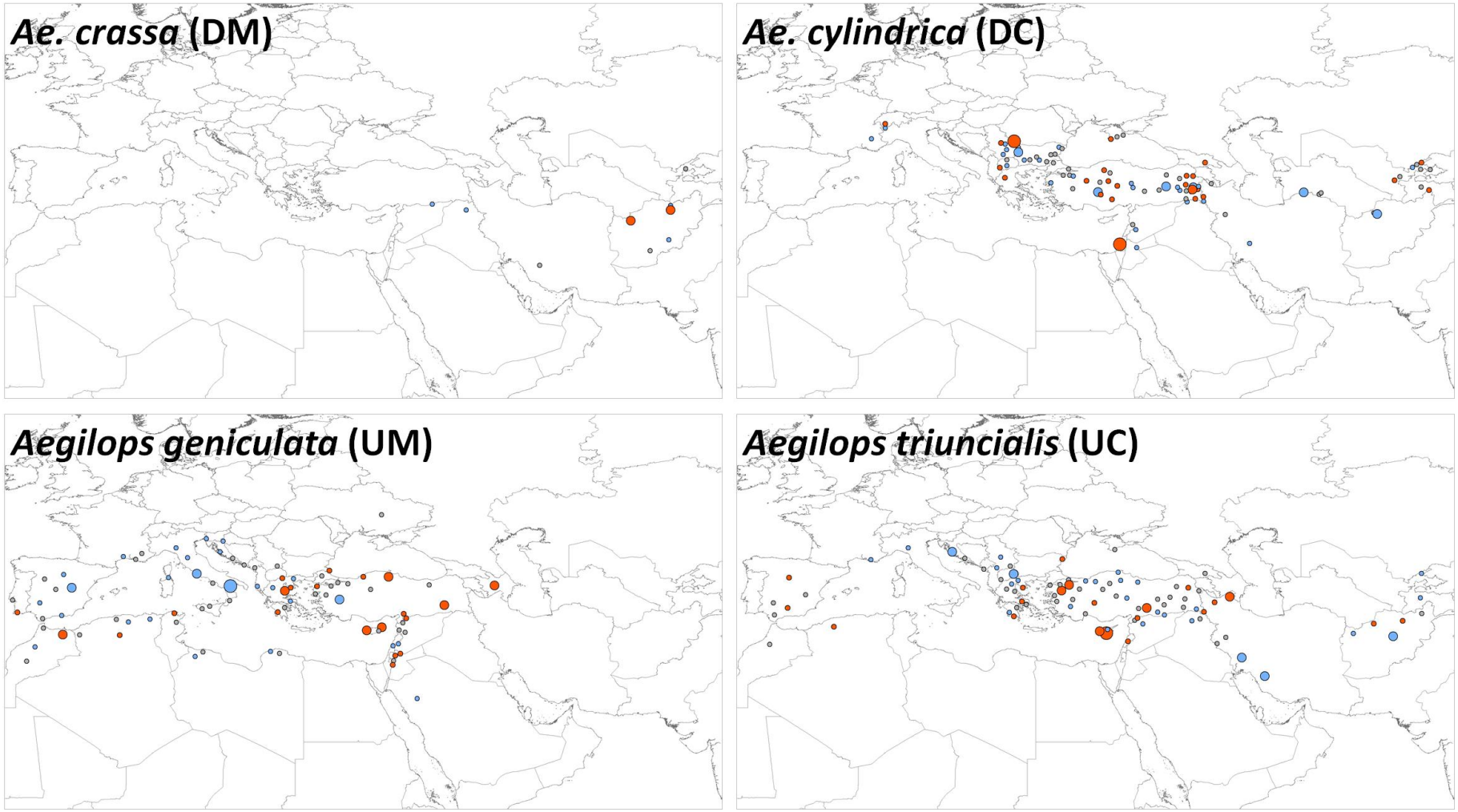


Figure S7a: Distribution of individual *Xalax* dynamics in terms of new bands (N) within allopolyploid *Aegilops* species. *Xalax* dynamics is expressed here as the departure of proportions (noted as 'z-scores') of new bands of *Xalax* from genome-wide variation as assessed by AFLP. Blue circles represent negative z-scores with increasing sizes for values lower than -0.5/-1.5/-2.5 standard deviation respectively. Red circles represent positive z-scores with increasing sizes for values higher than 0.5/1.5/2.5 standard deviation respectively. Values between -0.5 and 0.5 standard deviation are represented by small grey circles.

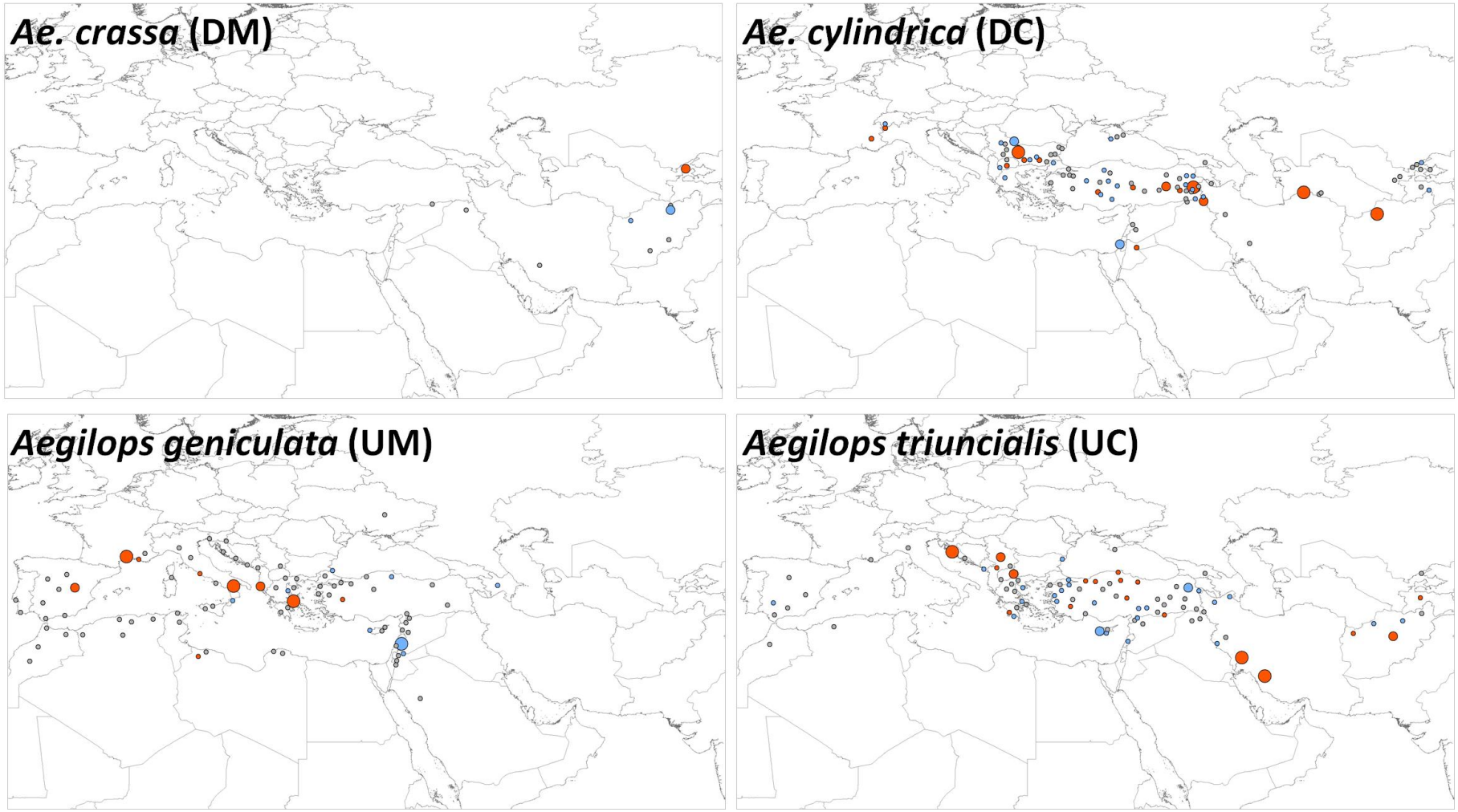


Figure S7b: Distribution of individual *Xalax* dynamics in terms of balance of lost bands over new bands (R) within allopolyploid *Aegilops* species. *Xalax* dynamics is expressed here as the departure (noted as 'z-scores') of balance of lost bands over new bands of *Xalax* from genome-wide variation as assessed by AFLP. Blue circles represent negative z-scores with increasing sizes for values lower than -0.5/-1.5/-2.5 standard deviation respectively. Red circles represent positive z-scores with increasing sizes for values higher than 0.5/1.5/2.5 standard deviation respectively. Values between -0.5 and 0.5 standard deviation are represented by small grey circles.

Table S1: Comparison of expected genetic additivity from progenitors of seven allopolyploid *Aegilops* accessions assessed through AFLP and SSAP in the present study and in Senerchia *et al.* (2014). Proportions of additive (A), lost (L) and new (N) bands have been assessed for the AFLP and the six LTR-RT families (SSAP): *Egug*, *Nusif*, *BARE1*, *Romani*, *Sabine* and *Xalax*.

			Results from Senerchia *et al.* (2014)			Results from present study		
Markers	**Species**	**Accessions**	**A**	**L**	**N**	**A**	**L**	**N**
AFLP	*Ae. crassa*	PI 542178	0.647	0.293	0.060	0.729	0.159	0.112
	Ae. cylindrica	PI 486235	0.725	0.227	0.048	0.847	0.116	0.037
	Ae. geniculata	PI 287737	0.685	0.273	0.042	0.806	0.121	0.073
		PI 487221	0.676	0.286	0.038	0.812	0.128	0.060
	Ae. triuncialis	PI 542345	0.708	0.236	0.056	0.841	0.134	0.025
		PI 491442	0.737	0.212	0.051	0.844	0.115	0.041
		PI 487246	0.736	0.207	0.056	0.847	0.118	0.034
Egug	*Ae. crassa*	PI 542178	0.687	0.269	0.044	0.704	0.189	0.107
	Ae. cylindrica	PI 486235	0.800	0.166	0.033	0.843	0.125	0.032
	Ae. geniculata	PI 287737	0.712	0.242	0.046	0.751	0.153	0.097
		PI 487221	0.772	0.184	0.044	0.743	0.153	0.104
	Ae. triuncialis	PI 542345	0.811	0.156	0.033	0.847	0.114	0.039
		PI 491442	0.814	0.152	0.034	0.843	0.120	0.037
		PI 487246	0.803	0.167	0.030	0.844	0.102	0.054
Nusif	*Ae. crassa*	PI 542178	0.675	0.238	0.088	0.788	0.149	0.063
	Ae. cylindrica	PI 486235	0.795	0.178	0.027	0.845	0.090	0.064
	Ae. geniculata	PI 287737	0.734	0.213	0.053	0.791	0.122	0.087
		PI 487221	0.727	0.218	0.055	0.810	0.085	0.105
	Ae. triuncialis	PI 542345	0.773	0.181	0.046	0.874	0.086	0.040
		PI 491442	0.782	0.181	0.037	0.832	0.129	0.038
		PI 487246	0.772	0.189	0.039	0.827	0.112	0.061
BARE1	*Ae. crassa*	PI 542178	0.648	0.262	0.091	0.728	0.179	0.093
	Ae. cylindrica	PI 486235	0.765	0.163	0.072	0.823	0.131	0.046
	Ae. geniculata	PI 287737	0.642	0.239	0.119	0.798	0.148	0.054
		PI 487221	0.580	0.306	0.114	0.754	0.137	0.109
	Ae. triuncialis	PI 542345	0.701	0.203	0.096	0.820	0.125	0.054
		PI 491442	0.710	0.198	0.092	0.830	0.122	0.047
		PI 487246	0.704	0.198	0.098	0.822	0.095	0.083
Romani	*Ae. crassa*	PI 542178	0.569	0.269	0.162	0.814	0.123	0.063
	Ae. cylindrica	PI 486235	0.795	0.144	0.061	0.832	0.106	0.062
	Ae. geniculata	PI 287737	0.667	0.207	0.125	0.811	0.095	0.094
		PI 487221	0.669	0.201	0.130	0.835	0.089	0.076
	Ae. triuncialis	PI 542345	0.784	0.137	0.080	0.846	0.129	0.025
		PI 491442	0.780	0.140	0.080	0.853	0.110	0.037
		PI 487246	0.773	0.153	0.074	0.856	0.092	0.051
Sabine	*Ae. crassa*	PI 542178	0.703	0.252	0.045	0.808	0.162	0.031
	Ae. cylindrica	PI 486235	0.813	0.126	0.061	0.772	0.163	0.065
	Ae. geniculata	PI 287737	0.669	0.308	0.023	0.796	0.159	0.045

		PI 487221	0.669	0.308	0.023	0.813	0.105	0.081
	Ae. triuncialis	PI 542345	0.761	0.183	0.055	0.761	0.213	0.026
		PI 491442	0.746	0.198	0.057	0.757	0.183	0.060
		PI 487246	0.753	0.194	0.053	0.714	0.160	0.126
Xalax	*Ae. crassa*	PI 542178	0.646	0.292	0.061	0.801	0.135	0.064
	Ae. cylindrica	PI 486235	0.774	0.164	0.061	0.891	0.076	0.034
	Ae. geniculata	PI 287737	0.720	0.205	0.074	0.815	0.107	0.077
		PI 487221	0.737	0.188	0.075	0.788	0.121	0.091
	Ae. triuncialis	PI 542345	0.620	0.360	0.020	0.859	0.086	0.055
		PI 491442	0.752	0.214	0.034	0.867	0.094	0.039
		PI 487246	0.729	0.236	0.035	0.848	0.094	0.058

Table S2: Sequences of adaptators (5' -> 3'), preselective and selective primers amplified fragment length polymorphism (AFLP) and sequence specific amplified polymorphic (SSAP) techniques. Sequences were taken from Senerchia *et al.* (2014) and error rates calculated following Bonin *et al.* (2004).

Primer type	**TE family**	**Primer sequence 5'-3'**	**MseI selective primer**	**Error rate %**
Selective	AFLP	E-AAG	M-CGA, M-CGC	3.31
		E-ATT	M-CAA, M-CCA	
		E-ACG	M-CAC, M-CTA	
	Egug	TGA AAG TTG TAT AAT ATT GGC ATG G	M-CGA, M-CTA, M-CAA	5.32
	Nusif	AGG CCG CAG GCA GGC AAG	M-CAC, M-CGA, M-CGC	5.84
	BARE1	GCC TCT AGG GCA TAT TTC CA	M-CCA, M-CTA, M-CAA	4.56
	Romani	TGA TGA TCT AAT GTA TAA CAT TCC AAA	M-CAC, M-CCA, M-CGC	4.76
	Sabine	GCA TGG AGC AAG CAG AAA TTA T	M-CAC, M-CTA, M-CGC	5.57 4.01
	Xalax	GGT TCT GAA ATC AAA AGA TAA ATC	M-CAC, M-CGA, M-CGC	
Adaptators EcoR1	all	CTC GTA GAC TGC GTA CC		
	all	AAT TGG TAC GCA GTC TAC		
Adaptators Mse1	all	GAC GAT GAG TCC TGA G		
	all	TAC TCA GGA GTC AT		
Preselective EcoR1	all	GAC TGC GTA CCA ATT CA		
Preselective Mse1	all	GAT GAG TCC TGA GTA AC		

Table S3: Mean number of loci in the *Aegilops* species for random sequences (AFLP) and the 6 TEs families (SSAP). The correlation coefficient *r* between fragment sizes and their associated frequencies was computed in AFLPsurv following Vekemans *et al.* (2002).

	Ae. caudata	***Ae. comosa***	***Ae. crassa***	***Ae. cylindrica***	***Ae. geniculata***	***Ae. tauschii***	***Ae. triuncialis***	***Ae. umbellulata***	**Total**	***r***
AFLP	266	292	238	254	349	284	266	237	522	-0.4341
Egug	95	158	111	93	146	131	113	105	287	-0.3782
Nusif	116	105	98	117	139	116	122	118	178	-0.6827
BARE1	91	131	84	77	147	132	134	113	257	-0.5383
Romani	50	66	41	57	83	62	41	39	190	-0.2960
Sabine	99	59	48	76	68	62	89	59	146	-0.5258
Xalax	74	99	55	80	118	86	93	70	248	-0.2772

Table S4: Expected progenitors' genetic additivity expected in the four allopolyploid *Aegilops* species. Mean proportions of shared bands among diploid progenitors (S) and comparisons with the expected additivity of the diploids (A, additive bands; L, lost bands; N, new bands; log(L/N) balance of lost bands over new bands) are presented for each allotetraploid species. The proportions of the A, L and N bands were compared between each TEs families and AFLP using a Tukey HSD test. AFLP or TE families not sharing a small letters have significantly different proportions.

	Ae. crassa **(genome DM)**				***Ae. cylindrica*** **(genome DC)**				***Ae. geniculata*** **(genome UM)**				***Ae. triuncialis*** **(genome UC)**			
	A	**L**	**N**	**R**	**A**	**L**	**N**	**R**	**A**	**L**	**N**	**R**	**A**	**L**	**N**	**R**
AFLP	0.722 ^{cd}	0.176 ^{ab}	0.102 ^{a}	0.544 ^{b}	0.836 ^{b}	0.129 ^{c}	0.035 ^{b}	1.474 ^{a}	0.795 ^{c}	0.136 ^{b}	0.070 ^{b}	0.691 ^{bc}	0.845 ^{c}	0.117 ^{c}	0.038 ^{bc}	1.125 ^{a}
Egug	0.690 ^{d}	0.196 ^{a}	0.113 ^{a}	0.558 ^{b}	0.837 ^{b}	0.128 ^{c}	0.034 ^{b}	1.352 ^{ab}	0.754 ^{e}	0.166 ^{a}	0.080 ^{a}	0.776 ^{b}	0.847 ^{bc}	0.113 ^{c}	0.039 ^{bc}	1.102 ^{a}
Nusif	0.745 ^{bc}	0.190 ^{a}	0.065 ^{b}	1.148 ^{a}	0.838 ^{b}	0.123 ^{c}	0.038 ^{b}	1.218 ^{b}	0.802 ^{bc}	0.121 ^{c}	0.077 ^{ab}	0.468 ^{c}	0.809 ^{d}	0.147 ^{b}	0.044 ^{b}	1.221 ^{a}
BARE1	0.724 ^{cd}	0.191 ^{a}	0.085 ^{ab}	0.826 ^{ab}	0.822 ^{c}	0.139 ^{b}	0.039 ^{b}	1.277 ^{ab}	0.773 ^{d}	0.156 ^{a}	0.072 ^{ab}	0.825 ^{b}	0.815 ^{d}	0.121 ^{c}	0.064 ^{a}	0.677 ^{b}
Romani	0.757 ^{abc}	0.158 ^{b}	0.086 ^{ab}	0.642 ^{b}	0.835 ^{b}	0.110 ^{d}	0.056 ^{a}	0.713 ^{c}	0.832 ^{a}	0.101 ^{d}	0.067 ^{b}	0.492 ^{c}	0.855 ^{ab}	0.111 ^{c}	0.033 ^{c}	1.271 ^{a}
Sabine	0.765 ^{ab}	0.148 ^{b}	0.087 ^{ab}	0.584 ^{b}	0.763 ^{d}	0.187 ^{a}	0.050 ^{a}	1.474 ^{a}	0.827 ^{a}	0.133 ^{b}	0.040 ^{c}	1.352 ^{a}	0.738 ^{e}	0.199 ^{a}	0.062 ^{a}	1.239 ^{a}
Xalax	0.790 ^{a}	0.149 ^{b}	0.061 ^{b}	0.931 ^{ab}	0.873 ^{a}	0.088 ^{e}	0.039 ^{b}	0.835 ^{c}	0.806 ^{b}	0.122 ^{c}	0.072 ^{ab}	0.605 ^{bc}	0.860 ^{a}	0.095 ^{d}	0.045 ^{b}	0.789 ^{b}

Table S5: Proportions of lost bands originating from the differential *vs.* the pivotal genome of each *Aegilops* allopolyploid species. Differences in proportions between the differential and the pivotal genomes were tested for each tetraploid species and TEs family through a proportion test and significance indicated as follows: *** < 0.001, ** < 0.01, * < 0.05, ns > 0.05.

Allopolyploid species	TE family	Nb of loci specific to differential genomal	Nb of loci specific to pivotal genome	% lost bands from the differential genome	% lost bands from the pivotal genome	p-value
Ae. crassa	AFLP	94	78	0.58	0.42	*
Ae. crassa	*Egug*	79	49	0.58	0.42	*
Ae. crassa	*Nusif*	24	35	0.65	0.35	ns
Ae. crassa	*BARE1*	61	63	0.53	0.47	ns
Ae. crassa	*Romani*	42	38	0.49	0.51	ns
Ae. crassa	*Sabine*	21	24	0.53	0.47	ns
Ae. crassa	*Xalax*	59	46	0.56	0.44	ns
Ae. cylindrica	AFLP	85	103	0.56	0.44	ns
Ae. cylindrica	*Egug*	35	71	0.52	0.48	ns
Ae. cylindrica	*Nusif*	24	24	0.52	0.48	ns
Ae. cylindrica	*BARE1*	29	70	0.62	0.38	***
Ae. cylindrica	*Romani*	29	41	0.56	0.44	ns
Ae. cylindrica	*Sabine*	54	17	0.54	0.46	ns
Ae. cylindrica	*Xalax*	39	48	0.54	0.46	ns
Ae. geniculata	AFLP	110	57	0.47	0.53	ns
Ae. geniculata	*Egug*	95	43	0.54	0.46	ns
Ae. geniculata	*Nusif*	19	32	0.51	0.49	ns
Ae. geniculata	*BARE1*	54	36	0.42	0.58	*
Ae. geniculata	*Romani*	46	19	0.59	0.41	*
Ae. geniculata	*Sabine*	24	24	0.44	0.56	ns
Ae. geniculata	*Xalax*	62	33	0.45	0.55	ns
Ae. triuncialis	AFLP	93	64	0.55	0.45	ns
Ae. triuncialis	*Egug*	40	49	0.50	0.50	ns
Ae. triuncialis	*Nusif*	21	23	0.53	0.47	ns
Ae. triuncialis	*BARE1*	32	54	0.57	0.43	ns
Ae. triuncialis	*Romani*	38	27	0.57	0.43	ns
Ae. triuncialis	*Sabine*	58	18	0.54	0.46	ns
Ae. triuncialis	*Xalax*	41	37	0.50	0.50	ns

Chapter 4

Allopolyploid ecological niche dynamics of wild wheats *Aegilops* species

Stella Huynh[1], Olivier Broenniman[2,3], François Felber[1,4], Antoine Guisan[2,3] and Christian Parisod[4]

[1] Institute of Biology, University of Neuchâtel, Switzerland

[2] Department of Ecology and Evolution, University of Lausanne, Switzerland

[3] Institute of terrestrial surface dynamics, University of Lausanne, Switzerland

[4] Musée et Jardins botaniques cantonaux de Lausanne et Pont-de-Nant, Switzerland

[5] Institute of Plant Sciences, University of Bern, Switzerland

List of figures and tables:

Abstract

Polyploidy represents one of the major mechanisms of speciation among flowering plants. In particular, allopolyploid species generally show a larger distribution than their diploid relatives, as exemplified by wild wheats, *Aegilops* species. The broader range of the allopolyploids might be explained by higher colonization abilities as compared to their diploid progenitors or by ecological niche shifts of the allopolyploids. To test the allopolyploid niche shift hypothesis, the ecological niches of four diploid *Aegilops* species and four of their derivative allopolyploid species were modelled in the environmental space. The ecological niche dynamics of the allopolyploids was decomposed into additive niche (i.e. shared niche with progenitor's niches) and non-additive niche (niche retraction or expansion compared to the addition of progenitors' niches). The ecological niches of the allopolyploid species revealed mostly additive to their diploid progenitors. Overall, diploids showed stronger patterns of isolation-by-ecology, potentially precluding local adaptation, than the allopolyploids. The wider distribution of allopolyploid species thus appeared mainly as the result of higher colonization abilities rather than ecological niche shifts driven by minority cytotype disadvantage. Additionally, allopolyploids niches showed a slight expansion toward more temperate climates, especially located the species margins. Allopolyploid range expansion into novel ecological niches may reflect the potential of allopolyploids to colonize new habitats, which may be due to their high genetic diversity and genome plasticity. The role of genetic novelties into the evolution of ecological novelties remains however obscure.

Keywords: niche modeling, ecological niche shift, allopolyploid niche dynamics, allopolyploid additivity, *Aegilops*

Introduction

Polyploidization, or whole-genome duplication (WGD), has been recognized as an important mechanism of speciation and diversification in plants (Stebbins, 1971; Levin, 2002; Wood *et al.*, 2009). Polyploid individuals can form within species (autopolyploidy) or among genetically divergent species (allopolyploidy). Allopolyploidy have both immediate consequences (i.e. the revolutionary phase) with drastic chromosomal rearrangements and epigenetic repatterning (Wang *et al.*, 2004, p. 201; Feldman *et al.*, 2012), and long-term effects (i.e. the evolutionary phase) on species evolution such as genetic mutations inactivating gene duplicates (pseudogenization) or modify the initial function of the gene (neofunctionalization) (Feldman 2009). Allopolyploidy can confer disadvantages such as changes in cellular architecture or size and meiotic or epigenetic instability, but also evolutionary advantages such as heterosis (hybrid vigor) or masking of deleterious alleles through gene redundancy (Comai, 2005). Allopolyploids generally arise in sympatry with their progenitors (Ramsey & Schemske, 1998, 2002) and thus may face difficulties to establish due to competition for resources or reproductive disadvantage in case of allogamous breeding system, diploid progenitors representing the major cytotypes (the 'minority cytotype disadvantage' hypothesis; Levin 1975). However, in case of selfing reproductive regime, the allopolyploid may circumvent the minority cytotype disadvantage (MCD). Allopolyploids commonly show larger geographical distribution than their progenitor species and possibly reflect wider ecological tolerances. It has also been proposed that nascent allopolyploids may avoid MCD and expand their distribution by differentiating their ecological niches (Levin, 1975), also termed as 'niche shift hypothesis', or due to long-distance dispersal, i.e. without the need to shift their ecological niche (Linder and Barker 2014), though their relative importance remains obscure. Allopolyploids harbor divergent genomic background, that may confer greater ecological tolerance than their progenitors, thus allowing them to expand into stressful habitats (Stebbins, 1950; te Beest *et al.*, 2012). For instance, higher tolerance to drought in allopolyploids has been associated to increased cell size subsequent to genome doubling, resulting in lower density of guard cells and thus reduced transpiration rates (Li *et al.*, 1996; te Beest *et al.*, 2012). However, association between allopolyploid expansion and ecological niche shift of allopolyploids relative to that of their progenitors have not yet been clearly established.

An allopolyploid species can expectedly represent the genetic additivity of its diploid progenitors. However, allopolyploid genomics rather reported shifts from the expected

progenitor genetic backgrounds, through irreversible structural changes generally leading to decreased genome size (Leitch & Bennett, 2004; Bento *et al.*, 2010) or through asymmetrical epigenetic changes (Adams *et al.*, 2003; Eckardt, 2010; Madlung & Wendel, 2013). To what extent similar hypothesis of ecological additivity holds or whether allopolyploidy typically shift from their progenitors' ecological requirements has been under-investigated. The improvement of georeferencing species occurrences and of climatic factors databases at a finer scale, along with the rapid development of methodological tools to model and compare species ecological niches, makes now feasible to explore the ecological niche dynamics of polyploid species (Theodoridis *et al.*, 2013; Anacker & Strauss, 2014; Glennon *et al.*, 2014; Casazza *et al.*, 2017). In particular, ecological niche modelling techniques can be used to investigate the ecological niche shift of allopolyploids from the addition of that of their progenitors, such as the ordination technique (e.g. PCA; Broennimann *et al.*, 2012). Compared to the Species Distribution Model (SDM), the ordination technique has the advantage to model species niches in a common environment space, therefore enabling direct comparison of modelled niches (Broennimann *et al.* 2012). Though performant statistical tools to model and compare species ecological niches are developing, ecological studies have been essentially conducted on autopolyploid species (Ramsey & Ramsey, 2014), as the null hypothesis (i.e. addition of diploids of the same species) is simpler to test for these organisms. However, most of these supposedly autopolyploid species revealed in fact being allopolyploids, highlighting the importance of confidently identifying the progenitor species of polyploids. To date, very few studies have investigated the ecology of allopolyploids relative to their progenitors, among which Marchant *et al.* (2016) have compared the ecological niches of 13 allopolyploid systems with their diploid progenitors, however without making a clear hypothesis of the allopolyploid niche expected under the additivity of progenitors' niches. Proper study of allopolyploid niche dynamics investigating the hypothesis of ecological additivity of allopolyploid's progenitors is thus still needed.

The *Aegilops* species. (Poaceae) are annual and selfing grasses mainly occurring in open or disturbed habitats and composed of diploid species that differentially combined into allotetraploids and allohexaploids (van Slageren, 1994; Badaeva *et al.*, 2002, 2004). While allopolyploids are widespread across the Mediterranean and Anatolia regions, the distributions of diploid species are more restricted (in the Balkan, Anatolian or Caspian region). The evolutionary history of the allopolyploids have previously been investigated (see chapter 2) and their diploid progenitors well identified (Badaeva *et al.*, 2002, 2004), thus

representing an ideal model to test the hypothesis of ecological additivity in allopolyploid systems.

The evolutionary history of four diploid species (*Ae. caudata* (CC), *Ae. comosa* (MM), *Ae. tauschii* (DD) and *Ae. umbellulata* (UU)) and four of their derivative allopolyploid species (*Ae. cylindrica* (DDCC), *Ae. crassa* (DDMM), *Ae. geniculata* (UUMM) and *Ae. triuncialis* (UUCC)) highlighted cycles of species range contraction-expansion subsequent to glaciation-deglaciation periods and rapid postglacial colonization of the allopolyploid species (see chapter 2). Whether the broader distribution of the allopolyploids compared to their diploid progenitors results from higher dispersal capacities of the allopolyploids, which may leave a trace of isolation-by-distance, or from allopolyploid lineages shifting their ecological preferences, which would be reflected by isolation-by-ecology, remains to be deciphered and is the main objective of the present study. In particular, whether the allopolyploids share the same ecological niches as their diploid progenitors and to what extent allopolyploids' distribution may be driven by ecological niches of the species was investigated.

We thus assessed the ecological niches of these four diploid and four allopolyploid *Aegilops* species and tested the hypothesis that the niche of each allopolyploid species consisted in the addition of their diploid progenitors' niches. The ecological niches, which showed contrasted niche centroid and niche breadth among species, were compared by measuring pairwise niche overlaps among diploids and their derivative allopolyploids. To assess whether the distribution of the species are driven by migration or by ecology, patterns of isolation by ecology were analyzed jointly with patterns of isolation by distance using partial Mantel tests.

Material and Methods

Species distribution and occurrence data

Occurrences of the eight *Aegilops* species were collected from the GBIF database (http://www.gbif.org/) and filtered out from replicates to keep a total of 3265 occurrences georeferenced under the WGS84 coordinates system, with at least 4-decimal coordinates corresponding to a geolocalization precision of 1 km (Figure 1, Supplemental Material Table S1). The 402 *Aegilops* accessions previously genotyped in Chapter 1, 2 and 3 were additionnally included in the present dataset (Figure 1), i.e. for a total of 3667 accessions used for inferring the ecological niches of *Aegilops* species. An additional factor with similar high resolution (30 arc sec, i.e. 1 km precision) has been added from CGIAR database (Trabucco & Zomer, 2010): the 'soil water balance', a hydrological dataset based on WorldClim and global potential evapo-transpiration (Global-PET; CGIAR) that highlights the climatic influence on hydrological dimensions regulating vegetation suitability (evapo-transpiration and soil water deficit). The global 'soil water balance' raster data are available monthly. Previous analyses revealing high correlation among the 12 months (data not shown), we thus reduced the soil water balance dataset to only one of the months (here the 11th).

Ecological niche modeling

The 20 environmental variables were sampled across the natural climatic biomes, as defined by the World Wildlife Fund (WWF; Olson *et al.*, 2001), where occurrences were recorded for at least one *Aegilops* species (Supplemental Material Figure S1). Occurrences of *Ae. triuncialis* and *Ae. cylindrica* recorded on the American continent were excluded from the sampling area as they actually occurred as invasive species there (Lyons *et al.*, 2010). The high resolution of the 20 bioclimatic datasets (1 km) enabled to model the ecological niches precisely. To further improve the modeling precision, large geographical areas where no occurrences were reported (e.g. Russia, the desert biome in North Africa; Supplemental Material S1) were excluded from the geographical area used to model the environmental space (E-space) common to the eight *Aegilops* species.

A total of 1'000'000 raster cells were randomly sampled across the geographical area delimited for the eight species and their environmental values extracted for the 20 bioclimatic variables, each raster cell corresponding to the bioclimatic map resolution (i.e. 1 km). The E-space was defined through ordination technique, which allowed for direct comparisons of species niches modeled in the E-space (Broennimann *et al.*, 2012). The common E-space was

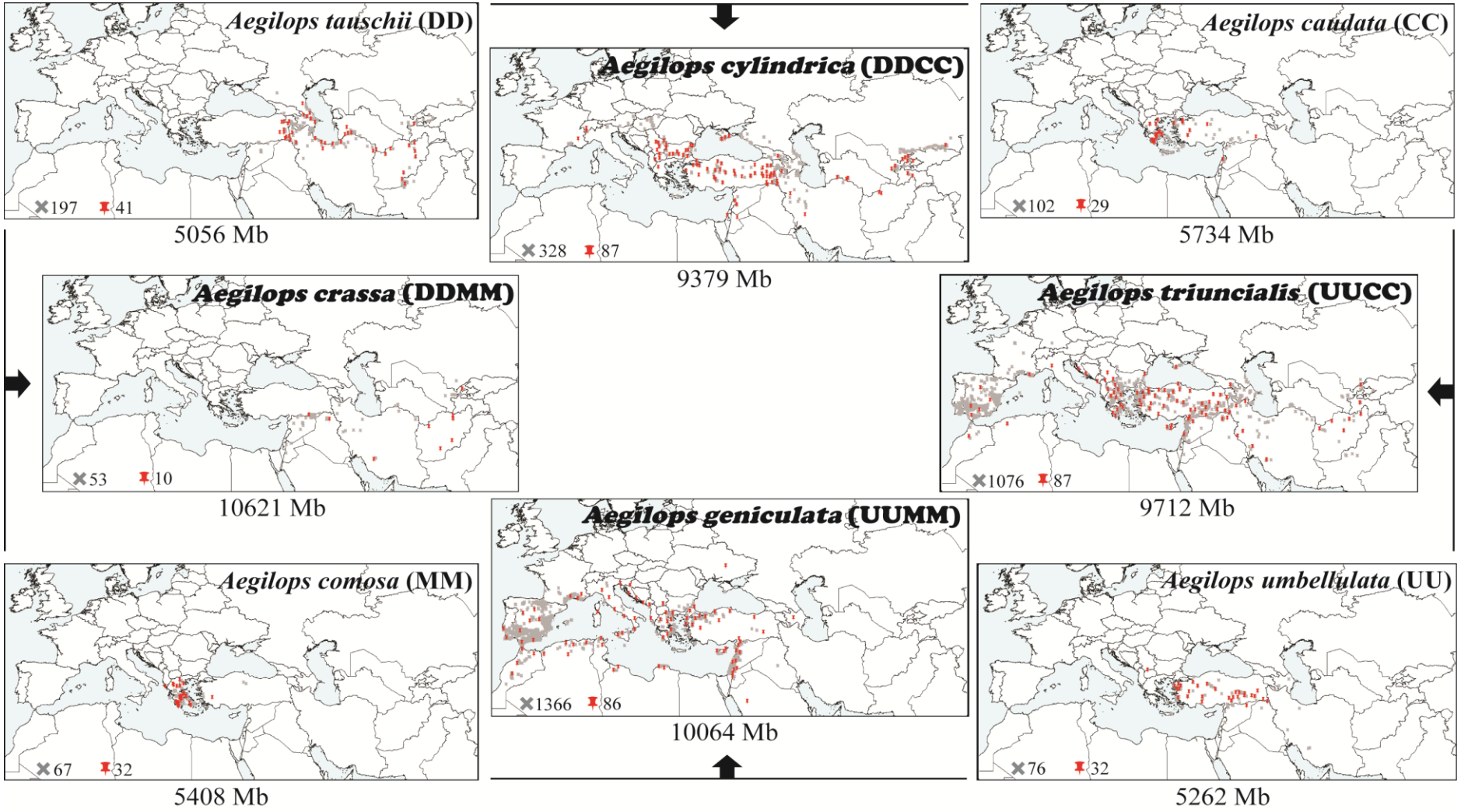


Figure 1: Distribution range of the four diploid *Aegilops* species and their four derivated allopolyploids. Species distributions are represented with grey circles for occurrences reported in GBIF database (https://www.gbif.org/) and red thumbtacks for genotyped accessions. The genomic formula of each species is indicated into brackets following van Slageren (1994) and the maternal origin(s) of polyploids underlined based on inferred plastid haplotype network in Chapter 1, except *Ae. crassa* whose progenitors remain unclear.

modelled based on the 20 bioclimatic variables extracted for 1'000'000 raster cells and projected in a two-dimensional space defined by the first two axes of a principal component analysis (PCA) using the R-package *ecospat* (Di Cola *et al.*, 2017), following the framework of Broennimann *et al.* (2012).

The realized ecological niche of each *Aegilops* species (i.e. the ecological niche where each species currently occur, as opposed to where the species could theoretically occur or 'fundamental niche') was modelled in the common E-space using Gaussian kernel density plots to smooth the density of occurrences, thus ensuring the analyses to be independent from the sampling effort and the resolution of the E-space (Broennimann *et al.*, 2012). The niche breadth was estimated for each species as the dispersal around the centroid position of the modelled niche for each principal component. As the ecological niche of each species was modelled in a common E-space, the inferred ecological niches could be straightforwardly compared among species and their overlap calculated using the metric D (Schoener, 1968). Schoener's D measured the percentage of similarity between two species niches modelled in the E-space based on species probability scores in each raster cell where the species occurs. The significance of niche overlaps was tested through similarity tests following Warren *et al.* (2008), by randomly reallocating niche centroid of one or both species over the common E-space and recalculating their niche overlap through 100 repetitions. The similarity test thus assessed whether the ecological niche occupied by one species was more similar to the niche occupied by the other species than would be expected chance (5% threshold). Similarity tests of niche overlap among diploids and among allopolyploids were performed by randomizing niche centroid of both species, as no reference niche was expected for these comparisons. Similarity tests for diploid-allopolyploid niche overlaps were similarly performed, but by randomly reallocating only the centroid of the allopolyploid niche in the E- space and using the centroid the progenitor niche as the reference.

Association between ecological and genetic distances among Aegilops species

To test for genetic patterns coherent with isolation-by-distance or isolation-by-environment within each species, correlations between genetic, ecological and geographic pairwise distances between genotyped individuals were assessed through Mantel and partial Mantel tests. The pairwise genetic distances among the genotyped individuals were previously inferred from nuclear genotype frequencies in Chapter 2 and measured as Rousset's *a*, a statistic analogous to $F_{ST}/(1-F_{ST})$ but based on the probability of identity of genes within and among individuals (Rousset, 2000). The geographical pairwise distances were computed as

the shortest flight-distances between geographical coordinates of each accessions, and accounting for the Earth ellipsoid in the WGS84 system, using the R package *geosphere* (Hijmans, 2017). Pairwise ecological distances were computed based on the individual scores along the first two PCs of the common E-space (explaining 70.57% of the total inertia). Genetic pairwise distances were first correlated to the geographical and the ecological distances separately within each species, using Mantel tests through 1000 permutations. Genetic pairwise distances were then correlated to ecological pairwise distances while accounting for the effect of geographical pairwise distances using partial Mantel tests, with significance tested through 1000 permutations. Mantel and partial Mantel tests were both performed using the R package *vegan* (Oksanen *et al.*, 2018). For each species, positive correlations between geographical and genetic pairwise distances of genotyped individuals were interpreted as patterns of 'isolation-by-distance' (IBD), reflecting species range expansion driven by migration. Positive correlations between ecological and genetic pairwise distances, and accounting for the geographical pairwise distances, among genotyped individuals within each species were however interpreted as patterns of 'isolation-by-environment' (IBE), which may reflect species range expansion driven instead by specific ecological requirements, independently from the geographic distances (Wang *et al.*, 2013a; Wang & Bradburd, 2014).

Allopolyploid niche dynamics

To test the hypothesis of allopolyploid niche additivity (i.e. being the sum of its progenitors' niches, thereafter the expected allopolyploid niche), we compared the observed allopolyploid niche to the expected niche modeled from combined occurrences of both its progenitors. Niche overlaps between the observed and expected niches of the allopolyploid species were similarly measured with the Schoener's D metric and their significance assessed through the similarity test in 'ecospat' as before, but using here the expected niche as the reference. Allopolyploid niche dynamics, characterized by the difference between the observed allopolyploid niche and the expected combined niche from its progenitors, was decomposed into niche stability and niche shift. Niche stability was defined as the fraction of the observed allopolyploid niche overlapping the addition of both progenitors' niches. The non-additive fraction of the observed allopolyploid niche (i.e. niche shift) was decomposed into niche retraction (i.e. proportion of E-space unoccupied by the allopolyploid as compared to additive predictions) and niche expansion (i.e. proportion of E-space newly occupied by the allopolyploid as compared to additive predictions).

Results

Ecological niche modeling

Modeling the realized ecological niche of each *Aegilops* species, the first two principal components (PCs) extracted 70.57% of the bioclimatic variation within the selected E-space (Figure 2). Temperature and precipitation variables were mostly associated with the first PC. High precipitation seasonality (PS) and high temperatures during the warmest months (i.e. T, Twarm, Twarm3) were positively correlated with this PC, whereas variables related to high precipitations across seasons (i.e. P, Pdry3, Pwarm3) as well as high soil moisture (swc11) were negatively correlated. Positive values along the first PC thus likely characterized ecoregions with hot and dry summer typical to Mediterranean climates, whereas negative values identified cool and humid summer and winter corresponding to temperate climates. The second PC was essentially associated with temperature seasonality (TdV, TS, TaV), distinguishing high annual temperature amplitudes characteristic to continental climates.

As expected, modeling of the E- space occupied by the different species showed that the ecological niche of each *Aegilops* species was mostly Mediterranean, although with slightly different optima (Figure 2). In particular, *Ae. tauschii* and *Ae. umbellulata* appeared as the diploid species with the most continental niches, although showing contrasted niche breadths along PC1 and PC2 (Table 1). The former indeed occurred over a range of conditions along the temperate – Mediterranean gradient, whereas the latter mostly extended across the gradient of continentality. *Ae. tauschii* showed the largest niche breadth, contrasting with other diploid species such as *Ae. comosa* and *Ae. caudata* showing similar niche optima and equally small niche breadths. Allopolyploids generally occupied slightly divergent niche optima, but noticeably larger niche breadths (ranging from 1.17 to 4.05) than diploids (ranging from 0.47 to 2.49; Table 1). Accordingly, niche overlap was relatively low among diploids (ranging from 0.14 to 0.49) whereas relatively high between allopolyploid species, reaching 0.72 between *Ae. geniculata* and *Ae. triuncialis* (Figure 2). *Ae crassa* appeared as an exception to this pattern as indicated by its relatively small niche breadth and limited overlap with other allopolyploids.

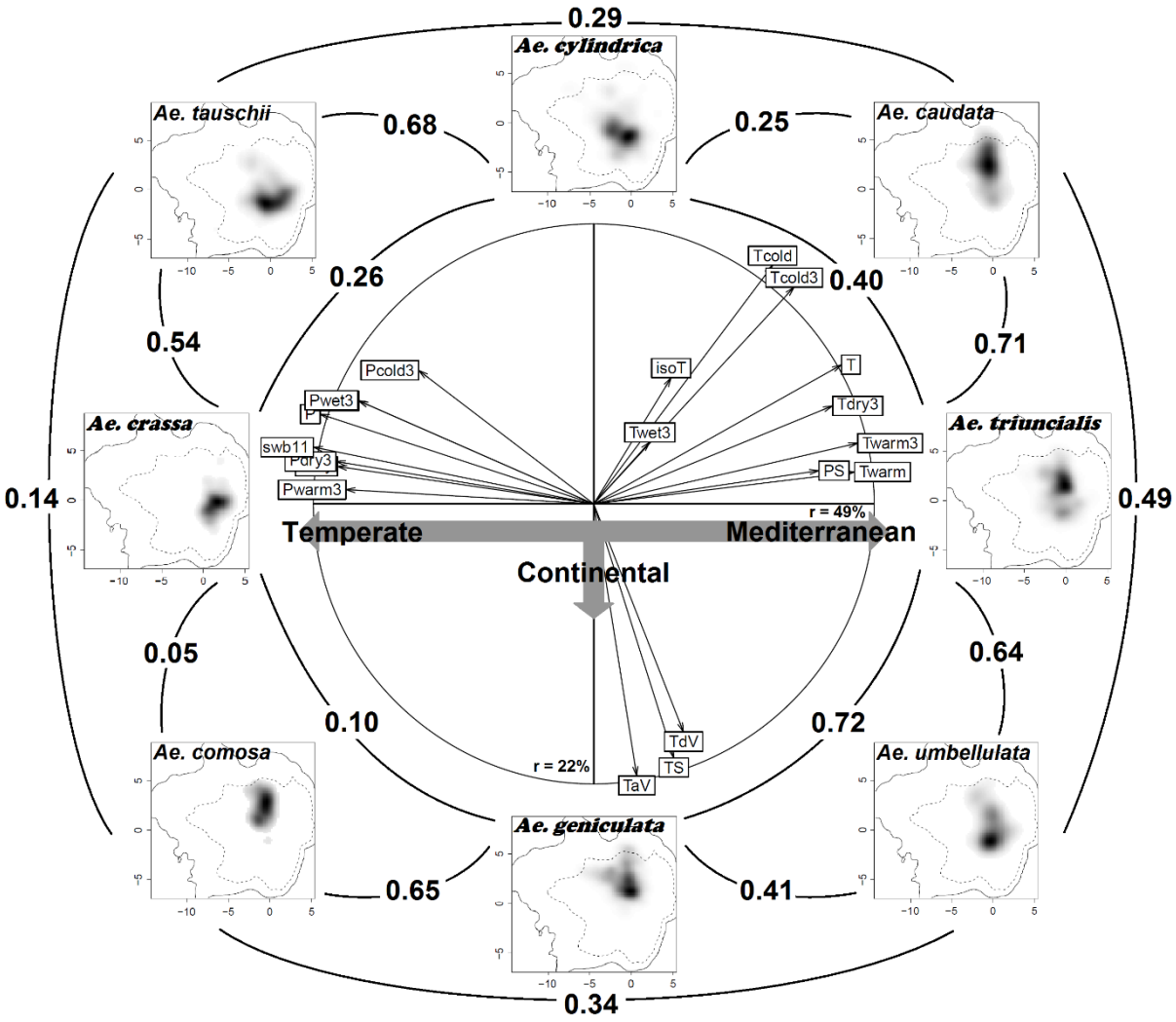


Figure 2: Ecological niches of diploid *Aegilops* species and their derivated allopolyploid species. Species niches were modelled on the first two axes of PCA analyses. The density of occurrences are modelled using kernel density plot, providing a standardized density of each species occurrences. The relative contribution of the 20 selected bioclimatic factors to the first two principal components (~71% of total variance) used here to display ecological niches of surveyed species is shown in the central correlation circle that points to a gradient along Temperate and Mediterranean climates and towards a more Continental climate according to the grey arrows. The bioclimatic factors are coded as follows: (T) annual mean temperature, (TdV) mean diurnal range, (isoT) isothermality, (TS) temperature seasonality, (Twarm) maximum temperature of the warmest month, (Tcold) minimum temperature of the coldest month, (TaV) temperature annual range, (Twet3) mean temperature of the wettest quarter, (Tdry3) mean temperature of the driest quarter, (Twarm3) mean temperature of the warmest quarter, (Tcold3) mean temperature of the coldest quarter, (P) annual precipitation, (PS) precipitation seasonality, (Pwet) precipitation of the wettest month, (Pdry) precipitation of the driest month, (Pwet3) precipitation of the wettest quarter, (Pdry3) precipitation of the driest quarter, (Pwarm3) precipitation of the warmest quarter, (Pcold3) precipitation of the coldest quarter and (swc11) soil water content. Pairwise ecological niche overlap among species, indicated by black lines, are presented as the proportion of environmental space shared among species.

Table 1: Modelled centroids and breadths of ecological niches and isolation by distance and by environment in *Aegilops* species.

Ploidy	Species	Niche centroids PC1 / PC2[a]	Niche breadths PC1 / PC2 [a]	G ~ Geo [b]	G ~ E [b]	G ~ E | Geo [b]
2n	*Ae. caudata*	-0.28 / 2.21	0.92 / 1.56	0.455 ***	0.421 ***	0.221 *
	Ae. comosa	-0.32 / 2.32	0.55 / 0.47	0.226 ***	- 0.008 ns	- 0.153 ns
	Ae. tauschii	0.06 / -0.82	2.49 / 2.02	-0.004 ns	0.313 ***	0.315 ***
	Ae. umbellulata	-0.07 / -0.46	0.78 / 1.81	0.483 ***	0.587 ***	0.447 ***
4n	*Ae. geniculata*	-0.56 / 1.90	4.05 / 2.85	0.042 ns	- 0.007 ns	- 0.017 ns
	Ae. triuncialis	-0.27 / 0.09	2.30 / 2.90	0.055 ns	0.223 ***	0.217 ***
	Ae. cylindrica	-0.77 / 1.82	1.97 / 1.82	0.145 *	0.141 *	0.094 ns
	Ae. crassa	1.67 / 0.94	1.17 / 0.94	0.172 ns	- 0.080 ns	- 0.085 ns

[a] The ecological niche centroid is defined as the centroid of the niche density along principal components (PC1 and PC2), while niche breadth is the dispersal from this centroid. [b] Mantel and Partial Mantel tests associating genetic (G; based on Rousset's *a*), ecological (E) and geographical (Geo) distances among individuals within each species, with significance tested through 1000 permutations (p-values: (***) < 0.001, (**) < 0.01, (*) < 0.5 and (ns) non-significant). G ~ E | Geo presents tests of isolation by environment while taking isolation by distance is taken into account.

Association between ecologic and genetic distances among Aegilops species

Mantel tests assessed to what extent genetic variation was significantly associated with either geographical or ecological differentiation. In overall, diploids showed genetic variations associated with both geographic and ecological differentiation. Allopolyploids presented comparatively broader ecological niches but overall no significant evidence of association between genetic and ecological variations within species. However, all diploid species (except *Ae. tauschii*) and the allopolyploid *Ae. cylindrica* (DDCC) showed significant patterns of genetic isolation by geographical distance across their respective ranges, indicating that genetic structure within these species had a limited spatial component (Table 1 and chapter 2). Considering spatial autocorrelation through partial Mantel tests, all diploids (except *Ae. comosa*) and the allopolyploid *Ae. triuncialis* (UUCC) showed a strong pattern of genetic isolation by environment, which may indicate genetic differentiation going along with ecological differentiation.

Allopolyploid niche dynamics

Niche overlap was generally higher when comparing allopolyploid niches to their diploid progenitors' niches than when comparing diploid niches among them (Figure 2), reaching up to 0.71 between the allopolyploid *Ae. triuncialis* and its progenitor *Ae. caudata*. Allopolyploid niche overlap appeared particularly asymmetric between diploid progenitors for *Ae. cylindrica* (DDCC; 0.68 with *Ae. tauschii* (DD) *vs.* 0.25 with *Ae. caudata* (CC)) and for *Ae. crassa* (DDMM; 0.54 with *Ae. tauschii* (DD) *vs.* 0.05 with *Ae. comosa* (MM)). However, the similarity test revealed allopolyploid niches of *Ae. triuncialis* and *Ae. cylindrica* being significantly similar to the ecological niches of both their respective diploid progenitors (Supplemental Material Table S2).

Comparing the environmental space occupied by the addition of diploid progenitors to that of their derived allopolyploids (Figure 3), *Ae. geniculata*, *Ae. triuncialis* and *Ae. cylindrica* presented relatively additive ecological niches (0.88 – 0.99), occupying most of the environmental gradients also occupied by their progenitors. Only a small fraction of their expected niches was indeed lost (0.01 – 0.12), along with a slight niche expansion toward temperate climates in these three allopolyploids (0.03 – 0.12). The expansion of the allopolyploid species niche appeared specifically located at the northern margins of the species ranges for all three allopolyploids (Supplemental Material Figure S2), but also at the southern margins (i.e. North Africa) and in the Levant for *Ae. geniculata*.

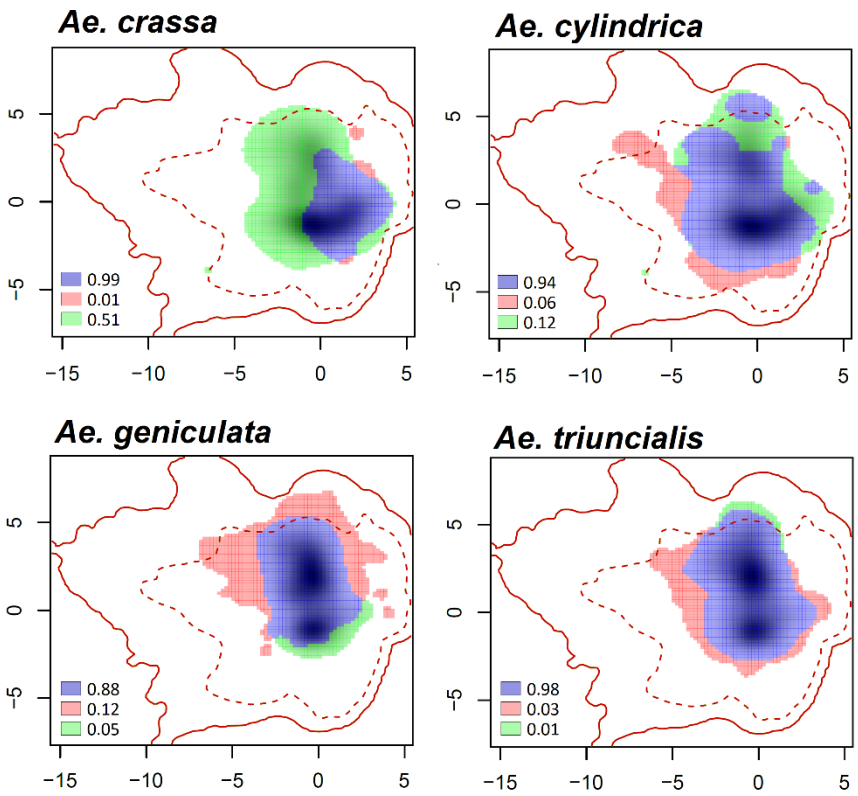


Figure 3: Dynamics of the ecological niche in *Aegilops* polyploid species of the MCDU complex. Comparison of the observed niche of a polyploid to the expected sum of niches of its diploid progenitors. Niche stability (in blue) is characterized by the additive fraction of this expected niche that is effectively occupied by the polyploid. Niche retraction (in green) corresponds to the proportion of the expected niche that is unoccupied by the polyploid, whereas niche expansion (in red) corresponds to the fraction of the realized polyploid niche that is new compared to the sum of its diploid progenitors. Estimated proportions are indicated next to corresponding colors.

On the opposite, the narrow distribution range of the allopolyploid *Ae. crassa* was associated with a considerable shrinkage of its occupied ecological niche, with 51 % of the expected niche based on its diploid progenitors being lost (Figure 3). Regarding the uncertainties underlying the donor of the premised 'M' genome of *Ae. crassa* (Badaeva *et al.*, 1998; Chapter 1 and 2), which may be further supported by the extremely low niche overlap (0.5) with *Ae. comosa* (M), the ecological niche patterns observed for *Ae. crassa* should be carefully interpreted and will thus not be further discussed.

Discussion

Ecological niche additivity of the allopolyploid species

The allopolyploid *Aegilops* species globally share similar ecological requirements with their respective diploid progenitors as shown by their high proportion of niche stability (Figure 3) and the overlap of their niches (Figure 2). In particular, the ecological niche of the allopolyploid *Ae. triuncialis* greatly overlap both its progenitors' niches with only 1 % of niche retraction (Figure 3), matching the expected ecological additivity of progenitors. On the contrary, the observed additivity is nuanced for the allopolyploids *Ae. geniculata* and *Ae. cylindrica*, whose ecological niches asymmetrically overlap that of their respective progenitors. However, the observed asymmetrical overlap appear significant only for *Ae. geniculata* (Supplemental Table S2). The ecological niche breadth of *Ae. cylindrica* as well as of its progenitors being relatively large, the overlap of their niches may remained equally high despite randomization of the allopolyploid's niche centroid.

According to the allopolyploid species trees inferred in Chapter 2, *Ae. cylindrica* and *Ae. triuncialis* appeared as the youngest species with an estimated age of about 0.77 – 0.65 million years ago (Mya). Their relatively young age might not have given enough time to shift or enlarge their ecological tolerances from their diploid progenitors, whereas the older allopolyploid *Ae. geniculata* (whose origin was dated around 1.38 Mya) show a shift from the ecological additivity of its progenitors with the highest ecological niche expansion (about 12%). The slight niche expansion observed for most allopolyploids suggests that the species have first occupied similar ecological niches as their progenitors before progressively enlarging their ecological tolerance toward more temperate climates.

Regarding the location of expanding niches (Supplemental Material Figure S2), *Ae. geniculata* appear to expand its ecological niche in the Levant, unlike the co-occurring allopolyploid *Ae. triuncialis*. The origin of the *Aegilops* species is commonly thought to be in Transcaucasia, with the center of genetic diversity along the Crescent Fertile (van Slageren, 1994). These two allopolyploid species, which share a common U genome, have thus likely evolved in sympatry in the Levant from the time of their origin, and hybrid zones currently exist there (Senerchia *et al.*, 2016). *Ae. geniculata* thus seemingly avoided gene flow and competition for available ecological niche with *Ae. triuncialis* in the Levant by partly shifting its ecological preferences. Both allopolyploids have their ecological niche greatly overlapping (Figure 2), and the fact that only *Ae. geniculata* expand into a new niche in the Levant suggest

that the allopolyploid may have had enough time to shift its ecological niche. Senerchia *et al.* (2016) suggested that differential introgression of retrotransposons in the hybrid zone between the two allopolyploid species in Levant may be the key incompatibilities shaping species boundaries. Both allopolyploid species also co-occur in the Iberian Peninsula, where *Ae. geniculata* also expand into new niches but only in restricted regions of the peninsula (Supplemental Material Figure S2). Whether intrinsic mechanisms similar to that observed by Senerchia *et al.* (2016) in the Levant may also explain that these species maintain their integrity without shifting their ecological niches across most of the Iberian Peninsula need to be further investigated.

Migration as the main driver of allopolyploids range expansion

According to Chapter 2, the *Aegilops* species range evolved through repeated cycles of retraction into glacial refugia followed by post-glacial recolonization of deglaciated habitats. Regarding the current species distribution, allopolyploids must have demonstrated higher re-colonization abilities to achieve such broad distribution compared to their diploid progenitors. Polyploidy is indeed often associated with greater colonization abilities. The allopolyploids may differentially express their numerous gene copies, conferring them a wider plastic response to adapt to new or rapidly changing environments (Dong & Adams, 2011; te Beest *et al.*, 2012). After the Last Glacial Maximum (LGM) about 15 000 years ago (Clark *et al.*, 2009), the allopolyploid *Aegilops* species have likely colonized the available suitable habitats more effectively than their diploid relatives, as also suggested by Brochmann *et al.* (2004), without necessarily shifting their ecological preferences. The wider ecological niche breadth and geographical distribution exhibited by allopolyploids compared to their diploid progenitors, and the mostly additive ecological niche of the allopolyploids, seems to support this scenario. Diploid species such as *Ae. caudata* and *Ae. comosa* thus likely face competition for available niche with the allopolyploid species (now the major cytotype), preventing them to successfully extend their distribution outside the Balkans (their putative glacial refugia during the LGM, see Chapter 2).

Nevertheless, *Ae. triuncialis* is the only allopolyploid species showing a pattern of isolation-by-ecology (IBE), when taking into account genetic isolation-by-distance. The genetic structure of *Ae. triuncialis* inferred in Chapter 2 showed complex with at least three genetic lineages having probably separately evolved into different refugia during the LGM and that further recolonized from either the Balkans, the Iberian peninsula, Anatolia or Central Asia. The maternal lineages of *Ae. triuncialis*, which maternally originated from both its

progenitors, both co-occur across the species range (see chapter 2), did not show a pattern of IBE (data not shown) and thus unlikely evolved into different ecological niches. However, whether the nuclear genetic lineages have evolved into distinct sets of ecological requirements from the combined ecological requirements of their diploid progenitors constitute a plausible explanation that should be further tested.

Role of pivotal genomes in the evolution of allopolyploid ecological niches

Interestingly, most diploid species reveal significant patterns of IBE (Table 1) whereas only the allopolyploid *Ae. triuncialis* similarly show a strong IBE pattern. However, the pattern of IBE should be carefully interpreted as it does not necessarily result from a local speciation, but it nevertheless suggest that the conditions for ecological speciation may exist (Wang & Bradburd, 2014). In particular, the diploid species *Ae. caudata* and *Ae. umbellulata* both show a strong pattern of IBD and IBE, suggesting that ecological conditions may vary with the geographical distance in the Balkans and West of Anatolia, both of which shape the distribution of intra-specific genetic variation in these diploids. The two other diploids, *Ae. tauschii* and *Ae. umbellulata*, show distinct geographical distribution but significantly similar ecological niches with a niche overlap of 0.69 (data not shown). *Ae. tauschii* has long been reported to be especially tolerant to high drought conditions (van Slageren, 1994; Kilian *et al.*, 2011). Although similar tolerance to intense drought has not been reported for *Ae. umbellulata*, further investigation on the ecological tolerance of this diploid species would be relevant.

The diploids *Ae. tauschii* (DD) and *Ae. umbellulata* (UU) are known to be at the origin of most allopolyploid *Aegilops* species by recurrently combining with other diploid species (Zohary & Feldman, 1962; Badaeva *et al.*, 2002, 2004; Meimberg *et al.*, 2009). The D and U genomes showed mainly unaltered (hence called 'pivotal genome') whereas the other recombined genomes usually show structural rearrangements (hence called 'differential genome') (Zohary & Feldman, 1962; Feldman *et al.*, 2012). The pivotal genome hypothesis from Zohary and Feldman (1962) indeed stated that allopolyploids sharing common pivotal genome recurrently hybridize between them. The differential genomes should thus mainly evolving through homeologous recombinations leading to subsequent genomic restructuring under this hypothesis, which was supported to some extent by the lower genome restructuring observed in the pivotal genomes compared to the differential genome (see chapter 3). As the pivotal genome remains relatively intact, ecological adaptations that may be linked to the unaltered pivotal genome are expected to be similar between the allopolyploid and its pivotal

progenitor, thereby contributing to the observed asymmetrical niche overlap for *Ae. cylindrica*. However, the niche of *Ae. triuncialis* similarly overlap both its progenitors'niches whereas the allopolyploid show higher restructuring of its differential genome, and the niche of *Ae. geniculata* largely overlap that of its differential progenitor *Ae. comosa* whereas the allopolyploid show no asymmetrical genome restructuring. This suggests that other mechanisms or processes may also drive the evolution of allopolyploid's ecological niches, which remain to be identified.

References

Adams KL, Cronn R, Percifield R, Wendel JF. **2003**. Genes duplicated by polyploidy show unequal contributions to the transcriptome and organ-specific reciprocal silencing. *Proceedings of the National Academy of Sciences* **100**: 4649–4654.

Anacker BL, Strauss SY. **2014**. The geography and ecology of plant speciation: range overlap and niche divergence in sister species. *Proceedings of the Royal Society B: Biological Sciences* **281**: 20132980–20132980.

Badaeva ED, Amosova AV, Muravenko OV, Samatadze TE, Chikida NN, Zelenin AV, Friebe B, Gill BS. **2002**. Genome differentiation in *Aegilops*. 3. Evolution of the D-genome cluster. *Plant Systematics and Evolution* **231**: 163–190.

Badaeva ED, Amosova AV, Samatadze TE, Zoshchuk SA, Shostak NG, Chikida NN, Zelenin AV, Raupp WJ, Friebe B, Gill BS. **2004**. Genome differentiation in *Aegilops*. 4. Evolution of the U-genome cluster. *Plant Systematics and Evolution* **246**: 45–76.

Badaeva ED, Friebe B, Zoshchuk SA, Zelenin AV, Gill BS. **1998**. Molecular cytogenetic analysis of tetraploid and hexaploid *Aegilops crassa*. *Chromosome Research 1998* **6**: 629–637.

te Beest M, Le Roux JJ, Richardson DM, Brysting AK, Suda J, Kubesova M, Pysek P. **2012**. The more the better? The role of polyploidy in facilitating plant invasions. *Annals of Botany* **109**: 19–45.

Bento M, Gustafson P, Viegas W, Silva M. **2010**. Genome merger: from sequence rearrangements in triticale to their elimination in wheat–rye addition lines. *Theoretical and Applied Genetics* **121**: 489–497.

Brochmann C, Brysting AK, Alsos IG, Borgen L, Grundt HH, Scheen A-C, Elven R. **2004**. Polyploidy in arctic plants. *Biological Journal of the Linnean Society* **82**: 521–536.

Broennimann O, Fitzpatrick MC, Pearman PB, Petitpierre B, Pellissier L, Yoccoz NG, Thuiller W, Fortin M-J, Randin C, Zimmermann NE, *et al.* 2012. Measuring ecological niche overlap from occurrence and spatial environmental data: Measuring niche overlap. *Global Ecology and Biogeography* **21**: 481–497.

Casazza G, Boucher FC, Minuto L, Randin CF, Conti E. **2017**. Do floral and niche shifts favour the establishment and persistence of newly arisen polyploids? A case study in an Alpine primrose. *Annals of Botany* **119**: 81–93.

Clark PU, Dyke AS, Shakun JD, Carlson AE, Clark J, Wohlfarth B, Mitrovica JX, Hostetler SW, McCabe AM. **2009**. The Last Glacial Maximum. *Science* **325**: 710–714.

Comai L. **2005**. The advantages and disadvantages of being polyploid. *Nature Reviews Genetics* **6**: 836–846.

Di Cola V, Broennimann O, Petitpierre B, Breiner FT, D'Amen M, Randin C, Engler R, Pottier J, Pio D, Dubuis A, *et al.* 2017. ecospat: an R package to support spatial analyses and modeling of species niches and distributions. *Ecography* **40**: 774–787.

Dong S, Adams KL. **2011**. Differential contributions to the transcriptome of duplicated genes in response to abiotic stresses in natural and synthetic polyploids. *New Phytologist* **190**: 1045–1057.

Eckardt NA. **2010**. A Double Lock on Polyploidy-Associated Epigenetic Gene Silencing. *The Plant Cell* **22**: 3–3.

Feldman M. 2009. Genome evolution in allopolyploid wheat – a revolutionary reprogramming followed by gradual changes. Journal of Genetics and Genomics **6**:511-518.

Feldman M, Levy AA, Fahima T, Korol A. **2012**. Genomic asymmetry in allopolyploid plants: wheat as a model. *Journal of Experimental Botany* **63**: 5045–5059.

Glennon KL, Ritchie ME, Segraves KA. **2014**. Evidence for shared broad-scale climatic niches of diploid and polyploid plants (R Bardgett, Ed.). *Ecology Letters* **17**: 574–582.

Hijmans RJ. **2017**. geosphere: Spherical trigonometry. R package version 1. 5-5. https://CRAN.R-project.org/package=geosphere.

Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A. **2005**. Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* **25**: 1965–1978.

Karger DN, Conrad O, Böhner J, Kawohl T, Kreft H, Soria-Auza RW, Zimmermann NE, Linder HP, Kessler M. **2017**. Climatologies at high resolution for the earth's land surface areas. *Scientific Data* **4**: 170122.

Kilian B, Mammen K, Millet E, Sharma R, Graner A, Salamini F, Hammer K, Özkan H. **2011**. *Aegilops*. In: Kole C, ed. Wild Crop Relatives: Genomic and Breeding Resources. Berlin, Heidelberg: Springer Berlin Heidelberg, 1–76.

Leitch IJ, Bennett MD. **2004**. Genome downsizing in polyploid plants. *Biological Journal of the Linnean Society* **82**: 651–663.

Levin DA. 1975. Minority cytotype exclusion in local populations. Taxon **24**:35-43.

Levin DA. **2002**. *The role of chromosomal change in plant evolution*. Oxford ; New York: Oxford University Press.

Li W-L, Berlyn GP, Ashton PMS. **1996**. Polyploids and their Structural and Physiological Characteristics Relative to Water Deficit in *Betula papyrifera* (Betulaceae). *American Journal of Botany* **83**: 15–20.

Linder HP and Barker NP. 2014. Does polyploidy facilitate long-distance dispersal?. Annals of Botany **113**:1175-1183.

Lyons KG, Shapiro AM, Schwartz MW. **2010**. Distribution and ecotypic variation of the invasive annual barb goatgrass (*Aegilops triuncialis*) on serpentine soil. *Invasive Plant Science and Management* **3**: 376–389.

Madlung A, Wendel JF. **2013**. Genetic and Epigenetic Aspects of Polyploid Evolution in Plants. *Cytogenetic and Genome Research* **140**: 270–285.

Marchant DB, Soltis DE, Soltis PS. **2016**. Patterns of abiotic niche shifts in allopolyploids relative to their progenitors. *New Phytologist* **212**: 708–718.

Meimberg H, Rice KJ, Milan NF, Njoku CC, McKay JK. **2009**. Multiple origins promote the ecological amplitude of allopolyploid *Aegilops* (Poaceae). *American Journal of Botany* **96**: 1262–1273.

Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, Minchin PR, O'Hara RB, Simpson GL, Solymos P, *et al.* 2018. vegan: Community Ecology Package. R package version 2.4-4. https://CRAN.R-project.org/package=vegan.

Olson DM, Dinerstein E, Wikramanayake ED, Burgess ND, Powell GVN, Underwood EC, D'amico JA, Itoua I, Strand HE, Morrison JC, *et al.* 2001. Terrestrial Ecoregions of the World: A New Map of Life on Earth. *BioScience* **51**: 933.

Ramsey J, Ramsey TS. **2014**. Ecological studies of polyploidy in the 100 years following its discovery. *Philosophical Transactions of the Royal Society B: Biological Sciences* **369**: 20130352–20130352.

Ramsey J, Schemske DW. **1998**. Pathways, mechanisms, and rates of polyploid formation in flowering plants. *Annual Review of Ecology and Systematics* **29**: 467–501.

Ramsey J, Schemske DW. **2002**. Neopolyploidy in Flowering Plants. *Annual Review of Ecology and Systematics* **33**: 589–639.

Rousset. **2000**. Genetic differentiation between individuals. *Journal of Evolutionary Biology* **13**: 58–62.

Schoener TW. **1968**. The *Anolis* lizards of Bimini: Resource partitioning in a complex fauna. *Ecology* **49**: 704–726.

Senerchia N, Felber F, North B, Sarr A, Guadagnuolo R, Parisod C. **2016**. Differential introgression and reorganization of retrotransposons in hybrid zones between wild wheats. *Molecular Ecology* **25**: 2518–2528.

van Slageren MWSJM. **1994**. *Wild wheats: a monograph of Aegilops L. and Amblyopyrum (Jaub. & Spach) Eig (Poaceae)*. Wageningen, The Netherlands : Aleppo, Syria: Wageningen Agricultural University : International Center for Agricultural Research in the Dry Areas.

Stebbins GL. **1950**. *Variation and evolution in plants*. New York, NY.

Stebbins GL. **1971**. Chromosomal evolution in higher plants. *Chromosomal evolution in higher plants.*

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

Trabucco A, Zomer RJ. **2010**. Global soil water balance geospatial database. *CGIAR Consortium for Spatial Information. Published online, available from the CGIAR-CSI GeoPortal at: http://www.cgiar-csi.org.*

Wang IJ, Bradburd GS. **2014**. Isolation by environment. *Molecular Ecology* **23**: 5649–5662.

Wang IJ, Glor RE, Losos JB. **2013**. Quantifying the roles of ecology and geography in spatial genetic divergence. *Ecology Letters* **16**: 175–182.

Wang J, Madlung A, Lee H-S, Chen M, Lee JJ, Watson B, Kagochi T, Comai L, Chen ZJ. **2004**. Stochastic and epigenetic changes of gene expression in *Arabidopsis* polyploids. *Genetics* **167**: 1961–1973.

Warren DL, Glor RE, Turelli M. **2008**. Environmental niche equivalency versus conservatism: quantitative approaches to niche evolution. *Evolution* **62**: 2868–2883.

Wood TE, Takebayashi N, Barker MS, Mayrose I, Greenspoon PB, Rieseberg LH. **2009**. The frequency of polyploid speciation in vascular plants. *Proceedings of the National Academy of Sciences* **106**: 13875–13879.

Zohary D, Feldman M. **1962**. Hybridization between amphidiploids and the evolution of polyploids in the wheat (*Aegilops-Triticum*) group. *Evolution* **16**: 44-61.

Supplementary Material

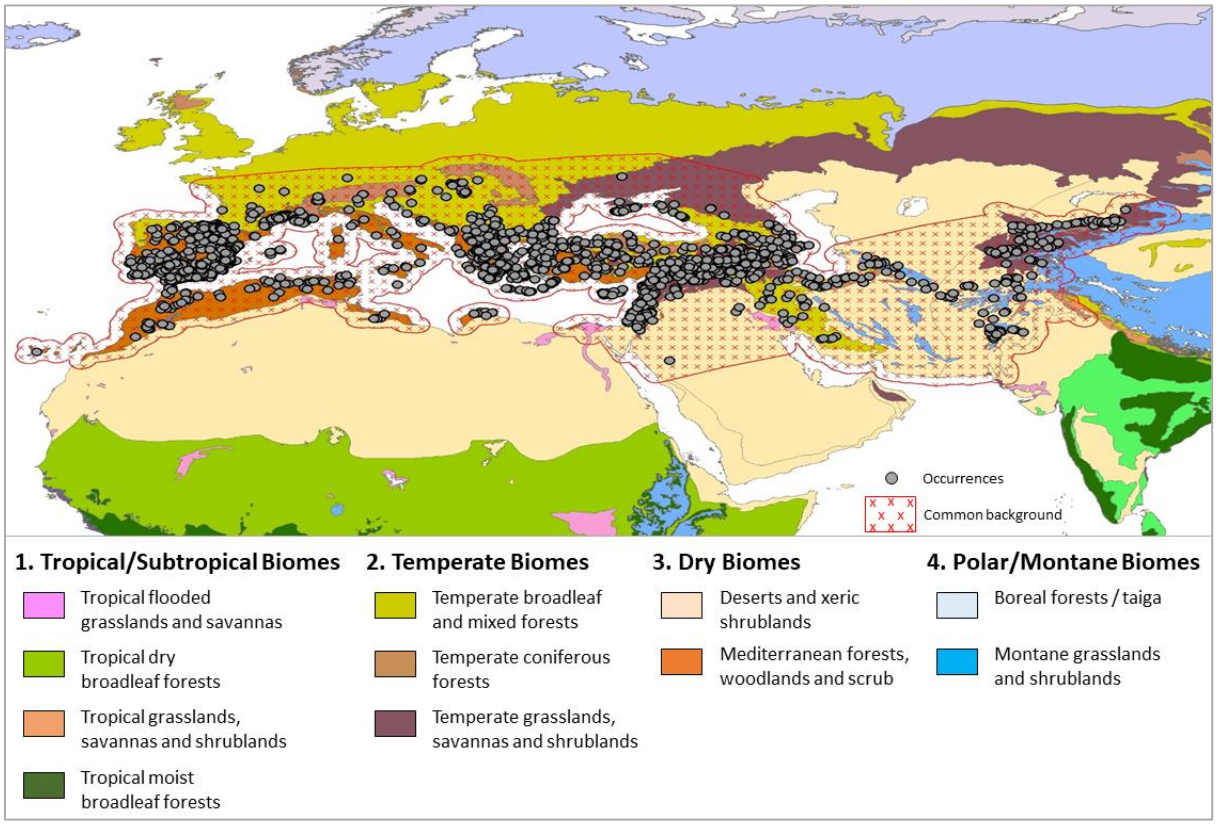


Figure S1: Delimitation of the common study area for ecological niche modeling of the *Aegilops* species. The study area used for modeling the ecological niches of *Aegilops* species is delimited by the red area with crosses. Occurrences of the eight studied species are represented by grey dots. The biomes as defined by the WWF (Olson *et al.*, 2001) are highlighted with different colors and listed in the above legend.

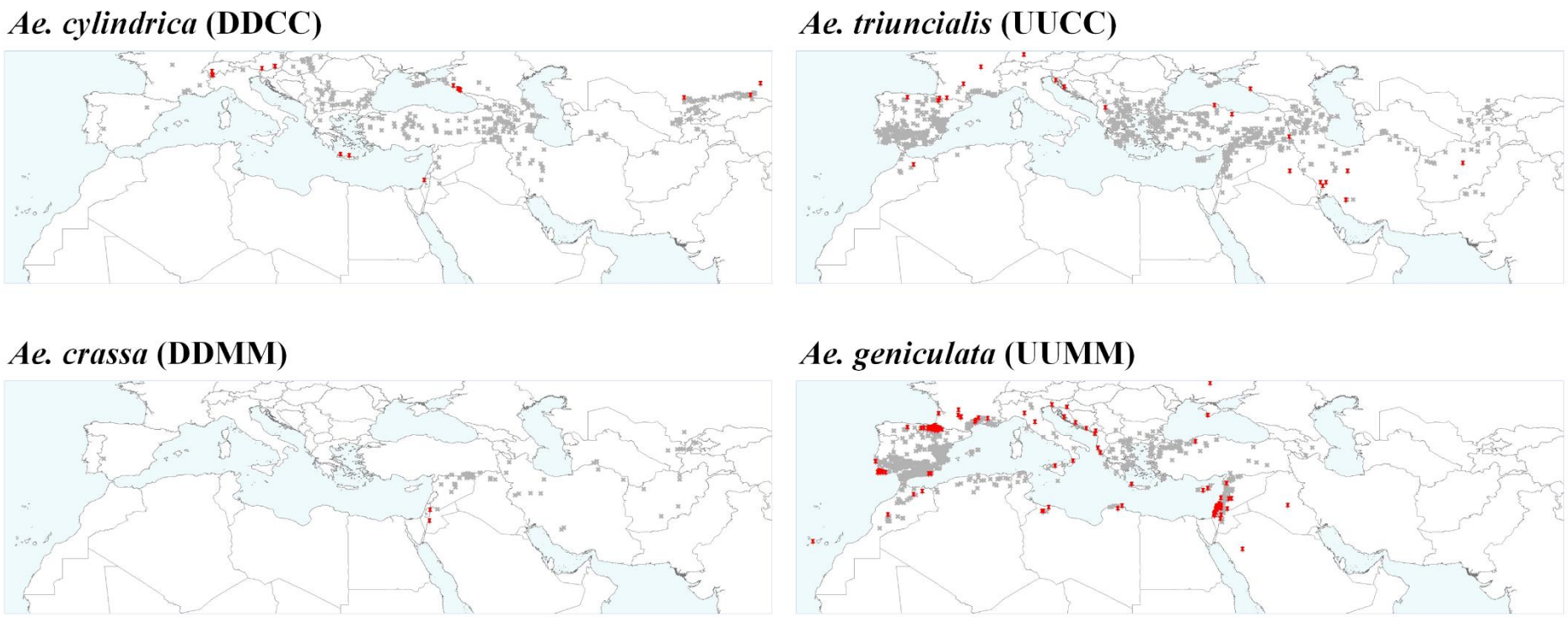


Figure S2: Distribution of accessions into niche expansion among the allopolyploid *Aegilops* species. Allopolyploid accessions inferred as in niche stability (i.e. sharing similar ecological requirements as one or both their diploid progenitors) or niche expansion (i.e. having distinct ecological requirements from both their diploid progenitors) are represented by grey crosses and red thumbtacks respectively. The allopolypoid species are indicated above their corresponding distribution map and their genomic formula reported into brackets following van Slageren (1994).

Table S1: List of *Aegilops* occurrences (Electronic material). The supplemental material can be found under the following Excel file: "StellaHuynh_PhD_Chapter4_SuppMat_TableS21-ListOfGBIFAccessions.xlsx". Each row corresponds to a specific accession and each column respectively indicates the species name, the accessions name, the latitude and longitude in decimal values defined in the WGS84 coordinates system, the country where it has been collected, the altitude extrapolated from the GPS coordinates and the gene bank institute (or personal collection) where the accessions were deposited.

Table S2: Pairwise niche overlap among *Aegilops* species. Niche overlap for pairwise comparisons was estimated using metric D from Schoener (1968) and tested through similarity tests following Warren *et al.* (2008) with 100 permutations and their associated p-value reported in the table. [a] Similarity tests were performed by iteratively reallocating niche centroid of both species for pairwise comparisons among diploids and among polyploids. [b] For pairwise comparisons between polyploid and each of its diploid progenitor species, only the centroid of polyploid niche was iteratively shifted whereas the diploid niche centroid was kept fixed.

Pairwise comparisons			Niche overlap	Similarity test
Ae. tauschii (DD)	x	*Ae. caudata* (CC)	0.29	0.19 [a]
Ae. tauschii (DD)	x	*Ae. comosa* (MM)	0.14	0.09 [a]
Ae. umbellulata (UU)	x	*Ae. caudata* (CC)	0.49	0.07 [a]
Ae. umbellulata (UU)	x	*Ae. comosa* (MM)	0.34	0.04 [a]
Ae. crassa (DDMM)	x	*Ae. tauschii* (DD)	0.54	0.05 [b]
Ae. crassa (DDMM)	x	*Ae. comosa* (MM)	0.05	0.42 [b]
Ae. cylindrica (DDCC)	x	*Ae. tauschii* (DD)	0.68	0.03 [b]
Ae. cylindrica (DDCC)	x	*Ae. caudata* (CC)	0.25	0.03 [b]
Ae. geniculata (UUMM)	x	*Ae. umbellulata* (UU)	0.41	0.20 [b]
Ae. geniculata (UUMM)	x	*Ae. comosa* (MM)	0.65	0.05 [b]
Ae. triuncialis (UUCC)	x	*Ae. umbellulata* (UU)	0.64	0.05 [b]
Ae. triuncialis (UUCC)	x	*Ae. caudata* (CC)	0.71	0.02 [b]
Ae. crassa (DDMM)	x	*Ae. cylindrica* (DDCC)	0.26	0.11 [a]
Ae. crassa (DDMM)	x	*Ae. geniculata* (UUMM)	0.10	0.30 [a]
Ae. triuncialis (UUCC)	x	*Ae. cylindrica* (DDCC)	0.40	0.11 [a]
Ae. triuncialis (UUCC)	x	*Ae. geniculata* (UUMM)	0.72	0.10 [a]

General conclusions and Perspectives

Conclusions

Evolutionary advantages of allopolyploidy in species range expansion

Most ancient whole-genome duplication (WGD) events appears clustered around the K/Pg boundary, a mass extinction event occurring about 65 Mya (Soltis & Burleigh, 2009; Van de Peer *et al.*, 2017). Correlation between polyploidization events and periods of drastic environmental changes is here further confirmed by the origin of the allopolyploid *Aegilops* species likely around climatic transitions, here interglacial-glacial periods of the late Quaternary. The evolutionary history of diploids and allopolyploids *Aegilops* have thus been punctuated by cyclic species range expansion and contraction, which affected their current distribution and genomic structure.

The hypothesis of allopolyploids being the genetic and ecological additivity of their diploid progenitors was mainly confirmed from the chapters 2, 3 and 4. Allopolyploids not only cover both the ecological niches of their progenitors but also both their genetic background. The larger distribution of the allopolyploids compared to that of the diploids thus mainly results from higher dispersal abilities of the allopolyploids enabling them to quickly expand their distribution after the last glacial maximum (LGM). The allopolyploids can indeed find suitable niches easily, as they have a broader range of ecological tolerances covering their progenitors' tolerances and that further contributed to their quick expansion. In addition, their high amount of gene copies combined with fixed heterozygosity may confer high genome plasticity contributing to the potential of allopolyploids in colonizing new ecological niches. Such large ecological breadth and genome plasticity of the allopolyploids may thus facilitate their invasions, as observed for *Spartina anglica* (Ainouche *et al.*, 2009) or *Ae. cylindrica* and *Ae. triuncialis* in North America (van Slageren, 1994; Lyons *et al.*, 2010), and thereby supporting the invasive potential of allopolyploids as suggested by te Beest *et al.* (2012).

Though allopolyploids are mainly the additivity of their progenitors, genetic novelties tend to accumulate in older allopolyploid such as *Ae. geniculata* (42% of derived haplotypes; chapter 2) as a result of species evolution. The chapter 4 also showed the ecological niches of allopolyploids seemingly expanding toward more temperate climate, with *Ae. geniculata* having the largest expansion toward this new niche (12% of its niche). Expansion of allopolyploids into new ecological niches can possibly be promoted by genetic novelties. All allopolyploids showed a high proportion of genetic novelties except *Ae. cylindrica* (chapter 2), as well as substantial genome restructuring associated with TEs (chapter 3), which is

thereby consistent with this hypothesis. However, the young species *Ae. cylindrica* also expanded into new ecological niches (6% of its niche; chapter 4) despite having less genetic novelties than the other allopolyploids (3.6% of derived haplotypes; chapter 2), thereby contradicting this hypothesis especially when comparing to the relative proportions of genetic and ecological novelties exhibited by *Ae. geniculata*. To what extent the observed genetic changes affecting low-copy genes or TEs may contribute to the evolution of new ecological niches need to be further tested.

Long-term evolution of retrotransposons in allopolyploid populations

The present thesis shed light on the long-term evolution of retrotransposons in allopolyploids (chapter 3), which showed mainly driven by random processes across species ranges. The higher sampling across species range appeared necessary to investigate TE copies segregating within the species and hence accurately estimate the number of TE copies expected to segregate in the allopolyploids under the hypothesis of progenitor additivity. Recent TE activity was highlighted along the Balkan-Anatolian suture zone identified in chapter 2 and at species ranges. This suture zone highlights the secondary contacts occurring between previously isolated genetic lineages within allopolyploid species. Genetic divergence between the allopolyploid lineages might thus be high enough to generate genomic conflicts re-activating TE copies. However, to what extent specific LTR-RT families are re-activated while other remain quiescent is still obscure. The relative importance of intrinsic and extrinsic processes acting on the long-term evolution TE copies, especially at the species range margins as observed in the allopolyploids *Ae. triuncialis* and *Ae. cylindrica* (chapter 3), remains to be further investigated.

The "pivotal-differential genome" hypothesis seem relatively consistent with the asymmetrical TE-restructuring pattern observed among allopolyploid species, but functional asymmetry as well as the underlying molecular processes driving these asymmetrical patterns are not yet resolved.

Methodological considerations of genome evolution of allopolyploid species

High-throughput amplicon sequencing proved as a powerful method to produce large amount of data in a short time (chapter 1). One Illumina MiSeq run provided enough coverage for accurate amplification and identification of homeologous sequences in 448 allopolyploid samples and 48 markers (chapter 2). Nevertheless, the quantification of actual dosage of the homeologs remained challenging due PCR-biased errors sequencing technique. Allelic variants occurring in less than 10 % of the reads amplified at a specific marker in a specific

sample were considered as sequencing errors and removed during the cleaning of reads. Sequencing errors can potentially occur at the early steps of PCR cycles and be further amplified, thereby reaching a substantial frequency of reads containing this error that might wrongly be considered as true variants. The presence of paralogs is one of the most challenging issues when producing data aiming at inferring phylogenies. Paralogs were indeed present in our dataset but could be detected with reasonable confidence due to the high sequencing coverage, which further highlighted rare variants giving doubting gene tree topologies (supplemental material S2, chapter 2). High-throughput sequencing methods requiring only raw DNA material (i.e. PCR-free) such as PacBio, which also has the advantage to produce longer reads enabling better phylogenetic resolution, might represent an valuable alternative for sequencing of polyploid genomes (Rothfels *et al.*, 2017).

Phylogenetic inferences of allopolyploid species also revealed challenging due to the limited computational tools specifically handling polyploid data. The highly reticulate evolution and incomplete lineage sorting characterizing the *Aegilops* clade induced conflicting gene topologies. Computational methods that can distinguish the effect of reticulation from incomplete lineage sorting need to be further developed. Phylogenetic tools enabling to input homeolog sequences for a single allopolyploid individual is also still needed. Despite of 30 nuclear loci being accurately genotyped, our dataset seem lacking information to clearly detect the reticulation signal associated to the homoploid hybrid of the clade D. Additional analyses including the nuclear dataset used by Nakhleh *et al.* (unpublished data), which was originally taken from Marcussen *et al.* (2014), could however successfully recover the homoploid hybrid of the clade D. Thorough phylogenetic analyses are still in process, which will potentially manage resolve the complex phylogenetic relationships of *Ae. speltoides* and *Ae. mutica*.

Phylogenetic relationships among the wheats

The first chapter also confirmed conflicting topologies involving the phylogenetic position of *Ae. speltoides* (SS), the representative of the B clade, and *Ae. mutica* (TT). In particular, the taxonomy of the latter species is still debated and the species is sometimes referred to as *Amblyopyrum muticum*. While the *Aegilops* mainly reproduce through self-fertilizing, these two species are the only that are (partially) outcrossers (Sakamoto, 1982). This difference in reproductive regime have likely affected the estimation of population sizes while inferring the species tree through Bayesian method, which might have highly impacted the branch lengths and topologies of *Ae. speltoides* and *Ae. mutica*. Estimating the population sizes for each

ancestors (i.e. tree nodes) can potentially improve the phylogenetic inferences and needs to be further tested.

The phylogenetic analyses in Chapter 2 confirmed former uncertainties formulated by Badaeva *et al.* (1998) about the premised progenitors of the allopolyploid *Ae. crassa* (DDMM), the pivotal genome donor *Ae. tauschii* (DD) and the differential genome donor *Ae. comosa* (MM). The allopolyploid is thought to first originate from diploid progenitors to form the tetraploid individuals, and later on a second allopolyploidization occurred that combined one tetraploid individual and its pivotal progenitor *Ae. tauschii*.

References

Ainouche ML, Fortune PM, Salmon A, Parisod C, Grandbastien M-A, Fukunaga K, Ricou M, Misset M-T. **2009**. Hybridization, polyploidy and invasion: lessons from *Spartina* (Poaceae). *Biological Invasions* **11**: 1159–1173.

Badaeva ED, Friebe B, Zoshchuk SA, Zelenin AV, Gill BS. **1998**. Molecular cytogenetic analysis of tetraploid and hexaploid *Aegilops crassa*. *Chromosome Research 1998* **6**: 629–637.

te Beest M, Le Roux JJ, Richardson DM, Brysting AK, Suda J, Kubesova M, Pysek P. **2012**. The more the better? The role of polyploidy in facilitating plant invasions. *Annals of Botany* **109**: 19–45.

Marcussen T, Sandve SR, Heier L, Spannagl M, Pfeifer M, The International Wheat Genome Sequencing Consortium, Jakobsen KS, Wulff BBH, Steuernagel B, Mayer KFX, *et al.* 2014. Ancient hybridizations among the ancestral genomes of bread wheat. *Science* **345**: 1250092–1250092.

Rothfels CJ, Pryer KM, Li F-W. **2017**. Next-generation polyploid phylogenetics: rapid resolution of hybrid polyploid complexes using PacBio single-molecule sequencing. *New Phytologist* **213**: 413–429.

Sakamoto S. **1982**. The Middle East as a cradle for crops and weeds. In: Geobotany. Biology and ecology of weeds. Springer, Dordrecht, 97–109.

Soltis DE, Burleigh JG. **2009**. Surviving the K-T mass extinction: New perspectives of polyploidization in angiosperms. *Proceedings of the National Academy of Sciences of the United States of America* **106**: 5455–5456.

Van de Peer Y, Mizrachi E, Marchal K. **2017**. The evolutionary significance of polyploidy. *Nature Reviews Genetics* **18**: 411–424.

General Supplemental Material

Table S1: List of *Aegilops* and outgroups accessions. Each row corresponds to one individual and each column respectively indicates the species name, the accessions identifier, the latitude and longitude in decimal values defined in the WGS84 coordinate system, the country where it has been found, the altitude extrapolated from the GPS coordinates (in meters), the gene bank institute (or personnal collection) where the accessions were obtained and the relative contribution of each accession to the chapters in the present PhD thesis. Individuals from personal collection are field-collected samples from Arrigo *et al.* (2010), Senerchia *et al.* (2014) or Senerchia *et al.* (2015).

Species	Accessions	X	Y	Countries	Altitude	GeneBank	Chapters
Ae. caudata	PI 551132	22.9333333	37.5833333	Greece	176	USDA ARS-GRIN	1,2,3,4
Ae. caudata	PI 551143	22.2	36.9666667	Greece	370	USDA ARS-GRIN	2,3,4
Ae. caudata	PI 551145	22.0166667	37.1833333	Greece	36	USDA ARS-GRIN	2,3,4
Ae. caudata	PI 551126	23.95	38	Greece	-50.2	USDA ARS-GRIN	2,3,4
Ae. caudata	PI 551121	21.6166667	39.6833333	Greece	450	USDA ARS-GRIN	2,3,4
Ae. caudata	PI 203431	32.85	39.6166667	Turkey	1101	USDA ARS-GRIN	1,2,3,4
Ae. caudata	PI 551137	22.7333333	37.9	Greece	180	USDA ARS-GRIN	2,3,4
Ae. caudata	PI 551129	23.3833333	38.4333333	Greece	193	USDA ARS-GRIN	2,3,4
Ae. caudata	PI 551148	22.4666667	37.0166667	Greece	190	USDA ARS-GRIN	2,3,4
Ae. caudata	IG 48080	35.897181	33.95	Lebanon	0	ICARDA	2,3,4
Ae. caudata	PI 554194	29.05	37.85	Turkey	180	USDA ARS-GRIN	2,3,4
Ae. caudata	PI 298888	27.9666667	40.0333333	Turkey	44	USDA ARS-GRIN	2,3,4
Ae. caudata	PI 551134	23	37.8666667	Greece	20	USDA ARS-GRIN	2,3,4
Ae. caudata	PI 551135	22.9833333	38.0666667	Greece	81	USDA ARS-GRIN	2,3,4
Ae. caudata	PI 298889	26.3325	39.79	Turkey	101	USDA ARS-GRIN	2,3,4
Ae. caudata	PI 560731	42.35	37.6333333	Turkey	1410	USDA ARS-GRIN	2,3,4
Ae. caudata	PI 551127	23.5833333	38.3666667	Greece	200	USDA ARS-GRIN	2,3,4
Ae. caudata	PI 551146	22.3666667	37.0833333	Turkey	-653.8	USDA ARS-GRIN	2,3,4
Ae. caudata	PI 298887	27.9666667	40.35	Turkey	37	USDA ARS-GRIN	2,3,4
Ae. caudata	PI 551125	23.9833333	38.0333333	Greece	100	USDA ARS-GRIN	1,2,3,4
Ae. caudata	PI 551141	22.2833333	36.8333333	Greece	53	USDA ARS-GRIN	2,3,4
Ae. caudata	PI 551124	23.15	40.3833333	Greece	220	USDA ARS-GRIN	2,3,4
Ae. caudata	PI 551131	22.8166667	37.5666667	Greece	64	USDA ARS-GRIN	1,2,3,4
Ae. caudata	PI 551136	22.9666667	37.9666667	Greece	0	USDA ARS-GRIN	2,3,4
Ae. caudata	PI 263554	29.4166667	38.6833333	Turkey	923	USDA ARS-GRIN	2,3,4
Ae. caudata	IG 47695	28.0333	40.1333	Turkey	981.2	ICARDA	2,3,4
Ae. caudata	PI 551139	22.5666667	36.75	Greece	124	USDA ARS-GRIN	2,3,4
Ae. caudata	IG 107317	22.25	37.0833	Greece	-661.9	ICARDA	2,3,4
Ae. caudata	PI 551120	22.8166667	39.7	Greece	90	USDA ARS-GRIN	2,3,4
Ae. comosa	PI 551029	23.95	38	Greece	150	USDA ARS-GRIN	2,3,4
Ae. comosa	PI 551030	23.5833333	38.3666667	Greece	200	USDA ARS-GRIN	1,2,3,4
Ae. comosa	PI 551037	22.5666667	36.75	Greece	124	USDA ARS-GRIN	2,3,4
Ae. comosa	AE 1260	24.6663889	37	Greece	-332.8	IPK	2,3,4
Ae. comosa	PI 551047	23.0666667	37.6333333	Greece	370	USDA ARS-GRIN	2,3,4
Ae. comosa	PI 551075	21.9833333	37.1833333	Greece	59	USDA ARS-GRIN	2,3,4

Ae. comosa	PI 551021	21.3666667	39.9833333	Greece	560	USDA ARS-GRIN	1,2,3,4
Ae. comosa	PI 551055	22.9833333	38.0666667	Greece	81	USDA ARS-GRIN	2,3,4
Ae. comosa	PI 551040	23.4833333	40.05	Greece	11	USDA ARS-GRIN	2,3,4
Ae. comosa	PI 551026	23.9166667	38.1166667	Greece	308	USDA ARS-GRIN	2,3,4
Ae. comosa	PI 542176	29.05	37.85	Turkey	180	USDA ARS-GRIN	2,3,4
Ae. comosa	PI 551019	22.1333333	39.6166667	Greece	100	USDA ARS-GRIN	1,2,3,4
Ae. comosa	PI 551033	23.3833333	38.45	Greece	339	USDA ARS-GRIN	2,3,4
Ae. comosa	PI 551045	22.7333333	37.75	Greece	110	USDA ARS-GRIN	1,2,3,4
Ae. comosa	PI 551043	22.8833333	37.9	Greece	100	USDA ARS-GRIN	2,3,4
Ae. comosa	PI 551057	22.7333333	37.9	Greece	-920	USDA ARS-GRIN	2,3,4
Ae. comosa	PI 551046	22.8666667	37.5833333	Greece	72	USDA ARS-GRIN	2,3,4
Ae. comosa	PI 551080	22.5333333	36.9166667	Greece	279	USDA ARS-GRIN	2,3,4
Ae. comosa	PI 551032	23.5	38.4833333	Greece	99	USDA ARS-GRIN	2,3,4
Ae. comosa	PI 551022	22.7	40.7	Greece	150	USDA ARS-GRIN	2,3,4
Ae. comosa	PI 551069	22.1666667	36.9833333	Greece	206	USDA ARS-GRIN	1,2,3,4
Ae. comosa	PI 551020	21.55	39.8333333	Greece	401.1	USDA ARS-GRIN	2,3,4
Ae. comosa	PI 551060	22.4	36.55	Greece	0	USDA ARS-GRIN	2,3,4
Ae. comosa	AE 783	19.8866667	39.6252778	Greece	0	IPK	2,3,4
Ae. comosa	PI 551017	22.7666667	39.6833333	Greece	140	USDA ARS-GRIN	2,3,4
Ae. comosa	PI 551038	22.3833333	36.6666667	Greece	261	USDA ARS-GRIN	2,3,4
Ae. comosa	AE 1258	24.8997222	36.6330556	Greece	-916.2	IPK	1,2,3,4
Ae. comosa	PI 551071	22.2833333	36.8333333	Greece	53	USDA ARS-GRIN	2,3,4
Ae. comosa	PI 551041	23.48333333	40.05	Greece	10	USDA ARS-GRIN	2,3,4
Ae. comosa	PI 551052	23.05	37.8	Greece	430	USDA ARS-GRIN	2,3,4
Ae. comosa	PI 551062	22.4833333	36.5	Greece	172	USDA ARS-GRIN	2,3
Ae. comosa	PI 551077	22.3666667	37.08333333	Greece	345	USDA ARS-GRIN	2,3
Ae. crassa	PI 276972	69.6666667	41	Uzbekistan	-171.4	USDA ARS-GRIN	2,3,4
Ae. crassa	PI 227342	52.5866667	29.7202778	Iran	1525	USDA ARS-GRIN	2,3,4
Ae. crassa	PI 219863	44	36.1666667	Iraq	458	USDA ARS-GRIN	2,3,4
Ae. crassa	PI 245725	40.0475	36.8508333	Turkey	367	USDA ARS-GRIN	2,3,4
Ae. crassa	PI 298892	67.9	36.7	Afghanistan	458	USDA ARS-GRIN	2,3,4
Ae. crassa	PI 487286	35.9	30.7	Syria	850	USDA ARS-GRIN	2,3,4
Ae. crassa	PI 219958	65.5	31.4166667	Afghanistan	915	USDA ARS-GRIN	2,3,4
Ae. crassa	PI 220274	67.7	32.7	Afghanistan	1954	USDA ARS-GRIN	2,3,4
Ae. crassa	PI 317400	67.8833333	36.1833333	Afghanistan	1098	USDA ARS-GRIN	2,3,4
Ae. crassa	PI 317393	63.25	34.9333333	Afghanistan	1068	USDA ARS-GRIN	2,3,4
Ae. cylindrica	IG 47901	32.2666666	38.2666666	Turkey	744	ICARDA	2,3,4
Ae. cylindrica	IG 48584	36.8	31.7833333	Jordan	328	ICARDA	2,3,4
Ae. cylindrica	PI 639286	76.6075	43.1388889	Kazakhstan	-221.4	USDA ARS-GRIN	2,3,4
Ae. cylindrica	IG 46621	36.6725	33.8844	Syria	1428	ICARDA	2,3,4
Ae. cylindrica	IG 48376	22.95	42.9166666	Bulgaria	972	ICARDA	2,3,4
Ae. cylindrica	PI 276976	50	32.26667	Iran	115	USDA ARS-GRIN	2,3,4
Ae. cylindrica	PI 560508	41.583333	38.8166667	Turkey	0	USDA ARS-GRIN	2,3,4
Ae. cylindrica	PI 486246	43.36667	40.15	Turkey	619	USDA ARS-GRIN	2,3,4
Ae. cylindrica	PI 639317	70.0597222	38.8719444	Tajikistan	-190.3	USDA ARS-GRIN	2,3,4
Ae. cylindrica	IG 47783	25.4166667	42	Bulgaria	832	ICARDA	2,3,4

Ae. cylindrica	IG 48325	23.65	42.0166667	Bulgaria	942	ICARDA	2,3,4
Ae. cylindrica	IG 47754	27.05	41.6666667	Turkey	1111	ICARDA	2,3,4
Ae. cylindrica	PI 172683	44.033333	38.933333	Turkey	0	USDA ARS-GRIN	2,3,4
Ae. cylindrica	IG 48311	26.2666666	41.8	Bulgaria	874	ICARDA	2,3,4
Ae. cylindrica	PI 554200	33.4833333	39.55	Turkey	860	USDA ARS-GRIN	2,3,4
Ae. cylindrica	PI 614624	34.4761111	44.7438889	Ukraine	141.4	USDA ARS-GRIN	2,3,4
Ae. cylindrica	PI 639332	70.9861111	38.4888889	Tajikistan	1343	USDA ARS-GRIN	2,3,4
Ae. cylindrica	PI 554201	37.7	38.36667	Turkey	544	USDA ARS-GRIN	2,3,4
Ae. cylindrica	PI 554202	39.406667	38.486667	Turkey	1416	USDA ARS-GRIN	2,3,4
Ae. cylindrica	IG 47830	27.25	42.7	Bulgaria	814	ICARDA	2,3,4
Ae. cylindrica	620280	7.38	46.2413888	Spain	515.7	personnal collection	2,3,4
Ae. cylindrica	IG 47894	32.5833333	37.9666666	Turkey	660	ICARDA	2,3,4
Ae. cylindrica	PI 574461	45.5	39.25	Azerbaijan	29	USDA ARS-GRIN	2,3,4
Ae. cylindrica	PI 554222	40.216667	38.9166667	Turkey	1732	USDA ARS-GRIN	2,3,4
Ae. cylindrica	IG 47770	25.0666667	42.3833333	Bulgaria	818	ICARDA	2,3,4
Ae. cylindrica	PI 560734	43.63333	37.483333	Turkey	-2263.8	USDA ARS-GRIN	2,3,4
Ae. cylindrica	IG 48569	71.1	40.8833333	Uzbekistan	-155.7	ICARDA	2,3,4
Ae. cylindrica	PI 314185	69.55	41.46667	Uzbekistan	-178.7	USDA ARS-GRIN	2,3,4
Ae. cylindrica	IG 47931	36.1666667	39.2833333	Turkey	329	ICARDA	2,3,4
Ae. cylindrica	IG 47888	33.9166667	37.4333333	Turkey	942	ICARDA	2,3,4
Ae. cylindrica	PI 486235	43.16667	38.3	Turkey	-2266.5	USDA ARS-GRIN	2,3,4
Ae. cylindrica	PI 276977	34.8333333	32.1666667	Israel	44	USDA ARS-GRIN	2,3,4
Ae. cylindrica	PI 573364	32.9833333	40.8166667	Turkey	564	USDA ARS-GRIN	2,3,4
Ae. cylindrica	IG 48789	36.3333	34.4667	Lebanon	0	ICARDA	2,3,4
Ae. cylindrica	PI 344778	20.916667	44.016667	Serbia	785	USDA ARS-GRIN	2,3,4
Ae. cylindrica	PI 554220	34.53333	38.983333	Turkey	384	USDA ARS-GRIN	2,3,4
Ae. cylindrica	IG 47707	28.95	40.25	Turkey	837	ICARDA	2,3,4
Ae. cylindrica	F1107A	5.77425	44.50898	France	358.6	personnal collection	2,3,4
Ae. cylindrica	PI 176853	28.826667	40.9636111	Turkey	769.3	USDA ARS-GRIN	2,3,4
Ae. cylindrica	PI 374348	20.8	41.1166667	Macedonia	1533	USDA ARS-GRIN	2,3,4
Ae. cylindrica	IG 47699	28.2	40.23333333	Turkey	945	ICARDA	2,3,4
Ae. cylindrica	PI 486241	43.93333	38.583333	Turkey	-1460.8	USDA ARS-GRIN	2,3,4
Ae. cylindrica	PI 378188	21.166667	42.66667	Serbia	950	USDA ARS-GRIN	2,3,4
Ae. cylindrica	PI 374345	21.616667	41.358333	Macedonia	1352	USDA ARS-GRIN	2,3,4
Ae. cylindrica	PI 298893	64.516667	35.85	Afghanistan	0	USDA ARS-GRIN	2,3,4
Ae. cylindrica	PI 374377	21.5880556	43.2341667	Serbia	843.3	USDA ARS-GRIN	2,3,4
Ae. cylindrica	IG 48547	67.5833333	40.0833333	Uzbekistan	247	ICARDA	2,3,4
Ae. cylindrica	PI 560509	42.55	37.483333	Turkey	-2145.4	USDA ARS-GRIN	2,3,4
Ae. cylindrica	IG 48493	58.1166667	38	Turkmenistan	142	ICARDA	2,3,4
Ae. cylindrica	PI 573363	30.933333	39.583333	Turkey	835	USDA ARS-GRIN	2,3,4
Ae. cylindrica	IG 47778	24.2833333	42.0333333	Bulgaria	948	ICARDA	2,3,4
Ae. cylindrica	IG 48495	58.3666667	38.1666667	Turkmenistan	97	ICARDA	2,3,4
Ae. cylindrica	IG 48537	69.0333333	41.1666667	Uzbekistan	-93	ICARDA	2,3,4
Ae. cylindrica	PI 542179	26.75	39.35	Turkey	1171	USDA ARS-GRIN	2,3,4
Ae. cylindrica	PI 573369	32.4955556	39.431111	Turkey	878.8	USDA ARS-GRIN	2,3,4
Ae. cylindrica	PI 172357	40.25	40.266667	Turkey	1829	USDA ARS-GRIN	2,3,4

Ae. cylindrica	PI 560520	42.53333	39.13333	Turkey	-1951.7	USDA ARS-GRIN	2,3,4
Ae. cylindrica	PI 560517	44.6	37.7166667	Turkey	0	USDA ARS-GRIN	2,3,4
Ae. cylindrica	PI 228333	47.15	35.63333	Iran	67	USDA ARS-GRIN	2,3,4
Ae. cylindrica	PI 298891	64.9	35.716667	Afghanistan	-8.6	USDA ARS-GRIN	2,3,4
Ae. cylindrica	IG 47831	26.766666	42.6333333	Bulgaria	900	ICARDA	2,3,4
Ae. cylindrica	IG 48541	66.9333333	39.65	Uzbekistan	232	ICARDA	2,3,4
Ae. cylindrica	PI 614620	33.7588889	44.467778	Ukraine	39.6	USDA ARS-GRIN	2,3,4
Ae. cylindrica	PI 614622	35.2205556	44.9169444	Ukraine	402	USDA ARS-GRIN	2,3,4
Ae. cylindrica	Turk5	43.29732	38.56323	Turkey	-2170.1	personnal collection	2,3,4
Ae. cylindrica	PI 314593	77.4333333	43.4666667	Kazakhstan	669	USDA ARS-GRIN	2,3,4
Ae. cylindrica	PI 374322	22.4652778	44.18889	Serbia	982.8	USDA ARS-GRIN	2,3,4
Ae. cylindrica	PI 374353	21.6	42.05	Macedonia	1095	USDA ARS-GRIN	2,3,4
Ae. cylindrica	PI 551082	21.3666667	39.95	Greece	803	USDA ARS-GRIN	2,3,4
Ae. cylindrica	IG 48580	70	40.9166666	Uzbekistan	-103.9	ICARDA	2,3,4
Ae. cylindrica	IG 48372	22.9666667	43.4833333	Bulgaria	980	ICARDA	2,3,4
Ae. cylindrica	IG 47824	27.7	43.5	Bulgaria	687	ICARDA	2,3,4
Ae. cylindrica	PI 254864	42.6833333	37.1166667	Irak	-2135.8	USDA ARS-GRIN	2,3,4
Ae. cylindrica	PI 639301	78.3461111	44.6694444	Kazakhstan	-258.3	USDA ARS-GRIN	2,3,4
Ae. cylindrica	PI 560521	41.83333	38.466667	Turkey	0	USDA ARS-GRIN	2,3,4
Ae. cylindrica	120506	7.390555556	45.75	Spain	588.1	personnal collection	2,3,4
Ae. cylindrica	IG 47926	36.3667	38.8	Turkey	373	ICARDA	2,3,4
Ae. cylindrica	IG 47820	28.0666667	43.3666667	Bulgaria	660	ICARDA	2,3,4
Ae. cylindrica	PI 486237	44.61667	37.2	Turkey	0	USDA ARS-GRIN	2,3,4
Ae. cylindrica	PI 486249	42.633333	40.1833333	Turkey	-1548	USDA ARS-GRIN	2,3,4
Ae. cylindrica	IG 48529	56.3333333	38.25	Turkmenistan	202	ICARDA	2,3,4
Ae. cylindrica	PI 554212	43.4166667	38.816667	Turkey	-2180	USDA ARS-GRIN	2,3,4
Ae. cylindrica	Turk11	29.3108	38.67512	Turkey	726.2	personnal collection	2,3,4
Ae. cylindrica	PI 378190	21.407778	43.8608333	Serbia	877.7	USDA ARS-GRIN	2,3,4
Ae. cylindrica	PI 486250	41.8	39.8333333	Turkey	1670	USDA ARS-GRIN	2,3
Ae. cylindrica	PI 573365	33.6833333	40.4833333	Turkey	700	USDA ARS-GRIN	2,3
Ae. cylindrica	PI 314406	44.7833333	41.7166667	Georgia	441	USDA ARS-GRIN	2,3
Ae. cylindrica	PI 568161	70.1	41.7	Uzbekistan	1200	USDA ARS-GRIN	2,3
Ae. cylindrica	PI 554225	42.6	38.4	Turkey	1750	USDA ARS-GRIN	2,3
Ae. cylindrica	IG 47705	29.3833333	40.1333333	Turkey	454	ICARDA	2,3
Ae. geniculata	5516	-3.07	41.57	Spain	-2749	personnal collection	2,3,4
Ae. geniculata	AE 606	12.5	41.67	Italy	219.6	IPK	2,3,4
Ae. geniculata	AE 681	16.44	40.24	Italy	-12.8	IPK	2,3,4
Ae. geniculata	AE 710	15.58	45.47	Croatia	1034.4	IPK	2,3,4
Ae. geniculata	AE 771	9.18	41.23	France	1924	IPK	2,3,4
Ae. geniculata	22739	-2.12	40.05	Spain	34	personnal collection	2,3,4
Ae. geniculata	AE 863	14.36	40.58	Italy	135.3	IPK	2,3,4
Ae. geniculata	I0720A	13.59665	45.78348	Italy	994.2	personnal collection	2,3,4
Ae. geniculata	AE 925	16.34	38.52	Italy	1842.1	IPK	2,3,4
Ae. geniculata	IG 107276	23.7883	41.0833	Greece	985	ICARDA	2,3,4
Ae. geniculata	PI 287737	-4	39.83333	Spain	0	USDA ARS-GRIN	2,3,4
Ae. geniculata	PI 573376	39.6833333	40.3666667	Turkey	370	USDA ARS-GRIN	2,3,4

Ae. geniculata	IG 46684	-9.05	38.59	Portugal	265.5	ICARDA	2,3,4
Ae. geniculata	IG 46699	-8.48	37.15	Portugal	255.3	ICARDA	2,3,4
Ae. geniculata	IG 47700	28.2	40.2333	Turkey	944.8	ICARDA	2,3,4
Ae. geniculata	PI 573411	32.8166667	39.85	Turkey	850	USDA ARS-GRIN	2,3,4
Ae. geniculata	IG 47038	38.37	39.04	Turkey	1605	ICARDA	2,3,4
Ae. geniculata	TA 2899	NA	NA	NA	NA	WGGRC	2,3
Ae. geniculata	IG 47090	38.26	27.1	Turkey	140	ICARDA	2,3,4
Ae. geniculata	IG 47337	35.5	31.52	Jordan	-2630.2	ICARDA	2,3,4
Ae. geniculata	IG 48173	9.88333	37.0667	Tunisia	2466.7	ICARDA	2,3,4
Ae. geniculata	PI 573406	34.9333333	41.3333333	Turkey	350	USDA ARS-GRIN	2,3,4
Ae. geniculata	PI 542181	40.0833333	38.3833333	Turkey	990	USDA ARS-GRIN	2,3,4
Ae. geniculata	PI 573383	30.9333333	39.5833333	Turkey	900	USDA ARS-GRIN	2,3,4
Ae. geniculata	IG 47389	35.41	31.04	Jordan	-2262.8	ICARDA	2,3,4
Ae. geniculata	22747	-5.3	41.05	Spain	-2312.2	personnal collection	2,3,4
Ae. geniculata	IG 47529	36.3	32.33	Syria	63	ICARDA	2,3,4
Ae. geniculata	IG 47602	34.12	35.43	Cyprus	0	ICARDA	2,3,4
Ae. geniculata	IG 47649	26.37	39.32	Turkey	1151	ICARDA	2,3,4
Ae. geniculata	IG 47836	36.8444	34.8172	Syria	4.2	ICARDA	2,3,4
Ae. geniculata	PI 487221	36.9972222	36.4694444	Syria	400	USDA ARS-GRIN	2,3,4
Ae. geniculata	IG 48008	7.01667	36.3667	Algeria	1812.8	ICARDA	2,3,4
Ae. geniculata	IG 47994	3.48	34.51	Algeria	878	ICARDA	2,3,4
Ae. geniculata	PI 491428	5.35	43.35	France	573	USDA ARS-GRIN	2,3,4
Ae. geniculata	IG 48166	10.1	36	Tunisia	1009	ICARDA	2,3,4
Ae. geniculata	C0723C	16.68795	43.47021	Croatia	2021.4	personnal collection	2,3,4
Ae. geniculata	PI 573401	33.4666667	40.8666667	Turkey	860	USDA ARS-GRIN	2,3,4
Ae. geniculata	PI 551110	22.75	37.7	Greece	260	USDA ARS-GRIN	2,3,4
Ae. geniculata	PI 573397	31.9666667	41.3166667	Turkey	510	USDA ARS-GRIN	2,3,4
Ae. geniculata	PI 170195	27.1	41.4333333	Turkey	47	USDA ARS-GRIN	2,3,4
Ae. geniculata	IG 48190	22.15	32.37	Lybia	372	ICARDA	2,3,4
Ae. geniculata	47M	-5.45	35.32	Marocco	1084	INRA Montpellier	2,3,4
Ae. geniculata	IG 48383	-1.2	34.52	Algeria	1456	ICARDA	2,3,4
Ae. geniculata	PI 483009	33.7666667	34.9833333	Cyprus	30	USDA ARS-GRIN	2,3,4
Ae. geniculata	PI 483026	32.3666667	35.05	Cyprus	20	USDA ARS-GRIN	2,3,4
Ae. geniculata	IG 48468	12.31	32.01	Lybia	427	ICARDA	2,3,4
Ae. geniculata	PI 487199	36.68333333	37	Syria	600	USDA ARS-GRIN	2,3,4
Ae. geniculata	PI 554290	18.05	42.7	Croatia	20	USDA ARS-GRIN	2,3,4
Ae. geniculata	PI 551118	21.9833333	37.1833333	Greece	59	USDA ARS-GRIN	2,3,4
Ae. geniculata	PI 524952	13.1333333	37.5666667	Italy	700	USDA ARS-GRIN	2,3,4
Ae. geniculata	C0722A	15.20407	44.22463	Croatia	2383.9	personnal collection	2,3,4
Ae. geniculata	IG 48471	13.21	32.53	Lybia	385	ICARDA	2,3,4
Ae. geniculata	PI 374337	19.2833333	42.3666667	Montenegro	42	USDA ARS-GRIN	2,3,4
Ae. geniculata	IG 48624	3.09	36.33	Algeria	361	ICARDA	2,3,4
Ae. geniculata	IG 48632	4.49	36.04	Algeria	361	ICARDA	2,3,4
Ae. geniculata	PI 374338	21.9002778	42.5513889	Serbia	463	USDA ARS-GRIN	2,3,4
Ae. geniculata	PI 551083	22.7833333	39.6833333	Greece	250	USDA ARS-GRIN	2,3,4
Ae. geniculata	IG 48068	NA	NA	Lebanon	NA	ICARDA	2,3

Ae. geniculata	PI 551095	23.4833333	40.05	Greece	10	USDA ARS-GRIN	2,3,4
Ae. geniculata	I0626A	10.07316	44.69324	Italy	755.6	personnal collection	2,3,4
Ae. geniculata	PI 573373	28.9833333	40.6333333	Turkey	75	USDA ARS-GRIN	2,3,4
Ae. geniculata	PI 573392	30.17	40.5086111	Turkey	25	USDA ARS-GRIN	2,3,4
Ae. geniculata	Turk9	29.18967	38.66942	Turkey	768.6	personnal collection	2,3,4
Ae. geniculata	88 E 177	28	42.04	Bulgaria	760	INRA Montpellier	2,3,4
Ae. geniculata	PI 170191	27.6088889	39.1883333	Turkey	163	USDA ARS-GRIN	2,3,4
Ae. geniculata	PI 564185	26.55	40.2	Turkey	70	USDA ARS-GRIN	2,3,4
Ae. geniculata	IG 48807	36.07	33.48	Syria	-138.6	ICARDA	2,3,4
Ae. geniculata	PI 551107	23.4333333	38.4666667	Greece	170	USDA ARS-GRIN	2,3,4
Ae. geniculata	IG 49042	35.45	33.26	Lebanon	-683.5	ICARDA	2,3,4
Ae. geniculata	F1207A	6.05761	44.04616	France	549.9	INRA Montpellier	2,3,4
Ae. geniculata	PI 551086	21.3833333	40.0333333	Greece	520	USDA ARS-GRIN	2,3,4
Ae. geniculata	96/3 E	-3.3	36.8	Spain	1980	INRA Montpellier	2,3,4
Ae. geniculata	IG 48178	21.1	32.5833	Lybia	540	ICARDA	2,3,4
Ae. geniculata	PI 574475	24.9333333	42.1333333	Bulgaria	250	USDA ARS-GRIN	2,3,4
Ae. geniculata	PI 560741	41.45	38.0166667	Turkey	1000	USDA ARS-GRIN	2,3,4
Ae. geniculata	IG 48199	-7.41667	31.4667	Morocco	838	ICARDA	2,3,4
Ae. geniculata	IG 46578	36.5847	35.9428	Syria	0	ICARDA	2,3,4
Ae. geniculata	F070606D	3.9062251	43.679541	France	509.5	personnal collection	2,3,4
Ae. geniculata	AE 1155	19.58	40.2	Albania	-13.8	IPK	2,3,4
Ae. geniculata	PI 388756	-6.4333333	33.1166667	Morocco	770	USDA ARS-GRIN	2,3,4
Ae. geniculata	PI 487229	36.1666667	35.1	Syria	600	USDA ARS-GRIN	2,3,4
Ae. geniculata	I0623A	11.41591	43.53616	Italy	-8.4	personnal collection	2,3,4
Ae. geniculata	AE 271	47.3	40.3	Azrbaïdjan	6	IPK	2,3,4
Ae. geniculata	AE 311	34.1	48.58	Ukraine	1897	IPK	2,3,4
Ae. geniculata	19422	-5.52	36.44	Spain	1484	personnal collection	2,3,4
Ae. geniculata	TA 1800	NA	NA	NA	NA	WGGRC	2,3
Ae. geniculata	PI 524956	14.0166667	37.8666667	Italy	1430	USDA ARS-GRIN	2,3
Ae. geniculata	28M	-3.17	34.58	Morocco	406	INRA Montpellier	2,3
Ae. geniculata	PI 487283	35.7	32.1	Jordan	800	USDA ARS-GRIN	2,3
Ae. geniculata	E0615B	-5.87117	38.26001	Spain	578	Personal collection	2,3
Ae. geniculata	PI 374374	22.5025	41.1391667	Macedonia	62	USDA ARS-GRIN	2,3
Ae. tauschii	CIae 26	49.4622222	37.4711111	Iran	-27	USDA ARS-GRIN	2,3,4
Ae. tauschii	PI 511367	69.18	34.53	Aghanistan	1843	USDA ARS-GRIN	2,3,4
Ae. tauschii	CIae 5	68.9833333	35.4166667	Afghanistan	212	USDA ARS-GRIN	2,3,4
Ae. tauschii	AE 525	52.8833333	36.6166667	Iran	40	INRA Montpellier	1,2,3,4
Ae. tauschii	PI 662066	55.6252778	39.2486111	Turkmenistan	68	USDA ARS-GRIN	2,3,4
Ae. tauschii	PI 486274	44.3666667	40.15	Armenia	428	USDA ARS-GRIN	2,3,4
Ae. tauschii	PI 486277	42.9333333	40.0833333	Turkey	1300	USDA ARS-GRIN	2,3,4
Ae. tauschii	PI 662060	56.5775	38.3716667	Turkmenistan	882	USDA ARS-GRIN	2,3,4
Ae. tauschii	PI 603232	48.6166667	40.6333333	Azerbaijan	630	USDA ARS-GRIN	2,3,4
Ae. tauschii	CIae 22	50.6825	36.9002778	Iran	-1	USDA ARS-GRIN	2,3,4
Ae. tauschii	PI 560535	43.8	37.7166667	Turkey	-2126.5	USDA ARS-GRIN	2,3,4
Ae. tauschii	PI 603233	49.4	40.0833333	Azerbaijan	780	USDA ARS-GRIN	2,3,4
Ae. tauschii	PI 603237	47.1666667	41.2	Azerbaijan	595	USDA ARS-GRIN	2,3,4

Ae. tauschii	PI 662078	61.4461111	35.71583333	Turkmenistan	963	USDA ARS-GRIN	2,3,4
Ae. tauschii	PI 317392	63.1833333	34.9666667	Afghanistan	976	USDA ARS-GRIN	2,3,4
Ae. tauschii	PI 554321	44.5666667	37.1666667	Turkey	1250	USDA ARS-GRIN	2,3,4
Ae. tauschii	PI 511363	64.25	35.8	Afghanistan	947	USDA ARS-GRIN	2,3,4
Ae. tauschii	PI 317398	67.8333333	36.3166667	Afghanistan	1128	USDA ARS-GRIN	2,3,4
Ae. tauschii	PI 220326	68.9	36.7	Afghanistan	518	USDA ARS-GRIN	1,2,3,4
Ae. tauschii	CIae 68	44.1	39.8	Turkey	1197	USDA ARS-GRIN	2,3,4
Ae. tauschii	PI 486273	43.3666667	40.15	Armenia	619	USDA ARS-GRIN	2,3,4
Ae. tauschii	PI 603234	47.8333333	40.9833333	Azerbaijan	600	USDA ARS-GRIN	2,3,4
Ae. tauschii	PI 662112	68.7927778	39.7577778	Tajikistan	1513	USDA ARS-GRIN	2,3,4
Ae. tauschii	PI 560537	43.3	38.8166667	Turkey	-2245.1	USDA ARS-GRIN	2,3,4
Ae. tauschii	PI 554313	43.3	38.4166667	Turkey	-2194	USDA ARS-GRIN	2,3,4
Ae. tauschii	CIae 15	60.8094444	35.4369444	Iran	99	USDA ARS-GRIN	2,3,4
Ae. tauschii	PI 603235	48.45	41.45	Azerbaijan	750	USDA ARS-GRIN	2,3,4
Ae. tauschii	PI 662069	55.2891667	38.07	Turkmenistan	108	USDA ARS-GRIN	2,3,4
Ae. tauschii	Ciae 1	67.0302778	30.2147222	Pakistan	1688	USDA ARS-GRIN	2,3,4
Ae. tauschii	PI 431600	47	43	Dagestan	59	USDA ARS-GRIN	1,2,3,4
Ae. tauschii	PI 511378	44.95	38.4	Iran	1190	USDA ARS-GRIN	2,3,4
Ae. tauschii	PI 560753	43.5333333	37.4833333	Turkey	1080	USDA ARS-GRIN	2,3,4
Ae. tauschii	PI 486270	44.3333333	37.7833333	Turkey	1670	USDA ARS-GRIN	1,2,3,4
Ae. tauschii	PI 511368	50.9666667	35.8	Iran	1296	USDA ARS-GRIN	2,3,4
Ae. tauschii	PI 662070	55.8761111	38.3208333	Turkmenistan	159	USDA ARS-GRIN	2,3,4
Ae. tauschii	PI 560538	42.5	38.4666667	Turkey	1670	USDA ARS-GRIN	2,3,4
Ae. tauschii	CIae 3	66.7191667	31.9708333	Afghanistan	1448	USDA ARS-GRIN	2,3,4
Ae. tauschii	PI 511380	51.42	36.65	Iran	39	USDA ARS-GRIN	2,3,4
Ae. tauschii	CIae 27	45.7216667	36.7647222	Iran	-12	USDA ARS-GRIN	2,3,4
Ae. tauschii	PI 603227	54.4833333	36.8666667	Iran	62	USDA ARS-GRIN	2,3,4
Ae. tauschii	CIae 2	66.4463889	30.925	Pakistan	1296	USDA ARS-GRIN	2,3,4
Ae. tauschii	PI 662105	70.5366667	39.8611111	Kirghizistan	1307.3	USDA ARS-GRIN	2,3,4
Ae. tauschii	PI 662095	70.6847222	40.8216667	Ouzbekistan	661.6	USDA ARS-GRIN	2,3,4
Ae. triuncialis	PI 254860	43.23	36.18	Irak	-2187.6	USDA ARS-GRIN	2,3,4
Ae. triuncialis	Turk4	29.3108	38.67512	Turkey	719.5	personnal collection	2,3,4
Ae. triuncialis	PI 542345	37.95	37.6667	Turkey	900	USDA ARS-GRIN	2,3,4
Ae. triuncialis	PI 374344	21.9922222	42.555	Serbia	521	USDA ARS-GRIN	2,3,4
Ae. triuncialis	PI 639313	69.9536111	38.8591667	Tjikistan	1144	USDA ARS-GRIN	2,3,4
Ae. triuncialis	PI 486279	35.6333333	38.8166667	Turkey	1090	USDA ARS-GRIN	2,3,4
Ae. triuncialis	PI 204853	39.2333333	38.6833333	Turkey	1104	USDA ARS-GRIN	2,3,4
Ae. triuncialis	PI 614635	35.2322222	44.9238889	Ukraine	320	USDA ARS-GRIN	2,3,4
Ae. triuncialis	PI 560550	42.43333333	38.73333333	Turkey	1690	USDA ARS-GRIN	2,3,4
Ae. triuncialis	IG 46945	40	38.05	Turkey	2209	ICARDA	2,3,4
Ae. triuncialis	PI 573494	33.6333333	40.5166667	Turkey	680	USDA ARS-GRIN	2,3,4
Ae. triuncialis	IG 47613	33.25	34.7333	Cyprus	-1412.8	ICARDA	2,3,4
Ae. triuncialis	PI 573462	30.1	39.8666667	Turkey	900	USDA ARS-GRIN	2,3,4
Ae. triuncialis	PI 220329	64.5166667	35.85	Afghanistan	915	USDA ARS-GRIN	2,3,4
Ae. triuncialis	PI 573512	32.8166667	39.85	Turkey	850	USDA ARS-GRIN	2,3,4
Ae. triuncialis	PI 573454	28.9	40.3666667	Turkey	70	USDA ARS-GRIN	2,3,4

Ae. triuncialis	PI 486299	42.8333333	40.05	Turkey	1300	USDA ARS-GRIN	2,3,4
Ae. triuncialis	PI 486298	44.05	39.6833333	Turkey	1470	USDA ARS-GRIN	2,3,4
Ae. triuncialis	PI 554365	41.4333333	40.1666667	Turkey	1620	USDA ARS-GRIN	2,3,4
Ae. triuncialis	PI 170192	27.183889	39.116944	Turkey	1076.3	USDA ARS-GRIN	2,3,4
Ae. triuncialis	PI 170197	27.0930556	41.4325	Turkey	51	USDA ARS-GRIN	2,3,4
Ae. triuncialis	PI 317390	62.1166667	34.6833333	Afghanistan	1434	USDA ARS-GRIN	2,3,4
Ae. triuncialis	E0618A	-3.87889	41.22293	Spain	-1962.5	personnal collection	2,3,4
Ae. triuncialis	PI 554329	34.3833333	39.0666667	Turkey	1090	USDA ARS-GRIN	2,3,4
Ae. triuncialis	PI 542299	40.03333	36.85	Turkey	1390	USDA ARS-GRIN	2,3,4
Ae. triuncialis	PI 227437	51.7833333	29.7166667	Iran	2135	USDA ARS-GRIN	2,3,4
Ae. triuncialis	PI 268206	46.2	33.5333333	Iran	575	USDA ARS-GRIN	2,3,4
Ae. triuncialis	PI 551197	22.1666667	40.4666667	Greece	470	USDA ARS-GRIN	2,3,4
Ae. triuncialis	PI 542305	37.0166667	37.6166667	Turkey	800	USDA ARS-GRIN	2,3,4
Ae. triuncialis	E0610E	-1.94307	39.21795	Spain	228.3	personnal collection	2,3,4
Ae. triuncialis	PI 542309	31.8333333	38.2833333	Turkey	1200	USDA ARS-GRIN	2,3,4
Ae. triuncialis	IG 48210	-6.11667	33.4333	Morocco	1224.9	ICARDA	2,3,4
Ae. triuncialis	PI 542297	39.2333333	37.2	Turkey	625	USDA ARS-GRIN	2,3,4
Ae. triuncialis	Turk1	27.44197	38.42193	Turkey	973.3	personnal collection	2,3,4
Ae. triuncialis	PI 487246	36.8780556	36.4852778	Syria	350	USDA ARS-GRIN	2,3,4
Ae. triuncialis	E0615A	-5.69634	38.26326	Spain	497.6	personnal collection	2,3,4
Ae. triuncialis	PI 180793	28.8266667	40.9636111	Turkey	20	USDA ARS-GRIN	2,3,4
Ae. triuncialis	PI 483027	32.4666667	34.95	Cyprus	140	USDA ARS-GRIN	2,3,4
Ae. triuncialis	PI 374366	18.9333333	42.2	Montenegro	2192	USDA ARS-GRIN	2,3,4
Ae. triuncialis	PI 170212	27.9855556	39.7108333	Turkey	338	USDA ARS-GRIN	2,3,4
Ae. triuncialis	PI 314407	44.7833333	41.7166667	Georgia	441	USDA ARS-GRIN	2,3,4
Ae. triuncialis	PI 568163	70.1	41.7	Turkey	1000	USDA ARS-GRIN	2,3,4
Ae. triuncialis	IG 46571	36.5528	36.2306	Syria	678.6	ICARDA	2,3,4
Ae. triuncialis	PI 491442	5.71667	43.33333	France	600	USDA ARS-GRIN	2,3,4
Ae. triuncialis	PI 226499	49.0666667	31.8666667	Iran	61	USDA ARS-GRIN	2,3,4
Ae. triuncialis	PI 542323	29.05	37.85	Turkey	180	USDA ARS-GRIN	2,3,4
Ae. triuncialis	862578	-5.57823	36.86204	Spain	1312.8	personnal collection	2,3,4
Ae. triuncialis	PI 221899	66.7833333	34.3666667	Afghanistan	2257	USDA ARS-GRIN	2,3,4
Ae. triuncialis	PI 483036	33.3833333	35.15	Cyprus	160	USDA ARS-GRIN	2,3,4
Ae. triuncialis	PI 171468	36.9333333	40.6833333	Turkey	1095	USDA ARS-GRIN	2,3,4
Ae. triuncialis	IG 107359	21.56	39.885	Greece	387	ICARDA	2,3,4
Ae. triuncialis	PI 573506	34.6333333	41.8333333	Turkey	310	USDA ARS-GRIN	2,3,4
Ae. triuncialis	PI 344793	22.3833333	41.6333333	Macedonia	448	USDA ARS-GRIN	2,3,4
Ae. triuncialis	PI 250696	45.9	38.3333333	Iran	1563	USDA ARS-GRIN	2,3,4
Ae. triuncialis	E0611C	-4.01224	37.69606	Spain	973.9	personnal collection	2,3,4
Ae. triuncialis	PI 564219	30.8	36.9833333	Turkey	10	USDA ARS-GRIN	2,3,4
Ae. triuncialis	PI 564225	26.6666667	40.3333333	Turkey	10	USDA ARS-GRIN	2,3,4
Ae. triuncialis	PI 573479	30.8166667	40.8333333	Turkey	290	USDA ARS-GRIN	2,3,4
Ae. triuncialis	PI 574474	28.0833333	43.3833333	Bulgaria	20	USDA ARS-GRIN	2,3,4
Ae. triuncialis	PI 374357	20.8875	43.609722	Serbia	849.3	USDA ARS-GRIN	2,3,4
Ae. triuncialis	PI 551244	21.9	37.15	Greece	284	USDA ARS-GRIN	2,3,4
Ae. triuncialis	PI 317399	67.8833333	36.1833333	Afghanistan	1098	USDA ARS-GRIN	2,3,4

Ae. triuncialis	PI 215781	70.1166667	37.0333333	Afghanistan	836	USDA ARS-GRIN	2,3,4
Ae. triuncialis	PI 486295	44.65	37.2333333	Turkey	1400	USDA ARS-GRIN	2,3,4
Ae. triuncialis	PI 560764	42.25	37.733333	Irak	-2088.2	USDA ARS-GRIN	2,3,4
Ae. triuncialis	PI 573470	30.9333333	39.5833333	Turkey	900	USDA ARS-GRIN	2,3,4
Ae. triuncialis	PI 573481	31.9833333	40.7666667	Turkey	950	USDA ARS-GRIN	2,3,4
Ae. triuncialis	I0626A	10.07316	44.69324	Italy	755.6	personnal collection	2,3,4
Ae. triuncialis	C0723C	16.68795	43.47021	Croatia	2021.4	personnal collection	2,3,4
Ae. triuncialis	PI 374354	20.9166667	40.9833333	Macedonia	883	USDA ARS-GRIN	2,3,4
Ae. triuncialis	PI 564231	27.8333333	40.4166667	Turkey	15	USDA ARS-GRIN	2,3,4
Ae. triuncialis	PI 551223	22.75	37.7	Greece	260	USDA ARS-GRIN	2,3,4
Ae. triuncialis	PI 551199	22.9166667	40.8833333	Greece	80	USDA ARS-GRIN	2,3,4
Ae. triuncialis	PI 268207	47.216667	34.25	Iran	78	USDA ARS-GRIN	2,3,4
Ae. triuncialis	IG 48133	35.7667	33.7667	Lebanon	0	ICARDA	2,3,4
Ae. triuncialis	PI 250908	47.6666667	39	Iran	88	USDA ARS-GRIN	2,3,4
Ae. triuncialis	PI 173615	43.76667	37.53333	Turkey	458	USDA ARS-GRIN	2,3,4
Ae. triuncialis	C0722A	15.20407	44.22463	Croatia	2383.9	personnal collection	2,3,4
Ae. triuncialis	PI 551181	22.5166667	39.6166667	Greece	30	USDA ARS-GRIN	2,3,4
Ae. triuncialis	PI 573501	34.9333333	40.8833333	Turkey	460	USDA ARS-GRIN	2,3
Ae. triuncialis	PI 551241	22.4333333	36.7	Greece	237	USDA ARS-GRIN	2,3
Ae. triuncialis	IG 47999	1.3666667	35.4833333	Algeria	647	ICARDA	2,3
Ae. triuncialis	PI 614630	34.2333333	44.5108333	Ukraine	300	USDA ARS-GRIN	2,3
Ae. triuncialis	PI 219862	44.2	36.4	Iraq	935	USDA ARS-GRIN	2,3
Ae. triuncialis	PI 374350	20.4308333	42.3802778	Serbia	363	USDA ARS-GRIN	2,3
Ae. triuncialis	PI 551220	23.3833333	38.4333333	Greece	193	USDA ARS-GRIN	2,3
Ae. triuncialis	PI 554223	40.7	38.9166667	Turkey	1250	USDA ARS-GRIN	2,3
Ae. triuncialis	PI 551207	23.4833333	40.05	Greece	10	USDA ARS-GRIN	2,3
Ae. triuncialis	PI 487201	37.4927778	35.8141667	Syria	350	USDA ARS-GRIN	2,3
Ae. triuncialis	PI 551211	23.8666667	38.0666667	Greece	507	USDA ARS-GRIN	2,3
Ae. umbellulata	PI 298906	44.1333333	36.4	Iraq	709	USDA ARS-GRIN	2,3,4
Ae. umbellulata	PI 542366	37.8333333	37.7666667	Turkey	670	USDA ARS-GRIN	2,3,4
Ae. umbellulata	PI 573515	27.6833333	39.6666667	Turkey	952	USDA ARS-GRIN	2,3,4
Ae. umbellulata	PI 554404	31.8	37.0166667	Turkey	10	USDA ARS-GRIN	2,3,4
Ae. umbellulata	PI 486261	27.1666667	37.6666667	Turkey	559	USDA ARS-GRIN	1,2,3,4
Ae. umbellulata	PI 542381	37.4666667	37.1833333	Turkey	697	USDA ARS-GRIN	1,2,3,4
Ae. umbellulata	PI 542363	40.0833333	38.3833333	Turkey	990	USDA ARS-GRIN	2,3,4
Ae. umbellulata	PI 554405	29.05	37.85	Turkey	851	USDA ARS-GRIN	2,3,4
Ae. umbellulata	PI 560555	42.0166667	37.9166667	Turkey	1000	USDA ARS-GRIN	2,3,4
Ae. umbellulata	PI 564235	26.4833333	39.5833333	Turkey	1179	USDA ARS-GRIN	2,3,4
Ae. umbellulata	PI 554385	34.3833333	39.0666667	Turkey	1090	USDA ARS-GRIN	2,3,4
Ae. umbellulata	PI 542379	27.0166667	39.0666667	Turkey	30	USDA ARS-GRIN	2,3,4
Ae. umbellulata	PI 573436	30.9333333	39.5833333	Turkey	900	USDA ARS-GRIN	2,3,4
Ae. umbellulata	PI 554416	44.5333333	37.1333333	Turkey	1150	USDA ARS-GRIN	2,3,4
Ae. umbellulata	PI 542367	37.0166667	37.6166667	Turkey	800	USDA ARS-GRIN	2,3,4
Ae. umbellulata	PI 560557	42.6	38.7666667	Turkey	1670	USDA ARS-GRIN	2,3,4
Ae. umbellulata	PI 560773	43.65	37.4833333	Turkey	1150	USDA ARS-GRIN	2,3,4
Ae. umbellulata	PI 560775	41.45	38.0166667	Turkey	1000	USDA ARS-GRIN	2,3,4

Ae. umbellulata	PI 554282	30.1666667	41.0833333	Turkey	50	USDA ARS-GRIN	2,3,4
Ae. umbellulata	IG 46964	40	38.05	Turkey	2209	ICARDA	2,3,4
Ae. umbellulata	PI 554388	39.0333333	38.5333333	Turkey	1160	USDA ARS-GRIN	2,3,4
Ae. umbellulata	PI 542384	38.7666667	37.8	Turkey	850	USDA ARS-GRIN	2,3,4
Ae. umbellulata	PI 554410	31.8833333	37.8666667	Turkey	1250	USDA ARS-GRIN	2,3,4
Ae. umbellulata	PI 554413	38.3333333	38.55	Turkey	1300	USDA ARS-GRIN	2,3,4
Ae. umbellulata	PI 542377	27.4166667	39.0833333	Turkey	30	USDA ARS-GRIN	1,2,3,4
Ae. umbellulata	PI 573516	30.0166667	40.2333333	Turkey	120	USDA ARS-GRIN	1,2,3,4
Ae. umbellulata	IG 110801	40.4403	36.44	Syria	1256.8	ICARDA	2,3,4
Ae. umbellulata	PI 573420	27.2333333	39.9333333	Turkey	250	USDA ARS-GRIN	2,3,4
Ae. umbellulata	Turk12	29.27353	38.6746	Turkey	756.8	personnal collection	2,3,4
Ae. umbellulata	CIae66	21.2833333	42.15	Serbia	575	USDA ARS-GRIN	2,3
Ae. umbellulata	PI 554409	32.8166667	39.85	Turkey	850	USDA ARS-GRIN	2,3
Ae. bicornis	CIae 47	NA	NA	NA	NA	USDA ARS-GRIN	1,2
Ae. longissima	PI 604127	NA	NA	Israel	NA	USDA ARS-GRIN	1,2
Ae. longissima	PI 604108	34.7333333	31.8333333	Israel	50	USDA ARS-GRIN	1,2
Ae. searsii	PI 599168	35.1166667	31.35	Israel	850	USDA ARS-GRIN	1,2
Ae. searsii	PI 599149	35.1166667	31.35	Israel	850	USDA ARS-GRIN	1,2
Ae. sharonensis	PI 584396	34.6666667	31.8166667	Israel	50	USDA ARS-GRIN	1,2
Ae. sharonensis	PI 584412	34.8833333	32.4666667	Israel	20	USDA ARS-GRIN	1,2
Ae. speltoides subsp. Speltoides	PI 170204	27.4833333	41.3333333	Turkey	86	USDA ARS-GRIN	1,2
Ae. speltoides subsp. lingustica	PI 560528	42.1666667	37.75	Turkey	1100	USDA ARS-GRIN	1,2
Ae. uniaristata	PI 554418	NA	NA	NA	NA	USDA ARS-GRIN	1,2
Ae.uniaristata	PI 554420	27.0666667	38.6	Turkey	30	USDA ARS-GRIN	1,2
Aegilops mutica	PI 636562	34.1666667	38.0333333	Turkey	1340	USDA ARS-GRIN	1,2
Triticum monococcum subsp. monococcum	PI 94743	38.4	57.95	Russia	121	USDA ARS-GRIN	1,2
Triticum monococcum subsp. aegilopoides	PI 427990	35.8666667	33.5166667	Lebanon	1141	USDA ARS-GRIN	1,2
Triticum urartu	PI 428228	41.6833333	37.1166667	Turkey	519	USDA ARS-GRIN	1,2
Triticum urartu	PI 427328	44.3666667	36.4166667	Iraq	1100	USDA ARS-GRIN	2
Agropyron mongolicum	PI 499391	NA	NA	China	NA	USDA ARS-GRIN	1,2
Taeniatherum caput-medusae	PI 577709	NA	NA	Turkey	20	USDA ARS-GRIN	1,2
Secale cereale	PI 234656	NA	NA	Kazakhstan	NA	USDA ARS-GRIN	1,2
Secale cereale	PI 561793	27.9666667	38.5166667	Turkey	490	USDA ARS-GRIN	1,2
Hordeum vulgare subsp. *spontaneum*	PI 282585	NA	NA	Israel	NA	USDA ARS-GRIN	2
Thynopyrum junceum	PI 414667	NA	NA	Greece	NA	USDA ARS-GRIN	2
Dasypyrum villosum	PI 598396	NA	NA	Greece	NA	USDA ARS-GRIN	2
Bromus hordeaceus	PI 254877	NA	NA	Iraq	NA	USDA ARS-GRIN	2
Brachypodium distachyon	PI 317418	NA	NA	Afghanistan	830	USDA ARS-GRIN	2

Stella Huynh

stella.huynh@unine.ch

https://www.researchgate.net/profile/Stella_Huynh

Rue Emer-de-Vattel 30
2000 Neuchâtel
Suisse

+41 76 637 89 18

French citizen
Swiss residence permit (B)

Field work

- Field sampling
- Monitoring arthropods diversity
- Plant emasculation
- Interspecific crossings + monitoring

Lab work

- Sanger sequencing
- Illumina MiSeq
- AFLP, SSAP
- Seed growth on media

Bioinformatics

- R
- Bash, Python

Statistics & Tools

- Mantel tests, ANOVA
- BEAST2, PhyloNet, AlloppNET, PhyML, RAxML, Splitstree4, Dendroscope, DensiTree
- GeneMarker
- ArcGIS, ImageJ
- SPAGeDi, Arlequin, STRUCTURE

Language

- French (native)
- English (C1)

Scientific researcher

in species evolution and adaptation

Profile

- Recently graduated as PhD in evolutionary plant genomics
- Keen interests in population genomics, phylogenomics, adaptation, ecological modeling and plants
- Proficiency in molecular biology and bioinformatics to handle large-scale data.
- Facility to move for collaborations (e.g. Norway, Sweden) and congresses (e.g. China)

Achievements

NGS sequencing:
- Multiplexed PCR of 48 loci x 475 samples from 20 grass species, (fast and accurate technique, economically advantageous)
- Library preparation for paired-end Illumina MiSeq (3 runs, ~45M reads)

Scientific research:
- 1 paper published, 1 submitted, 1 to be submitted shortly, 1 in preparation
- 5 talks and 5 posters in international congresses and seminars
- 2 travel grants obtained

Research Experience

2018 – Scientific collaborator

Scientific collaborator in bioinformatic tool development
Pr. Nicolas Salamin, Univ. of Lausanne, Switzerland

- *First-phase development of a Python module for estimating the evolution of phenotypic traits*

2014 – 2018 PhD research

'Comparative phylogeography of diploid and polyploid wild wheats'
Pr. Christian Parisod, Dr. François Felber, Univ. of Neuchâtel, Switzerland

- *Inferred reticulate and polyploid MSC phylogenies showing strong support to complex origins of diploid and polyploid wheat species*
- *Tested hypothesis of progenitor's additivity in polyploid wheat, both at genetic and ecological level*
- *Main highlights: Increased genetic diversity and ecological tolerance promoting the fast expansion of polyploid populations*

2013 – 2014 Master research

'Detection of genomic regions underlying adaptation in heavy-metal contaminated soils in *Arabidopsis halleri*'
Pr. Xavier Vekemans, Dr. Pierre-Saumitou Laprade, Dr. Maxime Pauwels, Univ. of Lille 1, France

- *Identified 1 locus concentrating most SNPs, with a gene involved in zinc uptake (using 'Popoolation2' for whole-genome scan and 'Gorilla' for gene enrichment)*

2012 Research assistant

'Genetic variation of vegetative growth in natural conditions in the model plant *Arabidopsis thaliana*'
Dr. Romain Villoutreix, Univ. of Lille 1, France

- *Assessing seedlings' growth under different competitor's pressures*

2011 Research assistant

Ecoimmunology of the annelids *Nereis diversicolor*
Dr. Virginie Cuvillier, Univ. of Lille 1, France

- *Monitored feeding behavior of Nereis sp. under non- vs. contaminated food by phtalates*

Education

2018	PhD in Evolutionary Genetics	University of Neuchâtel, Switzerland
2014	Msc in Ecosystems and Biodiversity Management	University of Lille 1, France
2012	Bsc in Biology	University of Lille 1, France

Grants

2017	Subvention 'Egalité des chances' - *for research collaboration in Norway & Sweden* - *1500 CHF budget*	University of Neuchâtel
2017	Workshops and international congresses travel grant - *for IBC congress in China* - *1000 CHF budget*	University of Neuchâtel

Symposium and Congress Participations

2017 poster	'Polyploids are *not* the genetic and ecological additivity of their progenitors'	International Botanical Congress (IBC) 2018, China (~7000 participants)
2017 poster & flash-talk	'Are polyploids ecologically and genetically different from their diploid progenitors? The case of *Aegilops* spp.'	Biology17, Switzerland (~350 participants)
2016 poster	'Multiplex amplicon sequencing for genotyping and phylogenetic inference in polyploid wild wheats'	Biology16, Switzerland (~260 participants)
2016 talk	'Disentangling reticulate evolution in diploid-polyploid systems using high-throughput amplicon sequencing: the case of *Aegilops* species'	GDC symposium, Switzerland (~40 participants)
2015 poster	'Multiplex amplicon sequencing for genotyping and phylogenetic inference in polyploid wild wheats'	European Society for Evolutionary Biology (ESEB) 2015, Switzerland (~1400 participants)

Workshop and Seminar Participations

2018 talk	'Comparative phylogeography of diploid and allopolyploid wils wheats *Aegilops* spp.'	Internal seminar of the Department of Ecology and Evolution University of Lausanne, Switzerland
2017 talk	'Evolution of genomes and ecological niches in polyploid wild wheats'	Joint seminar in Plant Ecology University of Bern, Switerland
2017 talk	'Multiple origins and population structure of polyploid wild wheat species'	Phylogenomics and phylogeography workshop University of Gothenburg, Sweden

Publications

Bravo GA, Antonelli A, Bacon CD, Bartoszek K, Blom M, **Huynh S**, Jones G, Knowles LL, Lamichhaney S, Marcussen T, Morlon H, Nakhleh L, Oxelman B, Pfeil B, Schliep A, Wahlberg N, Werneck F, Wiedenhoeft J, Willows-Munro S and Edwards SV **(2019)**. Embracing heterogeneity: coalescing the Tree of Life and the future of phylogenomics. PeerJ **6**:e26449v3 https://doi.org/10.7287/peerj.preprints.26449v3 (link of accepted version to be coming soon)

Huynh S, Marcussen T, Felber F and Parisod C **(submitted)**. Hybridization preceded radiation in diploid wheats. New Phytologist

Huynh S, Felber F, Broennimann O, Guisan A and Parisod C (2019, ***in prep***). Genetic and ecological additivity in the evolutionary success of allopolyploid wheats. Ecology Letters

Huynh S, Felber F and Parisod C (2019, ***in prep***). Differential evolutionary dynamics of retrotransposons in the wild wheats. New Phytologist