Transposable element and genome evolution following hybridization in wild wheats

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

Par
Natacha Senerchia

Composition du jury :

PD Dr Parisod Christian, directeur de thèse
Prof Dr Kessler Félix, Co-directeur et rapporteur
PD Dr Felber François, Co-directeur et rapporteur
Dr Kashkush Khalil, rapporteur
Dr Wicker Thomas, rapporteur

Université de Neuchâtel, Switzerland
2014

UNIVERSITÉ DE NEUCHÂTEL

Faculté des sciences
Secrétariat-décanat de Faculté
Rue Emile-Argand 11
2000 Neuchâtel - Suisse
Tél: + 41 (0)32 718 2100
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 Natacha SENERCHIA

Titre:

“Transposable element and genome evolution following hybridization in wild wheats”

sur le rapport des membres du jury composé comme suit:

- PD Dr Christian Parisod, Université de Neuchâtel, directeur de thèse
- Prof. Felix Kessler, co-directeur de thèse, Université de Neuchâtel
- PD Dr François Felber, co-directeur de thèse, Université de Neuchâtel
- Dr Khalil Kashkush, Ben-Gourion University of the Nevev, Israël
- Dr Thomas Wicker, Université de Zürich

Neuchâtel, le 12 juin 2014

Le Doyen, Prof. P. Kropf

Table of content

Abstract

Genome dynamics is an essential process of eukaryote genome evolution. Hybridization and inter-species gene flow result in new interactions among divergent genomes and may reveal genetic incompatibilities having accumulated after the origin of species. Being highly mutagenic and repressed by various epigenetic mechanisms, transposable elements (TEs) are postulated to play a central role in fuelling genome reorganization following inter-genomic conflicts after hybridization. However, we are still far from understanding mechanisms and forces of such genome dynamics.

The main aim of this was to test the implication of multiple TE families on genome reorganization, hybridization success and introgression. Four *Aegilops* allotetraploid species with interconnected genomes belonging to the wheats group, *Aegilops crassa* (DDMM), *Ae. cylindrica* (CCDD), *Ae. geniculata* (MMUU) and *Ae. triuncialis* (CCUU) that derived from the diploids *Ae. caudata* (CC), *Ae. comosa* (MM), *Ae. tauschii* (DD) and *Ae. umbellulata* (UU) were used as model. These species have complex genomes with 80% of repetitive elements that have differentially diverged under the influence of TEs and hybridization.

As a first step, the TE composition was assessed using high-throughput sequencing in selected model species and TE families were classified as recently active or quiescent. Based on these results, fingerprint assays were designed and evaluated restructuring in 17 active TE families by comparing genome wide restructuring in diploid and derived tetraploid species and highlighted different TE specific evolutionary trajectories following polyploidy. Restructuring was TE-specific and species-specific, but consistently correlated with TE divergence between progenitors. Using this knowledge, levels of restructuring and methylation changes around insertions of nine TE families were assessed in artificial F1 hybrids between the tetraploid species and, again, correlated with TE divergence between parents. Asymmetrical patterns of genome reorganization paralleling patterns of reproductive isolation among tetraploids, together with nonrandom loss and methylation changes in hybrids suggest that certain events are necessary to produce viable plants. In a last study, hybridization and backcrossing was assessed in natural hybrid zones between the *Ae. geniculata* and *Ae. triuncialis*, relating important restructuring of which sequences lost was dominant and that three TE families among the six selected ones were differently exchanged.

Aegilops as model offered unparalleled opportunities to address the evolutionary trajectories of multiples candidate TE families and reported evidence of important genome reorganization following various independent hybridization, with multiple TE families clearly playing a central role in host genome evolution and revealing that the progenitors divergence between TE families impacts on the reorganization level. This work reported potential proximate and ultimate factors of genome reorganization driven by conflicts between intragenomic parasites.

Key word

LTR retrotransposon evolutionary trajectories, allopolyploidization, 454 pyrosequencing, AFLP, SSAP, MSAP, MSTD, *Aegilops*, F1 hybrids, intra-genomic conflicts, non-random genome reorganization, methylation changes, introgression, reproductive isolation

Abbreviations

AFLP	amplified fragment length polymorphism
DNA	deoxyribonucleic acid
LTR	long terminal repeat
MSAP	methyl sequence amplified polymorphism
MSTD	methyl-sensitive transposon display
RNA	ribonucleic acid
SSAP	sequence-specific amplified polymorphism
siRNA	small inverted ribonucleic acid
SSR	simple sequence repeat
TREP	Triticeae repeat sequence database

Acknowledgments

I would like sincerely to thank

Christian Parisod for giving me confidence to carry out this work, brought me into this transposable element adventure. Thank you for all the advice and support during all steps of this thesis doctoral and for having trusted my organization

François Felber for his constant support and for giving me the chance to do this research.

Giorgina Bernasconi for giving me the opportunity to do this research work.

Dr Kashkush Khalil, Prof Dr Kessler Felix and Dr Wicker Thomas as members of the Jury for all useful comments and corrections to this thesis.

People of the Laboratory of evolutionary botany and more generally of the University of Neuchâtel.

The Botanical garden team of Neuchâtel for their help during the set of culture and for having taking care of the plants during months.

Prof Dr Moshe Feldman for the crucial and extraordinary help in sampling and recognizing hybrids in Israel.

Prof Dr Avraham Levy for useful discussions which guided my work.

Dr Thomas Wicker and his team for bioinformatics supports.

Eloise Tranchand, Alexandre Ruiz and Léa Blanchard for their help in producing cross data and molecular analyses.

R. A. Slobodeanu for statistical advices.

People at the Genetic Diversity Centre, ETH Zurich, where some important data were generated.

Particular thanks to my family and friends for having supported and listened to me throughout this adventure.

Anouk Sarr for her essential advises, her valuable friendship and for the nice office sharing time.

Thanks to Cylia for those long hours of debriefing.

Livio.

Introduction

Genome reorganization following hybridization

Plant genomes are reputed highly labile over evolutionary scales (Walbot & Cullis, 1985; Kejnovsky et al., 2009). The emergence of high throughput data is confirming such dynamics and calls for a better understanding of underlying mechanisms (Lister et al., 2009; Brenchley et al., 2012). Genome reorganization occurs at the structural and the functional levels, and was mainly associated with prevalent processes such as polyploidy and retrotransposition (Kejnovsky et al., 2009; Murat et al., 2012).

Polyploidization (hybridization between more or less divergent genomes associated to whole genome duplication) is now postulated as a prevalent evolutionary process in plants, particularly among angiosperms (Jiao *et al.*, 2011). Whether polyploidization is a route to evolutionary success however remains an ongoing debate (Stebbins, 1971; Wendel, 2000; Mayrose *et al.*, 2011; Arrigo & Barker, 2012). Genome reorganization is commonly observed in the first few generations following hybridization and sometimes as early as in newly formed hybrids (Parisod *et al.*, 2009, see chapter 1). Focusing on hybrids, often ignored by evolutionary and phylogenetic thinking, provides a direct access on changes of an organism at the confrontation of divergent genomes, the causes and the consequences of revealed conflicts (detailed in chapter 1) allowing a better understanding of some driving forces and mechanisms driving such genomes evolution. High genome reorganization of hybrids was classified according to two types: revolutionary changes acting immediately after hybridization and stabilizing the neo plant type vs. evolutionary changes happening during the life of the hybrid and bringing genetic variability that might enhance adaptive potential (Feldman & Levy, 2005b; Feldman & Levy, 2009). Revolutionary changes can induce

genome restructuring, such as reproducible elimination of DNA sequences at early stage on polyploids (Feldman *et al.*, 1997; Ozkan *et al.*, 2001). Eilam's work (Eilam *et al.*, 2008; Eilam *et al.*, 2009) shows that allopolyploidization in Triticeae leads to rapid elimination of DNA sequences whether they are high-copy, low-copy, coding or non-coding sequences. Such elimination can increase divergence between homeologous chromosomes avoiding intergenomic recombination. Changes in gene expression could be due to genetic or epigenetic modifications (Feldman & Levy, 2005b). Such reorganization was found in wheat and other grasses (Galili *et al.*, 1986; Liu *et al.*, 1998; Kashkush *et al.*, 2002). Some revolutionary changes can affect both structure and gene expression. For instance, new insertion of transposable elements might affect gene expression or the genome structure right after hybridization (Feldman & Levy, 2005). The other class of changes, named evolutionary changes, occurs along species lifetime: it includes mutations, exchanges of chromosome sections between genomes, hybridization and introgression between polyploids, that may be to develop new abilities (Feldman & Levy, 2005b).

Hybridization is postulated to induce a genome shock (Genome Shock hypothesis) (McClintock, 1984; Josefsson *et al.*, 2006) initiating rapid genome modifications that probably restore: (i) genome harmony (ii) proper dosage of transcripts (iii) proper pairing of homologous and avoid homeologous genomes pairing at meiosis (Levy & Feldman 2002). The exact causes and consequences of reorganization following merging genomes remain still poorly understood.

Hybridization and reproductive isolation

Hybridization associated or not to whole genome doubling is known to induce drastic genome reorganization having sometimes positive consequences but most often negative effects on the newly formed hybrids (Burke & Arnold, 2001; Doyle *et al.*, 2008). The merging of distinct genomes at the origin of polyploid species may reveal conflicts leading to negative consequences, attributed to genetic incompatibilities accumulated in genomes that previously diverged from a common ancestor (Doyle *et al.*, 2008; Tayalé & Parisod, 2013). On one hand, genome reorganization following hybridization having negative consequence may be postulated as a mechanism maintaining reproductive isolation. On the other hand, genome reorganization may be necessary to resolve conflicts revealed at genomes merging and thus is a key mechanism to produce viable plants. Genetic incompatibilities may affect directly the embryo or indirectly its nutrient support, the endosperm (Soltis & Soltis, 2000). Symmetrical reproductive isolation between reciprocal crosses is expected under nuclear genes incompatibilities and Mendelian inheritance or chromosomal rearrangements, but asymmetry was shown to be common and taxonomically widespread among plant suggesting that other mechanisms than the cited are responsible (Tiffin *et al.*, 2001). One possible cause explaining asymmetrical reproductive isolation is transposable elements and the maternally inherited factors (Tiffin *et al.*, 2001; Liu & Wendel, 2003; Josefsson *et al.*, 2006; Parisod *et al.*, 2009). Understanding genome incompatibilities and mechanisms resolving conflicts may foster our understandings on the model ‘diverge, merge and diverge’ (Tayalé & Parisod, 2013) and the implications of TEs.

Complex genome and transposable elements

In addition to important genome changes, recent technologies expose the high complexity of plant genomes. In particularly, polyploid genomes show high genome plasticity, multiple polyploid level, important genome size variation. Moreover, significant proportion of repetitive fraction of which transposable elements (TEs) are often a majority (Leitch & Leitch, 2008). First discovered in plants (McClintock, 1984), TEs were documented in most organisms and play a key part of organism evolution leading to differential outcomes that will not be addressed in this thesis (Bennetzen, 2000; Bennetzen & Wang, 2013; Bonchev & Parisod, 2013; Lisch, 2013). TEs may be central actors in genome reorganization (Fedoroff, 2012).

TEs have extreme diversity and were classified in classes, subclasses, orders, superfamilies and families (Figure 1), following Wicker *et al.* (2007). In plant, LTR retrotransposons, mobile genetic elements that transpose via reverse transcription of a RNA intermediate, are ubiquitous (Kumar & Bennetzen, 1999). LTR retrotransposons display often high copy numbers, extensively heterogeneous populations having large TE families and chromosomal dispersion patterns (Kumar & Bennetzen, 1999; Wicker & Buell, 2009).

As comprehensively reviewed in chapter 1, mechanisms controlling TE activity are expected to be affected by hybridization and/or polyploidy. Allopolyploidy is thus postulated as a genome shock leading to TE-related genome reorganization. This is main theme of the following work.

Genus Aegilops: model species

Fast-evolving polyploids, such as species of the Triticeae clade, represent promising model to investigate genome reorganization and the role of TEs in driving genome evolution (Comai *et al.*, 2003; Levy, 2013). Triticeae is composed of different genera, hotspot of allopolyploid speciation, and encompasses wild taxa including genus *Aegilops* as well as important crop species such as wheat, rye and barley. In particular, the genus *Aegilops* shows rich species diversity with multiple ploidy level (di, tetra, hexaploid) that recurrently formed through multiple origins (Meimberg *et al.*, 2009). These species mostly co-occur in the Middle East (van Slageren, 1994; Kilian *et al.*, 2011) and interspecific genetic exchanges were identified by investigating *Aegilops* species' morphological and genome pairing (Feldman & Levy, 2005a; Kilian *et al.*, 2011).

These gene exchanges are probably responsible for the poorly dated evolutionary relationships among *Aegilops* taxa inferred by genetic and cytogenetic analyses (reviewed in Baum *et al.*, 2012). Tetraploid *Aegilops* species derived from a limited number of diploid ancestors (pivotal genomes) combined with several other diploids (differential genomes) (Zohary & Feldman, 1962). Among *Aegilops*, only the D and U genomes are identified as pivotal genomes, respectively inherited from *Ae. tauschii* (DD) and *Ae. umbellulata* (UU), suggesting that a limited number of combinations may produce successful polyploids in nature (Zohary & Feldman, 1962; Levy & Feldman, 2002). Within such polyploid clusters, cytogenetic evidence early showed that pivotal genome (chromosomes shared among polyploids) remain largely unaltered supporting hybrid fertility as compared to their diploid donor, whereas the other differential genome presented considerable modifications (Zohary & Feldman, 1962; Feldman, 1965a; Feldman, 1965b; Feldman, 1965c). Zohary and Feldman (1962) early postulated that this pattern results from gene flow among established

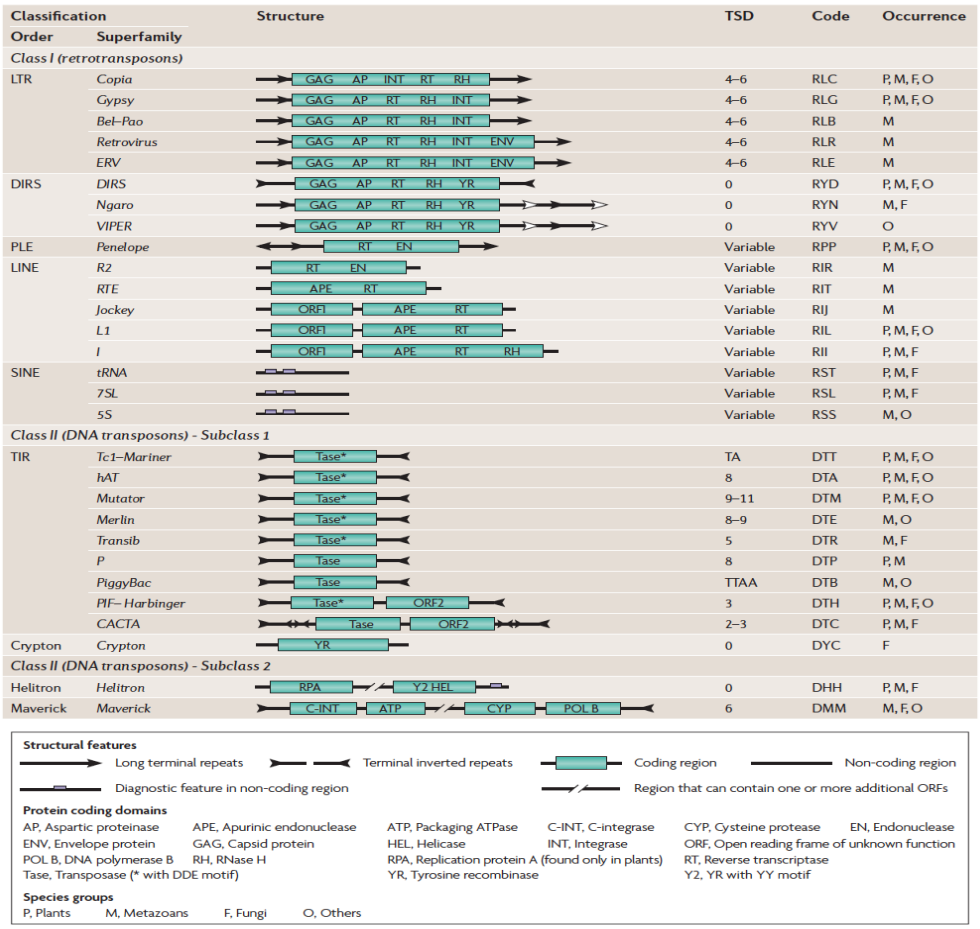


Figure 1 Classification of transposable element proposed by Wicker *et al.* (2007) hierarchized in classes, subclasses, orders and superfamilies. TSD: target site duplication. Figure from Wicker *et al.* (2007).

allopolyploids, with common pivotal genomes providing substrates for homologous recombination and buffering the initial loss of hybrid fertility, while the differential genomes were accumulating restructuring events through homeologous recombination. The stability of pivotal genomes could explain better hybrid fertility, high faculty of gene exchanges and easier homologous chromosome pairing at meiosis (Zohary & Feldman, 1962; Feldman, 1965a; Feldman, 1965b; Feldman, 1965c; Feldman *et al.*, 1997). The overall *Aegilops* genome reorganization following hybridization supports this hypothesis to a certain extent (Kimber *et al.*, 1988; Badaeva *et al.*, 2002; Badaeva *et al.*, 2004b; Bento *et al.*, 2011).

Implications of transposable elements on *Aegilops* genome evolution has been postulated first by the fact that genome composition of the Triticeae represent more than 80% of TEs, specifically LTR retrotransposon (Sabot & Schulman, 2009), second that their large genome size is mainly due to TE transposition (Vitte & Panaud, 2005), third that morphological and chromosomal changes were linked to TE activities (Li *et al.*, 2004; Raskina *et al.*, 2004b; Raskina *et al.*, 2004a), and finally that TE dynamics lead to genome evolution through restructuring and methylation changes among these genomes (Yaakov & Kashkush, 2011). Such reorganization might promote new genome combinations that can, depending on the genome plasticity, lead to new species (Belyayev *et al.*, 2010).

Aegilops species having evolved through hybridization and polyploidy, they represent good models to investigate the responses of TEs and their impact on genome reorganization. This thesis focuses on four tetraploid *Aegilops* species (*Ae. crassa*, *Ae. cylindrica*, *Ae. geniculata*, *Ae. triuncialis*) and their parental species (*Ae. caudata*, *Ae. comosa*, *Ae. tauschii*, *Ae. umbellulata*) (Figure1). Tetraploids have partially overlapping distribution area in Middle East (Figure 2) and all except *Ae. crassa* colonized Europe and USA whereas diploid species showed restricted distribution in the Middle East with limited overlapping. Tetraploid interspecies hybrids were documented only when *Ae. crassa* or *Ae. cylindrica* received pollen from all others and when *Ae. geniculata* received pollen from *Ae. triuncialis* (Kilian *et al.*, 2011). The fact that only limited hybrids between these species were known in nature ask to what extent conflicts were revealed following hybridization and what was the role of TEs in newly formed hybrid genome reorganization affecting fitness. The selected four tetraploid species had different genome composition interconnected each by one common genome. Among the four genome sets, the D and U genomes were designated as pivotal genomes (Figure 2). Badaeva's papers showed fewer genetic modifications following hybridization of

the pivotal genomes than the differential genomes of *Ae. crassa* and *Ae. geniculata*. Moreover, the pivotal and the differential genomes of *Ae. cylindrica* and *Ae. triuncialis* were found to be similarly restructured (Badaeva *et al.*, 2002; Badaeva *et al.*, 2004a). No tetraploid species was assessed among the diploid species having pivotal genomes (Meimberg *et al.*, 2009).

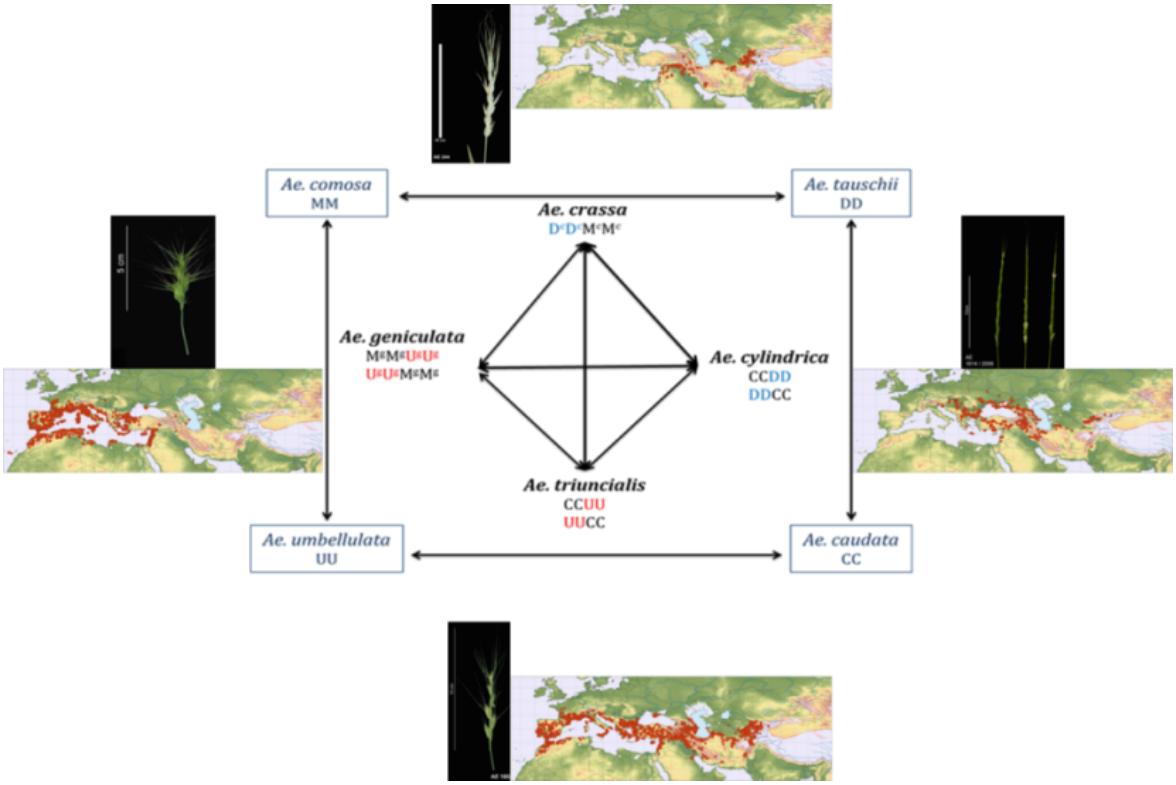


Figure 2 Natural *Aegilops* tetraploids (black) from four diploid species (blue). Black genomes designate differential genomes and colored (blue, red) pivotal genomes. Superscript design modified genome. Ear morphology and distribution maps based on Kilian *et al.* (2011).

Chapters outline and context

The present PhD thesis aims to better understand the role of TEs on genome reorganization following hybridization in selected wild wheats. It is organized in five complementary chapters.

The **chapter 1** is a literature review addressing the impact of TEs and mechanisms of genome reorganization (restructuring and epigenetic repatterning) in polyploids. Reviewed evidence stress on the importance of TEs on short-term and long-term changes in polyploid genomes, and highlight specific mechanisms controlling the activity of TEs and the evolutionary impact of TE-induced genome reorganization.

The **chapter 2** addressed the TEs content of the genomes of *Aegilops cylindrica* (CCDD) and *Ae. geniculata* (MMUU), species that contain all four genomes in the studied species. It developed a procedure inferring the evolutionary dynamics of TE families from low-coverage 454 pyrosequencing data. We identified and quantified different TEs at the family level and assessed the evolutionary dynamics of multiple abundant LTR retrotransposon families. We established that genomes were composed of a majority of LTR retrotransposons, showing high diversity of families with a couple of dominant ones belonging to a common TE pool of the wheat clade. Different evolutionary trajectories of LTR retrotransposons among related species at the family level distinguishing recently active from quiescent TE families were assessed.

The aim of the **chapter 3** was to investigate genome restructuring and to assess the evolutionary trajectories of abundant LTR retrotransposon families following independent allopolyploidy events. We used genome survey sequencing of chapter 2 to select 17 putatively active TEs. Then, we designed molecular fingerprint assays to track genome-wide TE insertions using sequence-specific amplified polymorphism (SSAP) in multiple accessions of species. Amplified fragment length polymorphism (AFLP) revealing genome-wide restructuring in random sequences was used to contrast TE-specific dynamics. Comparisons between progenitors and their respective polyploids showed important restructuring particularly in specific TE families. Furthermore, we identified that TE divergence among progenitors was strongly correlated with the level of restructuring in polyploid TE fractions. TEs rather presented family-specific and species-specific dynamics following polyploidy indicating an extensive impact of TEs on genomes restructuring after polyploidization.

The **chapter 4** addressed factors driving genome reorganization by TEs following hybridization. Using molecular fingerprint assays (AFLP and SSAP) developed in chapter 3 together with fingerprint assays analyzing methylation changes in random sequences (methyl-sensitive amplified polymorphism, MSAP) and genome-wide TE insertions (methyl-sensitive transposons display, MSTD), were evidenced going along with asymmetrical genome reorganization of nine selected TE fractions and random loci differential survival on reciprocal F1 hybrids between *Ae. cylindrica*, *Ae. geniculata* and *Ae. triuncialis*. We showed that divergence of merged TE fractions leads to reorganization following hybridization. We also assessed that new intergenomic interactions lead to repeated loss and methylation of TE sequences, indicating that repression of interspersed TE incompatibilities may be necessary to produce viable hybrids.

Chapter 5 analyzed natural hybrid zones between *Ae. geniculata* and *Ae. triuncialis* and tested potential implication of LTR retrotransposon families on patterns of interspecies gene flow. AFLP and SSAP tracking insertions of five selected active TE families assessed hybridization and genome restructuring. The multiple TE fractions were not differentially exchanged as compared to random sequences. No TE proliferation was identified but sequence losses were important among hybrids, indicating specific TE eliminations following hybridization.

References

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 Systematic and Evolution* **231**: 163-190.

Badaeva ED, Amosova AV, Samatadze TE, Zoshchuk SA, Shotsak NG, Chikida NN, Zelenin AV, Fiebe B, Gill BS. 2004a. Genome differentiation in *Aegilops*. 4. Evolution of the U-genome cluster. *Plant Systematic 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*). *Genome* **55**: 177-193.

Belyayev A, Kalendar R, Brodsky L, Nevo E, Schulman AH, Raskina O. 2010. Transposable elements in a marginal plant population: temporal fluctuations provide new insights into genome evolution of wild diploid wheat. *Mobile DNA* **1**: 2-16.

Bennetzen JL. 2000. Transposable element contributions to plant gene and genome evolution. *Plant Molecular Biology* **42**: 251-269.

Bennetzen JL, Wang H. 2013. The contributions of transposable elements to the structure,function, and evolution of plant genomes. *Annual review of Plant Biology*.

Bento M, Gustafson JP, Viegas W, Silva M. 2011. Size matters in *Triticeae* polyploids: larger genomes have higher remodeling. *Genome* **54**: 175-183.

Bonchev G, Parisod C. 2013. Transposable elements and microevolutionary changes in natural populations. *Molecular Ecology Resources* **13**: 765-775.

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 breadwheat genome using whole-genome shotgun sequencing. *Nature* **491**: 705-710.

Burke JM, Arnold ML. 2001. Genetics and the fitness of hybrids. *Annual Review of Genetics* **35**: 31-52.

Comai L, Madlung A, Josefsson C, Tyagi A. 2003. Do the different parental 'heteromes' cause genomic shock in newly formed allopolyploids? *Philosophical Transactions of the Royal Society B* **358**: 1149-1155.

Doyle J, Flagel L, Paterson A, Rapp R, Soltis D. 2008. Evolutionary genetics of genome merger and doubling in plants. *Annual Review of Genetics* **42**: 443-461.

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 *Aegilop*s and *Triticum*. *Genome* **51**: 616-627.

Eilam T, Anikster Y, Millet E, Manisterski J, Feldman M. 2009. Genome size in natural and synthetic autopolyploids and in natural segmental allopolyploid of several *Triticaceae* species. *Genome* **52**: 275-285.

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

Feldman M. 1965.a. Chromosome pairing between differential genomes in hybrids of tetraploid *Aegilops* species. *Evolution* **19**: 563-568.

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

Feldman M. 1965.c. Fertility of interspecific F1 hybrids and hybrid derivatives involving tetraploid species of *Aegilops* section *Pleionathera*. *Evolution* **19**: 556-562.

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 & Genomics* **36**: 511-518.

Feldman M, Segal B, Abbo Sl, Levy AA, Vega JM. 1997. Rapid elimination of low-copy DNA sequences in polyploid wheat: A possible mechanism for differentiation of homoeologous chromosomes. *Genetics* **147**: 1381-1387.

Galili G, Levy AA, Feldman M. 1986. Gene-dosage compensation of endosperm proteins in hexaploid wheat *Triticum aestivum*. *Proceedings of the National Academy of Sciences **83**: 6524-6528.*

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**: 97-113.

Josefsson C, Dilkes B, Comai L. 2006. Parent-dependent loss of gene silencing during interspecies hybridization. *Current Biology* **16**: 1322-1328.

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

Kejnovsky E, Leitch IJ, Leitch AR. 2009. Contrasting evolutionary dynamics between angiosperm and mammalian genomes. *Trends in Ecology & Evolution* **24**: 572-582.

Kilian B, Mammen K, Millet E, Sharma R, Graner A, Salamini F, Hammer K, O H 2011. *Aegilops*. In: Kole C ed. *Wild Crop Relatives: Genomic and Breeding Resources, Cereals*: Springer-Verlag Berlin Heidelberg, 1-17.

Kimber G, Sallee PJ, M.M. F. 1988. The interspecific and evolutionary relationships in *Triticum ovatum*. *Genome* **30**: 218-221.

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

Leitch AR, Leitch IJ. 2008. Genomic Plasticity and the Diversity of Polyploid Plants. *Science* **320**: 481-483.

Levy AA 2013. Transposons in plant speciation. In: Fedoroff NV ed. *Plant transposons and genome dynamics in evolution*: John Wiley and Sons Ltd, 165-176.

Levy AA, Feldman M. 2002. The impact of polyploidy on grass genome evolution. *Plant Physiology* **130**: 1587-1593.

Li WL, Zhang P, Fellers JP, Friebe B, Gill BS. 2004. Sequence composition, organization, and evolution of the core *Triticeae* genome. *Plant Journal* **40**: 500-511.

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

Lister R, Gregory BD, Ecker JR. 2009. Next is now: new technologies for sequencing of genomes, transcriptomes and beyond. *Current Opinion in Plant Biology* **12**: 107-118.

Liu B, Vega JM, Feldman M. 1998. Rapid genomic changes in newly synthesized amphiploids of *Triticum* and *Aegilops*. II. Changes in low-copy coding DNA sequences. *Genome* **41**: 535-542.

Liu B, Wendel JF. 2003. Epigenetic phenomena and the evolution of plant allopolyploids. *Molecular Phylogenetics and Evolution* **29**: 365-379.

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.

McClintock B. 1984. The significance of responses of the genome to challenge. *Science* **226**: 792-801.

Meimberg H, Rice K, Milan N, Njoku C, McKay J. 2009. Multiple origins promote the ecological amplitude of allopolyploid *Aegilops* (Poaceae). *American Journal of Botany* **96**: 1262-1273.

Murat F, Van de Peer Y, Salse J. 2012. Decoding plant and animal genome plasticity from differential paleo-evolutionary patterns and processes. *Genome Biology and Evolution* **4**: 917-928.

Ozkan H, Levy AA, Feldman M. 2001. Allopolyploidy-induced rapid genome evolution in the wheat (*Aegilops–Triticum*) group. *The Plant Cell* **13**: 1735-1747.

Parisod C, Salmon A, Zerjal T, Tenaillon M, Grandbastien M-A, Ainouche M. 2009. Rapid structural and epigenetic reorganization near transposable elements in hybrid and allopolyploid genomes in *Spartina*. *New Phytologist* **184**: 1003-1015.

Raskina O, Belyayev A, Nevo E. 2004a. Activity of the En/Spm-like transposons in meiosis as a base for chromosome repatterning in a small, isolated, peripheral population of *Aegilops speltoides* Tausch. *Chromosome Research* **12**: 153-161.

Raskina O, Belyayev A, Nevo E. 2004b. Quantum speciation in *Aegilops*: Molecular cytogenetic evidence from rDNA cluster variability in natural populations. *Proceedings of the National Academy of Sciences of the United States of America* **101**: 14818-14823.

Sabot F, Schulman AH 2009. Genomics of transposable elements in the *Triticeae*. In: Muehlbauer GJ, Feuille C eds. *Genetics and genomics of the Triticeae*: Springer US, 387-405.

Soltis PS, Soltis DE. 2000. The role of genetic and genomic attributes in the success of polyploids. *Proceedings of the National Academy of Sciences* **97**: 7051-7057.

Stebbins G 1971. Chromosomal evolution in higher plants. In: Arnold E ed. London.

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

Tiffin P, Olson MS, Moyle LC. 2001. Asymmetrical crossing barriers in angiosperms. *Proceedings of the Royal Society B: Biological Sciences* **268**: 861-867.

van Slageren MW. 1994. Wild wheats: a monograph of *Aegilops* L. and *Amblyopyrum* (Jaub. & Spach) Eig (*Poaceae*). *Wageningen Agricultural University Papers* **97**.

Vitte C, Panaud O. 2005. LTR retrotransposons and flowering plant genome size: emergence of the increase/decrease model. *Cytogenetic and Genome Research* **110**: 91-107.

Walbot, V. and C. A. Cullis. 1985. Rapid genomic change in plants. *Annual Review in Plant Physiology* **36**: 367-396.

Wendel JF. 2000. Genome evolution in polyploids. *Plant Molecular Biology* **42**: 225-249.

Wicker T, Buell CR 2009. Gene and repetitive sequence annotation in the *Triticeae*. In: Muehlbauer GJ, Feuillet C eds. *Genetics and genomics of the Triticeae*: Springer US, 407-425.

Wicker T, Sabot F, Hua Van A, Bennetzen J, 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.

Yaakov B, Kashkush K. 2011. Massive alterations of the methylation patterns around DNA transposons in the first four generations of a newly formed wheat allohexaploid. *Genome* **54**: 42-49.

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

Responses of transposable elements to polyploidy

Christian Parisod[1] and Natacha Senerchia[1]

[1]Laboratory of Evolutionary Botany, Institute of Biology, University of Neuchâtel, Switzerland

This chapter was published as a book chapter in: Plant Transposable Elements (Eds. Grandbastien & Casacuberta), 2013, Topics in Current Genetics, Springer Berlin Heidelberg 24:147-168

Abstract

Polyploidy (i.e. hybridization between more or less divergent genomes, associated with whole genome duplication) has been shown to result in drastic genome reorganization. Such changes involved drastic restructuring and epigenetic repatterning, mainly in transposable element (TE) fractions. Polyploidy thus is an adequate model to explore the mechanisms generating genome variation and their impact on evolution. In this chapter, we will review available evidence on the importance of TEs in the short-term and the long-term changes in polyploid genomes. We will argue that the study of polyploid systems not only offers the opportunity to highlight specific mechanisms controlling the activity of TEs, but also the evolutionary impact of TE-induced genome reorganization.

1 Polyploidy, a prominent evolutionary process

Polyploidy is a recurrent process in the evolutionary history of most organisms and can be understood as a major speciation mechanism (Wood et al. 2009). It is prominent in plants, but also commonly occurs in several animal taxa (Otto 2007; Mable et al. 2011). In particular, all Angiosperms have been demonstrated as having gone through one or more round of whole genome duplication (Jiao et al. 2011) and plant genomes thus contain considerable genetic redundancy (Figure 1). Two main types of polyploids, representing extreme cases of a continuum, have been traditionally recognized (Stebbins 1971). Autopolyploids are polyploids with chromosomes derived from two homologous genomes (AAAA) and are characterized by predominant polysomic inheritance at meiosis (Parisod et al. 2010a). Allopolyploids present chromosomes resulting from the merging of divergent (i.e. homeologous) genomes (AABB) and mostly show disomic inheritance (Leitch and Leitch 2008). The distinction between homologous and homeologous genomes is hardly clear-cut and there is a continuum between auto- and allo-polyploidy. It is thus important to realize that the evolutionary origin of all natural polyploids (i.e. both auto- and allo-polyploids) involves hybridization between variously related genomes.

2 Reorganization of polyploid genomes

Polyploid genomes are expected to be the addition of parental genomes, and departure from this additivity highlights genome reorganization. Recent studies revealed drastic polyploidy-induced genome reorganization, including reproducible structural and epigenetic alteration (Soltis and Soltis 1999; Comai 2000; Wendel 2000; Comai et al. 2000; Levin 2002; Adams and Wendel 2005; Comai 2005; Chen 2007; Doyle et al. 2008; Feldman and Levy 2009).

Such processes restore a secondary diploid-like genetics in polyploids and are commonly referred to as diploidization. Following Levy and Feldman (2004), genome reorganization after polyploidization can be conveniently classified as (i) short-term changes (or revolutionary changes), acting immediately after polyploidization, and (ii) long-term changes (or evolutionary changes), occurring during the lifetime of the polyploid lineage (Figure 1).

Genome reorganization is commonly observed in the first few generations following polyploidy and sometimes as early as in F1 hybrids (Parisod et al. 2009). Both intra- and intergenomic structural rearrangements have been reported and include (i) elimination of DNA sequences from homologous chromosomes and gene loss (Ozkan et al. 2001; Chantret et al. 2005), (ii) amplification or reduction of repetitive sequences (Zhao et al. 1998; Petit et al. 2010) and (iii) chromosomal repatterning (Pires et al. 2004; Udall et al. 2005). Genome downsizing after polyploidization appears to be a general trend (Leitch and Bennett 2004). In addition to restructuring, drastic epigenetic changes have been commonly reported in allopolyploids (Liu and Wendel 2003). These changes include (i) alteration of gene expression through alterations of cytosine methylation (Kashkush et al. 2002; Salmon et al. 2005) and through transcriptional activation of retroelements (Kashkush et al. 2003; Kashkush and Khasdan 2007), and (ii) chromatin remodelling due to modification of DNA methylation and acetylation (e.g. Wang et al. 2006). Polyploidy-induced epigenetic variation is certainly linked to intergenomic interactions and dosage compensation among subgenomes (Riddle and Birchler 2003). Methylation repatterning sometimes affects subgenomes equally (e.g. Song et al. 1995), but most often differentially affects the paternal (e.g. Shaked et al. 2001) or the maternal (e.g. Ainouche et al. 2009). Epigenetic changes were further associated with organ-specific silencing of coding genes in allopolyploids (Adams et al. 2003; Adams and Wendel 2005; Chen 2007). As a whole, diploidization could be a foster for new

phenotypes that could potentially be linked to the evolutionary outcomes of polyploidy (e.g. Levy and Feldman 2004; Doyle et al. 2008; Leitch and Leitch 2008; Parisod 2012). It could indeed be that genome reorganization in nascent polyploids leads to novel properties as compared to the addition of the parental genomes and may support the emergence of new species. Our knowledge on the causes and consequences of polyploidy-induced genome reorganization however remains elusive.

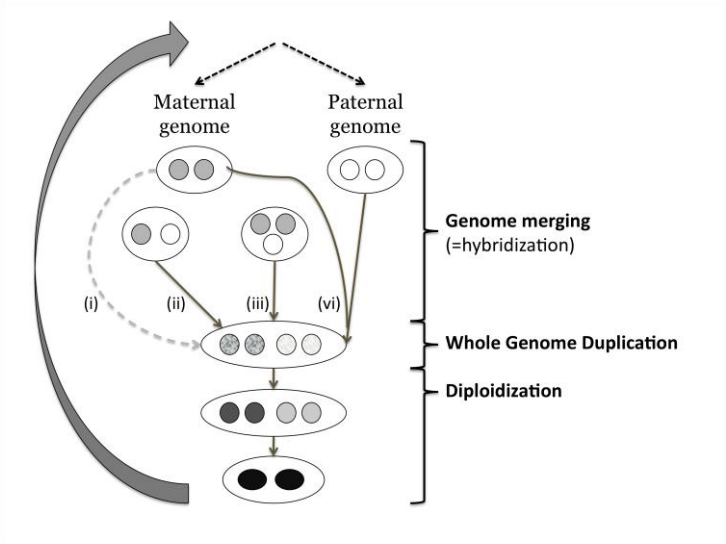


Figure 1 Evolution of natural polyploids. The merging (i.e. hybridization) of more or less diverged parental genomes associated with whole genome duplication leads to the formation of a nascent polyploid lineage. Autopolyploidy involves hybridization between closely related (i.e. homologous) genomes, while allopolyploidy is the merging of widely divergent parental genomes (i.e. homeologous). Genome changes occurring after the origin of the polyploid are referred to as diploidization, restoring a diploid-like genetic system. Seed plant genomes have evolved through successive rounds of polyploidy. The most common natural pathways to polyploidy are depicted: (i) spontaneous genome doubling, which is extremely rare under natural conditions; (ii) homoploid hybrid intermediate; (iii) triploid bridge through the union of an unreduced gamete with a reduced one, and (iv) one-step formation through the union of two unreduced gametes.

The confinement of divergent genomes in the single nucleus of nascent polyploids can induce troubles such as inaccurate pairing between hom(e)ologous sequences or dosage-dependent interactions (Doyle et al. 2008). Accordingly, quick sequence rearrangement (including DNA insertion/deletion) and epigenetic modifications could increase the divergence between

subgenomes, thereby further impeding the pairing of homeologous chromosome and thus indirectly facilitating proper homologous pairing at meiosis (Levy and Feldman 2002; Eilam et al. 2008), or could participate to the regulation of gene dosage, promoting intergenomic coordination (Rieseberg 2001). Reorganization targeted toward one of the parental subgenome is commonly interpreted as evidence that cytoplasmic-nuclear interactions represent crucial incompatibilities to be overcome after genome merging, but it has been noted that nuclear-nuclear interactions may be important as well (Josefsson et al. 2006). Although the exact cause of immediate genome reorganization after polyploidy deserve further work, a greater rate of genome reorganization is expected to be necessary to resolve conflicts in hybrids derived from genetically divergent parents. Accordingly, we can predict more changes to occur in allopolyploids than in autopolyploids. Evidence accumulated so far is coherent with this hypothesis (Parisod et al. 2010a), but we almost completely lack of knowledge about genome reorganization after autopolyploidy. Additional studies involving hybridization between closely related genomes may help to shed light on the mechanisms inducing immediate diploidization.

3 Reorganization of TE genome fractions after polyploidy

For those used to see TEs as major supporters of natural genetic engineering, it might be already clear that the plethora of mechanisms occurring after polyploidy can be related to TEs. In the formulation of the "Genome Shock" hypothesis, Barbara McClintock (1984) stated that challenges such as species cross may induce transposition bursts. This hypothesis, stating that transpositions should play a critical role in polyploidy-induced genome reorganization, has been repeatedly put forward (Matzke and Matzke 1998; Soltis and Soltis

1999; Comai et al. 2000; Wendel 2000). Although data showing an activation of TEs after hybridization and polyploidy have recently accumulated, conclusive evidence is still scarce and we are still far from understanding the mechanisms and the consequences of polyploid genome evolution under the influence of TEs.

Due to their prevalence in eukaryote genomes (Gaut and Ross-Ibarra 2008), it can be expected that TEs play a major role in the molecular events leading to the establishment of a viable polyploid genome. Furthermore, TEs can have a dual role in genome reorganization, affecting both structural features and epigenetic states of sequences throughout the host genome (Teixeira et al. 2009). In case of transposition, new TE insertions can promote proper pairing at meiosis, by triggering structural divergence between subgenomes through microchromosomal rearrangements. Transposition can also promote intergenomic coordination by disrupting genes or altering the epigenetic state of neighboring sequences, impacting on genome function by affecting chromatin structure and/or gene expression (Hollister et al. 2011). On the other hand, dispersed TE insertions might represent homologous substrate sustaining illegitimate recombination and fostering reorganization of TE fractions. Such changes without transposition can have similar consequences for subgenomes divergence and/or coordination.

The commonly anticipated proliferation of TEs in polyploid genomes can be explained by three non-mutually exclusive hypotheses. (i) Whole genome duplication may relax purifying selection against deleterious TE insertions (Matzke and Matzke 1998). In other words, gene redundancy may lead to an overall increase in the number of neutral sites available for TEs to insert and fix without strong selective constraints (the Redundancy hypothesis). Accordingly, under a constant transposition rate, TE insertions would accumulate neutrally in polyploids until such sites are all occupied. (ii) The origin of a polyploid lineage represents a transient

period with a low population size (i.e. bottleneck; Lynch 2007). As selection efficiency decreases when population size decreases, moderately deleterious TE insertions could be fixed in nascent polyploid genomes with a higher probability (the Bottleneck hypothesis; Parisod et al. 2010b). Accordingly, under a constant transposition rate, TE insertions would accumulate in nascent polyploids until the establishment of a large population. (iii) The merging of divergent genomes into a single nucleus would generate conflicts between the TEs and the host repressors (Box A; Figure 2 and 3), inducing a genome shock promoting TE activation and ultimately transposition (Genome Shock hypothesis; Comai et al. 2003). Accordingly, polyploidy would induce a change in the activity of TEs.

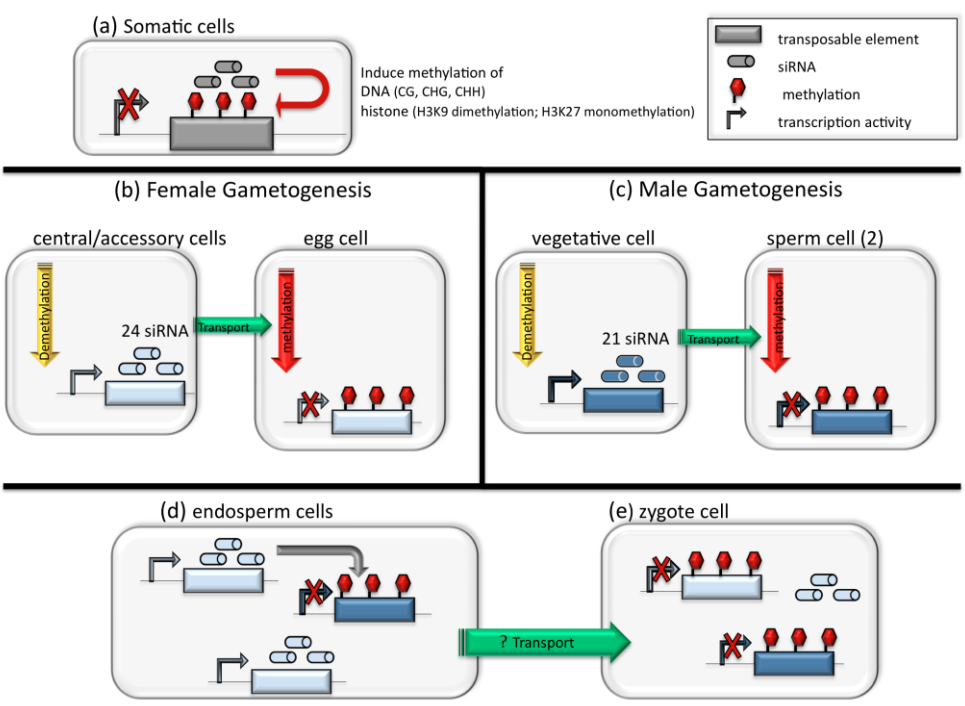


Figure 2 Transposable element (TE) silencing by siRNA during plant development and reproduction (after Feng et al. 2010; Bourc'his and Voinnet 2010). In somatic cells (a), siRNA derived from TEs recruit the methylation machinery in order to maintain the repression of TE transcription through methylated DNA or histones. During both female (b) and male (c) gametogenesis, 24-nucleotide-long and 21- nucleotide-long siRNAs are produced by the demethylated genomes of the central/accessory cells and from the vegetative cell, respectively. Those siRNA maintain or reinforce TE repression in the egg and sperm cells. During fecundation (d), the endosperm is demethylated and further produce siRNAs. Putative transport of siRNAs from the endosperm to the zygote might help to sustain TE methylation in the zygote. See Box A for details.

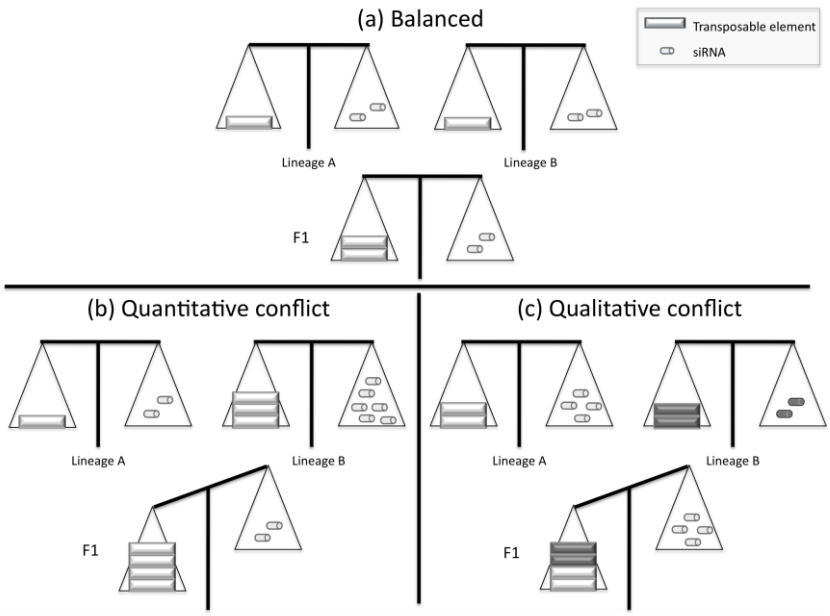


Figure 3 Conflicts between parental loads in transposable elements (TEs) during genome merging. (a) Balanced situation: parental TEs and siRNAs match, allowing an efficient control of TEs in F1. (b) Quantitative or (c) qualitative differences in TE loads between parents, potentially leading to insufficient or inefficient repression of TEs in F1 hybrids (modified from Bourchis and Voinnet 2010).

Mechanisms behind these three hypotheses are expected to result in different patterns of TE proliferation and may thus be distinguished by assessing TE activity, rate of accumulation during and after polyploidization and the parental genome divergence. Under the Redundancy hypothesis, no discrete burst of TE activity is expected and the rate of TE accumulation should be continuous until full diploidization is reached. A *bona fide* change in TE activity (transcriptional and, to a certain extent, transpositional) is postulated immediately after polyploidy under the Genome Shock hypothesis. Accordingly, both the Genome Shock and the Bottleneck hypothesis are expected to result in the accumulation of transposed TE copies during the first generations after polyploidy. However, genome merging should reveal genetic conflicts between specific TE families (see Box A) and only these TEs should be affected under the Genome Shock hypothesis, while a bottleneck would change the frequency of all polymorphic TE insertions. Noticeably, the Redundancy and the Bottleneck hypothesis could

explain TE dynamics in both auto- and allo-polyploids, while a genome shock is expected to result in reorganization of fewer TEs in hybrids between closely related genomes (i.e. autopolyploids) than in allopolyploids. While theory can help to predict the impact of polyploidy on TE activity, empirical data are still too scarce to test the different hypotheses. Accordingly, what follows remains a narrative review of the levels and timing of reorganization in TE genome fractions of polyploids.

4 Short-term reorganization of TE fractions

Short-term genome reorganization related to TEs can be straightforwardly evaluated by comparing the genome of experimental (i.e. resynthetized) or recent (i.e. less than a few hundred years old) polyploids to the expected addition of their parents (Figure 4). Several PCR-based fingerprint techniques can be used to assess reorganization throughout the genomes of both autopolyploids and allopolyploids (Parisod et al. 2010b; Kalendar et al. 2011). As different molecular methods allow focusing on either genome restructuring or methylation changes in TE fractions vs. random sequences (Box B), it is possible to assess the reorganization of TE genome fractions as compared to genome-wide changes (Table 1).

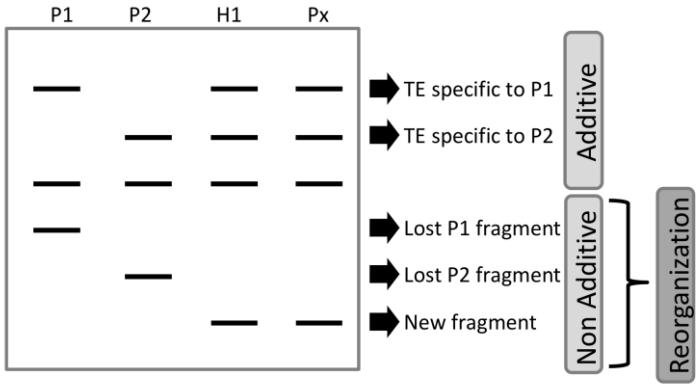


Figure 4 Principle of fingerprint analyses in polyploids. Genetic profiles in the hybrid (H1) and the polyploid (Px) are expected to be the addition of the parents (P1 and P2). Deviations from this additivity indicate genome reorganization in contrasted genome fractions. See Box B for detailed explanations.

Table 1 Summary of reviewed studies reporting evidence on the reorganization of the transposable element (TE) genome fraction after polyploidy.

Model species	TEs	Restruct-uring[1]	Trans-cription	Methy-lation changes[2]	References
Short term reorganization					
Arabidopsis thaliana x *A. lyrata* (F1/S0, S1, S2)	CAC, Ac-III	0 / -	.	+	Beaulieu et al. 2009
Arabidopsis thaliana x *A. arenosa* (F4)	En-Sp like	. / .	+	+	Madlung et al. 2005
Arabidopsis thaliana x *A. arenosa* (F1/F7)	Various TEs	. / .	.	+ (si)	Ha et al. 2009
Spartina alterniflora x *S. maritima* (F1)	Ins2, Cassandra, Wis-like	+ / --	.	++	Parisod et al. 2009
Nicotiana sylvestris x *N. tomentosiformis* (F1/S0)	Tnt1	0 / 0	.	.	Petit et al. 2010
Nicotiana sylvestris x *N. tomentosiformis* (S4)	Tnt1	++ / --	.	.	Petit et al. 2010
Aegilops sharonensis x *Triticum monoccucum* (F1/S1)	Wis2-1A	0 / .	+	+	Kashkush et al. 2002
Triticum turgidum x *Aegilops tauschi* (F1/S0)	Retrotransposons, CACTA	0 / 0	.	.	Mestiri et al. 2010
Triticum turgidum x *Aegilops tauschi* (S1-S4)	Balduin, Apollo and Thalos	+ / --	++	++	Yaakov et al. 2011
Triticum turgidum x *Aegilops tauschi* (S1-S5)	Veju	+ / -	+	+	Kraitshtein et al. 2010
Triticum turgidum x *Aegilops tauschi* (F1)	Veju and Wis2-1A	. / .	+	+ (si)	Kenan et al. 2010
Long term reorganization					
Arabidopsis arenosa	Ac-like	+ / -	.	.	Hazzouri et al. 2008
Arabidopsis suecica	Ac-like	0 / 0	.	.	Hazzouri et al. 2008
Nicotiana tabacum	Tnt1, Tnt2 and Tto1	+ / -	.	.	Petit et al. 2007
Nicotiana tabacum	Gypsy elements	. / -	.	.	Renny-Byfiled et al. 2011
Brassica napus, *B.carinata*, *B.juncea*	Retrotransposons	0 / .	.	.	Alix et al. 2004
Brassica napus	MITE BraSto	+ / 0	.	.	Sarilar et al. 2011
Brassica napus	CACTA Bot1	0 / .	.	.	Alix et al. 2008
Gossypium hirsutum	VariousTEs	+ / --	.	.	Grover et al. 2008
Gossypium hirsutum	Retrotransposon	0 / .	.	.	Hu et al. 2010
Gossypium hirsutum	Line	+ / .	.	.	Hu et al. 2010
Oryza ssp.	VariousTEs	+ / -	.	.	Lu et al. 2009
Triticum aestivum	Fatima	+ / .	.		Salina et al. 2011
Triticum aestivum	Various TEs	+ / -	.	.	Charles et al. 2008
Triticum aestivum	Various TEs	. / .	.	. (si)	Cantu et al. 2010
Triticum aestivum	Athila-like, Gypsy and Copia elements	0 / .	.	+	Bento et al. 2008
Sacharum ssp.	Various TEs	. / .	+	.	Garsmeur et al. 2010
Zea mays	CRM1	. / -	+	.	Sharma et al. 2008
Zea mays	Various TEs	. / -	.	.	Schnable et al. 2011
Arachis monticola	AhMITE1	. / -	+	.	Gowda et al. 2011
Coffea arabica	Copia elements	+ / .	.	.	Yu et al. 2011

[1]transposition / loss of TE sequences: 0, no evidence; + evidence of transposition (++, > 10%); - evidence of sequence loss in TE fractions (--, > 10%); ., not evaluated

[2](si) accounts for changes in siRNAs of the corresponding TEs

Box A: Dynamics of TE-repressing mechanisms during polyploidy

As a majority of transposition events are expected to have a deleterious effect, host genomes have evolved sophisticated mechanisms repressing the activity of functional TEs (Figure 2a). Recent studies have considerably improved our understanding of the various epigenetic pathways controlling TEs, but much remains to be done in order to decipher these overlapping mechanisms (Feng et al. 2010). Two main mechanisms are responsible for silencing of TEs: DNA methylation (in CG, CHG and CHH sequences contexts; H = C, T or A) and histone methylation (H3K9 dimethylation and H3K27 monomethylation). These pathways are triggered by repeat-derived small interfering RNA (siRNA) that target TE insertions through sequence homology and recruit the enzyme machinery responsible for DNA methylation and heterochromatinization (Martienssen 2010). Genomes typically contain specific TE sequences inducing the production of specific siRNA silencing corresponding TEs and thus assure genome stability during plant development.

While plants do not show a proper demethylated germ line, it seems that both female (Figure 2b) and male (Figure 2c) gametogenesis relaxes the repression of TEs in accessory cells, ensuring the massive production of siRNAs and reinforcing the silencing of TEs in the germ cells (i.e. consolidation; Bourc'his and Voinnet 2010). During male gametogenesis, post meiosis microspores develop into a vegetative cell and two sperm cells. Epigenetic pathways responsible for the maintenance of methylation are down-regulated and TEs are reactivated in pollen grains. It seems however that hypomethylation is exclusive to the vegetative cell and would serve the production of 21 nucleotide siRNAs mediating the repression of TEs in the adjacent sperm cells through CHG methylation. During female gametogenesis, post meiosis megaspore gives rise to one egg cell (participating to the zygote), one central cell with two nuclei (participating the endosperm) and other accessory cells. The genome of the central cell is specifically demethylated, leading to the expression of maternal alleles and TEs in the endosperm (i.e. imprinting; Figure 2d). Such soft reactivation of TEs in the central cell and the endosperm may serve the massive production of 24 nucleotides siRNAs to reinforce the silencing of TEs in the egg cell, and maybe the endosperm and the zygote.

The confrontation of paternal and maternal genomes (Figure 2e) presenting qualitative and/or quantitative mismatch in their respective TEs and siRNAs may result in the failure of the TE-siRNA system to reach equilibrium at fertilization (Figure 3). In other words, hybridization between lineages with incompatible TE loads is expected to results in conflicts between TEs and siRNAs. If siRNAs in the central cell do not match TE insertions in pollen, then corresponding TEs could be transcribed and could possibly transpose in the endosperm. Note that a massive proliferation of TEs in the endosperm is expected to have deleterious consequences such as seed failure. Similarly, if siRNAs in the egg cytoplasm do not match with TE insertions from the sperm cells, corresponding TEs could be activated and may proliferate in the zygote. The outcome of a cross thus depends on both the copy number of TEs and of siRNAs, but also on the dose of paternal and maternal genomes. Such a reactivation of TEs in F1 hybrids is similar to hybrid dysgenesis as described in *Drosophila* and may lead to strong incompatibility between gene pools (i.e. intrinsic postzygotic isolation; Josefsson et al. 2006; Martienssen 2010; Parisod et al. 2010b).

Recent hybrids are rare in nature and/or difficult to identify, and most studies used experimentally resynthetized hybrids. Massive reorganization in TE genome fractions has been documented during the first generations after polyploidization. In a few cases, the impact of genome merging (i.e. hybridization) vs. genome doubling has been experimentally contrasted and hybridization seems to induce most genomic changes (reviewed in Parisod et al. 2010b), but additional studies on autopolyploids are required before reaching conclusions.

Box B: Molecular fingerprint techniques to assess genomes reorganization in non-model polyploid species

Amplified Fragment Length Polymorphism (AFLP) is a high-resolution fingerprint technique generating markers following the digestion of genomic DNA with restriction enzymes, the ligation of adaptors and PCR amplifications of fragments. The resulting markers are dominant and anonymous, but are widely distributed throughout the genome (Meudt and Clarke 2007) and thus assess genome-wide variation in random sequences. Sequence-Specific Amplified Polymorphism (SSAP) is similar to AFLP, except that it is a TE-anchored PCR strategy (i.e. Transposon Display) allowing the simultaneous detection of multiple insertions (Waugh et al. 1997; Syed and Flavell 2006). Briefly, the amplification of digested genomic DNA, specifically targeting TEs insertions, generates a pool of labeled fragments containing the termini of inserted copies of a given TE and its flanking genomic region. As retrotransposons do not excise, particular insights concerning the molecular mechanisms underlying SSAP polymorphisms can be gathered: new bands are indicative of new TE insertions (i.e. transpositions), while lost bands point to restructuring in TE sequences (comprehensively described in Parisod et al. (2010b). Note that new SSAP bands should be cautiously interpreted as new transposition events, because they can result from other molecular events changing the band size of inserted TEs (Petit et al. 2010). As a whole, the comparison of AFLP versus SSAP profiles generated from the same individuals assesses the respective variation in random sequences versus specific TE fractions.

Methyl-sensitive derivative of multilocus fingerprint techniques can be exploited by using restriction enzymes with differential sensitivity to DNA methylation on the same samples. Methyl-sensitive AFLP is named Methyl-Sensitive Amplified Polymorphism (MSAP), while methyl-sensitive SSAP has been termed Methyl-Sensitive Transposon Display (MSTD; Parisod et al. 2009). The isoschizomers MspI and HpaII recognize the same tetranucleotide sequence (5'-CCGG-3'), but HpaII is sensitive to methylation of any cytosine at both strands (i.e. cuts 5'-CCGG-3'), while MspI cuts methylated internal cytosine (i.e. cuts 5'-C5mCGG-3'). These enzymes thus assess the methylation status of internal cytosine at restriction sites (CG methylated fractions of the genome). As a whole, comparing MSAP versus MSTD profiles, respectively, can assess CG methylation changes in random sequences versus TE fractions.

4.1 Structural changes in TE fractions

Beaulieu et al. (2009) analyzed genome reorganization in synthetic allotetraploids between *Arabidopsis thaliana* x *A. lyrata* subsp. *petrea*, and identified substantial restructuring. Changes assessed through various fingerprint techniques were mostly sequence deletions and no burst was revealed for the two DNA transposons surveyed (CAC and *Ac-III*). Another study on resynthesized *A. suecica* allopolyploids (*A. thaliana* x *A. arenosa*) used genomic microarray and fingerprint techniques to examine a region of the chromosome 4 (Madlung et al. 2005). This work highlighted transcriptional activation of En-Spm-like transposon in the allopolyploids and also identified chromosome abnormalities, suggesting possible polyploidy-induced restructuring at specific loci. These events may be related, but the exact role of TEs remains unknown. Similarly, the LTR retrotransposon WIS2-A was transcriptionally activated in experimental polyploids between *Aegilops sharonensis* and *Triticum monococcum* (Kashkush et al. 2002). However, new TE transcripts apparently did not increase the transposition rate. Accordingly, experimental F1 hexaploids of wheat were shown to be the addition of parental *T. turgidum* and *Ae. tauschii* at hundreds of loci (Mestiri et al. 2010). As many markers were targeting specific TE insertions, this work further indicates limited restructuring in TE fractions. In the 150 years old *Spartina* allopolyploids, Parisod et al. (2009) found limited evidence of immediate TE proliferation, with very few new SSAP bands revealed for Ins2 (hAT DNA transposon), Cassandra (terminal-repeat retrotransposon in miniature) and Wis-like (*copia* LTR retrotransposon) as compared to the addition of the parents. Moreover, the level of structural changes in TE fractions was comparable to random sequences, indicating no specific restructuring of TE fractions after genome merging or genome doubling. Noticeably, most structural changes occurred in F1 hybrids, suggesting that genome merging is inducing genome reorganization.

Contrasting with studies indicating limited transposition, young populations of the Tnt1 retrotransposon showed a transposition burst in early generations of synthetic allopolyploid tobacco (Petit et al. 2010). While newly synthesized polyploids were the addition of the parents, new insertion sites were detected at the fourth generation. Although the causes of Tnt1 transposition remain unclear, this work suggests that polyploidy may induce transposition of specific TEs in some cases.

While systematic and immediate transposition bursts seem to occur in specific polyploids only, the study of polyploidy-induced restructuring of TE genome fractions highlighted sequence elimination to a large extent. Studies on *Triticaceae* species (Feldman et al. 1997; Ozkan et al. 2001) showed that synthesized allopolyploids between *Triticum* and *Aegilops* have rapidly eliminated high-copy, low-copy, coding and non-coding DNA sequences. Allopolyploidy in *Spartina* was associated with a predominant loss of bands, principally from maternal origin, suggesting DNA elimination within or including TE insertions (Parisod et al. 2009). Petit et al. (2010) identified losses and indels around insertions of paternal Tnt1 sites in synthesized allotetraploid tobacco.

As a whole, the study of different polyploid systems revealed no evidence of immediate and systematic TE bursts after polyploidy, but suggest that TE genome fraction are affected by elimination of DNA sequences in the first generations after allopolyploidization.

4.2 Epigenetic modification in TE genome fraction

In synthetic allotetraploids (*Arabidopsis thaliana* x *A. lyrata* subsp. *petrea),* methylation changes at 25% of the genome-wide loci surveyed was assessed by MSAP (Beaulieu et al. 2009). Another study on newly synthetized allotetraploids *Arabidopsis suecica* identified that TE activation was correlated with sequence demethylation, but this was not associated with significantly higher rate of transposition (Madlung et al. 2005). Comparing reorganization of CG methylation in the whole genome vs. TE genome fractions, Parisod et al. (2009) revealed that most methylation changes occurred in the TE fraction of recent *Spartina* polyploid. The investigation of methylation changes around insertion sites of three DNA transposons (*Balduin*, *Apollo*, and *Thalos*) during the first four generations of newly formed allohexaploid wheats revealed that 54% of the sites have undergone CG methylation changes (Yaakov and Kashkush 2011). Noticeably, these epigenetic modifications were hypermethylation to a large extent and occurred mainly during the first two generations. Recently, study on newly formed wheat allohexaploids demonstrated substantial methylation changes around the TRIM *Veju* during the first four generations. Interestingly, hypomethylation was predominant in the first generation and quickly followed by hypermethylation (Kraitshtein et al. 2010). The study of 3072 transcripts in wheat allotetraploids (genome SSAA: *Aegilops sharonensis* (SS) x *Triticum monococcum* ssp. *aegilopoides* (AA)) showed that 12 transcripts, including retrotransposons, were activated at early stage after polyploidization probably in correlation with methylation changes (Kashkush et al. 2002). Such activation of TEs was shown to influence the expression of adjacent genes through methylation changes (Kashkush et al. 2003).

Twenty-four-nucleotide-long small interfering RNAs (siRNA) maintain DNA methylation and are enriched in and around TEs, suggesting that they play a major role in controlling transposition (Slotkin and Martienssen 2007; Teixeira et al. 2009; Bourc'his and Voinnet 2010). Comparisons of F1 and F7 generations of synthetic allotetraploids *Arabidopsis suecica* with the two parental diploids *A. thaliana* and *A. arenosa* showed that methylation changes were associated with variation in siRNAs (Ha et al. 2009). The expression of siRNA in the hybrids deviated from the additivity of the parents and presented drastic changes during the first generation (F1) before stabilizing in latter generations (F7). Accordingly, siRNA produced during interspecific hybridization seem to support a greater stability of the allopolyploid genome and may "serve as a buffer against the genome shock". Correspondingly, in a synthetic hexaploid wheat, the massive sequencing of siRNA revealed that the proportion of siRNA related to TEs decreased in allopolyploids compared to the parental lines or F1 hybrids, suggesting that TE regulation was destabilized in polyploids (Kenan Eichler et al. 2011). Detailed investigations of two *Copia*-like LTR retrotransposons (Veju and Wis2-1A) indicated that their transcription rate was higher in the polyploids, but no formal link was established between the levels of siRNA and transcription.

As a whole, polyploidy induces considerable reshuffling of epigenetic marks, mainly in TE fractions. This may change TE dynamics, but the formal link between these processes remains to be clarified. As genetic and epigenetic variation sit on top of each other, it is crucial to further understand the fuelling role of TEs on restructuring and epigenetic repatterning across the genome. Polyploidy seem to induce the transcriptional activation of specific TEs (although not necessarily transposition) and may help to shed light on the mechanisms underlying the control of such elements.

5 Long term restructuring in TE genome fraction

Long term genome reorganization underlying evolutionary changes during the species lifespan includes mutations, exchanges of chromosome sections, evolution of TE families in subgenomes and introgression between polyploids (Comai 2005; Doyle et al. 2008; Leitch and Leitch 2008; Feldman and Levy 2009). The properties of polyploid genomes as compared to diploids are not fully clear yet, but it seems that genetic redundancy might allow higher accumulation of mutations, which may be recruited by adaptive processes to improve the success of polyploids in nature (Feldman and Levy 2005; Otto 2007; Parisod et al. 2010a). Our knowledge of the causes and consequences of polyploid genomes evolution over thousands of years is still limited because it is experimentally impossible to reproduce and thus can only be indirectly analyzed (Table 1). Genome changes are indeed investigated by comparing established polyploids to extent diploids and, since both diploids and polyploids may have evolved since the polyploidy event, it remains hard to distinguish between changes due to allopolyploidy and those that occurred during the polyploid species lifespan. As the turnover of TE insertions is relatively high (Vitte and Panaud 2005), the study of TE dynamics in millions-year old polyploids is challenging.

Several studies on the polyploid wheats (*Triticum durum*; genome BA) and *T. aestivum*; genome BAD) investigated the TEs by sequencing large genomic regions and identified waves of TE insertions proliferation at different time and in different genomes. A detailed survey of parts of the chromosome 3B of hexaploid wheat highlighted more than 3000 TEs that evolved through several waves of transposition within the last four million years (Choulet et al. 2010). While fluorescent in situ hybridization revealed that the retrotransposon Fatima contributed to B-genome specific patterns (Salina et al. 2011), Charles et al. (2008) used BAC sequencing and assessed that 90% of the divergence between the A and B subgenomes was

due to restructuring of TE fractions. However, the inferred timing of transposition for *Athila*-like, other *Gypsy* and *Copia* retrotransposons was not matching the polyploidy events, indicating that their proliferation was related to the divergence of parental genomes before merging more than to genome merging. While significant transposition seems to rarely occur in polyploid wheats, evidence from transposon displays (Bento et al. 2008) and from the comparison of the *hardness* locus (Chantret et al. 2005) in diploid, tetraploid and hexaploid wheat species showed major rearrangements in repetitive fractions of polyploid genomes. TE insertions were indeed often truncated and/or presented large indels in the polyploids, suggesting that TEs sustain unequal or illegitimate recombination in response to polyploidy. Moreover, a recent study investigating the dynamic of siRNAs in natural hexaploids wheat confirmed their important role in repressing TE activation through methylation in the short-term, but also noticed an increased mutation rate in heavily methylated TEs (Cantu et al. 2010). Interestingly, this suggests that short-term repression might turn into a long-term mechanism of TE inactivation and genome evolution.

The sequencing of partial reverse transcriptase from six diploid and related allotetraploids of *Brassica* showed that most *copia* and *gypsy* sequences are shared by all species (Alix and Heslop Harrison 2004). No evidence of specific amplification in polyploids was revealed based on sequence similarity. More recently, Alix et al. (2008) provided evidence for several waves of amplification of a specific *CACTA* transposon (BOT1) in the diploid *Brassica oleracea* as compared to the allopolyploid *Brassica napa*. Accordingly, the transposition of BOT1 was responsible for the divergence between diploid species but no recent transposition activation was assessed in polyploids. While the BraSto MITE apparently amplified in the two parental genomes (*B. rapa* and *B. oleracea)* and their allotetraploid (*B. napus)*, no specific burst at allopolyploidization was inferred (Sarilar et al. 2011). Based on the

sequencing of reverse transcriptase in the allopolyploid cotton (*Gossypium hirsutum)* and its parental diploids (*Gossypium arboreum* and *G. raimondii*), different activity of *Copia*, *Gypsy*-like *Gorge3* long terminal repeat (LTR) retrotransposons and long interspersed nuclear elements (LINEs) was highlighted (Hu et al. 2010). While various proliferation periods were identified for the different TEs in the different species, bursts were apparently TE specific and hardly related to polyploidy. The comparison of sequences around the cellulose synthase locus (Grover et al. 2004) and the alcohol dehydrogenate locus (Grover et al. 2007) in the diploid progenitors and tetraploid cottons revealed a similar rate of TE activity, but a higher turnover in the polyploid TE fraction (Grover et al. 2008). Small deletions in TEs were indeed found to be extremely frequent in the polyploid, underlying genome contraction as compared to diploids. Corresponding conclusions were reached by comparing the MONOCULM1 region in diploids and tetraploids *Oryza* species (Lu et al. 2009). While different TE amplified in divergent species and were associated with different genome size, polyploid TE fractions were characterized by sequence elimination and, mostly, TE truncation.

A few studies provided circumstantial evidence of significant TE proliferation in polyploid genomes. BAC sequencing in diploid progenitors and allopolyploid coffeas (*Coffea caneophora, C. eugenioides* and the polyploid *C. arabica*) revealed differential transposition of specific TEs in the polyploid (Yu et al. 2011). In particular, a recent proliferation of *copia* retrotransposon was highlighted in *C. arabica* and participated to size variation of the corresponding subgenome as compared to its diploid state. Similarly, confronting hom(eo)ologous sequences of modern sugarcane (*Saccharum* spp.), breakdown of colinearity was specifically observed in the TE fraction, suggesting a dynamic of expansion of TEs (Garsmeur et al. 2011). Focusing on evolutionary dynamic of several *copia* retrotransposons (Tnt1, Tnt2 and Tto1) in allotetraploids *Nicotiana tabacum* and its two parental species (*N.*

sylvestris and *N. tomentosiformis*) with SSAP, Petit et al. (2007) inferred considerable turnover in TEs sequences, including several new bands suggestive of transposition as well as sequence loss. Recently, Renny-Byfield et al. (2011) used low-coverage 454 sequencing to investigate the dynamics of transposable elements in *N. tabacum* and the its progenitors. The high degree of similarity between *gypsy* sequences indicated a potential TE expansion in *N. sylvestris*, but not in *N. tomentosiformis* or in the allopolyploid *N. tabacum.* The characterization of a large number of TE insertions in a single analysis strongly suggests the observed pattern to be explained by TE expansion in *N. sylvestris* after the polyploidization, but cannot entirely rule out massive TE deletions in polyploids. Associated with rigorous statistical treatment still to be developed, new sequencing techniques will offer decisive insights on the impact of TEs on long-term polyploid genome evolution, because they enable the investigation of whole genome reorganization.

Some of the difficulties inherent to the inference of long-term evolutionary processes can be circumvented by population approaches surveying genome diversity and interpreting patterns within a reliable population genetics framework. Little work adopted this promising method. Investigation of a stress-inducible MITE (AhMITE1) transposon in polyploid peanuts showed that a specific insertion at the FST-1 locus was segregating within the allopolyploid lineages (Gowda et al. 2011). As the AhMITE1 insertion was absent from the primitive allopolyploids (*Arachis monticola*), but present in derived *Arachis hypogaea*, this may suggest TE activation after polyploidy. Hazzouri et al. (2008) compared the distribution of insertions of Ac-like transposon in populations of the allopolyploid *A. suecica* and of the autopolyploid *A. arenosa*. In stark contrast with expectations raised under the hypothesis of a polyploidy-induced burst of transposition, the allopolyploids had mostly fixed insertions (i.e. non-polymorphic and mainly inherited from the parents). Autopolyploids showed significant segregation of

polymorphic insertions, indicating that some TEs recently transposed and were not removed by selection. A similar approach was used in the 4.5 million years old polyploids of the monophyletic *Nicotiana* section *Repandae* and highlighted considerable restructuring in TE fractions (Parisod, Grandbastien et al. unpublished results; Lim et al. 2007). Although the exact timing of restructuring events was hardly assessed, most new and lost SSAP bands were shared by all polyploid species, suggesting that substantial genome changes occurred shortly after the polyploidy event. Noticeably, the different TEs showed contrasted segregation patterns in the different polyploid species, indicating that long-term genome turnover may depend on intrinsic properties of TE populations, but also on constraints imposed by host populations.

As a whole, insights on the impact of polyploidy on TE and on the long-term genome evolution of polyploid genomes remain hardly conclusive. Available evidence seems to suggest that polyploidy *per se* did seldom influence transposition rate on the long-term. However, the evolution of TE after genome merging has not been extensively addressed yet. Interestingly, Sharma et al. (2008) noticed recombination between centromeric TE family (CRM1) from the two parental subgenomes of maize and suggested that such novel recombinant TE might proliferate in relation to polyploidy. Although massive TE proliferation long after the polyploidy event seems not to be the rule, TE fractions show considerable restructuring and apparently foster genome evolution in the long-term. Polyploid TE insertions indeed reveal indels and truncation to a large extent, suggesting that TEs represent opportune substrate for recombination to actively shape genome architecture (Devos et al. 2002).

6 Conclusion

Polyploidy is a major evolutionary process leading to massive restructuring events and/or epigenetic modifications throughout the genome. Evidence is accumulating that TEs play a central role in fuelling such genome reorganization (Table 1). In contrast to a common belief, recent studies on several polyploids systems indicate that polyploidy-induced transposition bursts are far from being a general rule. Only few studies assessed an important burst of transposition from young and specific TE families (Parisod et al. 2010b). Available evidence however indicates that restructuring events associated with polyploidy are more frequent in TE genome fractions than in random sequences, but predominantly involve DNA sequence deletion rather than transposition. It suggests TE-specific mechanisms, but untargeted DNA lesions affecting the predominant fraction of genomes (i.e. TEs) cannot be ruled out.

Available data suggest that genome reorganization generally occurs in the first generations following the polyploidy event and involves epigenetic changes in the vicinity of TEs to a large extent. Such evidence matches the expectations of the Genome Shock hypothesis and suggests that hybridization reveals TE-specific incompatibilities. Genome merging is indeed prone to alter the balance between TEs and siRNAs and such conflict might thus induce the activation of TEs during polyploidy (Box A). It should however be noted that a massive transpositional activation of TEs could be strongly deleterious to the nascent hybrid genome. Accordingly, it is tempting to speculate that only polyploids having controlled transposition through substantial repatterning of epigenetic marks and/or having lost TE fragments could be viable.

7 Perspectives

Despite a growing number of examples illustrating the central role of transposable elements during genome evolution, many crucial issues remain unanswered. We are indeed far from understanding the molecular mechanisms or the evolutionary forces underlying genome reorganization. The race between host genomes and highly mutagenic TEs deserves additional work (Blumenstiel 2011) and polyploidy seem to represent a convenient process to further explore the mechanisms activating and repressing TEs in both the short and the long term.

Future studies shall address whether the necessary genome changes related to TEs could turn beneficial by improving the viability and fertility of the nascent polyploid genome. Although some cases of adaptive evolution through TE insertion have been assessed (Bennetzen 2005), the frequency of beneficial vs. neutral vs. deleterious insertions is still largely unknown. As polyploids often see the expression pattern of duplicated genes modified, such system may help to assess to what extent TEs trigger phenotypic evolution through non-functionalization, sub-functionalization or neo-functionalization (Walsh 2003). Moreover, nascent polyploids have to establish populations and form reproductively isolated lineages to persist in nature. Accordingly, it remains to be assessed to what extent (TE-induced) genome reorganization sustains ecological shifts associated with polyploid speciation (Parisod 2012).

Most studies reviewed here relied on allopolyploid species originating from the merging of widely divergent genomes. Accordingly, further comparison of autopolyploids vs. allopolyploids could be fruitful in order to better understand the impact of genome merging vs. genome doubling on the control of TEs and the evolutionary forces acting on the resulting variation. Furthermore, conflicts between subgenomes as put forward here to explain TE-induced reorganization after polyploidy is a process occurring at the fundamental level of the

genome, while evolutionary forces such as selection or genetic drift act at the level of populations. Accordingly, the Genome Shock, the Redundancy and the Bottleneck hypotheses are not mutually exclusive. Future work addressing the causes and consequences of TE activation on (polyploid) genome evolution shall integrate this full hierarchy (Tenaillon et al. 2010).

Acknowledgments

This work was funded by the National Centre of Competence in Research 'Plant Survival' and a grant (PZ00P3-131950 to CP), both from the Swiss National Science Foundation.

References

Adams KL, Cronn R, Percifield R and Wendel JF (2003) Genes duplicated by polyploidy show unequal contributions to the transcriptome and organ-specific reciprocal silencing. Proc Natl Acad Sci USA 100: 4649-4654.

Adams KL and Wendel JF (2005) Novel patterns of gene expression in polyploid plants. Tr Genet 21: 539-543.

Ainouche ML, Fortune PM, Salmon A and Parisod C (2009) Hybridization, polyploidy and invasion: lessons from *Spartina* (*Poaceae*). Biol Inv 11: 1159-1173.

Alix K and Heslop Harrison JS (2004) The diversity of retroelements in diploid and allotetraploid *Brassica* species. Plant Mol Biol 54: 895-909.

Alix K, Joets J, Ryder C, Moore J and Barker G (2008) The CACTA transposon Bot1 played a major role in *Brassica* genome divergence and gene proliferation. Plant J 56: 1030-1044.

Beaulieu J, Jean M and Belzile F (2009) The allotetraploid *Arabidopsis thaliana-Arabidopsis lyrata* subsp *petraea* as an alternative model system for the study of polyploidy in plants. Mol Genet Genom 281: 421-435.

Bennetzen JL (2005) Transposable elements, gene creation and genome rearrangement in flowering plants. Curr Op Genet Dev 15: 621-627.

Bento M, Pereira HS, Rocheta M, Gustafson P and Viegas W (2008) Polyploidization as a retraction force in plant genome evolution: sequence rearrangements in Triticale. PLoS ONE 3.

Blumenstiel J (2011) Evolutionary dynamics of transposable elements in a small RNA world. Tr Genet 27: 23-31.

Bourc'his D and Voinnet O (2010) A Small-RNA perspective on gametogenesis, fertilization, and early zygotic development. Science 330: 617-622.

Cantu D, Vanzetti L, Sumner A, Dubcovsky M and Matvienko M (2010) Small RNAs, DNA methylation and transposable elements in wheat. BMC Genomics 11.

Chantret N, Salse J, Sabot F, Rahman S and Bellec A (2005) Molecular basis of evolutionary events that shaped the hardness locus in diploid and polyploid wheat species (*Triticum* and *Aegilops*). Plant Cell 17: 1033-1045.

Charles M, Belcram H, Just J, Huneau C and Viollet A (2008) Dynamics and differential proliferation of transposable elements during the evolution of the B and A genomes of wheat. Genetics 180: 1071-1086.

Chen ZJ (2007) Genetic and epigenetic mechanisms for gene expression and phenotypic variation in plant polyploids. Annu Rev Plant Biol 58: 377-406.

Choulet F, Wicker T, Rustenholz C, Paux E and Salse J (2010) Megabase level sequencing reveals contrasted organization and evolution patterns of the wheat gene and transposable element spaces. Plant Cell 22: 1686-1701.

Comai L (2000) Genetic and epigenetic interactions in allopolyploid plants. Plant Mol Biol 43: 387-399.

Comai L, Madlung A, Josefsson C, Tyagu A (2003) Do the different parental 'heteronomes' cause genomic shock in newly formed allopolyploids? Phil Trans Roy Soc B 358: 1149-1155.

Comai L (2005) The advantages and disadvantages of being polyploid. Nature Rev Genet 6: 836-846.

Comai L, Tyagi AP, Winter K, Holmes Davis R and Reynolds SH (2000) Phenotypic instability and rapid gene silencing in newly formed *Arabidopsis* allotetraploids. Plant Cell 12: 1551-1567.

Devos KM, Brown JKM and Bennetzen JL (2002) Genome size reduction through illegitimate recombination counteracts genome expansion in *Arabidopsis*. Genome Res 12: 1075-1079.

Doyle J, Flagel L, Paterson A, Rapp R and Soltis D (2008) Evolutionary genetics of genome merger and doubling in plants. Annu Rev Genet 42: 443-461.

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

Feldman M and Levy A (2009) Genome evolution in allopolyploid wheat - a revolutionary reprogramming followed by gradual changes. J Genet Genom 36: 511-518.

Feldman M and Levy AA (2005) Allopolyploidy - a shaping force in the evolution of wheat genomes. Cytogen Genome Res 109: 250-258.

Feldman M, Liu B, Segal G, Abbo S and Levy AA (1997) Rapid elimination of low-copy DNA sequences in polyploid wheat: A possible mechanism for differentiation of homoeologous chromosomes. Genetics 147: 1381-1387.

Feng SH, Jacobsen SE and Reik W (2010) Epigenetic Reprogramming in Plant and Animal Development. Science 330: 622-627.

Garsmeur O, Charron C, Bocs S, Jouffe V, Samain S, Couloux A, Droc G, Zini C, Glaszmann JC, Van Sluys MA and D'Hont A (2011) High homologous gene conservation despite extreme autopolyploid redundancy in sugarcane. New Phytol 189: 629-642.

Gaut BS and Ross-Ibarra J (2008) Selection on major components of angiosperm genomes. Science 320: 484-486.

Gowda MVC, Bhat RS, Sujay V and Kusuma P (2011) Characterization of AhMITE1 transposition and its association with the mutational and evolutionary origin of botanical types in peanut (*Arachis* spp.). Plant Syst Evol 291: 153-158.

Grover C, Kim H, Wing R, Paterson A and Wendel J (2007) Microcolinearity and genome evolution in the AdhA region of diploid and polyploid cotton (*Gossypium*). Plant J 50: 995-1006.

Grover C, Yu Y, Wing R, Paterson A and Wendel J (2008) A phylogenetic analysis of indel dynamics in the cotton genus. Mol Biol Evol 25: 1415-1428.

Grover CE, Kim HR, Wing RA, Paterson AH and Wendel JF (2004) Incongruent patterns of local and global genome size evolution in cotton. Genome Res 14: 1474-1482.

Ha M, Lu J, Tian L, Ramachandran V and Kasschau K (2009) Small RNAs serve as a genetic buffer against genomic shock in *Arabidopsis* interspecific hybrids and allopolyploids. Proc Natl Acad Sci USA 106: 17835-17840.

Hazzouri R, Mohajer A, Dejak S, Otto S and Wright S (2008) Contrasting patterns of transposable-element insertion polymorphism and nucleotide diversity in autotetraploid and allotetraploid *Arabidopsis* species. Genetics 179: 581-592.

Hollister J, Smith L, Guo Y-L, Ott F and Weigel D (2011) Transposable elements and small RNAs contribute to gene expression divergence between *Arabidopsis thaliana* and *Arabidopsis lyrata*. Proc Natl Acad Sci USA 108: 2322-2327.

Hu G, Hawkins J, Grover C and Wendel J (2010) The history and disposition of transposable elements in polyploid *Gossypium*. Genome 53: 599-607.

Jiao YN, Wickett NJ, Ayyampalayam S, Chanderbali AS, Landherr L, Ralph PE, Tomsho LP, Hu Y, Liang HY, Soltis PS, Soltis DE, Clifton SW, Schlarbaum SE, Schuster SC, Ma H, Leebens-Mack J and dePamphilis CW (2011) Ancestral polyploidy in seed plants and angiosperms. Nature 473: 97-100.

Josefsson C, Dilkes B and Comai L (2006) Parent-dependent loss of gene silencing during interspecies hybridization. Curr Biol 16: 1322-1328.

Kalendar R, Flavell AJ, Ellis THN, Sjakste T, Moisy C and Schulman AH (2011) Analysis of plant diversity with retrotransposon-based molecular markers. Heredity 106: 520-530.

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

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

Kashkush K and Khasdan V (2007) Large-scale survey of cytosine methylation of retrotransposons and the impact of readout transcription from long terminal repeats on expression of adjacent rice genes. Genetics 177: 1975-1985.

Kenan Eichler M, Leshkowitz D, Tal L, Noor E and Melamed Bessudo C (2011) Wheat hybridization and polyploidization results in deregulation of small RNAs. Genetics 188: 263-279.

Kraitshtein Z, Yaakov B, Khasdan V and Kashkush K (2010) Genetic and epigenetic dynamics of a retrotransposon after allopolyploidization of wheat. Genetics 186: 801-U889.

Leitch AR and Leitch IJ (2008) Genomic plasticity and the diversity of polyploid plants. Science 320: 481-483.

Leitch IJ and Bennett MD (2004) Genome downsizing in polyploid plants. Biol J Linn Soc 82: 651-663.

Levin DA (2002) The role of chromosomal change in plant evolution. Oxford University Press, New York.

Levy AA and Feldman M (2002) The impact of polyploidy on grass genome evolution. Plant Physiol 130: 1587-1593.

Levy AA and Feldman M (2004) Genetic and epigenetic reprogramming of the wheat genome upon allopolyploidization. Biol J Linn Soc 82: 607-613.

Lim KY, Kovarik A, Matyasek R, Chase M and Clarkson J (2007) Sequence of events leading to near-complete genome turnover in allopolyploid *Nicotiana* within five million years. New Phytol 175: 756-763.

Liu B and Wendel JF (2003) Epigenetic phenomena and the evolution of plant allopolyploids. Mol Phylogen Evol 29: 365-379.

Lu F, Sanyal A, Zhang S and Song R (2009) Comparative sequence analysis of MONOCULM1-orthologous regions in 14 *Oryza* genomes. Proc Natl Acad Sci USA 106: 2071-2076.

Lynch M (2007) The origins of genome architecture. Sinauer Associates, Sunderland.

Mable BK, Alexandrou MA and Taylor MI (2011) Genome duplication in amphibians and fish: an extended synthesis. J Zool 284: 151-182.

Madlung A, Tyagi AP, Watson B, Jiang HM and Kagochi T (2005) Genomic changes in synthetic *Arabidopsis* polyploids. Plant J 41: 221-230.

Martienssen RA (2010) Heterochromatin, small RNA and post-fertilization dysgenesis in allopolyploid and interploid hybrids of *Arabidopsis*. New Phytol 186: 46-53.

Matzke MA and Matzke AJM (1998) Polyploidy and transposons. Tr Ecol Evol 13: 241-241.

McClintock B (1984) The significance of responses of the genome to challenge. Science 226: 792-801.

Mestiri I, Chague V, Tanguy A-M, Huneau C and Huteau V (2010) Newly synthesized wheat allohexaploids display progenitor-dependent meiotic stability and aneuploidy but structural genomic additivity. New Phytol 186: 86-101.

Meudt HM and Clarke AC (2007) Almost forgotten or latest practice? AFLP applications, analyses and advances. Tr Pl Sci 12: 106-117.

Otto SP (2007) The evolutionary consequences of polyploidy. Cell 131: 452-462.

Ozkan H, Levy AA and Feldman M (2001) Allopolyploidy-induced rapid genome evolution in the wheat (*Aegilops-Triticum*) group. Plant Cell 13: 1735-1747.

Parisod C (2012) Polyploids integrate genomic changes and ecological shifts. New Phytol 193: 297-300.

Parisod C, Alix K, Just J, Petit M and Sarilar V (2010b) Impact of transposable elements on the organization and function of allopolyploid genomes. New Phytol 186: 37-45.

Parisod C, Holderegger R and Brochmann C (2010b) Evolutionary consequences of autopolyploidy. New Phytol 186: 5-17.

Parisod C, Salmon A, Zerjal T, Tenaillon M, Grandbastien MA and Ainouche M (2009) Rapid structural and epigenetic reorganization near transposable elements in hybrid and allopolyploid genomes in *Spartina*. New Phytol 184: 1003-1015.

Petit M, Guidat C, Daniel J, Denis E and Montoriol E (2010) Mobilization of retrotransposons in synthetic allotetraploid tobacco. New Phytol 186: 135-147.

Petit M, Lim KY, Julio E, Poncet C and de Borne F (2007) Differential impact of retrotransposon populations on the genome of allotetraploid tobacco (*Nicotiana tabacum*). Mol Genet Genom 278: 1-15.

Pires JC, Zhao JW, Schranz ME, Leon EJ and Quijada PA (2004) Flowering time divergence and genomic rearrangements in resynthesized *Brassica* polyploids (*Brassicaceae*). Biol J Linn Soc 82: 675-688.

Renny-Byfield S, Chester M, Kovarik A, LeComber AC, Grandbastien MA, Deloger M, Nichols RA, Macas J, Novak P, Chase MW and Leitch AR (2011) Next generation sequencing reveals genome downsizing in allotetraploid Nicotiana tabacum, predominantly through the elimination of paternally derived repetitive DNAs. Mol Biol Evol 28: 2843-2853.

Riddle NC and Birchler JA (2003) Effects of reunited diverged regulatory hierarchies in allopolyploids and species hybrids. Tr Genet 19: 597-600.

Rieseberg LH (2001) Polyploid evolution: Keeping the peace at genomic reunions. Curr Biol 11: R925-R928.

Salina EA, Sergeeva EM, Adonina IG, Shcherban AB, Belcram H, Huneau C and Chalhoub B (2011) The impact of Ty3-gypsy group LTR retrotransposons Fatima on B-genome specificity of polyploid wheats. BMC Plant Biol 11: 99.

Salmon A, Ainouche ML and Wendel JF (2005) Genetic and epigenetic consequences of recent hybridization and polyploidy in *Spartina* (*Poaceae*). Mol Ecol 14: 1163-1175.

Sarilar V, Marmagne A, Brabant P, Joets J and Alix K (2011) BraSto, a Stowaway MITE from *Brassica*: recently active copies preferentially accumulate in the gene space. Plant Mol Biol 77: 59-75.

Shaked H, Kashkush K, Ozkan H, Feldman M and Levy AA (2001) Sequence elimination and cytosine methylation are rapid and reproducible responses of the genome to wide hybridization and allopolyploidy in wheat. Plant Cell 13: 1749-1759.

Sharma A, Schneider K and Presting G (2008) Sustained retrotransposition is mediated by nucleotide deletions and interelement recombinations. Proc Natl Acad Sci USA 105: 15470-15474.

Slotkin RK and Martienssen R (2007) Transposable elements and the epigenetic regulation of the genome. Nature Rev Genet 8: 272-285.

Soltis DE and Soltis PS (1999) Polyploidy: recurrent formation and genome evolution. Tr Ecol Evol 14: 348-352.

Song KM, Lu P, Tang KL and Osborn TC (1995) Rapid genome change in synthetic polyploids of *Brassica* and its implications for polyploid evolution. Proc Natl Acad Sci USA 92: 7719-7723.

Stebbins GL (1971) Chromosomal evolution in higher plants. Edward Arnold, London.

Syed NH and Flavell AJ (2006) Sequence-specific amplification polymorphisms (SSAPs): a multi-locus approach for analyzing transposon insertions. Nature Protocols 1: 2746-2752.

Teixeira FK, Heredia F, Sarazin A, Roudier F, Boccara M, Ciaudo C, Cruaud C, Poulain J, Berdasco M, Fraga MF, Voinnet O, Wincker P, Esteller M and Colot V (2009) A Role for RNAi in the Selective Correction of DNA Methylation Defects. Science 323: 1600-1604.

Tenaillon MI, Hollister JD and Gaut BS (2010) A triptych of the evolution of plant transposable elements. Tr Pl Sci 15: 471-478.

Udall JA, Quijada PA and Osborn TC (2005) Detection of chromosomal rearrangements derived from homeologous recombination in four mapping populations of *Brassica napus* L. Genetics 169: 967-979.

Vitte C and Panaud O (2005) LTR retrotransposons and flowering plant genome size: emergence of the increase/decrease model. Cytogen Genome Res 110: 91-107.

Walsh B (2003) Population-genetic models of the fates of duplicate genes. Genetica 118: 279-294.

Wang J, Tian L, Lee H-S and Chen ZJ (2006) Nonadditive regulation of FRI and FLC loci mediates flowering-time variation in *Arabidopsis* allopolyploids. Genetics 173: 965-974.

Waugh R, McLean K, Flavell AJ, Pearce SR, Kumar A, Thomas BBT and Powell W (1997) Genetic distribution of *Bare*-1-like retrotransposable elements in the barley genome revealed by sequence-specific amplification polymorphisms (S-SAP). Mol Gen Gen 253: 687-694.

Wendel JF (2000) Genome evolution in polyploids. Plant Mol Biol 42: 225-249.

Wood TE, Takebayashi N, Barker MS, Mayrose I, Greenspoon PB and Rieseberg LH (2009) The frequency of polyploid speciation in vascular plants. Proc Natl Acad Sci USA 106: 13875-13879.

Yaakov B and Kashkush K (2011) Massive alterations of the methylation patterns around DNA transposons in the first four generations of a newly formed wheat allohexaploid. Genome 54: 42-49.

Yu Q, Guyot R, de Kochko A, Byers A and Navajas Perez R (2011) Micro-collinearity and genome evolution in the vicinity of an ethylene receptor gene of cultivated diploid and allotetraploid coffee species (*Coffea*). Plant J 67: 305-317.

Zhao XP, Si Y, Hanson RE, Crane CF and Price HJ (1998) Dispersed repetitive DNA has spread to new genomes since polyploid formation in cotton. PCR Meth Appl 8: 479-492.

Evolutionary dynamics of retrotransposons assessed by high throughput sequencing in wild relatives of wheat

Natacha Senerchia[1], Thomas Wicker[2], François Felber[1,3], Christian Parisod[1]

[1]Laboratory of Evolutionary Botany, Institute of Biology, University of Neuchâtel, Switzerland

[2]Institute of Plant Biology, University of Zurich, Switzerland

[3]Musée et Jardins botaniques cantonaux, Lausanne, Switzerland

This chapter was published in Genome Biology and Evolution (2013) 5 (5): 1010-1020.

Abstract

Transposable elements (TEs) represent a major fraction of plant genomes and drive their evolution. An improved understanding of genome evolution requires the dynamics of a large number of TE families to be considered. We put forward an approach bypassing the required step of a complete reference genome to assess the evolutionary trajectories of high copy number TE families from genome snapshot with high throughput sequencing. Low coverage sequencing of the complex genomes of *Aegilops cylindrica* and *Ae. geniculata* using 454 identified more than 70% of the sequences as known TEs, mainly long terminal repeat retrotransposons. Comparing the abundance of reads as well as patterns of sequence diversity and divergence within and among genomes assessed the dynamics of 44 major LTR retrotransposon families of the 165 identified. In particular, molecular population genetics on individual TE copies distinguished recently active from quiescent families and highlighted different evolutionary trajectories of retrotransposons among related species. This work presents a suite of tools suitable for current sequencing data, allowing to address the genome-wide evolutionary dynamics of TEs at the family level and advancing our understanding of the evolution of non-model genomes.

Introduction

The sequencing of multiple plant genomes has dramatically improved our understanding of the impact of whole genome duplication and transposable elements (TEs) on the organization of angiosperm genomes (Kejnovsky, et al. 2009). Notably, single or low-copy sequences (e.g. functional genes) often comprise a modest fraction of genomes, whereas repetitive sequences form a major component (Hua-Van, et al. 2011). In particular, interspersed TEs are typically found at multiple copies in intergenic regions and form the most dynamic fraction of genomes (Gaut and Ross Ibarra 2008). The episodic activation of such intragenomic parasites as well as mutations associated with inserted copies plays a pivotal role in fuelling host genome reorganization and, ultimately, biological diversification (Feschotte, et al. 2002; Kazazian 2004). TEs have the ability to affect genome structure and function through transposition, illegitimate recombination and epigenetic repatterning (Bennetzen 2005; Fedoroff 2012; Parisod, et al. 2009; Slotkin and Martienssen 2007). To what extent TEs shall be considered "junk DNA" as compared to functional sequences under direct selection remains a matter of debate and their impact on genome evolution deserve further attention (Biemont and Vieira 2006).

TEs represent a very diverse community of sequences that fundamentally differ in their mechanism of transposition: Class I retrotransposons move via "copy and paste" mechanisms using RNA intermediates, whereas Class II DNA transposons move via "cut and paste" mechanisms through DNA intermediates (Wicker, et al. 2007). Lower levels of the TE classification are based on the number of DNA strands transferred from the original to the insertion site (i.e. subclasses), insertions mechanisms (i.e. orders) and DNA sequence similarity (i.e. families). Individual insertions (i.e. particular TE copies at specific chromosomal locations) represent the lowest level of this hierarchy (Le Rouzic, et al. 2007).

Related TE copies from a given family inhabiting the host genome are considered here as a population and such TE populations can show dramatically different evolutionary trajectories in different species (e.g. Parisod, et al. 2012). As the various TE families within a genome may show distinct trajectories (e.g. Baucom, et al. 2009; Choulet, et al. 2010), the dynamics of several TE families should be simultaneously addressed to further understand the evolution of this major genome fraction through comparative approaches (Brookfield 2005a).

Long terminal repeat (LTR) retrotransposons represent the predominant order of TEs in plants (Kumar and Bennetzen 1999). The life cycle of LTR retrotransposons involves the reverse-transcription of a RNA intermediate from a mother copy and such TEs can potentially amplify within the host genome by inserting several daughter copies at multiple sites (Sabot and Schulman 2006). The balance between genome expansion through TE proliferation and contraction through deletion of TE sequences drives variation in genome size and organization (Bennetzen and Kellogg 1997; Vitte and Panaud 2005; Sabot and Schulman 2006; Tenaillon, et al. 2010). Accordingly, the replicative proliferation of a single, active copy of LTR retrotransposon generates a population of closely related sequences, with most copies sharing high genetic similarity within a genome (Casacuberta, et al. 1997). In contrast, TE populations originating from older proliferation events are genetically heterogeneous due to accumulation of mutations. Most copies of a given TE family detected from genomic sequences are indeed defective, presenting premature stop-codons, indels or further rearrangements such as truncation or nested insertions (SanMiguel, et al. 1996). Populations of genetically heterogeneous copies are predicted for inactive TE families and patterns of sequence diversity assessed from large amounts of TE sequences distinguish recently active from quiescent TE lineages (Brookfield 2005b).

The central features of TEs, abundance and dynamics, result in challenges for comparative genomics (Treangen and Salzberg 2012). In particular, repetitive TEs create ambiguities in alignments and assemblies of short sequences (e.g. data typically produced from so called next generation sequencing), which prejudice meaningful interpretations and hinders large-scale sequencing of complex genomes (Berkman, et al. 2012). TEs are central to genome evolution and shall not be simply ignored, but most studies investigating TE genome fractions have had to focus on selected genome regions (Choulet, et al. 2010; Li, et al. 2004; Sabot, et al. 2005; SanMiguel, et al. 2002). Accordingly, only a subset of the significant TE diversity is potentially identified and the dynamics of TEs may be intermingled with the evolutionary history of the loci under scrutiny. Efficient approaches, assessing TE dynamics at a genome-wide scale without having to rely on complete reference genomes, are thus currently required.

Here, we put forward a procedure inferring the genome-wide evolutionary dynamics of TE families from low-coverage sequencing data, thus bypassing the required step of a fully assembled genome. We used the *Triticeae* clade of the grass family (i.e. the genus *Aegilops* that is closely related to cultivated wheat), because these species harbor complex genomes containing 80% of well annotated TEs, with a majority of LTR retrotransposons (Li, et al. 2004; Sabot and Schulman 2009; Wicker and Buell 2009). We produced genome snapshots of *Aegilops cylindrica* (genome DC, 9398Mb per 1C) and *Aegilops geniculata* (genome UM, 10074Mb per 1C) by 454 pyrosequencing with the aims to (i) identify and quantify the proportion of the genome occupied by the different TE families at the genome-wide scale and (ii) assess the evolutionary dynamics of the major LTR retrotransposons in species having different genome content, but close sizes (Eilam, et al. 2008; van Slageren 1994). We show that the genomes of wild wheats are composed of a majority of LTR retrotransposons from more than 160 families with a couple of dominant ones. In addition to differences in

abundance among host genomes, several TE families presented genetically divergent but homogeneous populations of individual copies, indicating species-specific proliferation. This approach thus offers appropriate tools to distinguish recently active from quiescent TE families and address the processes underlying the evolution of TE genome fractions from patterns of genetic diversity among copies. With the advent of short reads sequencing, it will advance the genomics of model and non-model species.

Materials and Methods

Plant material

Accessions of *Aegilops cylindrica* (TA 2204 = AE 719) collected in Armenia and of *Ae. geniculata* (TA 1800) collected in Turkey (Kirkareli) were obtained from the Wheat Genetics Resource Center (Kansas State University, USA). These accessions have been previously characterized at the cytogenetic level, assessing chromosomal rearrangement, amplification and deletion of repetitive sequences (Badaeva, et al. 2004; Linc, et al. 1999). Plants were grown under controlled conditions (18°C, 18h light) and leaves were collected on two weeks old seedlings. Genomic DNA was extracted from fresh leaves following a standard cetyltrimethylammonium bromide (CTAB) protocol (Chen and Ronald 1999).

The exact phylogeny of *Aegilops* remain poorly resolved and dated (Baum, et al. 2012), but diploid *Aegilops* and *Triticum* species apparently diverged within the last 2.5-4.5 MY (Huang, et al. 2002), whereas *Triticum* diverged from barley 11MY ago (Bossolini, et al. 2007), from *Brachypodium* ca. 35 MY ago and from rice and maize some 60 MY ago (Wolfe, et al. 1989).

Whole genome snapshot and sequence classification

Approximately 40ng of genomic DNA of one individual per accession was mechanically shotguned and random fragments were sequenced on half a plate of the Roche 454 GS FLX titanium platform (service provided by Microsynth, Switzerland, following manufacturer's instructions). The quality of reads was checked with PRINSEQ v0.20.1 (Schmieder and Edwards, 2011) and revealed normally distributed GC content (ranging from 30 to 70%) and Phred quality score (25 to 40). Less than 6% of the reads presented ambiguous bases occurring over less than 1% of the sequence. Accordingly, no further trimming of the 454 reads than the standard quality trim offered by sffTools was required. Reads showing strong bias of bases distribution compared to random expectations (i.e. classified as SSR using personal scripts described in Wicker, et al. 2009) were removed. Identical sequences starting within two nucleotides of one another were further identified with 454 Replicate Filter (Gomez-Alvarez, et al. 2009) and represented a negligible 1% of the reads that may be artificially replicated fragments during sequencing.

To determine the sequence composition, 454 reads (SSR excluded) were subjected to BLAST analyses against various databases and consecutively classified as repetitive element sequences (using complete TREP: wheat.pw.usda.gov/ITMI/Repeats/), organelle and coding sequences (using TIGR rice genome version 5: rice.plantbiology.msu.edu and RAP-DB: rapdb.dna.affrc.go.jp), using personal scripts described in Wicker *et al.* (2009) (Figure 1). To identify TEs, complete TREP was preferred, as this curated database includes known copies rather than consensuses of 584 TE families mostly from barley and cultivated wheat. Complete TREP included thus variation within TE families from species related to those investigated here. BLASTN searches identified hits showing 80% similarity with sequences from databases and selected hits with e-values < 10E-6. As BLASTN retrieved numerous hits,

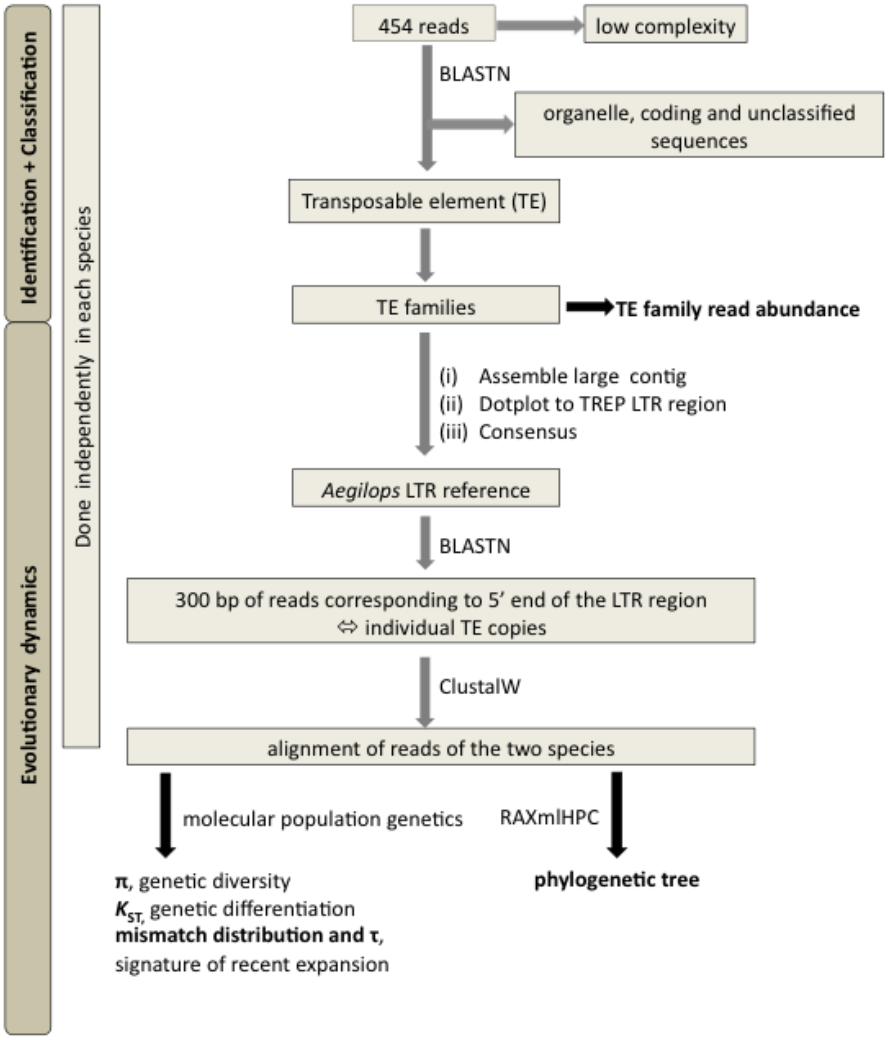


Figure 1 Overview of the approach used here to identify and classify 454 reads into families of transposable elements (TEs), and then investigate their evolutionary dynamics through molecular population genetics.

BLASTX search was not required. Classification of TE families into classes, orders, superfamilies and families was consistent with Wicker *et al.* (2007). Noticeably, rare or *a priori* unknown families may be overlooked with this approach, but it could be nicely complemented by clustering (e.g. Novák, et al. 2010) or assisted automated assembly (DeBarry, et al. 2008), allowing to potentially describe new TE families.

Abundance of the different TE families among the two species with close genome size was estimated as proportions of reads matching the different TE family out of the total number of reads. In order to assess possible sampling effects on estimated proportions, vectors with assignments of reads to TE families were resampled 999 times with replacement to estimate the distribution of proportions using the sample function in R Cran. In addition, 95% confidence intervals were further assessed under binomial distribution theory, using Wald's method for large number with the binCI function in the package binGroup in R Cran. Both estimates closely correlate, confirming that resampling estimates agree with the full multinomial distribution of all TE families considered together. Accordingly, resampling 95% confidence intervals around the mean proportions presented here allowed to identify TE family with non-overlapping intervals and thus with significantly different abundances.

Evolutionary genetics of TE populations

Reads corresponding to TE copies were retrieved and analyzed through a pipeline summarized in Figure 1. For each TE family, we have reconstructed LTR regions of our *Aegilops* species instead of using LTR regions from TREP. Accordingly, 454 reads that initially matched against entries from TREP were assembled in large contigs with phredPhrap (www.phrap.org/) and contigs corresponding to TREP LTR regions were identified using DOTTER in LINUX. Majority rule consensuses were edited to make *Aegilops* LTR references.

Portions of reads corresponding to the first 300bp at the 5'end of the LTR region and thus corresponding to individual TE copies for each family were identified by BLASTN against *Aegilops* LTR references. Given the large genome of *Aegilops* and the low coverage of our sequencing, the probability of retrieving the same TE copy among the reads is negligible. Sequences flanking the 5'end of LTR region were checked and were mostly different from

one another (flanking sequences that were nearly identical corresponded to the 3' end of the internal region) (data not shown). Accordingly, analyzed reads corresponding to a short portion (here, 300 bp) of the TE can be assumed to represent individual TE copies. Our 454 reads had an average length of 380bp. We selected portions of 300bp, because large reads guarantee better resolution for subsequent analyses. Individual reads from each species were then merged and aligned using ClustalW. Alignments were manually edited if necessary. Using own *Aegilops* LTR references instead of LTR regions available in databases minimized potential bias due to divergence between TEs from *Aegilops* and those described in TREP (i.e. mainly from *Triticum*) and allowed collecting a larger sample of reads.

To assess LTR retrotransposon families showing evidence of recent proliferation within genomes, patterns of genetic diversity among independent 454 reads corresponding to the 5' end of the LTR region were investigated for each TE family. The 5' end of the LTR region is a variable and diagnostic portion of retrotransposons that includes both autonomous and non-autonomous elements, but other regions such as the reverse transcriptase domain may also offer suitable resolution. Assuming that reads represent independent copies and considering the genomes of the different accessions as distinct TE populations, we used molecular population genetics approaches to estimate significant parameters highlighting the evolutionary dynamics of TE families. Our method is based on the identification of groups of TE copies sharing higher genetic similarity than otherwise observed among copies of quiescent TE families that accumulated genetic differences to highlight recent proliferation events.

Phylogenetic inferences

Phylogenetic relationships among copies of each TE families were investigated by inferring an unrooted maximum likelihood trees with high performance computing using RAXmlHPC (GTR model with gamma-distributed rate across sites; branch support assessed by 100 bootstraps; Stamatakis 2006). No specific gap coding was done because indels were mostly 1bp long.

Genetic differentiation between TE populations

Genetic differentiation among individual copies of the TE populations (i.e. host genomes) was assessed by the fixation index (K_{ST}), representing the proportion of genetic diversity that is observed between populations out of the total diversity (Holsinger and Weir 2009). A low K_{ST} indicates that the TE populations mostly share similar TE copies, while higher K_{ST} means greater differentiation between TE copies inserted in the different genomes. K_{ST} was estimated with the Tamura and Nei distance between sequences using Arlequin version 3.5 (Excoffier, et al. 2005). Whether K_{ST} was significantly different from zero was tested by permuting haplotypes between populations 100 times, giving the null distribution of pairwise K_{ST} values under the hypothesis of no difference between populations. The proportion of permutations leading to a K_{ST} value larger or equal to the observed one represents the P-value. In addition, the nucleotide diversity (π) among the sequences in the two species was evaluated using DnaSP v5 (Librado and Rozas 2009).

Molecular signature of TE expansion

Within TE families, the presence of groups of similar TE copies indicative of recent expansion can be highlighted through the analysis of the mismatch distribution of distance among sequences (i.e. the proportion of pairwise nucleotide differences; Schneider and

Excoffier 1999). TE copies after expansion indeed share genetic differences, resulting in unimodal distributions of pairwise differences. Given the biology of TEs, bimodal distributions may be expected following successive waves of expansion, with secondary peaks characterizing prior events.

Mismatch distributions were performed on species-specific alignments using MEGA version 5 (Tamura, et al. 2011) and statistically visualized with siZer (Chaudhui and Marron 1999) to mark significant increases or decreases of slopes around reliable peaks. In complement, the parameter τ (i.e. τ, 2.5 and 97.5 quantiles), representing time in mutational units since expansion, was evaluated using Arlequin 3.5 (Excoffier, et al. 2005) on species-specific alignments as well as alignments including the two species. Comparison between total τ distribution and the species-specific ones identified species with lower τ and thus more recent expansion.

Results

Reads classification

454 sequencing produced 667'485 reads with a mean size of 385.9 bp in *Ae. cylindrica,* representing up to 2.7% of 1C genome coverage, and 646'327 reads with a mean size of 388.8 bp in *Ae. geniculata*, covering of up to 2.5% of the 1C genome (table 1).

Table 1 Proportions of 454 reads corresponding to known transposable elements (TEs), organelle, coding and low complexity (SSR) sequences

	Ae. cylindrica (9398Mb 1C)	*Ae. geniculata* (10074Mb 1C)
TEs	71.3%	72%
Organelle	4.6%	4.7%
Coding sequences	1.4%	1.4%
SSR	0.3%	0.6%
Unclassified	22.4%	21.3%
Total number of 454 reads	667485	646327

BLASTN against the various databases classified around 80% of the reads in both species, whereas slightly more than 20% of the reads remained unclassified. 475'620 and 465'289 reads corresponded to known TEs in *Ae. cylindrica* and *Ae. geniculata* respectively, representing more than 70% of the sequences. In contrast, only 1.4% of the reads were identified as coding sequences.

Identified LTR retrotransposons

Among the 454 reads corresponding to the TE genome fraction, five TE orders were identified: the most abundant were LTR retrotransposons, representing 81.46% and 84.56% of the TEs in *Ae. cylindrica* and *Ae. geniculata*, respectively (table S1, Supplementary Material online). Terminal inverted repeat (TIR) transposons accounted for 17.01% and 13.92%, respectively. Much fewer reads were classified as long interspersed nuclear elements (LINEs) (0.56% for both species), short interspersed nuclear elements (SINEs) (0.07% and 0.06%, respectively), and Helitron (0.09% and 0.08% respectively).

More than 400 TE families were identified with 160 and 165 LTR retrotransposons as well as 177 and 155 TIR transposons families in *Ae. cylindrica* and *Ae. geniculata*, respectively (table S4, Supplementary Material online). Among the LTR retrotransposon families, 53% and 25 to 27% corresponded to Gypsy and Copia elements respectively (table S1, Supplementary Material online). 44 families were present at more than 0.1% of the reads, representing the most abundant families (Figure 2, table S2, Supplementary Material online) and including all families making at least 50% of the two *Aegilops* genomes.

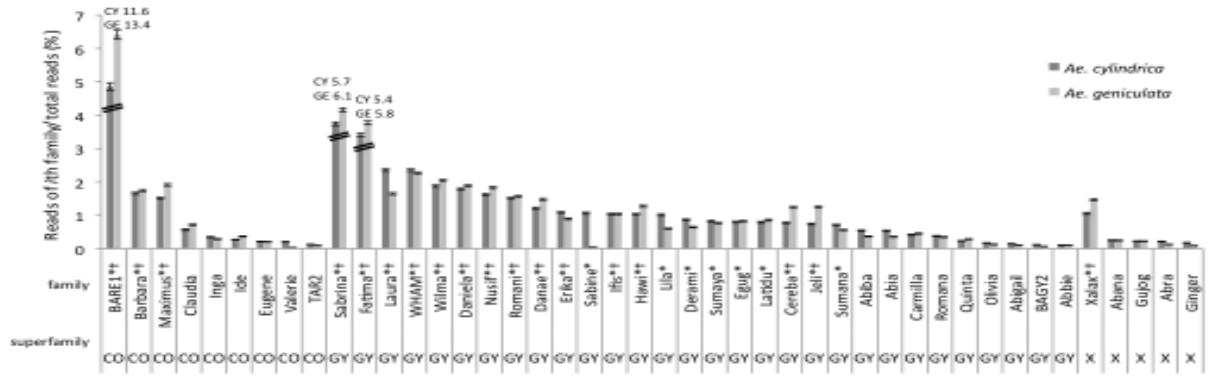


Figure 2 Contribution of 44 major LTR retrotransposon families to 454 reads in *Aegilops cylindrica* (CY, dark grey) and *Ae. geniculata* (GE, light grey) with 95% confidence interval estimated by resampling. Families are grouped according to their superfamilies: CO = copia, GY=gypsy, X =unclassified retrotransposon. *,†mark TE families making at least 50% of the genome complement of *Ae. cylindrica* and *Ae. geniculata* respectively. Proportions for *BARE1*, *Fatima* and *Sabrina* are not at scale and are indicated by values.

The *BARE1* clade (*BARE1*, *WIS* and *Angela*) contributed the most to the overall genome, with around 12% of the total number of reads, and was significantly more abundant in *Ae. geniculata* (13.4%) than in *Ae. cylindrica* (11.6%). *Sabrina* and *Fatima* represented as much as 5 to 6% of the total number of reads, showing higher proportions in *Ae. geniculata* than in *Ae. cylindrica*. Several TE families presented a similar pattern (e.g. *Maximus*, *Cereba*, *Hawi*, *Xalax*), whereas others showed significantly higher proportions of reads in *Ae. cylindrica* than in *Ae. geniculata* (e.g. *Laura*, *Derami*, *Erika*, *Lila*). The LTR retrotransposon family with the largest difference between species was *Sabine*, with more than 7000 reads (1.06%) in *Ae. cylindrica*, but only 274 (0.04%) in *Ae. geniculata*.

Twenty-seven LTR retrotransposon families presented sufficiently large amounts of reads matching the 5' end of the LTR region to be reliably aligned and analyzed through phylogenetics and molecular population genetics.

Phylogenetic inferences

Maximum likelihood trees evaluated to what extent TE families present species-specific TE copies sharing high genetic similarity and forming clades of insertions (table 2, fig S1, Supplementary Material online). Most trees resolved clades of TE copies from both species and, to a certain extent, species-specific clades of insertions with well-supported branches (bootstrap support > 60). Phylogenetic trees were classified into four main topologies (Figure 3): (I) trees with few, but unambiguous species-specific clades of insertions (e.g. *Daniela* and *Lila*, Figure 3a); (II) trees composed of several species-specific clades of insertions (e.g. *BARE1* and *Romani*, Figure 3b); (III) trees with a majority of clades encompassing the two species, although a couple of species-specific clades of copies were observed (e.g. *Maximus* and *WHAM*, Figure 3c); and (IV) trees displaying only mixed-species clades of insertions (*Egug*, *Sabrina* and *Hawi*, Figure 3d).

Genetic differentiation between TE populations

Genetic differentiation between TE populations (i.e. host genomes), as assessed by K_{ST}, varied dramatically depending on the TE family considered, ranging from 0 for *Hawi* to 0.79 for *Daniela* (table 2). Several TE families, such as *Daniela*, *Lila*, *Romani* or *Xalax* showed $K_{ST} > 0.2$, meaning that more than 20% of the total genetic diversity among inserted TE copies is due to differences between species. Several other TE families presented lower differentiation (K_{ST} from 0.2 to 0.025; e.g. *Fatima*, *Maximus*) or non-significant K_{ST} (e.g. *Sabrina*, *Egug*). Accordingly, most populations of LTR retrotransposons showed intermediate K_{ST} values indicative of similar TE copies shared among species together with species-specific groups of copies.

Table 2 Evolutionary genetics of copies of LTR retrotransposons in *Aegilops cylindrica* (CY) and *Ae. geniculata* (GE)

Name	Alignment[a]	K_{ST}	Tree[b]	CY mismatch distribution[c]	GE mismatch distribution[c]	Tau[d]	π CY[e]	π GE[e]
Daniela	300/23/27	0.7921*	I	1	2	-	0.04	0.10
Lila	180/47/35	0.6505*	I	1	2	-	0.14	0.18
Xalax	150/14/37	0.4449*	I	2	1	-	0.12	0.15
BARE1	300/286/133	0.3013*	II	1	2	GE	0.06	0.03
Carmilla	300/28/27	0.2166*	II	1	2	GE	0.14	0.12
Romani	300/23/59	0.2121*	II	2	1	CY	0.09	0.09
Fatima	300/187/147	0.1386*	II	1	2	GE	0.04	0.05
Danae	295/58/149	0.3162*	II	1	1	GE	0.19	0.16
Gujog	250/20/29	0.3126*	II	1	1	GE	0.12	0.06
Barbara	300/65/90	0.2056*	II	1	1	-	0.08	0.06
WHAM	300/71/71	0.2606*	III	1	1	-	0.11	0.10
Cereba	300/97/180	0.1117*	III	1	1	-	0.05	0.06
Maximus	300/162/204	0.0915*	III	1	1	-	0.04	0.06
Eugene	276/50/43	0.076*	III	1	1	-	0.10	0.11
Ginger	250/22/11	0.0528*	III	1	1	-	0.10	0.10
Derami	300/82/97	0.0376*	III	1	1	-	0.09	0.13
Abia	200/35/20	0.0362*	III	1	1	-	0.07	0.08
Romana	300/22/21	0.0306*	III	1	1	-	0.06	0.10
Quinta	300/43/75	0.0249*	III	1	1	-	0.06	0.06
Wilma	250/51/46	0.0238*	III	2	1	-	0.08	0.08
Jeli	285/24/44	0.0227*	III	1	1	-	0.09	0.12
Nusif	280/85/166	0.0194*	III	2	1	-	0.11	0.10
Claudia	300/32/51	0.0141*	III	2	1	-	0.08	0.08
Egug	288/48/83	0.0230*	IV	1	1	-	0.09	0.10
Sabrina	225/307/245	0.0078	IV	1	1	-	0.06	0.07
Hawi	195/40/11	0	IV	1	1	-	0.18	0.18
Sabine	300/128/5	0	IV	1	1	-	0.06	NA

*$p<0.05$

[a]Length of the sequence alignment from both species / number of reads in CY / number of reads in GE

[b]Topology of the maximum likelihood tree

[c]Number of significant peaks detected in the mismatch distribution identified in the species-specific alignments

[d]Species with a significantly lower time since expansion (τ) than the other species and the total alignment, non-significant (-)

[e]Nucleotide diversity (π) among copies within each species

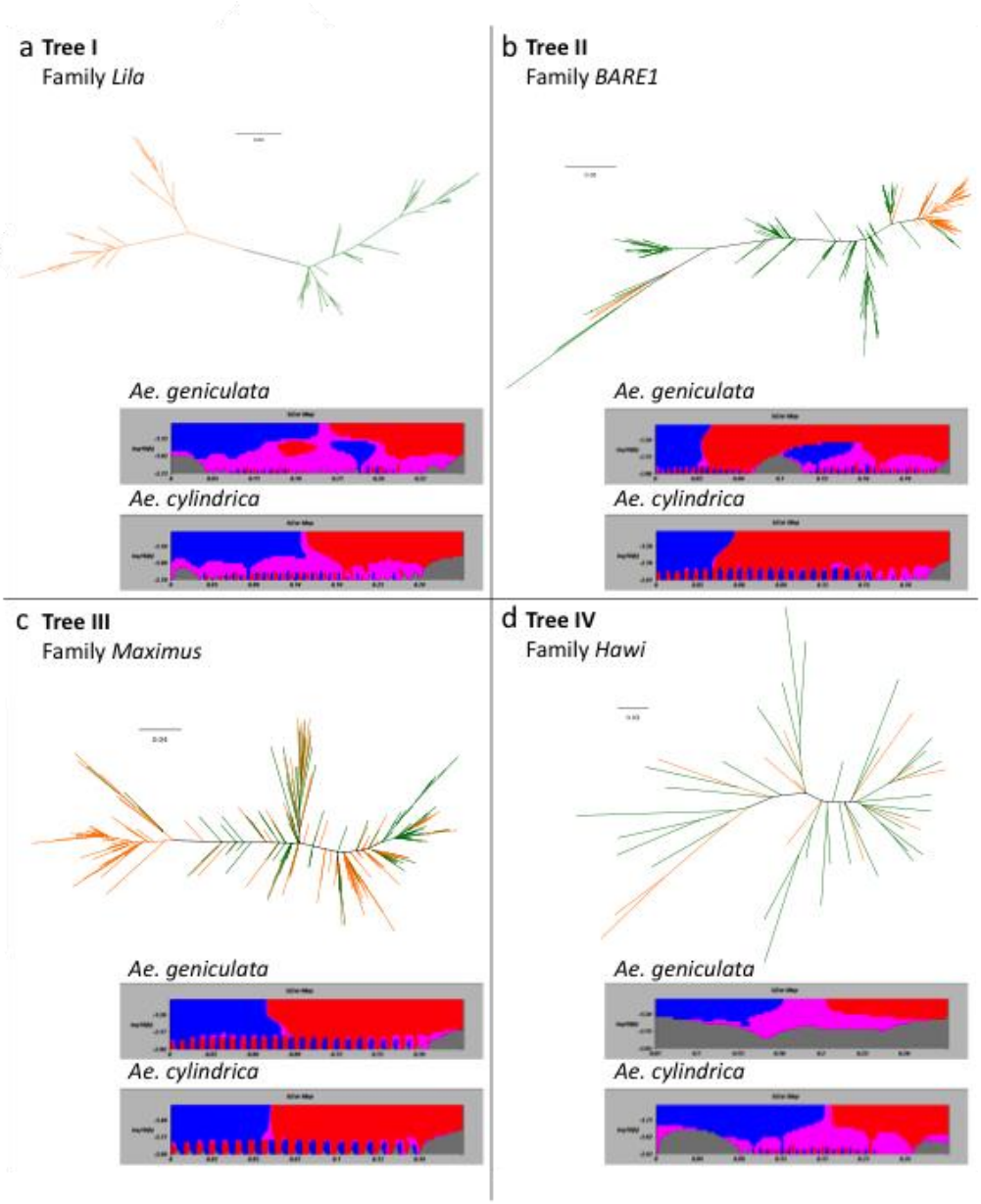


Figure 3 Examples of phylogenetic relationships among copies of LTR retrotransposon families in *Aegilops cylindrica* (green) and *Ae. geniculata* (orange). Unrooted accelerated maximum likelihood trees distinguish four main topologies: (a) trees with only few species-specific clades of insertions as shown by the *Lila* family (referred as Tree I), (b) Tree II are composed of several species-specific clades of insertions, such as the family *BARE1*, (c) Tree III with only few species-specific clades but preponderant mixed-species clades of insertions, as shown by the *Maximus* family, and (d) families like *Hawi* showing only mixed-species clades of insertions (Tree IV). Scale bar represents the branch lengths. In each panel, the distribution of distances among sequences from species-specific alignments (i.e. mismatch distribution) is shown using blue and red for significantly positive and negative slopes respectively.

Molecular signature of TE expansion

Analyses of mismatch distribution between TE copies (i.e. distance among sequences and time since expansion, τ) detected families with recently expanded clades of insertions (table 2). Most TE families revealed unimodal distributions of pairwise differences within species and non-significantly different τ (Figure S1, table S3, Supplementary Material online). Bimodal distribution and lower τ in one of the two species provided evidence of recent proliferation of specific families in *Ae. cylindrica* (e.g. *BARE1*, *Carmilla* and *Fatima*) or in *Ae. geniculata* (e.g. *Romani*).

Recently active vs. quiescent TE family

Evidence based on the ML tree topology as well as genetic differentiation (K_{ST}), distributions of pairwise differences and τ among TE populations congruently highlighted LTR retrotransposon families presenting TE copies with particularly high sequence similarity (table 2). Taken together, these parameters distinguish recently active from quiescent TEs. In particular, *Daniela*, *Lila* and *Xalax* presented ML trees with well-supported species-specific clades (Tree I), high K_{ST}, and bimodal mismatch distributions, offering convincing signs of proliferation during species divergence. Correspondingly, *BARE1*, *Carmilla*, *Romani* and *Fatima* presented signs of recent proliferation, although ML trees showed several independent species-specific clades (Tree II). Three families (*Claudia*, *Nusif* and *Wilma*) exhibited bimodal distributions of pairwise distance among sequences, but ML trees with mostly mixed-species clades of insertions (Tree III), intermediate K_{ST} and similar τ. Such a syndrome would be congruent with old TE proliferation overlaid by more recent expansion. Finally, three families, *Egug*, *Sabrina* and *Hawi*, presented ML trees with mixed-species clades only (Tree IV) and low K_{ST}, indicating quiescence during species divergence.

Discussion

LTR retrotransposons represent the prevalent component of most plant genomes in relation with ancient or recent transpositional activity, but much remains to be determined about the induction and the rate of transposition as well as other processes shaping the genome-wide TE landscape (Brookfield 2005a; Tenaillon, et al. 2010). In that context, low coverage sequencing of randomly distributed fragments across the genome represents an efficient strategy to address variation in TE fractions through bioinformatics methods (Xing, et al. 2013). Here, the sequencing of 2.5% of *Aegilops* genomes was sufficient to survey genome-wide copies of several abundant TEs, including characteristic families from pericentromeric regions (e.g. *Cereba*; Dvorak 2009). Our method not only identifies and quantifies TE families inhabiting genomes, as otherwise performed in various Eukaryote taxa (e.g. Macas, et al. 2007; Wicker, et al. 2009; Sun, et al. 2012; Estep, et al. 2013), but further infers the evolutionary dynamics of several retrotransposon families by comparing individual copies from multiple genomes through appropriate population genetics approaches. It can be straightforwardly extended to include inter-individual variance. Provided that detailed knowledge of TEs inhabiting genomes is accessible at the family level, this procedure can thus be applied to abundant TEs of any genome surveyed with short-read sequencing (i.e. so called next generation sequencing). It allows to further understanding the evolution of dynamic genomes by distinguishing recently active from quiescent TE families without having to produce complete genomes. Future work on fully assembled, high quality genomes may further confirm the accuracy of this procedure.

Composition of the TE genome fraction in wild wheats

The analysis of 454 reads from *Ae. cylindrica* and *Ae. geniculata* provided valuable information on the genome-wide composition of the TE fraction at the family level that is consistent with the one reported in closely related species from the *Triticeae* (Li, et al. 2004; Choulet, et al. 2010; Middleton, et al. 2012) or even maize (Schnable, et al. 2009). *Aegilops* indeed have complex genomes with a minor genic component encompassed with more than 70% of TEs, including at least 80% of LTR retrotransposons from up to 165 different families.

Most TE families such as *Barbara* or *Claudia* represented slightly distinct proportions of the genomes of the *Aegilops* investigated here, whereas a couple of TEs were very contrasted (Figure 2). In particular, *Sabine* (i.e. a moderately abundant LTR retrotransposon in *Hordeum vulgare*, Wicker, et al. 2009) presented more than 7000 reads in *Ae. cylindrica* as compared to only 274 in *Ae. geniculata*. Most LTR retrotransposon families were more abundant in either *Ae. cylindrica* or *Ae. geniculata*, representing one line of evidence suggesting that the different TEs from Triticeae followed contrasted amplification trajectories after the origin of species.

As a whole, the present snapshot offers a detailed description of the TE fractions, showing that only a small proportion of the LTR retrotransposon families (ca. 40) are abundant in wild wheats (Figure 2). In particular, 25 and 18 families sum up to more than 50% of the genome of *Ae. cylindrica* and *Ae. geniculata* respectively (Figure 2), as expected from other *Triticeae* (Sabot and Schulman 2009). In striking contrast to compact genomes such as rice or *Arabidopsis*, the genomes of *Aegilops* and apparently of other *Triticeae* thus present a moderate diversity of TEs, dominated by a couple of very abundant families (Baucom, et al. 2009; Tenaillon, et al. 2011).

Evolutionary dynamics of LTR retrotransposons in wild wheats

Replicative proliferation of LTR retrotransposons gives rise to a population of closely related TE sequences within genomes (Casacuberta, et al. 1997; Brookfield 2005b). Comparing variation of TE copies (here, 300bp at the 5'-end of LTR regions) among host genomes (i.e. TE populations) with molecular population genetics allows distinguishing TE families that remained quiescent or that recently proliferated.

Among the twenty-seven major LTR retrotransposon families studied here in details (table 2), only three (i.e. *Sabrina*, *Egug* and *Hawi*) presented copies with similar sequences in both host genomes, indicating low levels of transposition (i.e. quiescence) during the divergence of *Ae. cylindrica* and *Ae. geniculata*. In contrast, most retrotransposon families showed complementary evidence of recent proliferation with phylogenetic trees resolved in species-specific clades of insertions, high genetic differentiation among populations (K_{ST}) as well as molecular signature of expansion (mismatch distributions). Accordingly, independent transposition of several TE families likely occurred during the divergence of host genomes. In particular, *Lila*, *Daniela* and *Xalax* showed high K_{ST}, signature of recent expansion and trees with two distinct species-specific clades, suggesting that they proliferated from one or few master copies. Seven families (i.e. *BARE1, Barbara, Carmilla*, *Fatima*, *Gujog* and *Romani*) revealed similar patterns, but showed several species-specific clades of TE copies, suggesting proliferation from multiple insertions (i.e. transposon model; Brookfield and Johnson 2006).

Noticeably, abundant TE families such as *BARE1* or *Fatima* showed evidence of recent proliferation in wild wheats, but *Sabrina* was revealed here as quiescent. Congruently, *BARE1* and *Fatima* have been reported active in several related species, supporting continuous transposition during the divergence of host genomes, whereas *Sabrina* was reported as quiescent in *Aegilops speltoides* and probably proliferated at a more distant past (Belyayev, et al. 2010; Kalendar, et al. 2000; Vicient, et al. 2001; Vicient, et al. 1999).

Abundance of TE may thus be a misleading proxy for transpositional activity, even for recently active TEs. For instance, *BARE1* and *Fatima* are more abundant in *Ae. cylindrica*, but evidence reported here indicate that they proliferated more recently in *Ae. geniculata*.

Differential evolutionary trajectories of retrotransposons and host genome evolution

The present results indicate that the transpositional activity of TE families is species-specific to a large extent. Choulet et al. (2010) assessed that TE families amplified in cultivated wheat following different waves during the last 4 MY, with most bursts having occurred around 1.5 MY ago. Accordingly, *BARE1*, *Daniela*, *Fatima*, *Lila* and *Romani* (and *Barbara*, *Cereba*, *Gujog*, *Jeli*, *Maximus*, *Quinta*, *WHAM* and *Wilma* to a certain extent) present evidence of parallel amplifications in wheat and wild wheats. In contrast, *Egug* and *Sabrina* were detected as recently active in domesticated wheat, but quiescent in wild wheats, whereas *Derami* and *Nusif* amplified in *Aegilops* but not in wheat.

Most abundant LTR retrotransposon families investigated here in two *Aegilops* species present evidence of proliferation along with species differentiation, highlighting the importance of TE dynamic in shaping the diversity of *Triticeae* genomes (Bennetzen 2005; Brenchley, et al. 2012; Vitte and Panaud 2005). In particular, wild wheats and domesticated wheat show similar TE genome fractions, but contrasted abundances and evolutionary dynamics of several LTR retrotransposon families. Accordingly, this survey indicates that ancestral TE families followed independent evolutionary trajectories among related species, highlighting the evolution of TE populations as a key factor of genome differentiation. The mechanisms behind such differential dynamics of TE families among species deserve further attention. To what extent the incomplete sorting of a particularly diversified ancestral pool of TEs or the continuous diversification of TE families along with the evolutionary divergence

of host genomes explains differences in the LTR retrotransposon composition remains poorly known (Jurka, et al. 2011). Moreover, the relative importance of intrinsic properties of TEs and of mechanisms acting at the host level should be further considered (Tenaillon, et al. 2010). The effective approach described here uses information from short reads sequences to ultimately understand the forces shaping TE landscape and genome architecture. In particular, species of the *Triticeae* tribe evolved through series of hybridization and polyploidy events (Feldman and Levy 2012) and investigating to what extent TE dynamics is associated with the origin of polyploid lineages is a crucial issue (Parisod and Senerchia 2013).

Acknowledgements

We thank Maud Tenaillon, Julien Vieu and four anonymous reviewers for constructive comments on the manuscript. This work was supported by the Swiss National Science Foundation through the National Centre of Competence in Research 'Plant Survival' and a grant (PZ00P3-131950 to CP).

References

Badaeva ED, et al. 2004. Genome differentiation in *Aegilop*s. 4. Evolution of the U-genome cluster. Plant Syst Evol 246: 45-76.

Baucom RS, et al. 2009. Exceptional diversity, non-random distribution, and rapid evolution of retroelements in the B73 maize genome. PLOS Genet 5:e1000732.

Belyayev A, et al. 2010. Transposable elements in a marginal plant population: temporal fluctuations provide new insights into genome evolution of wild diploid wheat. Mobile DNA 1: 6.

Baum BR, et al. 2012. Phylogenetic relationships among the polyploid and diploid Aegilops species inferred from the nuclear 5S rDNA sequences (*Poaceae*: Triticeae). Genome 55: 177-193.

Bennetzen JL. 2005. Transposable elements, gene creation and genome rearrangement in flowering plants. Curr Opin Genet Dev15: 621-627.

Bennetzen JL, Kellogg EA. 1997. Do plants have a one-way ticket to genomic obesity? Plant Cell 9: 1509-1514.

Berkman P, Lai K, Lorenc M, Edwards D. 2012. Next-generation sequencing applications for wheat crop improvement. Am J Bot 99: 365-371.

Biemont C, Vieira C. 2006. Junk DNA as an evolutionary force. Nature 443: 521-524.

Bossolini E, Wicker T, Knobel P, Keller B. 2007. Comparison of orthologous loci from small grass genomes *Brachypodium* and rice: implications for wheat genomics and grass genome annotation. Plant J 49: 704-717.

Brenchley R, et al. 2012. Analysis of the bread wheat genome using whole-genome shotgun sequencing. Nature 491: 705-710.

Brookfield JFY. 2005a. The ecology of the genome - Mobile DNA elements and their hosts. Nat Rev Genet 6: 128-136.

Brookfield JFY. 2005b. Evolutionary forces generating sequence homogeneity and heterogeneity within retrotransposon families. Cytogen Gen Res 110: 383-391.

Brookfield JFY, Johnson L. 2006. The evolution of mobile DNAs: when will transposons create phylogenies that look as if there is a master gene? Genetics 173: 1115-1123.

Casacuberta JM, Vernhettes S, Audeon C, Grandbastien MA. 1997. Quasispecies in retrotransposons: a role for sequence variability in Tnt1 evolution. Genetica 100: 109-117.

Chaudhui P, Marron J. 1999. SiZer for exploration of structures in curves. J Am Stat Assoc 94: 807-823.

Chen DH, Ronald PC. 1999. A rapid DNA minipreparation method suitable for AFLP and other PCR amplifications. Plant Mol Biol Rep 17: 53-57.

Choulet F, et al. 2010. Megabase level sequencing reveals contrasted organization and evolution patterns of the wheat gene and transposable element spaces. Plant Cell 22: 1686-1701.

DeBarry JD, Liu R, Bennetzen JL. 2008. Discovery and assembly of repeat family pseudomolecules from sparse genomic sequence data using the assissted automated assembler of repeat families (AAARF) algorithm. BMC Bioinf 9: 235.

Dvorak J. 2009. Triticeae genome structure and evolution. In: Muehlbauer GJ, Feuillet C, editors. Genetics and genomics of the *Triticeae*: Springer US. p. 685–712.

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*. Genome 51: 616-627.

Estep MC, DeBarry JD, Bennetzen JL. 2013. The dynamics of LTR retrotransposon accumulation across 25 million years of panicoid grass evolution. Heredity 110: 194–204.

Excoffier L, Laval G, Schneider S. 2005. Arlequin (version 3.0): an integrated software package for population genetics data analysis. Evol Bioinform 1: 47-50.
Fedoroff NV. 2012. Transposable elements, epigenetics, and genome evolution. Science 338: 758-767.
Feldman M, Levy AA. 2012. Genome evolution due to allopolyploidization in wheat. Genetics 192: 763-774.
Feschotte C, Jiang N, Wessler SR. 2002. Plant transposable elements: Where genetics meets genomics. Nat Rev Genet 3: 329-341.
Gaut B, Ross Ibarra J. 2008. Selection on major components of angiosperm genomes. Science 320: 484-486.
Gomez-Alvarez V, Teal TK, Schmidt TM, 2009. Systematic artifacts in metagenomes from complex microbial communities 3: 1314–1317.
Holsinger K, Weir B. 2009. Genetics in geographically structured populations: defining, estimating and interpreting F(ST). Nat Rev Genet 10: 639-650.
Hua-Van A, Le Rouzic A, Boutin TS, Filee J, Capy P. 2011. The struggle for life of the genome's selfish architects. Biol Direct 6:19.
Huang S, Sirikhachornkit A, Su XJ, Faris J, Gill B. 2002. Genes encoding plastid acetyl-CoA carboxylase and 3-phosphoglycerate kinase of the *Triticum*/*Aegilops* complex and the evolutionary history of polyploid wheat. P Natl Acad Sci USA 99: 8133-8138.
Jurka J, Bao W, Kojima K. 2011. Families of transposable elements, population structure and the origin of species. Biol Direct 6:44.
Kalendar R, Tanskanen J, Immonen S, Nevo E, Schulman AH. 2000. Genome evolution of wild barley (*Hordeum spontaneum*) by *BARE-1* retrotransposon dynamics in response to sharp microclimatic divergence. P Natl Acad Sci USA 97: 6603-6607.
Kazazian HH. 2004. Mobile elements: drivers of genome evolution. Science 303: 1626-1632.
Kejnovsky E, Leitch IJ, Leitch AR. 2009. Contrasting evolutionary dynamics between angiosperm and mammalian genomes. Trends Ecol Evol 24: 572-582.
Kumar A, Bennetzen JL. 1999. Plant retrotransposons. Annu Rev Genet 33: 479-532.
Le Rouzic A, Dupas S, Capy P. 2007. Genome ecosystem and transposable elements species. Gene 390: 214-220.
Li WL, Zhang P, Fellers JP, Friebe B, Gill BS. 2004. Sequence composition, organization, and evolution of the core *Triticeae* genome. Plant J 40: 500-511.
Librado P, Rozas J. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25: 1451-1452.
Linc G, Friebe B, Kynast RG. 1999. Molecular cytogenetic analysis of *Aegilops cylindrica* Host. Genome 42: 497-503.
Macas G, Pavel N, Navratilova A. 2007. Repetitive DNA in the pea (*Pisum sativum* L.) genome: comprehensive characterization using 454 sequencing and comparison to soybean and *Medicago truncatula*. BMC Genomics 8:427
Middleton C, Stein N, Keller B, Kilian B, Wicker T. 2012. Comparative analysis of genome composition in *Triticeae* reveals strong variation in transposable element dynamics and nucleotide diversity. Plant J in press doi: 10.1111/tpj.12048
Novák P, Neumann P, Macas J. 2010. Graph-based clustering and characterization of repetitive sequences in next-generation sequencing data. BMC bioinformatics 11: 378.
Parisod C, et al. 2012. Differential dynamics of transposable elements during long-term diploidization of *Nicotiana* section *Repandae* (Solanaceae) allopolyploid genomes. PLOS one 7:e50352.
Parisod C, et al. 2009. Rapid structural and epigenetic reorganization near transposable elements in hybrid and allopolyploid genomes in *Spartina*. New Phytol 184: 1003-1015.

Parisod C, Senerchia N. 2013. Responses of transposable elements to polyploidy. In: Grandbastien MA, Casacuberta JM, editors. Plant transposable element: Springer Verlag, Current Topics in Genetics 24: 147-168.
Sabot F, et al. 2005. Updating of transposable element annotations from large wheat genomic sequences reveals diverse activities and gene associations. Mol Genet Genomics 274: 119-130.
Sabot F, Schulman AH. 2006. Parasitism and the retrotransposon life cycle in plants: a hitchhiker's guide to the genome. Heredity 97: 381-388.
SanMiguel P, Ramakrishna W, Bennetzen J, Busso C, Dubcovsky J. 2002. Transposable elements, genes and recombination in a 215-kb contig from wheat chromosome 5A(m). Funct Integr Genomics 2: 70-80.
SanMiguel P, et al. 1996. Nested retrotransposons in the intergenic regions of the maize genome. Science 274: 765-768.
Schmieder R and Edwards R. 2011. Quality control and preprocessing of metagenomic datasets. Bioinformatics 27:863-864.
Schnable PS, et al. 2009. The B73 maize genome: complexity, diversity, and dynamics. Science 326: 1112-1115.
Schneider S, Excoffier L. 1999. Estimation of past demographic parameters from the distribution of pairwise differences when the mutation rates very among sites: application to human mitochondrial DNA. Genetics 152: 1079-1089.
Slotkin RK, Martienssen R. 2007. Transposable elements and the epigenetic regulation of the genome. Nature reviews. Genetics 8: 272-285.
Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22: 2688-2690.
Sun G, et al. 2011. LTR retrotransposons contribute to genomic gigantism in Plethodontid Salamanders. Gen Biol Evol 4:168-183.
Tamura K, et al. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28: 2731-2739.
Tenaillon M, Hollister J, Gaut B. 2010. A triptych of the evolution of plant transposable elements. Trends Plant Sci 15: 471-478.
Tenaillon MI, Hufford MB, Gaut BS, Ross-Ibarra J. 2011. Genome size and transposable element content as determined by high-throughput sequencing in maize and *Zea luxurians*. Gen Biol Evol 3: 219-229.
Treangen TJ, Salzberg SL. 2012. Repetitive DNA and next-generation sequencing: computational challenges and solutions. Nat Rev Genet 13: 36-46.
van Slageren MW. 1994. Wild wheats: a monograph of *Aegilops* L. and *Amblyopyrum* (Jaub. & Spach) Eig (*Poaceae*). Wageningen Agricultural University Papers 97.
Vicient CM, Jaaskelainen MJ, Kalendar R, Jääskeläinen MJ, Schulman AH. 2001. Active retrotransposons are a common feature of grass genomes. Plant Physiol 125: 1283-1292.
Vicient CM, et al. 1999. Retrotransposon *BARE-1* and its role in genome evolution in the genus *Hordeum*. Plant Cell 11: 1769-1784.
Vitte C, Panaud O. 2005. LTR retrotransposons and flowering plant genome size: emergence of the increase/decrease model. Cytogenet Genome Res 110: 91-107.
Wicker T, Buell CR. 2009. Gene and repetitive sequence annotation in the *Triticeae*. In: Muehlbauer GJ, Feuillet C, editors. Genetics and genomics of the Triticeae: Springer US. p. 407-425.
Wicker T, et al. 2007. A unified classification system for eukaryotic transposable elements. Nat Rev Genet 8: 973-982.

Wicker T, et al. 2009. A whole-genome snapshot of 454 sequences exposes the composition of the barley genome and provides evidence for parallel evolution of genome size in wheat and barley. Plant J 59: 712-722.

Wolfe KH, Gouy ML, Yang YW, Sharp PM, Li WH. 1989. Date of the monocot-dicot divergence estimated from chloroplast DNA sequence data. P Natl Acad Sci USA 86: 6201-6205.

Xing J, Witherspoon DJ, Jorde LB. 2013. Mobile element biology: new possibilities with high-throughput sequencing. Tr Genet doi 10.1016/J.TIG.2012.12.002

Supplementary Material

Supplementary Table S1 Proportions of the different classes, orders (long terminal repeat retrotransposon, LTR; long interspersed nuclear element, LINEs; short interspersed nuclear element, SINE; terminal inverted repeat, TIR and Helitron) and superfamilies of transposable element in *Aegilops cylindrica* and *Ae. geniculata*

Order	Superfamily	*Ae. cylindrica*	*Ae. geniculata*
Class I retrotransposon			
LTR	Gypsy	53.35%	53.87%
	Copia	25.08%	27.30%
	RLX[a]	3.03%	3.39%
LINE	LINE	0.56%	0.56%
SINE	SINE	0.07%	0.06%
Class II DNA transposon			
TIR	CACTA	15.47%	12.48%
	Tc1-Mariner	0.85%	0.81%
	PIF-Harbinger	0.29%	0.27%
	hAT	0.01%	0.01%
	Mutator	0.39%	0.35%
Helitron	Helitron	0.09%	0.08%
Unclassified		0.83%	0.81%

a Unclassified LTR retrotransposon

Supplementary Table S2 Number of 454 reads (mean and 95% confidence interval) for the 44 major LTR retrotransposon families of *Ae. cylindrica* (CY) and *Ae. geniculata* (GE)

family	CY mean 454 reads	CY 0.025 quantile	CY 0.975 quantile	Ge mean 454 reads	GE 0.025 quantile	GE 0.975 quantile
Barbara	11082	10877	11280	11667	11476	11864
BARE1	79188	78466	79921	89573	88708	90456
Claudia	3841	3718	3955	4803	4663	4946
Eugene	1433	1359	1506	1366	1296	1436
Ide	1808	1721	1892	2473	2380	2568
Inga	2328	2231	2433	1949	1864	2038
Maximus	10118	9904	10314	12798	12579	13028
TAR2	871	814	928	649	600	699
Valerie	1364	1293	1434	219	191	249
Abbie	702	651	751	684	637	736
Abia	3579	3457	3701	2409	2314	2502
Abiba	3635	3520	3748	2476	2383	2571
Abigail	978	918	1040	745	693	794
BAGY2	815	760	871	490	448	533
Carmilla	2780	2675	2877	3044	2935	3151
Cereba	5212	5074	5352	8263	8093	8437
Danae	8129	7954	8294	9903	9714	10098
Daniela	12002	11803	12214	12587	12370	12806
Derami	5782	5641	5929	4319	4196	4454
Egug	5368	5230	5506	5538	5389	5689
Erika	7230	7066	7400	6010	5864	6163
Fatima	36125	35759	36506	38569	38229	38963
Hawi	6863	6698	7032	8544	8361	8718
Ifis	6954	6803	7106	6925	6751	7105
Jeli	4969	4835	5104	8336	8148	8510
Latidu	5319	5180	5466	5737	5589	5885
Laura	15728	15499	15962	11017	10814	11218
Lila	6779	6611	6948	4018	3894	4137
Nusif	10822	10621	11010	12200	11985	12415
Olivia	1148	1087	1220	893	836	953
Quinta	1567	1496	1642	1906	1821	1986
Romana	2538	2439	2641	2353	2257	2447
Romani	10101	9905	10284	10436	10242	10628
Sabine	7141	6973	7309	274	243	307
Sabrina	38358	38005	38700	41166	40797	41572
Sumana	4745	4610	4877	3765	3652	3887
Sumaya	5550	5400	5688	5097	4951	5240
WHAM	15715	15482	15965	15162	14923	15401
Wilma	12524	12302	12750	13645	13434	13860
Abana	1677	1596	1755	1650	1569	1734
Abra	1445	1368	1521	874	818	930

Ginger	1198	1131	1262	611	561	660
Gujog	1512	1439	1587	1520	1450	1597
Xalax	7090	6938	7247	9819	9628	10015

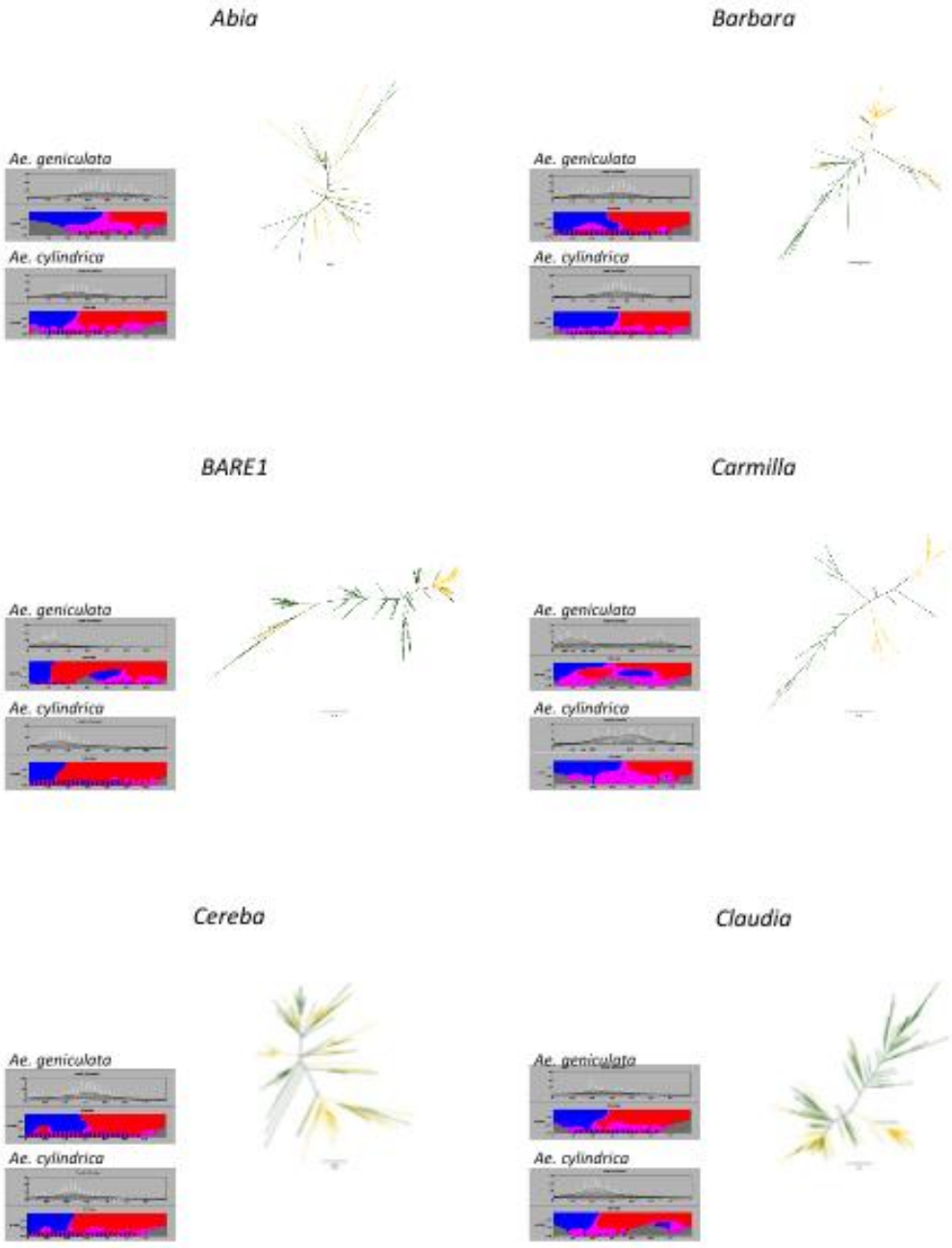
Supplementary Table S3 Time since expansion (τ) and quantiles for the total alignment and species-specific alignments containing the insertions of either *Aegilops cylindrica* (CY) or *Ae. geniculata* (GE)

TE family	Alignment	τ	τ 2.5%	τ 97.5%
Abia	Total	9.9	6.5	15.7
	CY	8.6	6.3	15.0
	GE	11.8	6.6	17.1
Barbara	Total	27.7	22.8	47.4
	CY	31.7	23.0	61.0
	GE	34.8	19.7	40.8
BARE1	Total	20.2	14.8	50.0
	CY	32.8	23.0	50.0
	GE	10.3	7.1	16.3
Carmilla	Total	43.2	34.6	57.4
	CY	39.8	30.8	70.1
	GE	6.6	2.85	43.0
Cereba	Total	15.4	9.9	36.6
	CY	19.0	8.9	36.6
	GE	25.0	13.0	30.1
Claudia	Total	22.0	17.4	27.7
	CY	26.0	18.7	28.6
	GE	23.7	20.3	34.8
Danae	Total	36.2	28.1	98.7
	CY	64.0	47.3	76.3
	GE	48.6	35.5	52.9
Daniela	Total	7.4	0	48.8
	CY	20.3	6.6	26.9
	GE	9.3	5.5	42.6
Derami	Total	21.7	16.5	50.7
	CY	19.2	13.8	32.3
	GE	32.5	22.8	53.4
Egug	Total	47.0	39.0	53.2
	CY	51.3	39.0	57.5
	GE	55.6	45.0	60.8
Eugene	Total	22.1	17.8	29.1
	CY	21.6	13.1	42.9
	GE	24.2	8.2	47.5
Fatima	Total	15.4	11.7	20.1
	CY	16.4	12.8	19.8
	GE	6.1	3.2	24.0
Ginger	Total	21.3	15.8	30.7
	CY	22.4	17.1	45.7
	GE	29.6	23.3	36.5
Gujog	Total	12.7	8	32.6

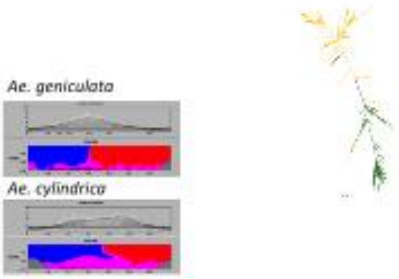
	CY	37.1	21.5	43.5
	GE	13.9	9.6	16.5
Hawi	Total	24.8	18.8	33.8
	CY	25.1	19.3	33.3
	GE	25.8	16.5	49.7
Jeli	Total	22.8	16.4	49.3
	CY	14.9	9.3	58.4
	GE	26.8	20.5	44.4
Lila	Total	12.6	8.2	51.4
	CY	16.6	11.8	36.0
	GE	37.8	16.9	44.0
Maximus	Total	15.5	11.4	77.4
	CY	16.8	10.6	34.4
	GE	34.2	22.0	48.2
Nusif	Total	21.1	16.0	29.0
	CY	17.8	13.0	28.3
	GE	23.9	18.4	42.0
Quinta	Total	15.4	11.1	23.3
	CY	17.8	11.0	32.2
	GE	23.9	15.6	32.0
Romana	Total	30.7	23.5	81.2
	CY	28.2	20.9	49.5
	GE	33.7	26.0	50.7
Romani	Total	27.9	21.3	41.8
	CY	6.6	2.6	30.1
	GE	30.6	22.4	48.9
Sabine	Total	16.3	11.7	28.4
	CY	19.9	13.7	38.6
	GE	19.2	12.6	35.0
Sabrina	Total	14.7	8.4	26.1
	CY	12.3	7.9	27.6
	GE	14.2	10.5	34.0
Wham	Total	24.0	18.7	75.2
	CY	24.6	18.25	45.0
	GE	22.2	17.0	43.0
Wilma	Total	11.7	8.1	19.9
	CY	11.0	7.5	19.9
	GE	14.5	9.54	22.2
Xalax	Total	52.3	39.5	98.0
	CY	190	0.8	402.0
	GE	62.0	45.3	71.0

Supplementary Table S4 is available at Genome Biology and Evolution online (http://www.gbe.oxfordjournals.org/).

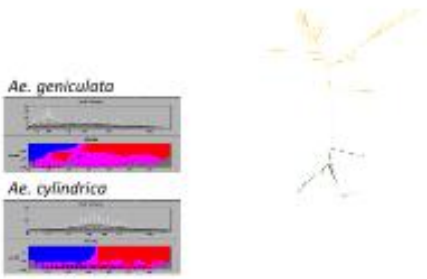
Supplementary Figure S1



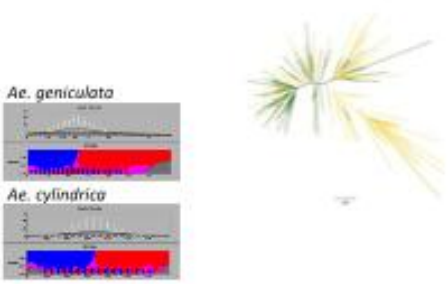
Danae



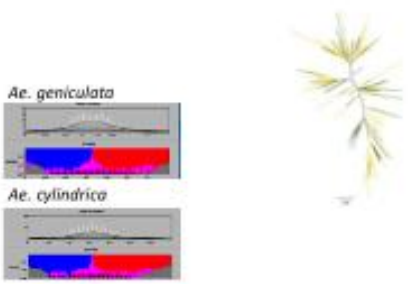
Daniela



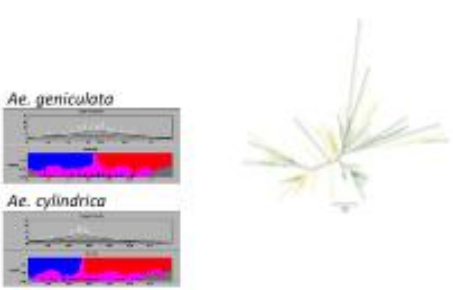
Derami



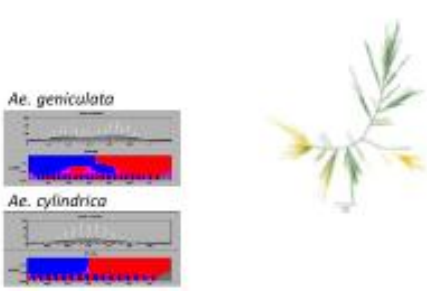
Egug



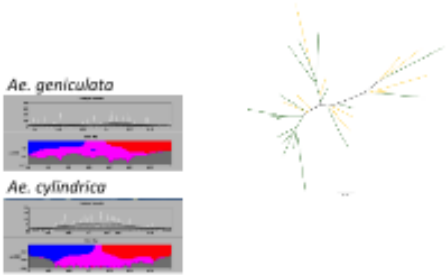
Eugene



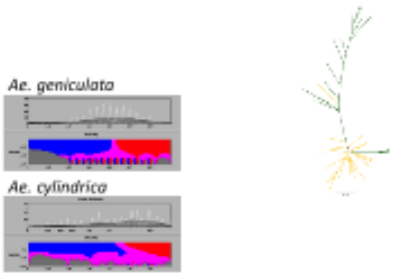
Fatima



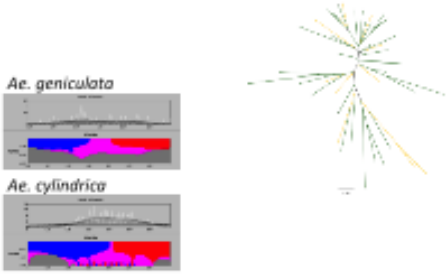
Ginger



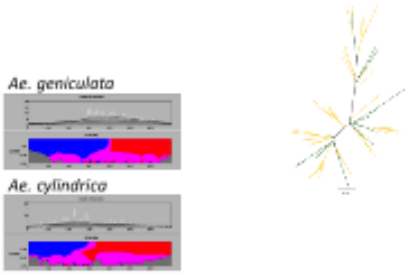
Gujog



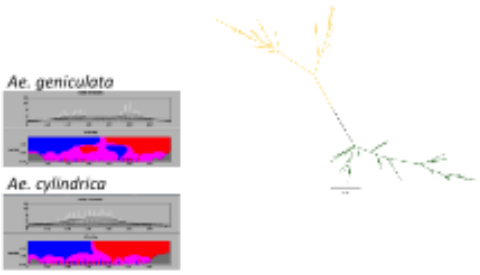
Hawi



Jeli

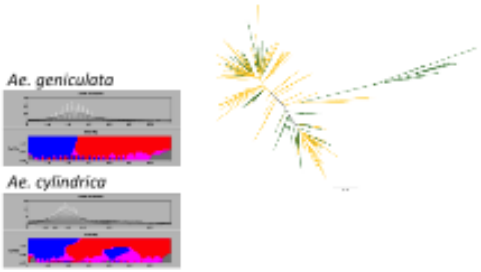


Lila

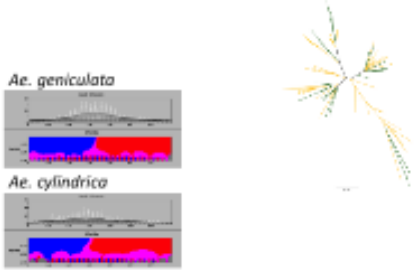


Maximus

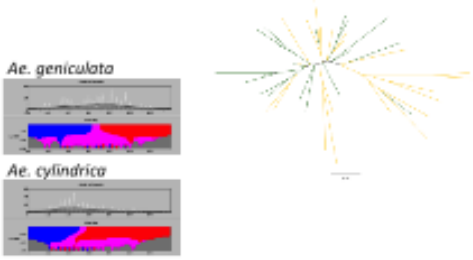
Nusif



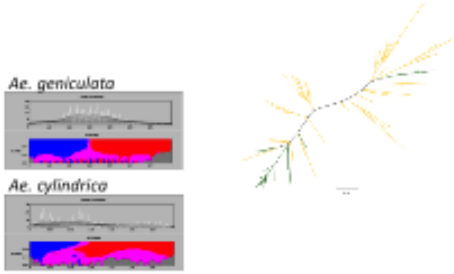
Quinta



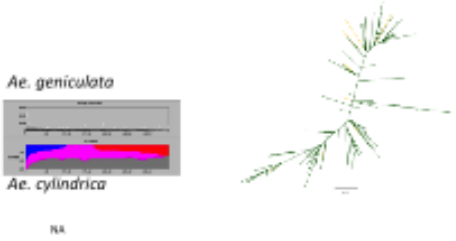
Romana



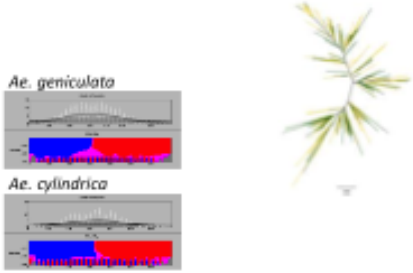
Romani

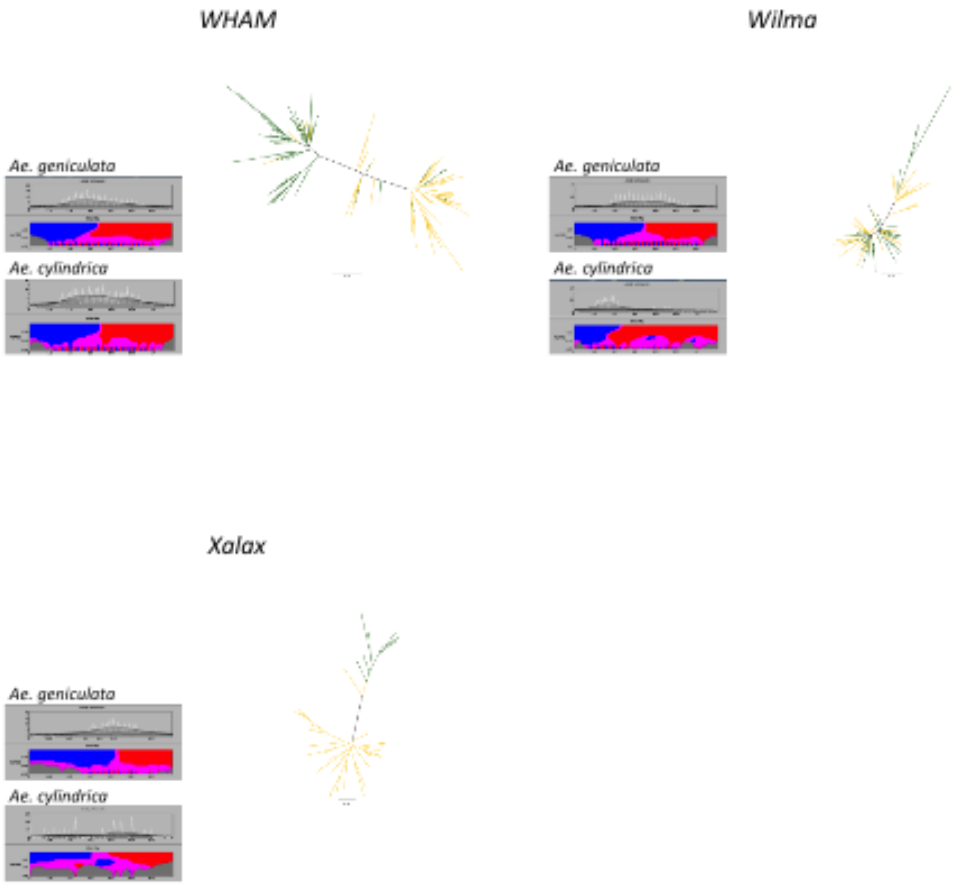


Sabine



Sabrina





Legend

For all retrotransposon family investigated by evolutionary genetics

left bimodal distribution of distance among sequences (i.e. mismatch distribution) done with MEGA version 5 (above) on species-specific alignments), and statistically visualized with the software siZer (below), siZer used blue color to indicate when a 95% confidence interval for the slope is <0, thus when the slope is significantly positive, and red when the same confidence interval is >0, the slope is significantly negative.

right phylogenetic tree inferred with an unrooted maximum likelihood trees with high performance computing using RAXmlHPC (GTR model with gamma-distributed rate across sites; branch support assessed by 100 bootstraps).

green: *Ae. cylindrica* individual inserted copies

yellow: *Ae. geniculata* individual inserted copies

Contrasting evolutionary trajectories of multiple retrotransposons following independent allopolyploidy in wild wheats

Natacha Senerchia[1], François Felber[1,2], Christian Parisod[1]

[1]Laboratory of Evolutionary Botany, Institute of Biology, University of Neuchâtel, 2000 Neuchâtel, Switzerland

[2]Musée et Jardins botaniques cantonaux, Avenue de Cour 14bis, 1007 Lausanne, Switzerland

This chapter was published in New Phytologist (2014) 202: 975–985.

Summary

- Transposable elements (TEs) are expectedly central to genome evolution. To assess the impact of TEs in driving genome turnover, we used allopolyploid genomes, showing considerable deviation from the predicted additivity of their diploid progenitors and thus having undergone major restructuring.
- Genome survey sequencing was used to select 17 putatively active families of long terminal repeat retrotransposons. Genome-wide TE insertions were genotyped with sequence specific amplified polymorphism (SSAP) in diploid progenitors and their derived polyploids, and compared to changes in random sequences to assess restructuring of four independent *Aegilops* allotetraploid genomes.
- Generally, TEs with different evolutionary trajectories than random sequences were identified. Thus, TEs presented family-specific and species-specific dynamics following polyploidy, as illustrated by *Sabine* showing proliferation in particular polyploids, but massive elimination in others. Contrasting with that, only few families (*BARE1* and *Romani*) showed proliferation in all polyploids. Overall, TE divergence among progenitors was strongly correlated with the level of restructuring in polyploid TE fractions.
- TE families present evolutionary trajectories that are decoupled from genome-wide changes after allopolyploidy and have a pervasive impact on their restructuring.

Introduction

The last decades highlighted the astonishing dynamics of plant genomes over evolutionary times, mainly in relation to prominent mechanisms such as retrotransposition, recombination and polyploidy (Kejnovsky *et al.*, 2009; Murat *et al.*, 2012). Polyploidy (i.e. hybridization between more or less divergent genomes, associated with duplication) is a recurrent process across angiosperms (Jiao *et al.*, 2011). Nascent and established polyploids show significant deviation from the expected additivity of their progenitor genomes, indicating major restructuring and epigenetic changes after the origin of polyploid lineages (Soltis & Soltis, 1999; Doyle *et al.*, 2008). Molecular mechanisms and evolutionary processes underlying such genome reorganization are not fully clear yet (Tayalé & Parisod, 2013), but the major fraction of plant genomes represented by transposable elements (TEs) seems to be specifically affected by polyploidy (McClintock, 1984; Fedoroff, 2012). "Genomic stress" such as polyploidy may indeed activate TEs and fast-evolving polyploids thus represent promising models to investigate the role of TEs in driving genome evolution (Comai *et al.*, 2003; Parisod & Senerchia, 2012; Levy, 2013).

Long terminal repeat (LTR) retrotransposons represent the predominant TEs in plants (Kumar & Bennetzen, 1999). Active LTR retrotransposons indeed increase their copy number through a copy-paste mode of transposition, affecting genome size and overall organization (Lynch & Conery, 2003; Ma & Bennetzen, 2004; Wang & Dooner, 2006). However, sequence elimination was highlighted as a major process in TE genome fractions (Vitte & Panaud, 2005), especially following polyploidy (Parisod *et al.*, 2010; Yaakov & Kashkush, 2011; Parisod & Senerchia, 2012). A balance between proliferation and partial elimination by ectopic and illegitimate recombination may thus lead to a high genome turnover (Bennetzen & Kellogg, 1997; SanMiguel *et al.*, 1998; Vitte & Panaud, 2005; El Baidouri & Panaud, 2013). Given that mutations accumulate among TE copies during and after transposition, the

different TE lineages form large populations of repetitive sequences that evolve within genomes (Wicker *et al.*, 2007; Jurka *et al.*, 2011; Wicker, 2013). TE families (i.e. TE copies sharing at least 80% nucleotide similarity) may thus show distinct evolutionary dynamics (Senerchia *et al.*, 2013). However, the processes underlying the particular trajectories of TEs remain unclear and may be influenced by intrinsic TE proprieties as well as processes acting at the host level (Tenaillon *et al.*, 2010; Esnault *et al.*, 2011).

The wheat group belongs to the *Triticeae* clade of the grass family and presents a great diversity of species that evolved through homoploid divergence, hybridization and allopolyploidy. In particular, wild relatives of cultivated wheat (i.e. *Aegilops* species) independently evolved multiple di-, tetra- and hexaploid species that mostly co-occur in the Middle East (van Slageren, 1994). The reticulate evolutionary relationships among *Aegilops* taxa have been thoroughly inferred by genetic and cytogenetic analyses, although they remain poorly dated so far (reviewed in Baum *et al.*, 2012). *Aegilops* allopolyploids are young entities that mostly formed by combinations of diploid species with D or U genomes (Zohary & Feldman, 1962). Within such polyploid clusters, cytogenetic evidence early showed that chromosomes shared among polyploids (i.e. pivotal genome) remain largely unaltered as compared to their diploid donor, whereas the other genome (i.e. differential genome) presented considerable alteration (Zohary & Feldman, 1962; Feldman, 1965a; Feldman, 1965b; Feldman, 1965c). Zohary and Feldman (1962) early postulated that this pattern results from gene flow among established allopolyploids, with common pivotal genomes providing substrates for homologous recombination and buffering the initial loss of hybrid fertility, while the differential genomes were accumulating restructuring events through homeologous recombination.

The overall genome organization of hybridizing *Aegilops* supports this hypothesis to a certain extent (Zohary & Feldman, 1962; Wang *et al.*, 2000), but hardly explains the balanced restructuring of both genomes reported in some species (Kimber *et al.*, 1988; Badaeva *et al.*, 2002; Badaeva *et al.*, 2004) or that natural hybrids between species with so called pivotal genomes are noticeably rare. Challenging the pivotal-differential hypothesis, recent studies used various molecular markers to show that asymmetrical restructuring of polyploid genomes may depend on other factors such as sequence types and genome sizes (reviewed in Bento *et al.*, 2011). Additional investigations are thus required to understand the processes underlying the evolution of genomes within species complexes such as *Aegilops*.

Aegilops allopolyploids present fast-evolving genomes composed of ca. 80% of TEs (Li *et al.*, 2004; Sabot & Schulman, 2009; Wicker & Buell, 2009; Senerchia *et al.*, 2013), thus offering promising models to assess the role of TEs in genomes restructuring and shedding light on the processes shaping genome architecture. Here, we focused on four diploid species, *Ae. caudata* (genome C), *Ae. comosa* (M), *Ae. tauschii* (D) and *Ae. umbellulata* (U) that recently combined into four natural allotetraploid species, *Ae. crassa* (DM), *Ae. cylindrica* (DC), *Ae. geniculata* (UM) and *Ae. triuncialis* (UC) (Figure 1).

Our aims were to (i) investigate genome restructuring and (ii) assess the evolutionary trajectories of abundant LTR retrotransposon families after independent allopolyploidy events. Based on genome survey sequence (GSS) data from two *Aegilops* polyploid genomes, we designed complementary molecular fingerprint assays tracking restructuring events in random sequences as well as 17 LTR retrotransposon families. Comparisons between TEs and random sequences highlighted proliferation of several TE families and predominant sequence deletion in others, indicating species-specific and TE-specific evolutionary trajectories

following allopolyploidy. In particular, TE dynamics was associated with divergent insertions between progenitor species, suggesting that genome shocks at the origin of polyploid lineages had a long lasting influence on genome organization.

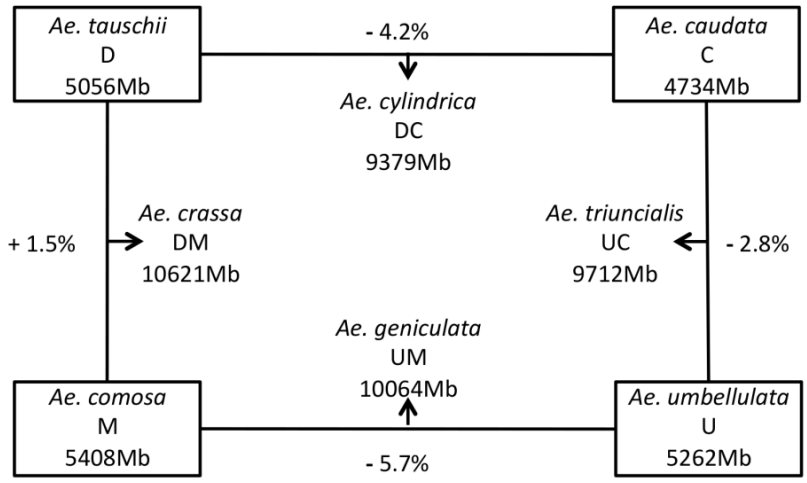


Figure 1 Genome composition of investigated *Aegilops* diploid and derived allopolyploid species. Genome composition followed van Slageren (1994). Deviations of observed tetraploid genome size from expectations based on the expected addition of diploid progenitors, according to Eilam *et al*. (2007, 2008, using 1 picogram of DNA = 978 Mb).

Materials and Methods

Plant material

Three accessions of the polyploids *Aegilops crassa (*PI 487286, PI 542178, TA 1880), *Ae. cylindrica* (TA 2204 = AE 719, PI 486235, PI 554221), *Ae. geniculata* (TA 1800, PI 487221, PI 287737), *Ae. triuncialis* (PI 487246, PI 542345, PI 491442) as well as of diploid progenitors species *Ae. caudata* (IG 47965, IG 48080, IG 107317), *Ae. comosa* (AE 1376, AE 1260, AE 1378), *Ae. tauschii* (IG 47219, IG 46847, AE 528), and 2 accessions of *Ae. umbellulata* (IG 48082, IG 46964) were obtained from the Institutes für Planzengenetik und Kulturplanzenforschung (AE), the United States Department of Agriculture (PI), the

International Center for Agricultural Research in the Dry Areas (IG) and the Wheat Genetic and Genomic Resources Center (TA). These accessions have been characterized as genetically variable in previous studies (Badaeva *et al.*, 2002; Badaeva *et al.*, 2004; Meimberg *et al.*, 2009). They have been maintained in germplasms by selfing and were thus assumed as inbred. Plants were grown under controlled conditions (18°C 18h light) and DNA of two weeks seedlings was extracted from fresh leaves disrupted in liquid nitrogen, following the standard DNeasy plant extraction mini kit protocol from Qiagen AG, Switzerland.

Genome Sequence Survey (GSS) and selection of TE families

40 ng of genomic DNA of one individual from both *Ae. cylindrica* (TA 2204 = AE 719) and of *Ae. geniculata* (TA 1800) was mechanically shotguned and sequenced on half a plate of the Roche 454 GS FLX titanium platform (service provided by Microsynth, Switzerland, following manufacturer's instructions). BLASTN of non-redundant reads against complete TREP database (wheat.pw.usda.gov/ITMI/Repeats/) classified TEs following Senerchia et al. (2013). Portions of reads matching to 300 bp at the 5'end of the LTR regions of 27 TE families were aligned using ClustalW and analyzed through molecular population genetics to distinguish active from quiescent families following Senerchia *et al* (2013). We selected here 16 LTR retrotransposon families showing evidence of recent transpositional activity (*BARE1*, *Barbara*, *Cereba*, *Claudia*, *Daniela*, *Danae*, *Derami*, *Fatima*, *Lila*, *Maximus*, *Nusif*, *Romani*, *Quinta*, *Sabine Wham*, and *Xalax)* as well as one quiescent TE family (*Egug*, selected as a negative control) and evaluated nucleotide diversity (π) along their LTRs within a sliding window of 5bp using DnaSP v5 (Librado & Rozas, 2009). For each TE family, a SSAP primer was designed across the most conserved region of the alignments using Primer 3 (Rozen & Skaletsky, 2000).

Molecular fingerprint techniques

Amplified fragment length polymorphism (AFLP) and sequence specific amplified polymorphism (SSAP) were carried out following the protocol of Parisod & 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 in plates with randomly positioned individual DNA. After inactivation of enzymes at 65°C for 20 min, primers corresponding to adaptors were used for preselective amplification with GoTaq®DNA polymerase (Promega). Twenty-times diluted PCR products were selectively amplified with a touch-down PCR using primers with three additional nucleotides. For AFLP, unlabelled *Mse*I primers were used together with fluorescently labeled *Eco*RI primers, whereas SSAP used fluorescently labeled TE family specific primers (Table S1) instead of *Eco*RI primers. SSAP thus mostly amplifies the region encompassing the termini of retrotransposon insertions and their flanking genomic sequences up to a restriction site cleaved by *Mse*I. To minimize experimental inconsistencies, digestion, ligation and preselective were performed simultaneously and polymerase chain reactions were carried out on thermocyclers with fixed ramp rate. Genotyping error rates were estimated based on the replication of the whole procedure on five samples (i.e. 22% of the dataset) as $E_b = M_{repl}/n_{bin}$, where M_{repl} is the total number of mismatches between replicated samples and n_{bin} is the total number of scored bins (Bonin *et al.*, 2004). Upon the sixty-four different *Eco*RI-*Mse*I combinations of AFLP selective primers tested and the eight *Mse*I selective primers combinations were tested for each TE family, only the most reproducible combinations were retained and fully analyzed (i.e. 17 AFLP combinations and a mean of 4.5 SSAP combinations per TE family).

PCR selective products amplified with FAM™, VIC®, NED™ fluorescent dye were pooled with GeneScan™ 500 LIZ™ Size Standard and separated with an ABI 3500 capillary sequencer. Resulting electropherograms were visualized and scored with GeneMapper 3.5 (Applied Biosystems) using AFLP default peak detection parameters. The scoring was manually checked and loci were recorded as present (1) or absent (0) in binary matrices.

AFLP and SSAP loci analysis

For each species, Nei's gene diversity was estimated, with standard error, for AFLP as well as SSAP for each TE family using the Bayesian method with non-uniform prior distribution of allele frequencies (assuming FIS=1) as implemented in AFLPsurv (Vekemans, 2002).

For each locus j, frequencies of band presence (1) and absence (0) were estimated among accessions of the diploid progenitors ($D_1(1)_j$, $D_1(0)_j$ and $D_2(1)_j$, $D_2(0)_j$, respectively) as well as accessions of the derived allopolyploid ($A(1)_j$, $A(0)_j$). The expected polyploid profile (i.e. the additivity of the progenitors) was estimated as probabilities of band presence and absence based on the frequencies in diploid progenitors. The probability of band presence in the polyploid was calculated as $E(1)_j = D_1(1)_j \times D_2(1)_j + D_1(1)_j \times D_2(0)_j + D_1(0)_j \times D_2(1)_j$, whereas the probability of band absence was calculated as $E(0)_j = D_1(0)_j \times D_2(0)_j$. Then, the observed polyploid profile was compared to the expected polyploid profile assessing the probability that the locus j was (i) additive (i.e. identity between expected and observed polyploid profiles) as $E(1)_j \times A(1)_j + E(0)_j \times A(0)_j$, (ii) a new band (i.e. presence of a band in the observed polyploid profile despite absence in the expected profile) as $E(0)_j \times A(1)_j$, or (iii) a lost band (i.e. band absence in the observed polyploid profile despite predicted presence in

the expected profile) as $(E(1)_j \times A(0)_j)$. Probabilities of additive, new and lost bands were calculated for each locus, summed across loci and divided by the total number of bands to obtain proportions of additive, new and lost band for AFLP and for each SSAP of each TE family. Proportions were calculated for each polyploid accession and averaged within species. The same procedure was applied to the four allopolyploids.

For each tetraploid species, one-way analysis of variance (ANOVA) followed by multiple comparisons with Tukey's honest significant differences post-hoc were carried out to assess TE families with proportions of (i) new bands and (ii) ratio of lost bands divided by new bands significantly different than AFLP at 5% levels, using the package 'agricolae' on R cran (cran.r-project.org).

Correlation between parental non-shared bands among progenitors and new bands in polyploids

For each pair of diploid progenitors, the proportion of parental non-shared bands among species, taking multiple accessions into account, was estimated for each locus as $D_1(0)_j \times D_2(1)_j + D_1(1)_j \times D_2(0)_j$. It was then summed over loci and divided by the total number of loci for AFLP as well as SSAP for each TE family.

Correlations between the proportions of non-shared bands among progenitors and the proportions of (i) new bands and (ii) non-additive bands in the derived polyploids were assessed with a mixed effect linear model using maximum likelihood method of the package 'lme4' and including TE families nested within polyploid species as random effects. R^2 was defined using a function correlating fitted values to observed values in R cran (Nakagawa & Schielzeth, 2013).

Origin of lost bands in polyploids

The origin of lost bands was assessed for each polyploid on loci presenting band presence in only one progenitor (i.e. either the pivotal or the differential genome in the polyploid). First, Kruskal Wallis tests tested significant difference in proportions of lost bands among pivotal and differential in each polyploid. Then, differences in proportions of lost bands between the pivotal and differential genome for each the 17 TE families were tested by a prop.test on R Cran.

Results

Candidate TE families

Genome survey sequence (GSS) of *Ae. cylindrica* (genome DC) and *Ae. geniculata* (genome UM) by 454 sequencing identified more than 70% of the genome as TEs, including more than 165 LTR retrotransposon families (Senerchia *et al.*, 2013). Comparison of 300 bp at the 5'end of LTR regions of the 17 selected TE families identified seven families (*BARE1, Danae, Daniela, Fatima, Lila, Romani, Xalax)* with clear evidence of recent activity, whereas nine (*Barbara, Cereba, Claudia, Derami, Maximus, Nusif, Quinta, Sabine, Wham)* showed signs of both old and recent transpositional activity (Table 1).

In contrast, *Egug* was considered as quiescent and selected as negative control. Based on GSS data, specific primers for each of these 17 candidate TE families were designed in conserved regions at the 5'end of the LTR (Table S1) and used to perform sequence specific amplified polymorphism (SSAP) assays representative of the diversity of TE copies within *Aegilops* genomes.

Genetic variation in Aegilops and genome restructuring in polyploids

Variable genetic diversity was assessed within *Aegilops* species, showing specific variation in AFLP as well as the different TE families (Table 2). Genetic diversity within diploid and polyploid species ranged from 0.104 for the TE *Romani* in *Ae. umbellulata* to 0.272 for *BARE1* in *Ae. comosa*, and from 0.099 for *Cereba* in *Ae. cylindrica* to 0.262 for *Derami* in *Ae. triuncialis*, respectively. Several TEs such as *BARE1* showed higher genetic diversity than genome-wide AFLPs, but no clear pattern differentiating diploid and tetraploid species was evident. Noticeably, *Sabine* showed a significantly high diversity in *Ae. cylindrica* than in *Ae. geniculata*. Genetic divergence among species for AFLP as well as the different TE families showed considerable variation depending on the genome fraction considered (Note S1). In particular, multiple Mantel tests correlating genetic distances between species as measured by AFLP and by SSAP of the different TEs were non-significant (after stringent Bonferroni correction for multiple testing), indicating TE specific patterns of divergence among species.

Genome restructuring in four *Aegilops* allotetraploids was assessed by comparing the fingerprint profiles of multiple accessions for random sequences (i.e. using amplified fragment length polymorphism; AFLP) and for the 17 candidate TE families (i.e. using SSAP) to the expected additivity of their respective diploid progenitor species. Comparison of AFLP vs. SSAP allows contrasting background genome restructuring and specific reorganization of TE fraction. Genotyping of 1224 AFLP loci (with an error rate of 3.36%) and 238 to 792 SSAP loci for each of the 17 TE families (with an error rate ranging from 1 to 6.6%, Table S2 and S3) from three accessions per species took genetic variation into account.

Table 1 Summary statistics of investigated LTR retrotransposon families based on the sequencing of 300 bp at the 5' end of the LTR region in *Aegilops cylindrica* (CY) and *Ae. geniculata* (GE).

TE family	Reads CY[a]	π CY[b]	Reads GE[a]	π GE[b]	Activity[c]
Barbara	11082 (1.66)	0.08	11677 (1.81)	0.06	+
BARE1	79188 (11.60)	0.06	89573 (13.41)	0.03	+
Cereba	5212 (0.78)	0.05	8263 (1.28)	0.06	+
Claudia	3841 (0.57)	0.08	4803 (0.74)	0.08	+
Danae	8129 (1.22)	0.19	9903 (1.53)	0.16	+
Daniela	12022 (1.80)	0.04	12587 (1.95)	0.1	+
Derami	5782 (0.87)	0.09	4319 (0.67)	0.13	+
Egug	5368 (0.81)	0.09	5538 (0.86)	0.1	-
Fatima	38357 (5.41)	0.04	38567 (5.97)	0.05	+
Lila	6779 (1.02)	0.14	4018 (0.62)	0.18	+
Maximus	10118 (1.52)	0.04	12798 (1.98)	0.06	+
Nusif	10822 (1.62)	0.11	12200 (1.89)	0.1	+
Quinta	1567 (0.23)	0.06	1906 (0.29)	0.06	+
Romani	10101 (1.51)	0.09	10436 (1.61)	0.09	+
Sabine	7141 (1.07)	0.06	274 (0.04)	NA	+
WHAM	15715 (2.35)	0.11	15162 (2.35)	0.1	+
Xalax	7090 (1.06)	0.12	9819 (1.52)	0.15	+

[a]Number of reads from 454 genome survey sequences (GSS) with proportion of reads out of the total of sequences in between parentheses

[b]Nucleotide diversity (π) among copies within genome

[c]Activity = TE evolutionary dynamics (+ recently active, - quiescent), inferred from molecular population genetics (details in Senerchia et al., 2013)

Combinational probabilities based on the frequencies of band presence and absence in the diploid progenitors were used to determine the expected polyploid profile at each locus and were then compared with the observed profile for each polyploid accession, assessing the probability that the locus was additive, lost or that a new locus appeared in the polyploid (Table 2). Averaged over loci, the mean proportion for additive AFLP bands across polyploids was 63.1%, close to the overall proportion of SSAP bands for the 17 TE families (62.9%). Proportion of additive AFLP bands however varied considerably among the four allopolyploid species, indicating species-specific level of background restructuring following the different polyploidy events. In particular, *Ae. crassa* showed higher level of restructuring (i.e. non-additive AFLP bands, 44.9%) than *Ae. cylindrica* (32.1%), *Ae. geniculata* (37.9%)

and *Ae. triuncialis* (32.9%). Proportions of additive SSAP bands showed considerable variation among TE families, ranging from 43.0% for *Romani* in *Ae. crassa* to 78.9% for *Sabine* in *Ae. cylindrica*.

New bands in random sequences and in multiple TE families

Comparing levels of restructuring from a large number of SSAP loci for each TE family to the level of restructuring in random sequences (AFLP) allowed to highlight TEs with specific evolutionary trajectories, while accounting for possible ambiguities of fingerprint methods (Petit *et al.*, 2010; Sarilar *et al.*, 2013). Tukey HSD tests assessed that 15 among 17 TE families showed significantly higher proportions of new bands than AFLP, indicating proliferation in at least one of the tetraploid species (Figure 2A). The number of TEs showing higher proportions of new bands than AFLP varied among the tetraploids (Table S3). Nine TE families in *Ae. crassa*, ten *in Ae. cylindrica* and 13 in *Ae. geniculata*, but only two TEs in *Ae. triuncialis* showed such a pattern indicative of TE proliferation. However, few TEs showed consistent patterns of new bands in all polyploids. Only *BARE1* and *Romani* presented significantly higher proportions of new bands than AFLP in the four tetraploids, whereas *Egug* and *Nusif* never exhibited significant differences. Most TE families rather showed species-specific patterns, with some TE families showing significantly more new bands in the three tetraploids (*Barbara*, *Claudia*, *Daniela*, *Fatima* and *Quinta* in *Ae. crassa*, *Ae. cylindrica* and *Ae. geniculata*), in two tetraploids (*Cereba* in *Ae. crassa* and *Ae. geniculata*, *Derami* in *Ae. crassa* and *Ae. cylindrica*, *Xalax* in *Ae. cylindrica* and *Ae. geniculata*) or in only one tetraploid (*Danae*, *Lila*, *Maximus* and *WHAM* in *Ae. geniculata* or *Sabine* in *Ae. cylindrica*).

Table 2 Distribution of loci within and among *Aegilops* species. Nei's gene diversity (Div) within diploid and allotetraploid species (genome composition following van Slageren, 1994) is presented for random sequences (AFLP) as well as the 17 TE families (SSAP) with standard error. Mean proportions of parental non-shared bands among diploid progenitors (i.e. shared bands, S) and comparisons to the expected additivity of the diploids (additive bands, A; lost bands, L; new bands, N) is presented for each allotetraploid species.

	Ae. caudata (C)	*Ae. comosa* (M)	*Ae. tauschii* (D)	*Ae. umbellulata* (U)	*Ae. crassa* (genome DM)					*Ae. cylindrica* (genome DC)					*Ae. geniculata* (genome UM)					*Ae. triuncialis* (genome UC)				
	Div	Div	Div	Div	Div	S	A	L	N	Div	S	A	L	N	Div	S	A	L	N	Div	S	A	L	N
AFLP	0.209± 0.006	0.191± 0.006	0.191± 0.006	0.180± 0.006	0.238± 0.007	63.57	55.11	38.34	6.46	0.161± 0.005	59.21	67.89	27.06	5.26	0.214± 0.006	69.61	62.06	33.14	4.65	0.215± 0.006	59.43	67.61	25.82	6.42
Barbara	0.226± 0.015	0.172± 0.015	0.203± 0.016	0.194± 0.015	0.189± 0.015	65.77	50.84	31.90	16.73	0.145± 0.012	54.75	66.47	26.20	6.80	0.173± 0.014	57.11	55.34	34.83	9.27	0.210± 0.015	58.29	59.87	32.84	6.76
BARE1	0.227± 0.008	0.272± 0.008	0.251± 0.008	0.203± 0.009	0.259± 0.008	59.47	58.07	30.71	11.09	0.149± 0.006	64.50	69.76	22.05	8.04	0.238± 0.008	55.55	57.60	29.06	13.20	0.216± 0.008	57.06	64.37	24.05	11.43
Cereba	0.180± 0.01	0.200± 0.010	0.201± 0.011	0.138± 0.008	0.164± 0.009	65.24	54.28	32.66	12.80	0.102± 0.007	61.90	66.54	27.11	6.07	0.146± 0.008	49.83	59.06	28.64	12.06	0.151± 0.009	52.94	63.55	28.98	7.19
Claudia	0.178± 0.009	0.180± 0.009	0.170± 0.008	0.115± 0.007	0.164± 0.008	54.95	52.13	35.51	12.16	0.113± 0.006	55.03	68.71	24.09	6.98	0.147± 0.007	46.21	64.58	26.54	8.67	0.151± 0.008	36.56	66.16	26.52	7.09
Danae	0.179± 0.009	0.151± 0.009	0.159± 0.009	0.139± 0.009	0.160± 0.009	49.39	48.72	40.61	10.41	0.129± 0.008	54.08	67.85	25.30	6.57	0.144± 0.008	46.19	59.36	27.17	13.20	0.130± 0.008	49.95	60.90	29.82	9.01
Daniela	0.193± 0.009	0.236± 0.009	0.177± 0.008	0.158± 0.007	0.173± 0.008	54.65	57.15	31.69	11.02	0.110± 0.006	55.97	67.80	23.21	8.73	0.209± 0.008	50.52	59.24	30.74	9.85	0.147± 0.007	48.44	65.39	27.88	6.55
Derami	0.198± 0.013	0.238± 0.013	0.132± 0.012	0.195± 0.013	0.204± 0.013	53.33	50.04	33.74	15.83	0.154± 0.012	50.56	54.72	33.26	11.57	0.202± 0.012	55.57	65.58	28.07	5.95	0.262± 0.015	50.62	56.59	38.77	4.21
Egug	0.173± 0.009	0.193± 0.01	0.183± 0.009	0.190± 0.01	0.184± 0.01	67.33	65.38	28.29	6.04	0.131± 0.007	73.30	76.15	19.91	3.64	0.171± 0.009	66.76	71.51	23.00	5.19	0.160± 0.009	78.71	77.28	18.63	3.78
Fatima	0.234± 0.009	0.192± 0.009	0.154± 0.008	0.147± 0.008	0.172± 0.008	56.32	46.78	40.14	12.91	0.108± 0.006	48.46	64.40	26.10	9.31	0.167± 0.008	48.61	57.54	31.80	10.48	0.160± 0.007	68.76	65.52	26.13	8.18
Lila	0.195± 0.009	0.170± 0.009	0.181± 0.009	0.156± 0.009	0.186± 0.009	60.17	60.98	28.65	10.14	0.111± 0.006	59.94	73.99	20.37	5.40	0.145± 0.008	55.10	66.02	24.86	8.76	0.161± 0.008	57.55	70.31	21.71	7.74
Maximus	0.174± 0.012	0.249± 0.013	0.187± 0.013	0.174± 0.012	0.220± 0.013	56.95	51.94	36.84	10.84	0.174± 0.012	48.53	63.73	30.93	4.91	0.205± 0.013	51.96	56.15	33.33	10.01	0.189± 0.012	54.43	64.70	27.36	7.52
Nusif	0.159± 0.010	0.169± 0.011	0.163± 0.011	0.189± 0.011	0.166± 0.011	67.35	58.14	30.81	10.70	0.099± 0.007	60.32	70.51	24.32	4.80	0.154± 0.009	55.40	68.36	24.89	6.42	0.170± 0.010	62.89	72.85	21.94	4.87
Quinta	0.186± 0.010	0.220± 0.011	0.225± 0.011	0.183± 0.100	0.178± 0.010	61.88	56.91	31.26	11.54	0.130± 0.008	63.38	71.09	21.06	7.59	0.152± 0.008	50.39	64.04	27.91	7.80	0.181± 0.009	55.46	66.87	24.70	8.17
Romani	0.118± 0.008	0.165± 0.009	0.158± 0.009	0.104± 0.008	0.172± 0.009	60.42	42.99	35.30	21.51	0.077± 0.006	55.28	63.89	24.97	10.87	0.125± 0.007	49.45	50.91	29.75	19.11	0.131± 0.008	47.07	59.43	26.10	14.18
Sabine	0.257± 0.013	0.212± 0.014	0.172± 0.013	0.190± 0.013	0.151± 0.012	67.35	62.18	30.96	6.34	0.142± 0.009	58.34	78.91	13.87	6.76	0.066± 0.005	59.47	59.27	37.51	2.69	0.187± 0.012	59.52	73.02	20.48	5.89
WHAM	0.163± 0.014	0.183± 0.015	0.188± 0.015	0.134± 0.011	0.161± 0.014	63.11	56.00	36.66	6.74	0.107± 0.008	60.75	76.85	16.65	5.56	0.173± 0.014	54.98	68.69	23.72	6.95	0.151± 0.013	64.54	70.51	23.64	5.18
Xalax	0.227± 0.012	0.161± 0.011	0.244± 0.013	0.179± 0.011	0.183± 0.012	63.34	58.36	33.36	7.93	0.108± 0.006	57.30	74.28	18.18	7.19	0.160± 0.010	60.60	66.99	23.49	9.15	0.203± 0.012	55.90	63.67	32.40	3.57

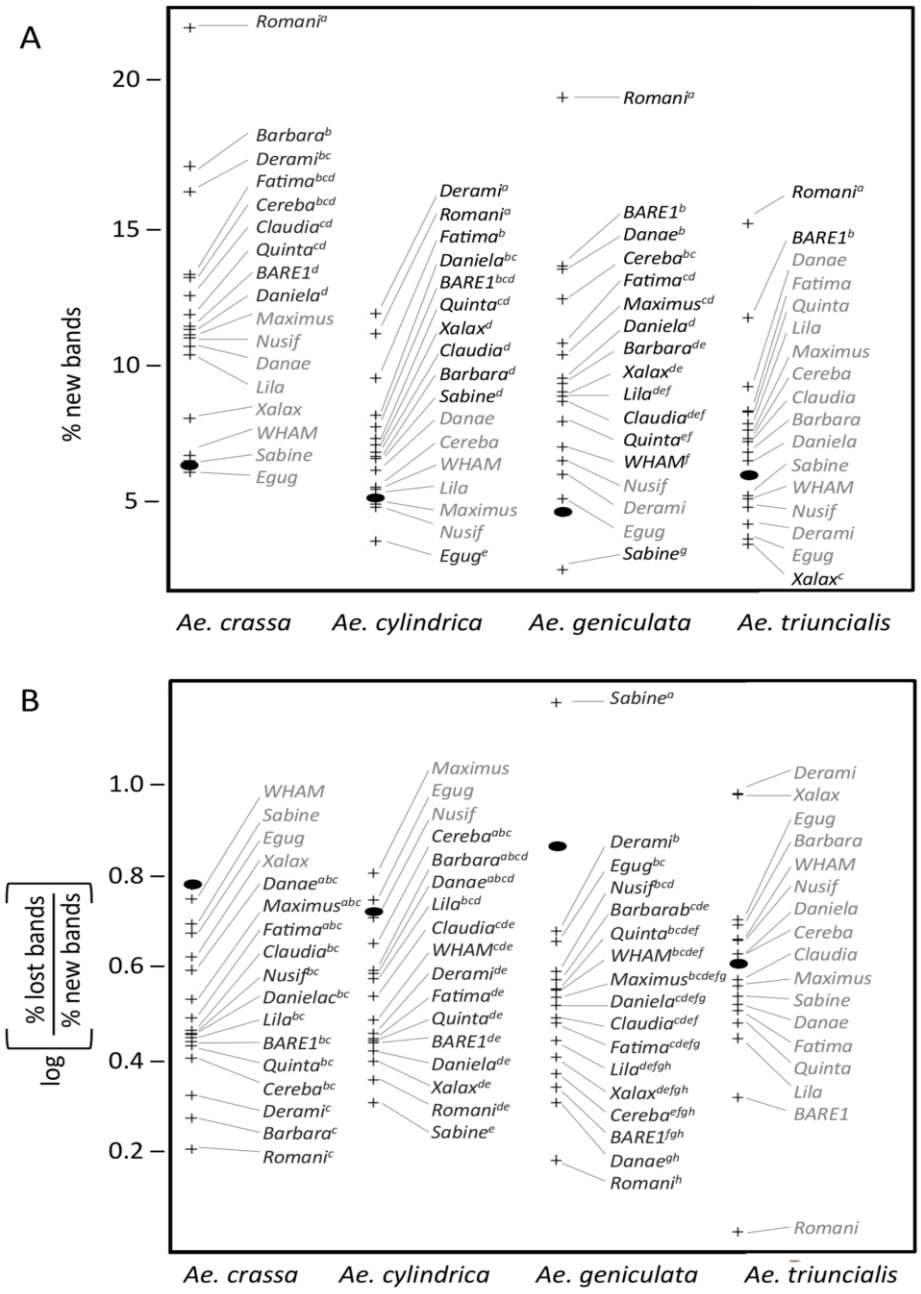


Figure 2. Deviation from the expected additivity of progenitors in *Aegilops* allopolyploid species. Proportions of deviating bands for the 17 TE families investigated with sequence specific amplified polymorphism (SSAP) and random sequences (amplified fragment length polymorphism, AFLP). (A) Proportions of new bands and (B) ratio between proportions of lost and new bands (log-scaled). TE families (crosses) presented in grey show non-significantly different proportions than AFLP (black circle), as assessed by Tukey tests. TE families differing significantly from AFLP (in black) and not sharing a small letters present significantly different proportions.

TE families with clear evidence of recent activity as assessed by 454 GSS (i.e. *Barbara, BARE1, Cereba, Claudia, Daniela, Danae, Derami, Fatima, Lila, Maximus, Quinta, Romani, Sabine, Wham* and *Xalax*) congruently showed higher proportions of new SSAP bands indicative of transposition after polyploidy in *Ae. cylindrica* and/or *Ae. geniculata* (see below). In contrast, molecular fingerprints detected non-significantly higher proportions of new SSAP bands than AFLP for *Nusif* and *Egug*. *Egug*, but also *Nusif* to a certain extent, indeed showed quiescence based on GSS data (also see Senerchia *et al*., 2013). Both high throughput sequencing and fingerprinting approaches showed congruent results, but their respective advantages deserve further attention.

Balance of lost/new bands in random sequences and in multiple TE families

The proportion of lost bands divided by the proportion on new bands (log-scaled) offered an integrative proxy of the evolutionary trajectories of TE families towards either sequence deletion (more than 1) or proliferation (less than 1) when compared to AFLP. Ratios for AFLP as well as TE family were usually larger than 1, indicating that sequence loss was the major restructuring process following polyploidy (Figure 2B). Tukey HSD tests assessed that all 17 TE families had significantly different ratio than AFLP in at least one of the four polyploids (Table S3). The number of TEs with significantly different ratio of lost/new bands proportions than AFLP varied considerably among polyploids. None of the TE family displayed significantly different ratio than the AFLP ratio in *Ae. triuncialis*, whereas twelve and ten TE families offered evidence of proliferation with significantly lower ratio than AFLP in *Ae. crassa* and *Ae. cylindrica*, respectively. In *Ae. geniculata*, all TEs presented significantly different ratio than AFLP with sixteen families showing lower ratio indicating proliferation, whereas *Sabine* presented a higher ratio indicating considerable sequence loss

after polyploidy. TE families thus showed species-specific long-term evolutionary trajectories following independent polyploidy events.

Most TEs with significantly different ratio of lost/new bands proportions than AFLP showed evidence of proliferation, but also revealed noticeable exceptions. In particular, *Sabine* showed considerable variation among tetraploid species, with clear evidence of proliferation in *Ae. cylindrica* (ratio of 2.05, significantly lower than AFLP), whereas considerable band loss was reported in *Ae. geniculata* (ratio of 13.94, significantly higher than AFLP). These results were firmly supported by random GSS that identified *Sabine* as the TE presenting the largest difference in number of reads, with more than 7000 out of 667'485 reads in *Ae. cylindrica* (i.e. 1.06% of the genome), whereas as few as 274 out of 646'327 reads (0.04%) were detected in *Ae. geniculata*. Semi-quantitative PCR using specific primers designed in the 300bp of the LTR of *Sabine* on the same DNA extracts further confirmed the contrasted abundance of TEs in the allopolyploid species (data not shown).

Correlation between progenitor divergence and restructuring after polyploidy

A linear mixed model nesting TE families within polyploids species as random-effects assessed that the proportion of non-shared bands between pairs of diploid progenitors was significantly associated with the proportion of new bands in the derived polyploid (Figure 3A, maximum likelihood linear mixed effect model t-value=4.18, $p<0.01$, $R^2=0.72$). Thus, for each TE family, the more divergent the arrangements of TE insertions among diploid genomes being merged, the more new SSAP bands indicative of transposition for that specific TE in the derived allotetraploid.

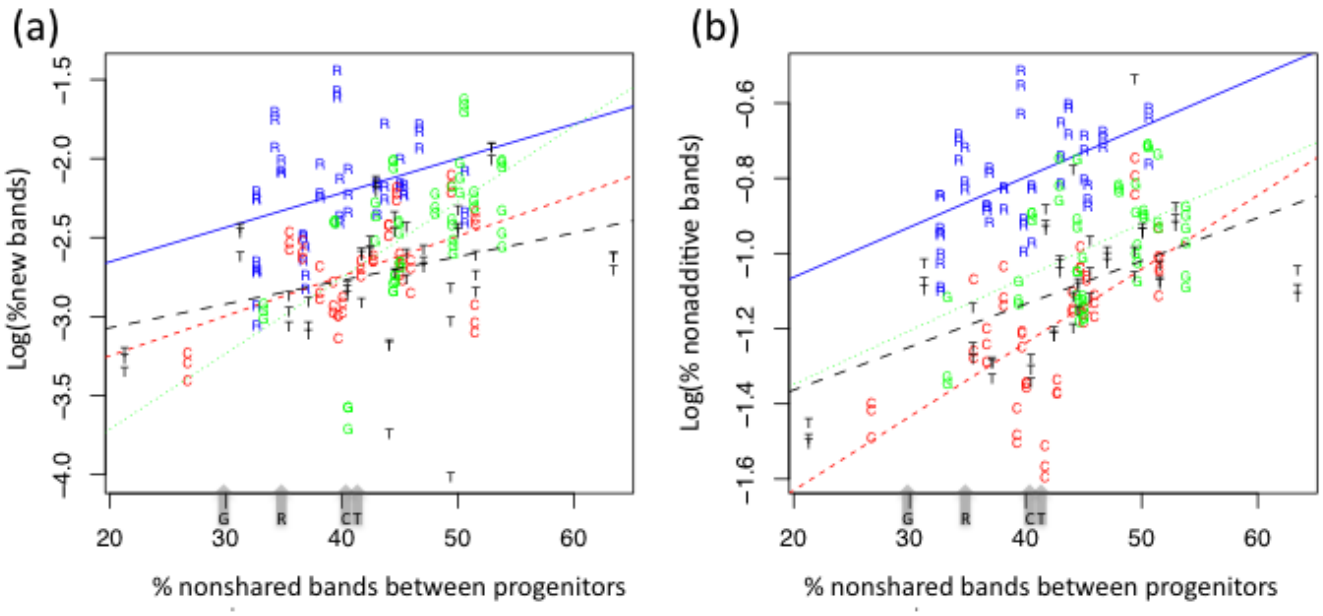


Figure 3. Association of divergence between progenitor species and restructuring in derived *Aegilops* polyploids. General linear mixed model associating proportions of parental non-shared bands between the progenitors with (A) the proportion of new bands or (B) the ratio between proportions of lost and new bands in the allopolyploids for each TE family. Grey arrows indicate proportions of non-shared bands between the progenitors species for random sequences (AFLP), whereas letters and lines show proportions and relationships, respectively, for the 17 TE families in *Ae. crassa* (R, solid), *Ae. cylindrica* (C, short dashed), *Ae. geniculata* (G, dotted) and *Ae. triuncialis* (T, long dashed).

A similar model assessed that the proportion of non-shared bands between pairs of diploid progenitors was significantly associated with the ratio between lost bands and new bands for each TE family (Figure 3B, linear mixed effect model t-value=-5.73, $p<0.01$, $R^2=0.74$). Thus, the higher the divergence between the respective diploid progenitors, the higher the proportion of non-additive bands (i.e. new bands and lost bands) in the derived polyploid.

Origin of lost bands following polyploidy

The proportion of lost bands originating from the pivotal vs. the differential genome was assessed on a subset of loci that were specific to each progenitor species (Table S4). The proportions of lost bands were significantly higher in the differential than in the pivotal genome in *Ae. crassa* ($p<0.001$, Kruskal-Wallis test chi-squared = 69.04) for all TEs except *Quinta* (Figure 4). Similarly, *Ae. triuncialis* ($p<0.001$, chi-squared = 29.77) and *Ae. cylindrica* ($p=0.036$, chi-squared = 4.39) showed significantly higher band loss from the differential genome, but only specific TEs significantly followed this pattern (*Claudia*, *Danae*, *Fatima*, *Nusif* and *Wham* for *Ae. triuncialis and BARE1*, *Maximus* and *Nusif* for *Ae. cylindrica*). In contrast, *Ae. geniculata* showed significantly higher proportions of lost bands from the pivotal genome ($p=0.001$, chi-squared = 50.81), with significant differences for *Cereba*, *Maximus*, *Romani*, *Sabine* and *Wham*.

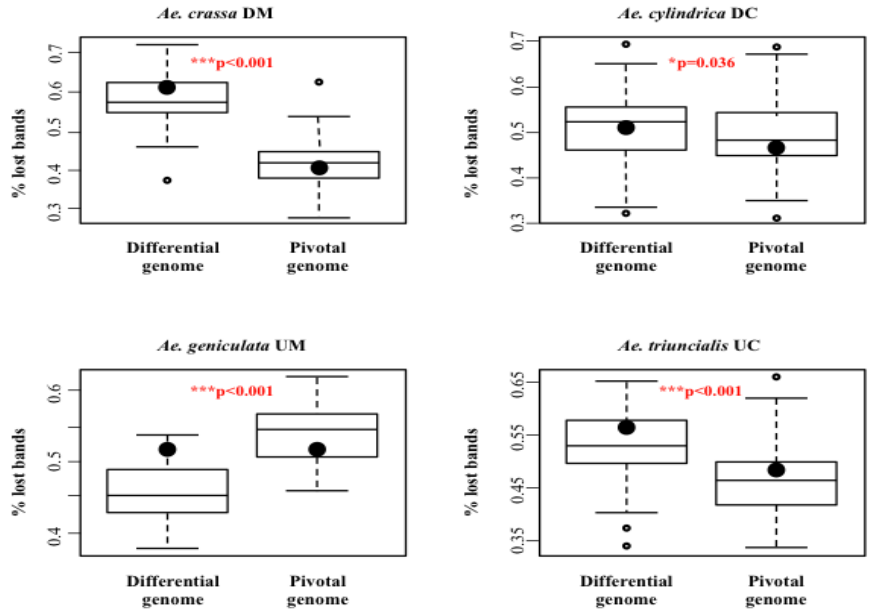


Figure 4 Proportion of lost bands originating from the differential and the pivotal genome. Boxplot of proportions of lost bands originating from the differential and the pivotal genome in the allotetraploid species *Aegilops crassa* (DM), *Ae. cylindrica* (DC), *Ae. geniculata* (UM) and *Ae. triuncialis* (UC). Significant differences were assessed by a Kruskal Wallis test. Medians and quartiles are shown as boxes, whereas whiskers show minimum and maximum values within 1.5 interquartile ranges. Open circles show outlier data. Black circles represent proportions of lost random sequences (AFLP).

Discussion

Restructuring of polyploid genomes

Combinations of different diploid genomes in four established *Aegilops* allotetraploids examined here with two complementary molecular fingerprint methods tracking either random (AFLP) or TE sequences (SSAP) revealed considerable departure from the expected additivity of their progenitors (Table 2). Despite a relatively low number of surveyed accessions per species, our quantitative handling of the presence/absence of hundreds of loci through a probabilistic approach sheds light on genome evolution in wild wheats, taking intraspecific variation into account. This approach offers significant novelties as compared to the traditional approach comparing single accessions through direct count and could be straightforwardly extended to large numbers of samples across the distribution ranges of species.

Genome restructuring highlighted here is congruent with GSS data (Senerchia *et al*., 2013) and coherent with the otherwise reported instability of the *Triticum-Aegilops* genomes over evolutionary times (Badaeva *et al.*, 2002; Badaeva *et al.*, 2004; Badaeva *et al.*, 2007; Brenchley *et al.*, 2012; Yaakov *et al.*, 2013). It further highlights allopolyploidy as a major process eliciting drastic genome changes (McClintock, 1984; Comai, 2000; Liu & Wendel, 2003) as otherwise reported in cultivated wheat (Kashkush *et al.*, 2003; Kraitshtein *et al.*, 2010; Feldman & Levy, 2012).

Aegilops is a radiating genus, whose polyploids have most likely originated during the Pleistocene (Baum *et al.*, 2012). Despite questionable dating of polyploidy events (Doyle & Egan, 2010), *Aegilops* allopolyploids were assumed of comparably recent origin. Accordingly, our results indicate variable levels of genome restructuring as illustrated by *Ae. crassa* presenting a higher proportion of non-additive bands than *Ae. cylindrica*, *Ae.*

geniculata or *Ae. triuncialis*. Noticeably, *Ae. crassa* presents an overall distinctive evolutionary trajectory as it is the only polyploid examined here showing genome upsizing (Eilam *et al.*, 2008). This species presents an atypically restricted distribution range for a polyploid *Aegilops* (Kilian *et al.*, 2011), but it remains unknown to what extent such particular genome dynamics is associated with processes acting at the level of natural populations (Bonchev & Parisod, 2013).

As expected under the pivotal-differential hypothesis proposed by Zohary and Feldman (1962) to explain chromosomal repatterning in polyploid species of the U-cluster, D and U genomes showed overall lower levels of band loss than differential genomes (here, C and M) in all polyploids except *Ae. geniculata*. Congruently, detailed cytogenetic works showed substantial modification of both parental genomes in *Ae. geniculata*, but also in *Ae. crassa* (Kimber *et al.*, 1988; Badaeva *et al.*, 2002; Badaeva *et al.*, 2004). The present results only partially match expectations raised by the pivotal-differential hypothesis and further indicate that restructuring does not consistently affect the larger parental genome as suggested by Bento *et al.* (2011). Accordingly, processes other than long-term gene flow between species and intergenomic recombination also drive genome reorganization in established polyploids (Feldman *et al.*, 2012). In particular, the specific trajectories of repeated sequences reported here suggest that the intrinsic TE dynamics played a significant role in restructuring polyploid genomes.

TE evolutionary trajectories following allopolyploidy

This study tracks several LTR retrotransposon families at a genome-wide scale, thus offering a powerful comparative framework to contrast their dynamics, particularly when compared to random sequences. Large numbers of SSAP loci were indeed scored for the 17 TE candidate families and systematically compared to AFLP proportions to minimize potential bias due to (i) polymorphic loci segregating among populations of the diploid species, (ii) non-additive bands appearing following molecular changes at insertion site and modifying the size of the amplified product rather than indicative of transposition events and (iii) the sensitivity of *Eco*RI restriction enzyme to rare cytosine methylation states (Cervera *et al*., 2003). Non-additive SSAP bands could indeed find their origin in more complex scenarios than transposition and deletions as shown by traditional studies that cloned and sequenced polymorphic SSAP and that reported new/lost bands corresponding to transposition/deletion events only to a certain extent (Petit *et al*., 2010; Sarilar *et al*., 2013). We provide here an original solution circumventing these limitations by relying on quantitative comparisons between SSAP (i.e. tracking TE insertions) and AFLP (i.e. tracking random sequences). As these techniques share very similar features, molecular events that are not TE specific (e.g. chromosomal rearrangements, deletions or introgression) result in non-additive bands in both SSAP and AFLP profiles, whereas TE-specific events are apparent in corresponding SSAP profiles only. Such a probabilistic approach does not designate specific bands, but highlight TEs showing significantly higher proportions of non-additive SSAP bands than AFLP profiles due to a predominance of TE-specific events (e.g. reorganization in TE-rich genome regions such as heterochromatin; Le Rouzic *et al*., 2007, Parisod *et al*., 2009). In loosely compartmentalized genomes such as *Aegilops*, SSAP vs. AFLP patterns can

thus be chiefly interpreted as evolutionary trajectories of the corresponding TEs (e.g. proliferation).

This study assessed TE families with significantly higher proportions of new SSAP bands than new AFLP bands and offered convincing support for proliferative transposition events after allopolyploidy. Noticeably, few TE families showed consistent proliferation in all polyploids. Only *BARE1*, a TE known to be active in *Triticeae* (Vicient *et al.*, 1999), and *Romani* always presented evidence of continual transpositional activity, whereas *Nusif* and *Egug* were seemingly quiescent. Other TE families presented evidence of proliferation after particular allopolyploidy events, but remained quiescent after others (Figure 2).

High proportions of lost bands highlighted sequence deletion as a major restructuring process following polyploidy. In particular, the balance between lost and new bands confirmed proliferation of a majority of TE families as compared to random sequences, but indicated that deletion of TE sequences was usually predominant. Such evolutionary trajectories are congruent with the general downsizing of polyploid genomes (Leitch & Bennett, 2004; Leitch *et al.*, 2008) as reported in *Ae. cylindrica*, *Ae. geniculata* or *Ae. triuncialis* (Ozkan *et al.*, 2003; Eilam *et al.*, 2007; Eilam *et al.*, 2008). Accordingly, the predominance of band loss reported here matches predictions raised by the increase/decrease model of genome size evolution as a balance between transposition and small deletions due to illegitimate recombination (Ma *et al.*, 2004; Parisod *et al.*, 2010; El Baidouri & Panaud, 2013). Consequently, TE genome fractions showed considerable turnover after independent polyploidy events, supporting polyploid genome divergence whose precise evolutionary causes and consequences remain not fully investigated.

The dynamics of multiple LTR retrotransposons revealed both species-specific and TE-specific dynamics in response to allopolyploidy as a common trigger. The remarkably contrasted evolutionary trajectory of *Sabine* in different polyploids is particularly illustrative of the variable evolutionary trajectories of TEs after polyploidy. Several lines of evidence indeed revealed that this retrotransposon presented significant proliferation in *Ae. cylindrica*, but massive elimination in *Ae. geniculata*, whereas it was apparently quiescent in the other polyploids. The molecular mechanisms underlying such deletion of *Sabine* sequences in *Ae. geniculata* remain elusive as illegitimate recombination may unlikely achieve such specific elimination of interspersed TE insertions (Devos *et al.*, 2002; El Baidouri & Panaud, 2013). The evolutionary processes driving contrasted evolutionary trajectories of TEs in different species deserve further attention.

Determinism of allopolyploid genome evolution

Despite species-specific and TE-specific evolutionary trajectories, TE dynamics following allopolyploidy seems non-random. Differences in arrangements of TE insertions between progenitors, rather than genome-wide differentiation, showed significant association with restructuring levels of corresponding TEs in the polyploids. In other words, for a given TE family, the higher the divergence between the diploid genome being merged, the higher the turnover of the corresponding TE fraction (i.e. new bands indicative of transposition and non-additive bands indicative of genome changes) in established polyploids examined here.

Several non-exclusive processes may account for TE dynamics after allopolyploidy (reviewed in Parisod *et al.*, 2010). The present results highlighting TE divergence among progenitors as an influential factor of the TE dynamics in allopolyploids largely support the genome shock hypothesis. Allopolyploidy indeed expectedly weakens key epigenetic

processes repressing TEs and lead to their activation (Bourc'his & Voinnet, 2010; Parisod & Senerchia, 2012). Consequently, the merging of specific parental genomes with divergent TEs at the origin of a particular allopolyploidy event could result in genomic conflicts would drive non-random restructuring of TE families (Tayalé & Parisod, 2013), as here reported in all polyploids examined.

Conclusions

The evolutionary trajectories of TEs do not necessarily match the dynamics of genomic backgrounds (e.g. Le Rouzic *et al.*, 2007; Parisod *et al.*, 2012). TEs can indeed show intrinsic dynamics through proliferation in response to various triggers, such as genome shocks, but their dynamics is also influenced by extrinsic factors such as sequence deletion (Kejnovsky *et al.*, 2009; Murat *et al.*, 2012; El Baidouri & Panaud, 2013) or evolutionary processes acting at the host level (Tenaillon *et al.*, 2010; Bonchev & Parisod, 2013).

TEs generally followed predictions raised by the pivotal-differential hypothesis in *Ae. crassa*, suggesting that extrinsic factors have shaped the evolutionary trajectories of TEs to a large extent. However, most TEs were evenly removed from both parental genomes and showed TE-specific trajectories as compared to random sequences in the majority of polyploids. Furthermore, the present study revealed long term restructuring of TE genome fractions that is remarkably coherent with expectations raised by intrinsic properties of TEs in response to the initial genome shock of allopolyploidy. Intrinsic and extrinsic processes may act in concert in driving the evolutionary trajectories of TEs (Ma & Gustafson, 2005), but their relative importance remains elusive. As polyploidy-induced genome shocks likely result in TE activation and genome restructuring during the first generations after the origin of new polyploid species (Ha *et al.*, 2009; Parisod & Senerchia, 2012), resynthesized polyploids

could be compared to established ones to distinguish processes acting after initial genome shock and gradual changes occurring throughout the lifespan of host species. In that respect, polyploid speciation is a promising model to investigate the multiple factors controlling TE dynamics.

Acknowledgements

We thank R. A. Slobodeanu for statistical advices, as well as M.-A. Grandbastien, A. Tayalé and anonymous reviewers for valuable comments on the manuscript. This work was funded by the Swiss National Science Foundation through the National Centre of Competence in Research 'Plant Survival' and a grant (PZ00P3-131950 to CP).

References

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

Badaeva ED, Dedkova OS, Gay G, Pukhalskyi VA, Zelenin AV, Bernard S, Bernard M. 2007. Chromosomal rearrangements in wheat: their types and distribution. *Genome* **50**: 907-926.

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*). *Genome* **55**: 177-193.

Bennetzen JL, Kellogg EA. 1997. Do plants have a one-way ticket to genomic obesity? *Plant Cell* **9**: 1509-1514.

Bento M, Gustafson JP, Viegas W, Silva M. 2011. Size matters in *Triticeae* polyploids: larger genomes have higher remodeling. *Genome* **54**: 175-183.

Bonchev G, Parisod C. 2013. Transposable elements and microevolutionary changes in natural populations. *Molecular Ecology Resources* **13**: 765-775.

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

Bourc'his D, Voinnet O. 2010. A small-RNA perspective on gametogenesis, fertilization, and early zygotic development. *Science* **330**: 617-622.

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 breadwheat genome using whole-genome shotgun sequencing. *Nature* **491**: 705-710.

Cervera MT, Ruiz-Garcia L, Martinez-Zapater JM. 2003. Analysis of DNA methylation in *Arabidopsis thaliana* based on methylation-sensitive AFLP markers. *Molecular Genetics and Genomics* **268**: 832-833.

Comai L. 2000. Genetic and epigenetic interactions in allopolyploid plants. *Plant Molecular Biology* **43**: 387-399.

Comai L, Madlung A, Josefsson C, Tyagi A. 2003. Do the different parental 'heteromes' cause genomic shock in newly formed allopolyploids? *Philosophical Transactions of the Royal Society B* **358**: 1149-1155.

Devos KM, Brown JKM, Bennetzen JL. 2002. Genome size reduction through illegitimate recombination counteracts genome expansion in *Arabidopsis*. *Genome Research* **12**: 1075-1079.

Doyle J, Flagel L, Paterson A, Rapp R, Soltis D. 2008. Evolutionary genetics of genome merger and doubling in plants. *Annual Review of Genetics* **42**: 443-461.

Doyle JJ, Egan AN. 2010. Dating the origins of polyploidy events. *New Phytologist* **186**: 73-85.

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*. *Genome* **51**: 616-627.

Eilam T, Anikster Y, Millet E, Manisterski J, Sag-Assif O, Feldman M. 2007. Genome size and genome evolution in diploid *Triticeae* species. *Genome* **50**: 1029-1037.

El Baidouri M, Panaud O. 2013. Comparative genomic paleontology across plant kingdom reveals the dynamics of TE-driven genome evolution. *Genome Biology and Evolution* **5**: 954-965.

Esnault C, Palavesam A, Pilitt K, O'Brochta DA. 2011. Intrinsic characteristics of neighboring DNA modulate transposable element activity in *Drosophila melanogaster*. *Genetics* **187**: 319-331.

Fedoroff NV. 2012. Transposable elements, epigenetics, and genome evolution. *Science* **338**: 758-767.
Feldman M. 1965a. Further evidence for natural hybridization between tetraploid species of *Aegilops* Section *Pleionathera*. *Evolution* **19**: 162-174.
Feldman M. 1965b. Fertility of interspecific F1 hybrids and hybrid derivatives involving tetraploid species of *Aegilops* Section *Pleionathera*. *Evolution* **19**: 566-562.
Feldman M. 1965c. Chromosome pairing between differential genomes in hybrids of tetraploid *Aegilops* species. *Evolution* **19**: 563-568.
Feldman M, Levy AA. 2012. Genome evolution due to allopolyploidization in wheat. *Genetics* **192**: 763-774.
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.
Ha M, Lu J, Tian L, Ramachandran V, Kasschau K. 2009. Small RNAs serve as a genetic buffer against genomic shock in *Arabidopsis* interspecific hybrids and allopolyploids. *Proceedings of the National Academy of Sciences of the United States of America* **106**: 17835-17840.
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**: 97-113.
Jurka J, Bao W, Kojima K. 2011. Families of transposable elements, population structure and the origin of species. *Biology Direct* **6**: 44.
Kashkush K, Feldman M, Levy AA. 2003. Transcriptional activation of retrotransposons alters the expression of adjacent genes in wheat. *Nature Genetics* **33**: 102-106.
Kejnovsky E, Leitch IJ, Leitch AR. 2009. Contrasting evolutionary dynamics between angiosperm and mammalian genomes. *Trends in Ecology and Evolution* **24**: 572-582.
Kilian B, Mammen K, Millet E, Sharma R, Graner A, Salamini F, Hammer K, O H 2011. *Aegilops*. In: Kole C ed. *Wild Crop Relatives: Genomic and Breeding Resources, Cereals*: Springer-Verlag Berlin Heidelberg, 1-17.
Kimber G, Sallee PJ, M.M. F. 1988. The interspecific and evolutionary relationships in *Triticum ovatum*. *Genome* **30**: 218-221.
Kraitshtein Z, Yaakov B, Khasdan V, Kashkush K. 2010. Genetic and epigenetic dynamics of a retrotransposon after allopolyploidization of wheat. *Genetics* **186**: 801-812.
Kumar A, Bennetzen JL. 1999. Plant retrotransposons. *Annual Review of Genetics* **33**: 479-532.
Le Rouzic A, Dupas S, Capy P. 2007. Genome ecosystem and transposable elements species. *Gene* **390**: 214-220.
Leitch IJ, Bennett MD. 2004. Genome downsizing in polyploid plants. *Biological Journal of the Linnean Society* **82**: 651-663.
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.
Levy AA 2013. Transposons in plant speciation. In: Fedoroff NV ed. *Plant transposons and genome dynamics in evolution*: John Wiley and Sons Ltd, Ames Iowa USA, p. 165-176.
Li WL, Zhang P, Fellers JP, Friebe B, Gill BS. 2004. Sequence composition, organization, and evolution of the core *Triticeae* genome. *Plant Journal* **40**: 500-511.
Librado P, Rozas J. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**: 1451-1452.
Liu B, Wendel JF. 2003. Epigenetic phenomena and the evolution of plant allopolyploids. *Molecular Phylogenetics and Evolution* **29**: 365-379.
Lynch M, Conery JS. 2003. The origins of genome complexity. *Science* **302**: 1401-1404.
Ma JX, Bennetzen JL. 2004. Rapid recent growth and divergence of rice nuclear genomes. *Proceedings of the National Academy of Sciences of the United States of America* **101**: 12404-12410.
Ma JX, Devos KM, Bennetzen JL. 2004. Analyses of LTR-retrotransposon structures reveal recent and rapid genomic DNA loss in rice. *Genome Research* **14**: 860-869.

Ma XF, Gustafson JP. 2005. Genome evolution of allopolyploids: a process of cytological and genetic diploidization. *Cytogenetics and Genome Research* **109**: 236-249.
McClintock B. 1984. The significance of responses of the genome to challenge. *Science* **226**: 792-801.
Meimberg H, Rice K, Milan N, Njoku C, McKay J. 2009. Multiple origins promote the ecological amplitude of allopolyploid *Aegilops* (Poaceae). *American Journal of Botany* **96**: 1262-1273.
Murat F, Van de Peer Y, Salse J. 2012. Decoding plant and animal genome plasticity from differential paleo-evolutionary patterns and processes. *Genome Biology and Evolution* **4**: 917-928.
Nakagawa S, Schielzeth H. 2013. A general and simple method for obtaining R^2 from generalized linear mixed-effects models. *Method in Ecology and Evolution* **4**: 133-142.
Ozkan H, Tuna M, Arumuganathan K. 2003. Nonadditive changes in genome size during allopolyploidization in the wheat (*Aegilops-Triticum*) group. *Journal of Heredity* **94**: 260-264.
Parisod C, Alix K, Just J, Petit M, Sarilar V, Mhiri C, Ainouche M, Chalhoub B, Grandbastien M-A. 2010. Impact of transposable elements on the organization and function of allopolyploid genomes. *New Phytologist* **186**: 37-45.
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, Mhiri C, Lim KY, Clarkson JJ, Chase MW, Leitch AR, Grandbastien M-A. 2012. Differential dynamics of transposable elements during long-term diploidization of *Nicotiana* section Repandae (*Solanaceae*) allopolyploid genomes. *PLoS ONE* **7**: e50352.
Parisod C, Senerchia N 2012. Responses of transposable elements to polyploidy. In: Grandbastien M-A, Casacuberta JM eds. *Plant Transposable Elements*: Springer-Verlag Berlin Heidelberg, 147-148.
Petit M, Guidat C, Daniel J, Denis E, Montoriol E, Bui QT, Lim KY, Kovarik A, Leitch AR, Grandbastien MA, *et al.* 2010. Mobilization of retrotransposons in synthetic allotetraploid tobacco. *New Phytologist* **186**: 135-147.
Rozen S, Skaletsky H. 2000. Primer3 on the WWW for general users and for biologist programmers. *Methods in Molecular Biology* **132**: 365-386.
Sabot F, Schulman AH 2009. Genomics of transposable elements in the *Triticeae*. In: Muehlbauer GJ, Feuille C eds. *Genetics and Genomics of the Triticeae*: Springer-Verlag, New-York USA, p. 387-405.
SanMiguel P, Gaut BS, Tikhonov A, Nakajima Y, Bennetzen JL. 1998. The paleontology of intergene retrotransposons of maize. *Nature Genetics* **20**: 43-45.
Sarilar V, Palacios PM, Rousselet A, Ridel C, Falque M, Eber F, Chevre A-M, Joets J, Brabant P, Alix K. 2013. Allopolyploidy has a moderate impact on restructuring at three contrasting transposable element insertion sites in resynthesized *Brassica napus* allotetraploids. *New Phytologist* **198**: 593-604.
Senerchia N, Wicker T, Felber F, Parisod C. 2013. Evolutionary dynamics of retrotransposons assessed by next generation sequencing in wild relatives of wheat. *Genome Biology and Evolution* **5**: 1010-1020.
Soltis DE, Soltis PS. 1999. Polyploidy: recurrent formation and genome evolution. *Trends in Ecology & Evolution* **14**: 348-352.
Tayalé A, Parisod C. 2013. Natural pathways to polyploidy in plants and consequences for genome reorganization. *Cytogenet Genome Research* **140**: 79-96**.**
Tenaillon M, Hollister J, Gaut B. 2010. A triptych of the evolution of plant transposable elements. *Trends in Plant Science* **15**: 471-478.
van Slageren MW. 1994. Wild wheats: a monograph of *Aegilops* L. and *Amblyopyrum* (Jaub. & Spach) Eig (*Poaceae*). *Wageningen Agricultural University Papers* **97**.
Vekemans X. 2002. AFLP-SURV version 1.0. Distributed by the author. *Laboratoire de Génétique et Ecologie Végétale, Université Libre de Bruxelles, Belgium.*

Vicient CM, Suoniemi A, Anamthawat-Jonsson 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.

Vitte C, Panaud O. 2005. LTR retrotransposons and flowering plant genome size: emergence of the increase/decrease model. *Cytogenetics and Genome Research* **110**: 91-107.

Wang JB, Wang C, Shi SH, Zhong Y. 2000. Evolution of parental ITS regions of nuclear rDNA in allopolyploid *Aegilops* (Poaceae) species. *Hereditas* **133**: 1-7.

Wang Q, Dooner HK. 2006. Remarkable variation in maize genome structure inferred from haplotype diversity at the bz locus. *Proceedings of the National Academy of Sciences of the United States of America* **103**: 17644-17649.

Wicker T 2013. So many repeats and so little time: how to classify transposable elements. In: Grandbastien M-A, Casacuberta JM eds. *Plant Transposable Elements*: Springer-Verlag Berlin Heidelberg, 1-15.

Wicker T, Buell CR 2009. Gene and repetitive sequence annotation in the *Triticeae*. In: Muehlbauer GJ, Feuillet C eds. *Genetics and Genomics of the Triticeae*: Springer-Verlag, New-York USA, p. 407-425.

Wicker T, Sabot F, Hua Van A, Bennetzen J, 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.

Yaakov B, Kashkush K. 2011. Massive alterations of the methylation patterns around DNA transposons in the first four generations of a newly formed wheat allohexaploid. *Genome* **54**: 42-49.

Yaakov B, Meyer K, Ben-David S, Kashkush K. 2013. Copy number variation of transposable elements in *Triticum-Aegilops* genus suggests evolutionary and revolutionary dynamics following allopolyploidization. *Plant Cell Reports* **32**: 1615-1624.

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

Supplementary Material

Supplementary Table S1 Sequences of adaptors (5'−> 3'), pre-selective and selective primers used for amplified fragment length polymorphism (AFLP), and sequence specific amplified polymorphism (SSAP) techniques. Error rates were evaluated on 5 replicates on the final binary matrix, following Bonin et al (2004; How to track and assess genotyping errors in population genetics studies. *Molecular Ecology* **13**: 3261-3273).

Primer type	TE family	Primer sequence 5'-3'	*Mse*I selective primer	Error rate %
Selective	*Barbara*	TTG TTA GGA ATT ACC AAA ACC AC	M-CAC, M-CCA, M-CTG	5.36
	BARE1	GCC TCT AGG GCA TAT TTC CA	M-CAC, M-CCA, M-CTA, M-CCG, M-CAA	6.01
	Cereba	GCG AGC TCT AGT AAT TGG TCC AG	M-CAC, M-CCA, M-CGC, M-CCG	4.93
	Claudia	ACC TGT GTT ATC CTG CAA TTC TCA CA	M-CAC, M-CCA, M-CGA, M-CTG, M-CTA, M-CCG, M-CAA	3.79
	Danae	TGT ACC CCT CTT TTT GTA CAC C	M-CAC, M-CCA, M-CGA, M-CTG, M-CCG, M-CAA	3.33
	Daniela	GGT CTT CTC TTG GCG CTT C	M-CAC, M-CCA, M-CGA, M-CTG, M-CGC, M-CCG	2.17
	Derami	TTG ATA ATA TAA CAG CAT GGA ACA A	M-CAC, M-CCA	5.00
	Egug	TGA AAG TTG TAT AAT ATT GGC ATG G	M-CAC, M-CCA, M-CGA, M-CTA, M-CAA	4.86
	Fatima	ATC TTG TAT CTT CAT AGT CCT GGA GTC	M-CAC, M-CCA, M-CTG, M-CTA, M-CCG, M-CAA	3.75
	Lila	CAG GGG TAG TCG GTT AGG CTA C	M-CAC, M-CCA, M-CGA, M-CTA, M-CCG, M-CAA	4.35
	Maximus	AAG TCT TCA ATA CAC CAA AAT CA	M-CCA, M-CTG, M-CCG	1.92
	Nusif	AGG CCG CAG GCA GGC AAG	M-CAC, M-CGA, M-CTG, M-CTA, M-CGC	4.80
	Quinta	TGC ATG TAT GCA ATA TTT TTG GTC GT	M-CCA, M-CTG, M-CTA, M-CCG, M-CAA	5.99
	Romani	TGA TGA TCT AAT GTA TAA CAT TCC AAA	M-CAC, M-CCA, M-CGA, M-CGC, M-CCG	1.14
	Sabine	GCA TGG AGC AAG CAG AAA TTA T	M-CAC, M-CGA, M-CTA, M-CGC	6.60
	Wham	AGC GTC AGT CCG GGT TAG TTC CA	M-CCA, M-CGA, M-CTG, M-CTA, M-CGC	6.36
	Xalax	GGT TCT GAA ATC AAA AGA TAA ATC C	M-CAC, M-CCA, M-CGA, M-CTG, M-CGC	4.41
	AFLP	E-AAG	M-CGA, M-CGC	
	AFLP	E-AAC	M-CGA, M-CTA	
	AFLP	E-AGG	M-CGA, M-CGC	
	AFLP	E-AGA	M-CAC, M-CGA	
	AFLP	E-ACG	M-CAC, M-CCA, M-CTA, M-CCG, M-CGC	
	AFLP	E-ATT	M-CAC, M-CCA, M-CCG, M-CAA	3.36
Adaptors *Eco*R1	all	CTC GTA GAC TGC GTA CC		
	all	AAT TGG TAC GCA GTC TAC		
Adaptors *Mse*1	all	GAC GAT GAG TCC TGA G		
	all	TAC TCA GGA GTC AT		
Pre-selective *Eco*R1 (E)	all	GAC TGC GTA CCA ATT CA		
Pre-selective *Mse*1 (M)	all	GAT GAG TCC TGA GTA AC		

Supplementary Table S2 Mean number of loci in the various diploid and tetraploid *Aegilops* species for random sequences (AFLP) and for each investigated LTR retrotransposon family (SSAP).

	Ae. tauschii D	*Ae. cylindrica* CD	*Ae. caudata* C	*Ae. triuncialis* CU	*Ae. umbellulata* U	*Ae. geniculata* UM	*Ae. comosa* M	*Ae. crassa* DM	total
AFLP	687	680	648	685	644	609	723	584	1224
Claudia	262	290	244	275	236	299	264	271	238
Danae	198	227	199	218	192	236	169	179	792
Quinta	214	245	207	250	229	244	210	223	501
Daniela	264	337	279	346	342	350	285	276	606
Barbara	86	105	100	98	106	90	74	93	503
Fatima	218	333	327	357	291	318	272	233	703
Maximus	132	132	134	151	144	151	147	133	280
BARE1	387	432	342	409	328	418	391	418	379
Egug	258	254	233	247	235	236	235	232	718
Lila	271	299	271	293	245	279	247	279	505
Cereba	181	196	195	218	214	242	193	194	298
Derami	93	119	122	118	153	157	152	142	354
Romani	175	200	114	196	160	232	160	223	471
Nusif	164	182	180	202	191	202	170	172	626
Sabine	114	172	163	160	129	103	127	109	238
WHAM	104	118	94	105	101	111	96	90	187
Xalax	166	205	172	163	180	191	164	157	329

Supplementary Table S3 Multiple comparisons of (A) proportions of new bands and (B) ratio between the proportions of lost and new bands (log-scaled) between random sequences (AFLP) and each LTR retrotransposon family (SSAP) by Tukey HSD. AFLP or TE families not sharing a small letters have significantly different proportions.

(A) New bands	*Ae. crassa*	*Ae. cylindrica*	*Ae. geniculata*	*Ae. triuncialis*
AFLP	ghi	ij	i	cdefgh
Barbara	b	defg	de	cdefg
BARE1	def	bcd	b	b
Cereba	bcd	fghij	bc	cdef
Claudia	cde	def	def	cdef
Danae	defghi	efghi	b	bc
Daniela	def	bc	d	cdefg
Derami	bc	a	ghi	ghi
Egug	i	k	hi	hi
Fatima	bcd	b	cd	cd
Lila	defghi	hij	def	cde
Maximus	defg	jk	cd	cdef
Nusif	defgh	jk	ghi	fghi
Quinta	cde	cde	efg	cd
Romani	a	a	a	a
Sabine	hi	defgh	j	defghi
Wham	fghi	ghij	fgh	efghi
Xalax	efghi	def	de	i
(B) ratio lost/new bands	*Ae. crassa*	*Ae. cylindrica*	*Ae. geniculata*	*Ae. triuncialis*
AFLP	a	abc	b	ab
Barbara	e	bcde	cdef	ab
BARE1	de	ef	ghi	b
Cereba	de	bcd	fghi	ab
Claudia	de	def	cdefgh	ab
Danae	bcd	bcde	hi	ab
Daniela	de	ef	defgh	ab
Derami	e	ef	c	a
Egug	abc	ab	cd	ab
Fatima	cde	ef	defgh	ab
Lila	de	cde	efghi	ab
Maximus	cde	a	cdefgh	ab
Nusif	de	abc	cde	ab
Quinta	de	ef	cdefg	ab
Romani	e	ef	i	a
Sabine	abc	f	a	ab
Wham	ab	def	cdefgh	ab
Xalax	abcd	ef	efghi	ab

Supplementary Table S4 Proportions of lost bands originating from the differential vs. the pivotal genome of each *Aegilops* tetraploid species.

Tetraploid species	TE family	% Lost bands from the differential genome	% Lost bands from the pivotal genome
Ae. crassa	*AFLP*	0.60	0.40
Ae. crassa	*Barbara*	0.61	0.39
Ae. crassa	*BARE1*	0.62	0.38
Ae. crassa	*Cereba*	0.58	0.42
Ae. crassa	*Claudia*	0.65	0.35
Ae. crassa	*Danae*	0.58	0.42
Ae. crassa	*Daniela*	0.61	0.39
Ae. crassa	*Derami*	0.44	0.56
Ae. crassa	*Egug*	0.57	0.43
Ae. crassa	*Fatima*	0.57	0.43
Ae. crassa	*Lila*	0.62	0.38
Ae. crassa	*Maximus*	0.60	0.40
Ae. crassa	*Nusif*	0.60	0.40
Ae. crassa	*Quinta*	0.54	0.46
Ae. crassa	*Romani*	0.58	0.42
Ae. crassa	*Sabine*	0.56	0.44
Ae. crassa	*Wham*	0.57	0.43
Ae. crassa	*Xalax*	0.56	0.44
Ae. cylindrica	*AFLP*	0.49	0.51
Ae. cylindrica	*Barbara*	0.55	0.45
Ae. cylindrica	*BARE1*	0.63	0.37
Ae. cylindrica	*Cereba*	0.54	0.46
Ae. cylindrica	*Claudia*	0.54	0.46
Ae. cylindrica	*Danae*	0.52	0.48
Ae. cylindrica	*Daniela*	0.48	0.52
Ae. cylindrica	*Derami*	0.47	0.53
Ae. cylindrica	*Egug*	0.43	0.57
Ae. cylindrica	*Fatima*	0.33	0.67
Ae. cylindrica	*Lila*	0.52	0.48
Ae. cylindrica	*Maximus*	0.62	0.38
Ae. cylindrica	*Nusif*	0.60	0.40
Ae. cylindrica	*Quinta*	0.46	0.54
Ae. cylindrica	*Romani*	0.52	0.48
Ae. cylindrica	*Sabine*	0.46	0.54
Ae. cylindrica	*Wham*	0.51	0.49
Ae. cylindrica	*Xalax*	0.57	0.43
Ae. geniculata	*AFLP*	0.50	0.50
Ae. geniculata	*Barbara*	0.51	0.49
Ae. geniculata	*BARE1*	0.43	0.57

Ae. geniculata	*Cereba*	0.41	0.59
Ae. geniculata	*Claudia*	0.46	0.54
Ae. geniculata	*Danae*	0.47	0.53
Ae. geniculata	*Daniela*	0.46	0.54
Ae. geniculata	*Derami*	0.47	0.53
Ae. geniculata	*Egug*	0.49	0.51
Ae. geniculata	*Fatima*	0.53	0.47
Ae. geniculata	*Lila*	0.50	0.50
Ae. geniculata	*Maximus*	0.42	0.58
Ae. geniculata	*Nusif*	0.45	0.55
Ae. geniculata	*Quinta*	0.44	0.56
Ae. geniculata	*Romani*	0.40	0.60
Ae. geniculata	*Sabine*	0.41	0.59
Ae. geniculata	*Wham*	0.43	0.57
Ae. geniculata	*Xalax*	0.50	0.50
Ae. triuncialis	*AFLP*	0.54	0.46
Ae. triuncialis	*Barbara*	0.53	0.47
Ae. triuncialis	*BARE1*	0.49	0.51
Ae. triuncialis	*Cereba*	0.51	0.49
Ae. triuncialis	*Claudia*	0.60	0.40
Ae. triuncialis	*Danae*	0.60	0.40
Ae. triuncialis	*Daniela*	0.47	0.53
Ae. triuncialis	*Derami*	0.51	0.49
Ae. triuncialis	*Egug*	0.51	0.49
Ae. triuncialis	*Fatima*	0.55	0.45
Ae. triuncialis	*Lila*	0.59	0.41
Ae. triuncialis	*Maximus*	0.56	0.44
Ae. triuncialis	*Nusif*	0.61	0.39
Ae. triuncialis	*Quinta*	0.51	0.49
Ae. triuncialis	*Romani*	0.54	0.46
Ae. triuncialis	*Sabine*	0.38	0.62
Ae. triuncialis	*Wham*	0.59	0.41
Ae. triuncialis	*Xalax*	0.51	0.49

Supplementary Note S1 Structure of genome-wide loci and TE genome fractions among four diploid and allotetraploid *Aegilops* species.

For AFLP as well as SSAP for the 17 different TE families, variance within and among species was partitioned by between-eigen analyses (BPCA) considering all axes and using the package 'adegenet' in R cran (Jombart, 2008). This procedure, analogous to *F*-statistics, is based on the frequencies of bands rather than alleles and maximizes between-group variance, thus offering adequate estimates for both diploids and polyploids (Parisod & Christin, 2008, and references therein). Genetic differentiation between pairs of species was estimated through multidimensional euclidian distances between centroids of species using the function 'dist'.

Correlation between matrices of pairwise distances among species for AFLP and the 17 SSAP datasets were further tested by multiple Mantel tests with 9'999'999 permutations using 'Manteln' (Ray & Excoffier, 2003). Significance was assessed with sequential Bonferroni correction (i.e. α starting at $7.6\ 10^{-7}$) to account for multiple comparison (Rice, 1989).
Associations were all non-significant after such a stringent test, indicating that all TEs and genome-wide loci reveal distinct patterns of differentiation among species.

References

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

Parisod C, Christin PA. 2008. Genome-wide association to fine-scale ecological heterogeneity within a continuous population of *Biscutella laevigata* (Brassicaceae). *New Phytologist* **178**: 436-447.
Ray N, Excoffier L. 2003. Manteln 1.1 Software for the computation of mantel tests for multiple matrices. *Genetics and biometry*, Laboratory of Geneva, Switzerland.
Rice WR. 1989. Analyzing tables of statistical tests. *Evolution* **43**:223-225.

AFLP

	Ae. caudata	*Ae. comosa*	*Ae. crassa*	*Ae. cylindrica*	*Ae. geniculata*	*Ae. tauschii*	*Ae. triuncialis*	*Ae. umbellulata*
Ae. caudata	-							
Ae. comosa	18.34	-						
Ae. crassa	18.84	18.65	-					
Ae. cylindrica	15.70	20.07	17.49	-				
Ae. geniculata	18.99	16.20	16.95	19.36	-			
Ae. tauschii	18.94	18.65	15.88	15.82	19.06	-		
Ae. triuncialis	15.85	18.50	18.30	18.01	16.98	18.99	-	
Ae. umbellulata	19.47	18.73	18.15	19.87	15.92	18.95	15.10	-

Barbara

	Ae. caudata	*Ae. comosa*	*Ae. crassa*	*Ae. cylindrica*	*Ae. geniculata*	*Ae. tauschii*	*Ae. triuncialis*	*Ae. umbellulata*
Ae. caudata	-							
Ae. comosa	8.03	-						
Ae. crassa	7.82	8.06	-					
Ae. cylindrica	6.79	8.69	8.01	-				
Ae. geniculata	8.02	7.21	8.51	8.86	-			
Ae. tauschii	8.13	7.38	6.83	6.36	7.45	-		
Ae. triuncialis	7.08	7.82	8.62	8.22	6.91	7.79	-	
Ae. umbellulata	8.13	8.34	8.97	9.17	6.82	8.24	6.56	-

BARE1

	Ae. caudata	*Ae. comosa*	*Ae. crassa*	*Ae. cylindrica*	*Ae. geniculata*	*Ae. tauschii*	*Ae. triuncialis*	*Ae. umbellulata*
Ae. caudata	-							
Ae. comosa	15.00	-						
Ae. crassa	15.12	14.83	-					
Ae. cylindrica	13.97	16.41	14.09	-				
Ae. geniculata	15.72	13.56	14.78	17.03	-			
Ae. tauschii	14.51	15.01	12.43	10.97	14.99	-		
Ae. triuncialis	13.14	15.22	14.76	15.47	14.55	14.97	-	
Ae. umbellulata	14.99	15.47	15.20	16.98	15.03	15.28	13.10	-

Cereba

	Ae. caudata	*Ae. comosa*	*Ae. crassa*	*Ae. cylindrica*	*Ae. geniculata*	*Ae. tauschii*	*Ae. triuncialis*	*Ae. umbellulata*
Ae. caudata	-							
Ae. comosa	10.83	-						
Ae. crassa	11.03	11.60	-					
Ae. cylindrica	9.38	12.27	10.66	-				
Ae. geniculata	11.79	10.94	12.61	13.31	-			
Ae. tauschii	10.54	11.10	9.74	8.74	12.18	-		
Ae. triuncialis	10.34	12.48	12.48	11.74	11.53	12.22	-	
Ae. umbellulata	12.23	13.06	13.15	13.76	11.95	12.96	9.75	-

Claudia

	Ae. caudata	*Ae. comosa*	*Ae. crassa*	*Ae. cylindrica*	*Ae. geniculata*	*Ae. tauschii*	*Ae. triuncialis*	*Ae. umbellulata*
Ae. caudata	-							
Ae. comosa	13.86	-						
Ae. crassa	14.22	14.31	-					
Ae. cylindrica	11.45	15.63	13.23	-				
Ae. geniculata	14.02	11.73	14.81	15.33	-			
Ae. tauschii	13.06	13.92	11.59	10.14	14.38	-		
Ae. triuncialis	12.16	15.43	15.75	13.49	13.46	14.54	-	
Ae. umbellulata	14.25	15.14	15.52	15.11	12.65	14.43	10.27	-

Danae

	Ae. caudata	*Ae. comosa*	*Ae. crassa*	*Ae. cylindrica*	*Ae. geniculata*	*Ae. tauschii*	*Ae. triuncialis*	*Ae. umbellulata*
Ae. caudata	-							
Ae. comosa	13.27	-						
Ae. crassa	12.02	12.52	-					
Ae. cylindrica	9.88	14.32	11.66	-				
Ae. geniculata	13.81	11.47	12.97	14.50	-			
Ae. tauschii	12.36	13.66	10.08	9.84	14.15	-		
Ae. triuncialis	11.90	14.33	13.54	13.35	13.25	14.24	-	
Ae. umbellulata	12.91	13.57	12.46	14.01	11.53	13.33	10.19	-

Daniela

	Ae. caudata	*Ae. comosa*	*Ae. crassa*	*Ae. cylindrica*	*Ae. geniculata*	*Ae. tauschii*	*Ae. triuncialis*	*Ae. umbellulata*
Ae. caudata	-							
Ae. comosa	14.94	-						
Ae. crassa	14.76	15.13	-					
Ae. cylindrica	12.14	16.10	15.21	-				
Ae. geniculata	14.81	13.50	15.04	16.34	-			
Ae. tauschii	14.35	14.35	12.66	12.91	14.75	-		
Ae. triuncialis	12.88	16.20	16.05	14.44	14.19	15.81	-	
Ae. umbellulata	15.76	16.05	15.99	16.67	13.48	16.33	12.30	-

Derami

	Ae. caudata	*Ae. comosa*	*Ae. crassa*	*Ae. cylindrica*	*Ae. geniculata*	*Ae. tauschii*	*Ae. triuncialis*	*Ae. umbellulata*
Ae. caudata	-							
Ae. comosa	9.27	-						
Ae. crassa	9.52	9.40	-					
Ae. cylindrica	8.17	10.08	8.97	-				
Ae. geniculata	9.22	7.67	9.67	10.66	-			
Ae. tauschii	9.59	10.13	10.25	8.79	10.42	-		
Ae. triuncialis	7.46	9.07	9.29	8.82	8.25	8.27	-	
Ae. umbellulata	9.76	9.64	10.31	10.46	7.99	10.26	7.43	-

Egug

	Ae. caudata	*Ae. comosa*	*Ae. crassa*	*Ae. cylindrica*	*Ae. geniculata*	*Ae. tauschii*	*Ae. triuncialis*	*Ae. umbellulata*
Ae. caudata	-							
Ae. comosa	9.50	-						
Ae. crassa	9.55	9.53	-					
Ae. cylindrica	8.22	10.50	10.07	-				
Ae. geniculata	9.49	8.37	9.72	10.56	-			
Ae. tauschii	9.99	9.92	8.57	8.71	9.80	-		
Ae. triuncialis	8.03	9.50	9.94	9.29	8.34	10.28	-	
Ae. umbellulata	9.91	9.71	10.08	10.69	8.41	10.02	7.62	-

Fatima

	Ae. caudata	*Ae. comosa*	*Ae. crassa*	*Ae. cylindrica*	*Ae. geniculata*	*Ae. tauschii*	*Ae. triuncialis*	*Ae. umbellulata*
Ae. caudata	-							
Ae. comosa	14.61	-						
Ae. crassa	13.78	14.53	-					
Ae. cylindrica	10.55	16.35	15.30	-				
Ae. geniculata	14.64	14.00	14.94	16.32	-			
Ae. tauschii	14.85	14.72	12.31	14.94	16.25	-		
Ae. triuncialis	12.81	15.87	15.12	14.45	13.54	17.04	-	
Ae. umbellulata	15.29	15.85	14.79	17.05	12.95	16.53	12.27	-

Lila

	Ae. caudata	*Ae. comosa*	*Ae. crassa*	*Ae. cylindrica*	*Ae. geniculata*	*Ae. tauschii*	*Ae. triuncialis*	*Ae. umbellulata*
Ae. caudata	-							
Ae. comosa	12.25	-						
Ae. crassa	11.52	12.38	-					
Ae. cylindrica	10.11	14.00	11.28	-				
Ae. geniculata	12.45	11.23	12.62	13.92	-			
Ae. tauschii	11.62	12.64	10.46	9.60	12.46	-		
Ae. triuncialis	10.72	13.34	13.20	12.43	11.11	13.00	-	
Ae. umbellulata	12.23	12.89	13.15	13.54	11.15	13.16	9.89	-

Maximus

	Ae. caudata	*Ae. comosa*	*Ae. crassa*	*Ae. cylindrica*	*Ae. geniculata*	*Ae. tauschii*	*Ae. triuncialis*	*Ae. umbellulata*
Ae. caudata	-							
Ae. comosa	9.42	-						
Ae. crassa	9.33	9.37	-					
Ae. cylindrica	8.04	9.70	8.10	-				
Ae. geniculata	9.06	8.49	9.21	9.54	-			
Ae. tauschii	9.46	9.23	7.71	6.66	9.78	-		
Ae. triuncialis	8.32	10.16	9.47	9.53	9.38	10.28	-	
Ae. umbellulata	9.78	10.28	9.71	10.38	9.39	10.32	7.48	-

Nusif

	Ae. caudata	*Ae. comosa*	*Ae. crassa*	*Ae. cylindrica*	*Ae. geniculata*	*Ae. tauschii*	*Ae. triuncialis*	*Ae. umbellulata*
Ae. caudata	-							
Ae. comosa	10.14	-						
Ae. crassa	10.32	10.01	-					
Ae. cylindrica	8.64	10.97	9.78	-				
Ae. geniculata	10.49	9.20	10.29	11.37	-			
Ae. tauschii	10.24	9.93	8.49	7.73	10.50	-		
Ae. triuncialis	9.24	11.18	11.15	10.94	9.33	11.37	-	
Ae. umbellulata	10.48	11.26	11.06	12.00	8.85	11.34	7.47	-

Quinta

	Ae. caudata	*Ae. comosa*	*Ae. crassa*	*Ae. cylindrica*	*Ae. geniculata*	*Ae. tauschii*	*Ae. triuncialis*	*Ae. umbellulata*
Ae. caudata	-							
Ae. comosa	10.54	-						
Ae. crassa	11.46	10.83	-					
Ae. cylindrica	9.07	11.66	11.69	-				
Ae. geniculata	12.31	10.39	12.07	12.96	-			
Ae. tauschii	10.69	10.89	10.39	9.45	11.80	-		
Ae. triuncialis	10.37	12.23	12.64	11.70	11.01	12.22	-	
Ae. umbellulata	12.24	12.70	12.63	12.99	10.81	12.53	9.75	-

Romani

	Ae. caudata	*Ae. comosa*	*Ae. crassa*	*Ae. cylindrica*	*Ae. geniculata*	*Ae. tauschii*	*Ae. triuncialis*	*Ae. umbellulata*
Ae. caudata	-							
Ae. comosa	12.11	-						
Ae. crassa	14.02	14.62	-					
Ae. cylindrica	11.55	14.90	14.11	-				
Ae. geniculata	14.59	12.27	15.72	15.88	-			
Ae. tauschii	12.66	13.70	12.71	10.19	14.54	-		
Ae. triuncialis	11.38	13.94	15.13	13.92	13.63	14.13	-	
Ae. umbellulata	12.79	13.76	14.61	15.10	13.53	13.46	9.94	-

Sabine

	Ae. caudata	*Ae. comosa*	*Ae. crassa*	*Ae. cylindrica*	*Ae. geniculata*	*Ae. tauschii*	*Ae. triuncialis*	*Ae. umbellulata*
Ae. caudata	-							
Ae. comosa	8.15	-						
Ae. crassa	8.84	6.61	-					
Ae. cylindrica	6.08	8.80	9.23	-				
Ae. geniculata	8.99	7.42	7.62	9.82	-			
Ae. tauschii	8.47	7.12	6.01	8.24	7.90	-		
Ae. triuncialis	6.45	8.74	9.39	8.05	8.83	8.99	-	
Ae. umbellulata	8.30	7.94	8.28	9.04	7.56	8.51	7.46	-

WHAM

	Ae. caudata	*Ae. comosa*	*Ae. crassa*	*Ae. cylindrica*	*Ae. geniculata*	*Ae. tauschii*	*Ae. triuncialis*	*Ae. umbellulata*
Ae. caudata	-							
Ae. comosa	7.83	-						
Ae. crassa	7.47	7.69	-					
Ae. cylindrica	6.45	8.23	7.22	-				
Ae. geniculata	7.37	6.19	7.42	7.86	-			
Ae. tauschii	7.62	7.90	5.93	6.03	7.77	-		
Ae. triuncialis	6.39	8.15	7.94	7.48	7.00	7.73	-	
Ae. umbellulata	7.82	8.26	7.50	8.47	6.98	7.80	5.52	-

Xalax

	Ae. caudata	*Ae. comosa*	*Ae. crassa*	*Ae. cylindrica*	*Ae. geniculata*	*Ae. tauschii*	*Ae. triuncialis*	*Ae. umbellulata*
Ae. caudata	-							
Ae. comosa	9.38	-						
Ae. crassa	9.65	8.96	-					
Ae. cylindrica	8.52	10.61	10.15	-				
Ae. geniculata	9.66	8.71	10.14	10.39	-			
Ae. tauschii	9.57	9.60	8.49	8.03	9.84	-		
Ae. triuncialis	7.81	9.47	9.80	9.60	8.49	9.34	-	
Ae. umbellulata	10.12	9.74	9.87	10.97	8.58	10.03	8.19	-

Correlation / p-value	AFLP	*Barbara*	*BARE1*	*Cereba*	*Claudia*	*Danae*	*Daniela*	*Derami*	*Egug*	*Fatima*	*Lila*	*Maxmus*	*Nusif*	*Quinta*	*Romani*	*Sabine*	*WHAM*	*Xalax*
			0.7939	0.6978	0.8314	0.8361	0.8637	0.7013	0.8984	0.7980	0.8518	0.7919	0.8721	0.7463	0.6742	0.6061	0.91069	0.83859
AFLP	-	0.7851	85	99	35	84	23	93	02	94	4	78	95	72	3	58	8	2
	0.0000		0.7854	0.8013	0.8580	0.6965	0.8784	0.6501	0.8568	0.6819	0.8690	0.7021	0.8872	0.8320	0.7642	0.5650	0.80342	0.88601
Barbara	5290	-	25	62	97	61	74	99	01	06	47	15	08	43	86	59	5	7
	0.0000	0.0000		0.8407	0.8311	0.8217	0.7723	0.5101	0.7050	0.5592	0.8592	0.8429	0.8360	0.7554	0.7784	0.5257	0.83537	0.80150
BARE1	2420	2600	-	64	7	49	81	41	53	37	6	15	25	67	81	39	8	4
	0.0000	0.0000	0.0000			0.8123	0.8746	0.5289	0.6943	0.7052	0.8928	0.8754	0.8719	0.9225	0.8082	0.5300		0.75462
Cereba	7460	2260	2150	-	0.8834	71	22	84	18	52	3	94	3	66	3	16	0.79148	6
	0.0000	0.0000	0.0000	0.0000		0.8708	0.9236	0.6264	0.8190	0.7038	0.9418	0.8631	0.9218	0.8751	0.8674	0.5580	0.91702	
Claudia	2540	2350	2610	2690	-	08	91	57	38	99	77	46	5	82	56	93	6	0.80708
	0.0000	0.0001	0.0000	0.0000	0.0000		0.8286	0.4777	0.7312	0.7785	0.9058	0.8534	0.8723	0.7931	0.7408	0.5974	0.87313	
Danae	2470	2270	2460	2660	2650	-	04	67	93	85	76	53	44	41	61	83	3	0.70565
	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000		0.6288	0.8866	0.8648	0.9205	0.8200	0.9260	0.9108	0.7614	0.7052	0.89759	
Daniela	2390	2360	2330	2450	2550	2440	-	21	18	1	84	94	26	95	75	44	8	0.85374
	0.0001	0.0001	0.0034	0.0039	0.0003	0.0063	0.0005		0.7747	0.5709	0.6393	0.4561	0.5965	0.5908	0.6077	0.2937	0.63082	0.74867
Derami	1780	4790	7990	6300	0860	3300	2110	-	16	86	55	91	24	76	86	22	3	4
	0.0000	0.0000	0.0000	0.0003	0.0000	0.0001	0.0000	0.0000		0.8501	0.8477	0.6743	0.8547	0.7800	0.7437	0.6328	0.86204	0.91307
Egug	2480	2750	9970	0090	2460	1920	2550	2730	-	99	87	65	76	31	37	5	3	4
	0.0000	0.0001	0.0038	0.0000	0.0001	0.0000	0.0000	0.0020	0.0000		0.7898	0.6548	0.7849	0.7688	0.5855	0.6958	0.76208	0.74455
Fatima	2210	4020	1930	4500	4450	2520	4620	4380	2340	-	2	97	36	47	73	03	9	8
	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0005	0.0000	0.0000		0.8491	0.9342	0.8617	0.8178	0.5532	0.89545	0.83191
Lila	2720	2340	2940	2290	2520	2510	2590	0510	2590	2440	-	97	78	3	75	83	4	2
	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0069	0.0002	0.0002	0.0000		0.8898	0.7781	0.6838	0.4460	0.88164	0.67566
Maximus	2500	2760	2570	2460	2570	2670	2550	7920	4300	7710	2490	-	9	78	71	33	6	9
	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0004	0.0000	0.0000	0.0000	0.0000		0.8969	0.7796	0.5960	0.90916	0.83294
Nusif	2530	2450	2660	2240	2530	2510	2380	5250	2250	2450	2360	2440	-	45	56	2	9	5
	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0007	0.0000	0.0000	0.0000	0.0000	0.0000		0.8091	0.6740	0.77317	
Quinta	0320	2570	2520	2260	2350	2620	2780	1990	4840	5090	3000	2820	2720	-	33	4	8	0.81124
	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0007	0.0000	0.0010	0.0000	0.0001	0.0000	0.0000		0.4912	0.74007	0.76608
Romani	7360	2540	2560	2510	2500	2700	2780	3350	4940	4550	2610	2410	2660	2440	-	88	3	3
	0.0019	0.0050	0.0072	0.0061	0.0054	0.0030	0.0000	0.0658	0.0009	0.0001	0.0078	0.0218	0.0062	0.0008	0.0074		0.54764	0.62285
Sabine	1420	4360	9300	9830	5550	3540	7290	9800	7120	4580	1140	1910	1840	4610	3260	-	7	1
	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0049		0.82407
WHAM	2210	2610	2390	2520	2670	2380	2500	9870	2460	4910	2340	2560	2510	2630	5270	9550	-	3
	0.0000	0.0000	0.0000	0.0000	0.0000	0.0002	0.0000	0.0000	0.0000	0.0000	0.0000	0.0002	0.0000	0.0000	0.0000	0.0010	0.00002	
Xalax	2710	2570	2390	2180	2530	2180	2730	4920	2500	5060	2430	2820	2450	2360	2420	2450	680	-

Nonrandom genome reorganization after hybridization reveals a major role of transposable elements in reproductive isolation

Natacha Senerchia[1], François Felber[12], Christian Parisod[1]

[1]Laboratory of Evolutionary Botany, Institute of biology, University of Neuchâtel, 2000 Neuchatel, Switzerland

[2]Musée et Jardins botaniques cantonaux, Lausanne, Switzerland

This chapter is in preparation as a report for Current Biology

Summary

Interspecific hybridization leads to new interactions among divergent genomes and thus reveals the nature of genetic incompatibilities having accumulated after the origin of species (Rieseberg, 2001; Lowry *et al.*, 2008; Abbott *et al.*, 2013). Intergenomic conflicts associated with mis-regulation of transposable elements (TEs) expectedly lead to their activation in hybrids and result in genome-wide changes that may be key to species boundaries (McClintock, 1984; Josefsson *et al.*, 2006; Ha *et al.*, 2009; Bourc'his & Voinnet, 2010; Parisod *et al.*, 2010a; Fedoroff, 2012; Levy, 2013). The highly repetitive genomes of wild wheats have differentially diverged mainly under the influence of specific TEs and hybridization (Feldman *et al.*, 2012; Middleton *et al.*, 2012; Senerchia *et al.*, 2013; Senerchia *et al.*, 2014), offering unparalleled opportunities to address the proximate and ultimate factors of genome reorganization by selfish elements (or intragenomic parasites). Reciprocal F1 hybrids between three *Aegilops* species consistently show differential survival together with asymmetrical restructuring and epigenetic reorganization of TE sequences as compared to random loci. Consistent with growing conflicts and activity of TEs following hybridization (Parisod & Senerchia, 2012), increasing divergence of merged TEs leads to higher levels of genome changes in corresponding genome fractions. Repeated, nonrandom loss and methylation of TE sequences across the genome indicates that new interactions in F1 hybrids lead to the necessary repression of incompatible TE loci in order to produce viable hybrids.

Results and Discussion

Asymmetrical reproductive isolation between wild wheats

Artificial F1 hybrids between all combinations of *Ae. cylindrica* (genome CD), *Ae. geniculata* (MU) *and Ae. triuncialis* (CU) were manually produced in both directions. Seed set was relatively high for all interspecific hybrids, ranging from 17.44% to 39.18% (Table 1), but F1 hybrids showed considerable variation in germination rates, ranging from 0% to 84.44%. Accordingly, all crosses, except *Ae. geniculata* x *Ae. triuncialis*, produced viable F1 hybrids in proportions ranging between 4.54% and 24.64%. Such large hybridization success among wild wheats was significantly lower for interspecific than for intraspecific crosses, indicating incomplete postzygotic isolation and supporting the occurrence of spontaneous hybrids between wild wheat species reported in the wild (Kilian *et al.*, 2011). Consistent with the hypothesis that chromosomal divergence sustains postzygotic isolation (Feldman *et al.*, 2012), pollen viability was null in all F1 hybrids and backcrossing to the parents produced only few seeds indicative of low female fertility (data not shown). Mostly wild wheats sharing a common genome were expected to show permeable species boundaries (Feldman, 1965a; Vega *et al.*, 1994), but *Ae. cylindrica* x *Ae. geniculata* here showed non-significantly lower hybridization rates than parental crosses despite their differential genomes and species sharing a genome produced few hybrids (e.g. between *Ae. geniculata* and *Ae. triuncialis*). Variable viability of F1 hybrids thus indicates that complementary postzygotic mechanisms to chromosomal divergence control hybridization between wild wheats.

Proportion tests consistently showed significantly different hybridization rates between reciprocal crosses of each species pairs (Table 1), supporting the influence of unidirectionally inherited factors on the viability of F1 hybrids depending on the crossing direction (Tiffin *et al.*, 2001; Tayalé & Parisod, 2013).

Table 1 Hybridization success between *Aegilops cylindrica* (CY), *Ae. triuncialis* (TR) and *Ae. geniculata* (GE).

Cross	N	Seed set	Germination rate	Hybridization success	Comparison to mothers	Comparison to fathers	Reciprocal crosses
CY x CY	986	487 (49.39%)	429 (88.09%)	20.09% (208)	°	°	
GE x GE	418	189 (45.21%)	133 (70.37%)	30.86% (129)	°	°	
TR x TR	520	178 (34.23%)	141 (79.21%)	25.00% (130)	°	°	
CY x GE	138	45 (32.6%)	38 (84.44%)	24.64% (34)	$\chi^2=0.70^{NS}$	$\chi^2=1.65^{NS}$	
GE x CY	308	64 (20.78%)	17 (38.63%)	4.54% (14)	$\chi^2=75.98^{***}$	$\chi^2=44.07^{***}$	CYxGE vs GExCY: $\chi^2=37.30^{***}$
CY x TR	490	192 (39.18%)	112 (58.33%)	17.55% (86)	$\chi^2=2.36^{NS}$	$\chi^2=7.89^{***}$	
TR x CY	484	146 (30.16%)	120 (82.19%)	23.34% (113)	$\chi^2=0.29^{NS}$	$\chi^2=0.29^{NS}$	CYxTR vs TRxCY: $\chi^2=4.68^{*}$
GE x TR	286	90 (31.47%)	0 (0%)	0% (0)	$\chi^2=102.98^{***}$	$\chi^2=80.57^{***}$	
TR x GE	516	90 (17.44%)	33 (36.66%)	5.23% (27)	$\chi^2=77.18^{***}$	$\chi^2=107.19^{***}$	GEvTR vs TRxGE: $\chi^2=11.61^{***}$

Seeds set and germination rate (i.e. number and proportion of seeds produced out of treated flowers (N) and of seeds growing a radicule out of F1 seeds, respectively) represent intermediate stages towards the production of F1 hybrids after manual crosses of species. Hybridization success (i.e. proportion of viable F1 hybrids out of treated flowers, with number of individuals between parentheses) is compared to the rate in intraspecific crosses involving the maternal and the paternal parents as well as between reciprocal crosses by tests of proportions (°: non-relevant; NS: non-significant, *** significant at $\alpha=0.01$)

In particular, *Ae. geniculata* x *Ae. triuncialis* hybrids showed a relatively high seed set, but none of the seemingly healthy seeds germinated, whereas several reciprocal hybrids were produced. Hybridization events involving *Ae. cylindrica* as a mother were consistently successful, reaching rates comparable to intraspecific crosses. Incompatible interactions between cytoplasmic

and nuclear factors likely cause such asymmetrical reproductive isolation, but not much is known about underlying mechanisms (Fishman *et al.*, 2001; Tiffin *et al.*, 2001; Lowry *et al.*, 2008; Leppälä *et al.*, 2013).

Differential genome restructuring in reciprocal F1 hybrids

Nucleocytoplasmic incompatibilities involving a few loci and supporting asymmetrical reproductive isolation have been documented in wheats and relatives (Wilson & Driscoll, 1983), but the reorganization of specific, paternally-inherited nuclear genes would expectedly be sufficient to restore F1 fertility (Abbott *et al.*, 2013; Tayalé & Parisod, 2013). In stark contrast, the genotyping of 38 F1 hybrids revealed considerable departure from the expected additivity of their respective parents (Table S1), indicating genome-wide restructuring.

New interactions between divergent selfish nuclear factors such as transposable elements (TEs) and cytoplasmic repressing small interfering RNAs (siRNA) expectedly result in their activation and eventually transposition in F1 hybrids (i.e. genome shock; (Blumenstiel, 2011; Parisod & Senerchia, 2012). Such dynamics potentially having deleterious consequences for the zygote and/or of the endosperm and thus differentially impacting the viability of reciprocal F1 hybrids, genome-wide changes specifically affecting conflicting TE fractions can be predicted (McClintock, 1984; Josefsson *et al.*, 2006; Parisod *et al.*, 2010b). Following a quantitative approach developed in Senerchia *et al.*, 2014), amplified fragment length polymorphism (AFLP; 701 to 792 loci representative of genome-wide random sequences) were accordingly compared to sequence specific amplified polymorphism (SSAP assays targeting from 136 to 307 genome-wide insertions from nine recently active families of LTR retrotransposons (Senerchia *et al.*, 2013) with

different evolutionary trajectories in wild wheats (Senerchia *et al.*, 2014), assessing patterns of genome restructuring coherent with the genome shock hypothesis in the 38 F1 hybrids. New loci (i.e. unexpectedly detected in F1 hybrids) represented between 4.60% and 8.78 % of AFLP loci and between 3.09% and 9.65% of SSAP loci, whereas lost loci (i.e. unexpectedly absent in F1 hybrids) were predominant in F1 hybrids of all crosses, ranging from 27.94% to 33.68% for AFLP and from 21.38% to 39.12% for SSAP (Table S3). Loci of paternal origin being nonsignificantly more frequently lost than those of maternal origin, surviving hybrids revealed considerable restructuring of both parental genomes, supporting expectations based on the genome shock hypothesis rather than incompatible interactions between particular nuclear and cytoplasmic loci. Tukey HSD indeed compared proportions of new and lost SSAP loci for each TE family to corresponding proportions of random AFLP loci within cross types however highlighted limited TE-specific restructuring (Figure 1a). In particular, the level of new SSAP loci indicative of transposition events showed considerable variation among individual, but was never significantly higher than the level of new AFLP loci in any of the F1 hybrid type (Table S3), indicating soft transpositional activity in surviving hybrids. Although a possible outcome of conflictual interactions among TEs after genome merging, massive transposition of multiple TEs has never been reported in F1 hybrids between species (Parisod *et al.*, 2010a) most likely because of strongly deleterious effects in the zygote and/or the endosperm.

A linear mixed effects model including TE family as random effects revealed significantly different proportions of new loci (z-value = -4.69, $p < 0.001$) and of lost loci (z-value = -3.7, $p = 0.002$) among reciprocal hybrids between *Ae. cylindrica* and *Ae. triuncialis*, but no significant differences among others. In particular, *Claudia*, *Fatima* and *Nusif* showed lower proportions of

new and lost loci than AFLP in F1 hybrids having a cytoplasm from *Ae. cylindrica*, but no TE showed this pattern in the reciprocals (Figure 1a), indicating asymmetrical reorganization of specific TE genome fraction depending on the direction of the cross. Noticeably, *Romani* and *Xalax* presented significantly more frequent lost SSAP loci were than AFLP in the *Ae. triuncialis* cytoplasm. Consistent with the genome shock hypothesis, *Romani* presented evidence of proliferation in both species, whereas *Xalax* proliferated in *Ae. cylindrica* but revealed shrinkage in *Ae. triuncialis* (Senerchia *et al.*, 2014), suggesting that the expected conflict revealed by hybridization resulted in specific sequence deletion in corresponding TE fractions. Similarly, *Fatima* proliferated in both *Ae. cylindrica* and *Ae. geniculata* (Senerchia *et al.*, 2014) and here only showed significantly lower proportions of new and lost loci than AFLP in F1 hybrids involving *Ae. cylindrica* as a mother, further suggesting that this cytoplasm supports greater structural stability as regards to specific TEs than other maternal backgrounds.

Methylation changes in parental TE fractions of F1 hybrids

TE activity being mainly controlled by cytoplasmic siRNA (Mirouze *et al.*, 2009), incompatibilities revealed by hybridization (Bourchis & Voinnet, 2010) expectedly lead to drastic reorganization of DNA methylation in conflicting genome fractions (Parisod & Senerchia, 2012). These predictions were checked in the 38 F1 hybrids using methyl sensitive amplified polymorphism display (MSAP; 584 to 605 loci) and methyl-sensitive transposon display (MSTD (Parisod *et al.*, 2009); 151 to 225 loci for the same 9 TEs), assessing genome-wide CG and CCG methylation states of random sequences and sequences flanking insertions of each TE family, respectively (Parisod *et al.*, 2009). Loci presented considerable deviation from the expected

additivity of their respective parents, indicating significant CG and CHG methylation changes following hybridization. Between 6.7% and 8.76% of the random MSAP loci and from 7.25% to 17.75% of the MSTD loci indeed underwent CG methylation, whereas 5.45% to 9.93% of the MSAP loci and 4.44% to 11.50% of the MSTD loci showed CHG methylation (Figure 1a; Table S4). Mann-Whitney-Wilcoxon test showed significantly higher CG methylation of MSTD than MSAP loci ($W = 11210.5$, $p < 0.001$), but not CHG methylation ($W = 7336$, $p = 0.192$), indicating increased levels of stable methylation specifically targeting TE genome fractions after the merging of divergent wild wheat genomes. As compared to one century-old hybrids of *Spartina*, where similar patterns were reported (Parisod *et al.*, 2009), the present results support the genome shock hypothesis and highlight that stable CG rather than metastable CCG silencing of TE sequences is quickly reached in hybrid genomes.

A linear mixed effects model, with TE family as random effect, tested the effect of the cross type on proportions of CG as well as CHG methylation and revealed an overall impact of neither the species involved, nor the direction of the cross, indicating consistent methylation around TE insertions in response to hybridization. Tukey HSD assessed that selected TE families expected to be involved in conflictual interactions after genome merging presented significantly higher proportions of CG or CHG methylation as compared to random MSAP loci (Figure 1a, Table S4). *Barbara*, *Claudia* and *Fatima* proliferated in *Ae. cylindrica* and *Ae. geniculata*, whereas *Sabine* has been massively eliminated from the *Ae. geniculata* genome (Senerchia *et al.*, 2014), and congruently presented here significantly higher proportions of insertions showing CG methylation than random MSAP in corresponding F1 hybrids with *Ae. cylindrica* as the mother.

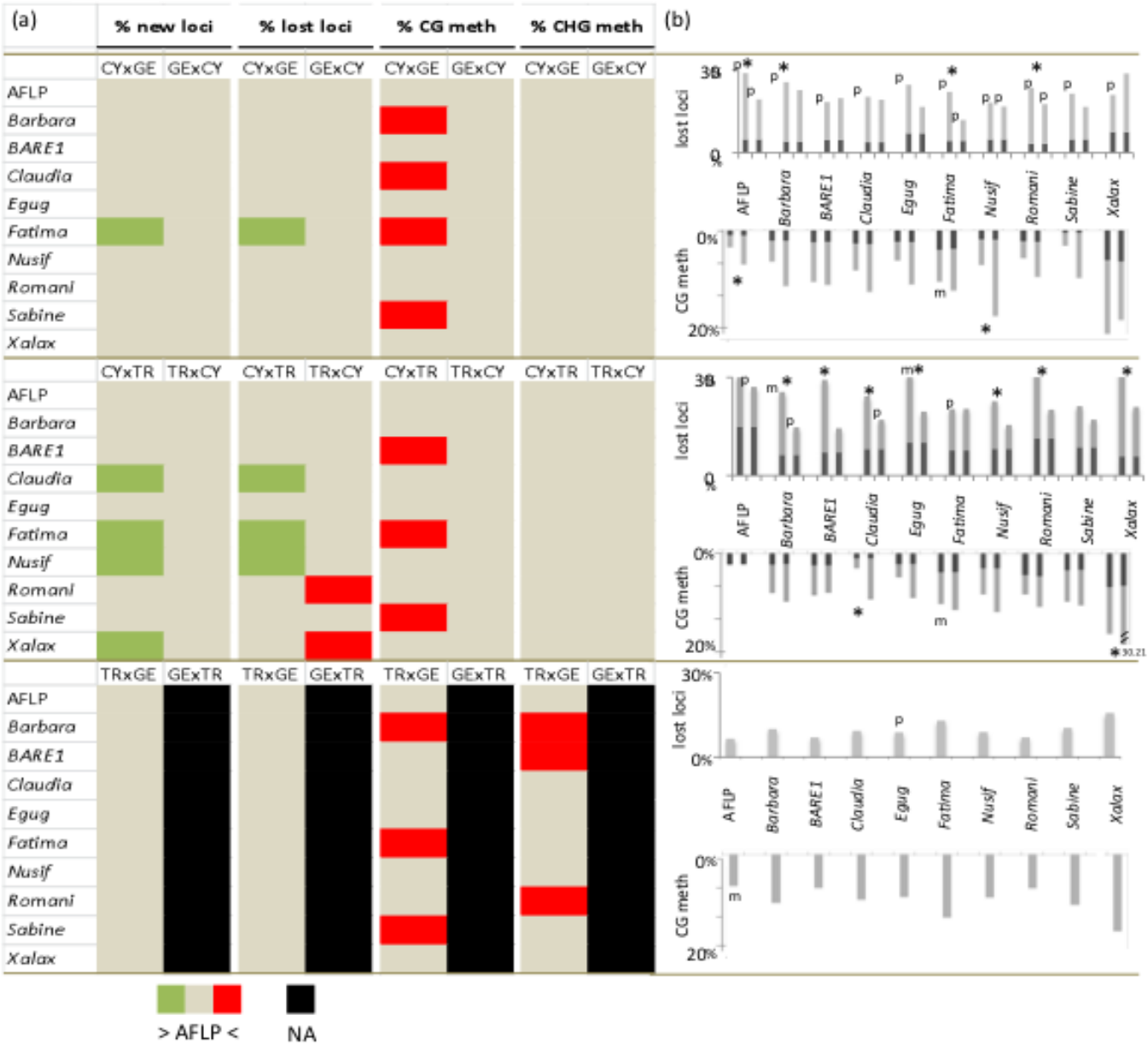


Figure 1 Genome restructuring (new and lost loci) and methylation changes (CG methylation and CHG methylation loci) in reciprocal F1 hybrids between *Aegilops cylindrica* (CY), *Ae. triuncialis* (TR) and *Ae. geniculata* (GE). (a) Distinctive colors show significantly different levels of genome reorganization in 9 families of transposable elements (TEs) as compared to random sequences (AFLP; green and read, significantly lower and higher, respectively; grey, not available). (b) Nonrandomly lost and CG-methylated loci for TE families and AFLP, with dark grey presenting the proportions of repeatedly reorganized loci in both directions (*, significantly different between reciprocal hybrids; significantly biased reorganization of nonrandom loci of either maternal (m) or paternal (p) origin).

Given that no TE showed this pattern in the reciprocal hybrids, epigenetic reorganization appears asymmetrical, but loci having undergone significant methylation changes have their origin from either the maternal or the paternal genome indifferently. As expected from their evolutionary trajectories in parental species, crosses between *Ae. cylindrica* and *Ae. triuncialis* presented a significantly higher proportions of CG methylation for *BARE1*, *Fatima* and *Sabine*, whereas crosses between *Ae. triuncialis* and *Ae. geniculata* showed this pattern for *Barbara*, *Fatima* and *Sabine*. Rapid methylation of sequences flanking TE insertions is coherent with increased silencing of potentially activated TEs following genome merging, as expected under the genome shock hypothesis. Coherent with *trans*-repression by siRNAs, methylation repatterning appears dependent of the cytoplasm of F1 zygotes, but affects loci coming from both subgenomes. Most TEs have indeed proliferated in the recent past in *Ae. cylindrica* (Parisod *et al.*, 2009; Senerchia *et al.*, 2013; Senerchia *et al.*, 2014), suggesting that siRNA-generating loci are abundant and support efficient repression of TE activity in this cytoplasm.

Nonrandom genome reorganization in F1 hybrids

Activation of TEs following the genome shock hypothesis expectedly affects the viability of F1 hybrids and thus predicts nonrandom genome reorganization in surviving F1 hybrids. Accordingly, a linear mixed model nesting TE families within reciprocal crosses as random-effects tested that proportions of non-shared SSAP loci between pairs of parents was significantly associated with proportions of new SSAP (t-value = 3.98, $p = 0.047$, $R^2 = 0.35$) and lost SSAP loci (t-value = 41.19, $p < 0.001$, $R^2 = 0.52$) in F1 hybrids (Figure 2). As expected, increasingly

divergent organization of TE insertions between parents leads to increasing levels of restructuring of the corresponding TEs after hybridization. Similarly, the proportions of non-shared MSTD loci between pairs of parents showed a positive association with proportions of CG methylation in F1 hybrids (t-value = 7.36, p = 0.007, R^2 = 0.3752), but a negative one with proportions of CHG methylation (t-value = 4.12, p = 0.043, R^2 = 0.3228; Figure 2). Paralleling observations for restructuring, increasing divergence of methylation states around TE insertions thus leads to increasing CG methylation in the corresponding TE fractions after hybridization. Consistent with the genome shock hypothesis, rising divergence in the organization of TEs being merged in F1 hybrids leads to increasingly incompatible interactions and results in rapid molecular events such as deletion or CG methylation repressing the deleterious effects of corresponding TEs (Levy & Feldman, 2004).

To further test that reproductive isolation is ruled by conflicting interactions among selected TEs (Tayalé & Parisod, 2013), we assessed to what extent repressive molecular events such as loss and CG methylation of loci were more repeated among F1 hybrids of each type than expected by chance. Between 8.4% and 25.3% (mean 16.4%±3.9) of loci were nonrandomly lost in viable F1 hybrids (Figure 1b, Table S5). Although a substantial proportion of loci (mean 6.95%±2.78) were repeatedly lost in both reciprocal hybrids, indicating substantial deletion independent of the cytoplasmic context, nonrandom loss of loci revealed asymmetrical, regularly showing different patterns between reciprocal hybrids and significant bias towards one of the parental genomes (i.e. mostly loci of paternal origin). Similarly, nonrandom CG methylation was reported for 3.1% to 28.8% of the loci (Figure 1b), but only a mean of 2.87%±1.56 of the loci showed repeated CG methylation in both directions. In contrast to sequence loss, methylation was

rarely significantly biased towards one the parental genomes, supporting nuclear-nuclear interactions as the overall trigger of genome reorganization in response to hybridization.

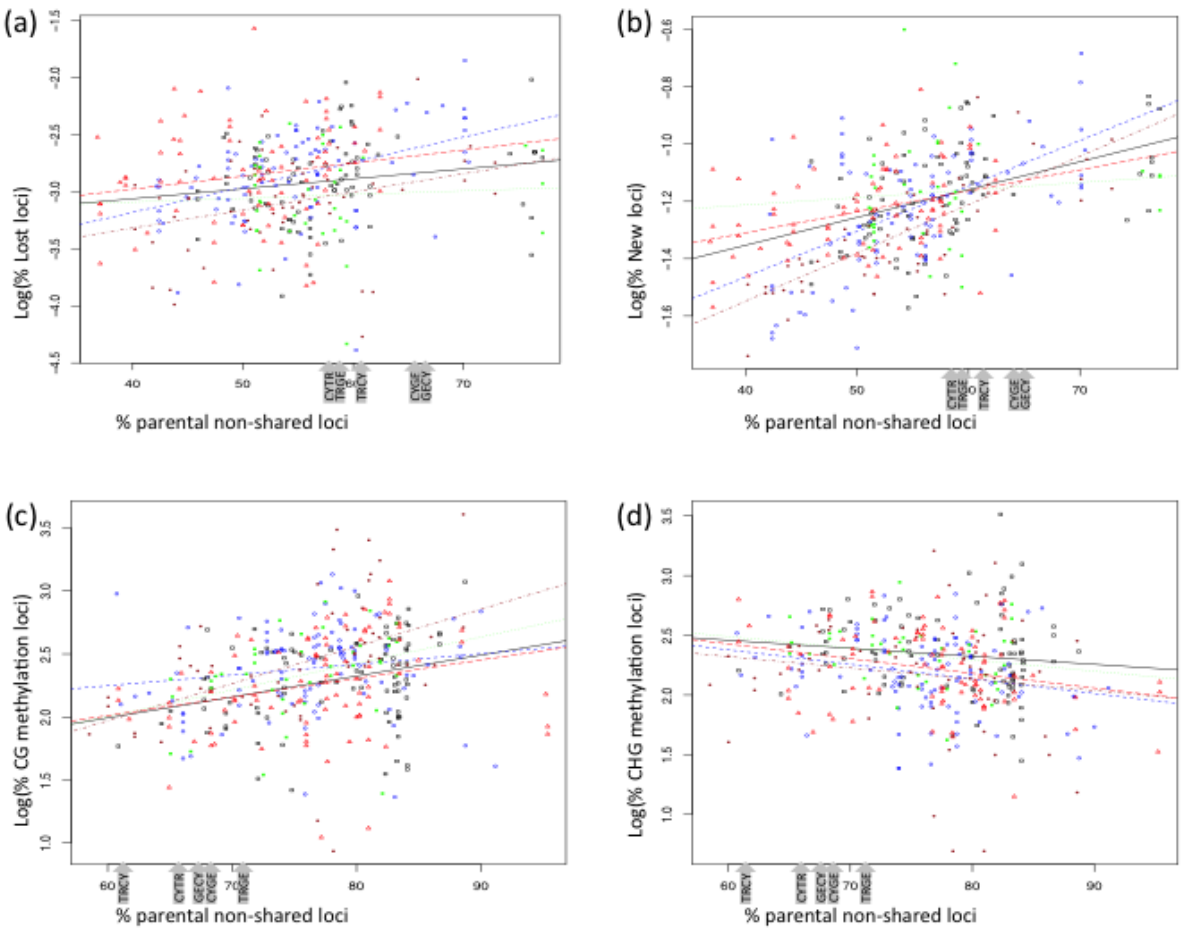


Figure 2 Association of divergence between parents and restructuring of TE genome fractions in resulting F1 hybrids. General linear mixed model associating proportions of non-shared loci between parents with (a) the proportion of lost SSAP loci, (b) the proportion of new SSAP loci, (c) the proportion of MSTD loci showing CG methylation and (d) the proportion of MSTD loci showing CHG methylation for each TE family within each F1 hybrids. Grey arrows indicate mean proportions of parental non-shared parental loci for random sequences (AFLP in A and B and MSAP in C and D), whereas lines show proportions and relationships for the 9 TE families in *Ae. cylindrica* x *Ae. geniculata* (square, solid line), *Ae. geniculata* x *Ae. cylindrica* (filled light circle, dashed line), *Ae. cylindrica* x *Ae. triuncialis* (circle, dotted line), *Ae. triuncialis* x *Ae. cylindrica* (filled dark circle, dotdashed line) and *Ae. triuncialis* x *Ae. geniculata* (triangle, longdashed line).

Genome shock and a role for TEs in reproductive isolation

Surveyed TEs followed differential evolutionary trajectories in the species being merged (i.e. *BARE1, Fatima* and *Romani* proliferated in all parental species, *Barbara* and *Claudia* showed quiescence in *Ae. triuncialis* only, whereas *Sabine* and *Xalax* were massively eliminated from *Ae. geniculata* and *Ae. triuncialis*, respectively) and matched expectations based on the genome shock hypothesis. Recently active TE families showing divergent arrangements between parents revealed conflicting interactions and considerable genome reorganization coherent with activation in F1 hybrids, whereas quiescent ones (i.e. *Egug* and *Nusif*) remained so following hybridization. Although particularly abundant TEs such as *BARE1* and *Romani* matched this expected pattern to a lower extent, suggesting that they may be less strictly controlled, conflicting interactions between interspersed TEs (i.e. genome shock) appears as a chief driver of genome evolution in response to hybridization.

Reorganization of TEs in F1 hybrids revealed nonrandom and predictable to a large extent, supporting the hypothesis that sequence deletion and methylation of specific loci from active TEs is necessary to produce viable hybrids. Despite limited evidence of substantial transposition in viable hybrids, considerable levels of asymmetrical and nonrandom molecular events coherent with repression of TE activity paralleled patterns of reproductive isolation between wild wheat species. Consistent with the chief availability of maternal *trans*-acting siRNAs in the cytoplasm to repress active TEs through methylation (Mari-Ordonez *et al.*, 2013 {Bourc'his, 2010 #191), F1 hybrids indeed revealed balanced epigenetic repatterning of subgenomes. In contrast, biased loss of sequences from the paternal donor was apparent, indicating that TEs of paternal origin may be mainly involved in new interactions and potentially activated. Accordingly, only paternal genomes

having undergone significant TE loss early after zygote formation would produce viable hybrids. Such a hypothesis would be further tested by including unviable hybrids, but available evidence of rapid reorganization of selected loci of active and potentially harmful TEs in F1 hybrids already confirms the overlooked impact of selfish TEs on the maintenance of species boundaries.

Material and methods

Plant material and crossing experiments

During May-June 2011 and 2012, reciprocal F1 hybrids were produced by manually crossing individuals from six accessions of the selected wild wheat species presenting different genome compositions (van Slageren, 1994). Parental accessions of *Aegilops cylindrica*, *Ae. geniculata* and *Ae. triuncialis* (table S1) were maintained by selfing in germplasms and were considered highly inbred. All intraspecific and interspecific combinations were crossed using randomly selected accessions and evaluated, following the same protocol.

Parental accessions were grown in individual pots and overwintered in a common garden at the botanical garden of Neuchâtel, Switzerland. Spikes were emasculated one week before maturation, bagged and then fertilized with selected pollen during the following week. Non-fertilized spikes did not produce any seed. Treated F1 seeds were harvested, counted and randomly germinated in watered petri dishes under controlled conditions (18°C, 18h light). The development of radicle was observed for up to two weeks and seeds were then sown in individual pots under controlled conditions (18°C 18h light). After two weeks, leaf tissue was collected and disrupted in liquid nitrogen for DNA extraction using a standard DNeasy plant extraction mini kit

protocol from Qiagen AG, Switzerland. The additivity of diagnostic EST-SSR markers (Parisod *et al.*, 2013) confirmed the hybrid nature of all F1 seedlings (data not shown). Seedlings were raised to maturity and pollen viability was assessed by counting the proportion of pollen grains stained in cotton blue and lacto phenol on a subset of individuals. Female fertility was also checked through manual backcrossing of emasculated F1 hybrids with the parents. Tests of proportion compared hybridization success to the maternal and paternal one as well as between reciprocal crosses, using R Cran.

Fingerprint techniques and patterns of genome reorganization

F1 hybrids and their respective parents were genotyped using amplified fragment length polymorphism (AFLP) and sequence specific amplified polymorphism (SSAP) following similar protocols in 96-well plates with randomly positioned individuals following {Parisod, 2009 #135} (Senerchia *et al.*, 2014). Nine combinations of selective *Eco*RI-*Mse*I primers were used for PCR amplification of AFLP profiles, whereas four combinations of TE-specific and *Mse*I selective primers were used for SSAP.

Methyl sensitive amplified polymorphism (MSAP) and methyl-sensitive transposon display (MSTD) used the isoschizomers *Hpa*II and *Msp*I in parallel, following (Parisod *et al.*, 2009). Protocols of MSAP/MSTD and AFLP/SSAP are similar, except that DNA restriction is performed with *Msp*I and *Hpa*II that recognize the same nucleotide sequence (5′-CCGG-3’), but have differential sensitivity to cytosine methylation. *Msp*I is indeed sensitive to methylation of the external cytosine, but cuts whether the internal cytosine are methylated on one or both stands, whereas *Hpa*II is sensitive to either methylation of any cytosine on both strands or internal

cytosine on one strand and thus does not cleave methylated restriction sites unless hemi-methylated (McClelland et al., 1994). Combining patterns resulting from parallel restriction of the same samples with both *Msp*I and *Hpa*II thus assess the methylation state of corresponding loci. Here, nine different combinations of selective primers were used for MSAP, whereas two combinations of TE-specific and selective primers were used for each of the SSAP assay (i.e. for each of the selected TE family; Table S2).

The whole procedure was replicated twice on five samples (i.e. 13% of the dataset) for both AFLP and SSAP to estimate their respective error rate following (Bonin 2004) (Table S2). MSAP and MSTD were similarly replicated, but the binary matrix combining *Mse*I and *Hpa*II was used to estimate the error rate in assessing the methylation state. PCR products amplified with FAM™, VIC®, NED™ fluorescent dye were pooled with GeneScan™ 500 LIZ™ Size Standard and separated with a 3730xl DNA Analyzer (ABI) capillary sequencer. Resulting electropherograms were visualized and scored with GeneMapper 4.0 (Applied Biosystems) using AFLP default peak detection parameters. The scoring was manually checked and loci were recorded as present (1) or absent (0) in binary matrices.

For each F1 hybrid, the expected AFLP/SSAP profile represented by the additivity of parental profiles was compared to the observed profile. Each locus was considered as either (i) additive (i.e. similarity between the expected and observed profiles) (ii) lost (i.e. locus absent in the observed F1 hybrid profile despite predicted as present in the expected profile), or (iii) new (i.e. presence of locus in the F1 observed profile despite absence in the expected one). In each F1 hybrid, the proportion of additive, new and lost loci was estimated for AFLP and for each SSAP, and then averaged for hybrid types.

The MSAP/MSTD patterns, considering observations of both profiles amplified from assays with either *Msp*I or *Hpa*II (see above), were used to compare the reported methylation state at each locus of F1 hybrids with the expected additivity of their methylation states in respective parents. These comparisons assessed the proportions of loci (i) that were additive (i.e. similarity between the observed and expected methylation state), (ii) that underwent CG methylation (i.e. the observed locus has internally methylated cytosine, whereas it was expected to be either fully non-methylated or hemi-methylated on external cytosines), (iii) that underwent CHG methylation (i.e. the observed locus has externally hemi-methylated cytosines, whereas it was expected to be either non-methylated or internally methylated). Loci showing specific non-additive patterns as compared to parents were observed (i.e. the observed locus is non-methylated or externally hemi-methylated, whereas it was expectedly internally methylated, or the observed locus was non-methylated locus or internally methylated, whereas its was expectedly externally hemi-methylated, or the locus was observed in neither *Msp*I and *Hpa*II profiles, although predicted from the parental additivity). These loci were not further considered for detailed analyses as the processes underlying such non-additivity were untraceable. Proportions of additive, CG methylation and CHG methylation events estimated for each F1 hybrid were then averaged within cross type.

Statistical inferences on levels of genome reorganization

Quantitative comparisons between fingerprint techniques tracking large numbers of random sequences and of TE insertions (i.e. AFLP / SSAP for genome restructuring and MSAP / MSTD for methylation changes) minimize potential bias in the interpretation of banding patterns as

processes of genome reorganization, highlighting TE-specific events (Parisod *et al.*, 2009; Senerchia *et al.*, 2014). Whether individual proportions of loci showing CG methylation and CHG methylation in random sequences (MSAP) was significantly different than MSTD proportions for all TE families merged into a single dataset was assessed within each hybrid type using a Mann-Whitney-Wilcoxon test on R cran (cran.r-project.org). Considering each TE family independently, TE-specific proportions of new SSAP loci and lost SSAP loci were compared to AFLP proportions within each hybrid type. Whether reciprocal hybrids presented significantly different restructuring levels (i.e. proportions of lost and new loci) or levels methylation changes (i.e. CG and CHG methylation) was assessed using one-way analysis of variance with Tukey HSD post-hoc tests (α at the 0.05 level) in the package 'agricolae' (de Mendiburu, 2013) and the package 'multcomp' (Hothorn *et al.*, 2008). Differences in the proportions of lost and loci showing CG methylation from either the maternal or the paternal origin were assessed with tests of proportion on R cran.

Proportions of non-shared loci between parents for AFLP as well as SSAP from each TE family was correlated with corresponding levels of new and lost loci in resulting F1 hybrids using a mixed effect linear model, with maximum likelihood method and including TE families as random effects, in the package 'lme4' (Bates *et al.*, 2012). R squared was defined using a function correlating fitted values to observed values in R cran (Nakagawa & Schielzeth, 2013). Proportions of loci with different methylation states between parents (i.e. % parental shared methylation) were estimated for MSAP and MSTD of each TE family and similarly correlated with proportions of loci showing either CG methylation or CHG methylation in resulting F1 hybrids.

Nonrandom loss or CG methylation of loci was inferred by estimating the proportion of hybrids showing a particular event and comparing it to randomly expected proportions. For each F1 hybrid within a hybrid type, loci having undergone a change were recoded as '1' and others were assigned a '0'. Random permutation of loci 999 times by a bootstrap procedure using the package "vegan" in R (Nakagawa & Schielzeth, 2013) assessed the proportion of loci obtaining a higher sum of '1' among hybrids than observed. This estimates the probability of obtaining the observed value by chance and loci lower than 5% were considered as nonrandomly reorganized. Tests of proportion (prop.test in R cran) assessed whether the number of (i) nonrandomly lost loci and (ii) the number of nonrandomly CG methylated loci was significantly different between reciprocal crosses. Differences in proportions of repeatedly reorganized loci from either the maternal or the paternal parent were assessed by chi-square tests on R Cran.

Acknowledgments

Data analyzed in this paper were generated at the Genetic Diversity Centre, ETH Zurich. We thank R. A. Slobodeanu for statistical advices. This work was funded by the Swiss National Science Foundation through the National Centre of Competence in Research 'Plant Survival' and a grant (PZ00P3-131950 to CP).

References

Abbott R, Albach D, Ansell S, Arntzen JW, Baird SJE, Bierne N, Boughman J, Brelsford A, Buerkle CA, Buggs R, *et al.* 2013. Hybridization and speciation. *Journal of Evolutionary Biology* **26**: 229-246.

Bates D, Maechler M, Bolker B 2012. lme4: Linear mixed-effects models using S4 classes. R package version 0.999999-0. http://CRAN.R-project.org/package=lme4.

Blumenstiel J. 2011. Evolutionary dynamics of transposable elements in a small RNA world. *Trends in Genetics* **27**: 23-31.

Bourc'his D, Voinnet O. 2010. A small-RNA perspective on gametogenesis, fertilization, and early zygotic development. *Science* **330**: 617-622.

Bourchis D, Voinnet O. 2010. A small-RNA perspective on gametogenesis, fertilization, and early zygotic development. *Science* **330**: 617-622.

de Mendiburu F 2013. agricolae: statistical procedures for agricultural research. R package version 1.1-4.
Fedoroff NV. 2012. Transposable elements, epigenetics, and genome evolution. *Science* **338**: 758-767.

Feldman M. 1965a. Further evidence for natural hybridization between tetraploid species of *Aegilops* Section *Pleionathera*. *Evolution* **19**: 162-174.

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.

Fishman L, Kelly AJ, Morgan E, Willis JH. 2001. A genetic map in the *Mimulus guttatus* species complex reveals transmission ratio distortion due to heterospecific interactions. *Genetics* **159**: 1701-1716.

Ha M, Lu J, Tian L, Ramachandran V, Kasschau K. 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.

Hothorn T, Bretz F, Westfall P. 2008. Simultaneous inference in general parametric models. *Biometrical Journal* **50**: 346-363.

Josefsson C, Dilkes B, Comai L. 2006. Parent-dependent loss of gene silencing during interspecies hybridization. *Current Biology* **16**: 1322-1328.

Kilian B, Mammen K, Millet E, Sharma R, Graner A, Salamini F, Hammer K, O H 2011. *Aegilops*. In: Kole C ed. *Wild Crop Relatives: Genomic and Breeding Resources, Cereals*: Springer-Verlag Berlin Heidelberg, 1-17.

Leppälä J, Bokma F, Savolainen O. 2013. Investigating incipientspeciation in *Arabidopsis lyrata* from patterns of transmission ratio distortion. *Genetics* **194**: 697-708.

Levy AA 2013. Transposons in plant speciation. In: Fedoroff NV ed. *Plant transposons and genome dynamics in evolution*: John Wiley and Sons Ltd, 165-176.

Levy AA, Feldman M. 2004. Genetic and epigenetic reprogramming of the wheat genome upon allopolyploidization. *Biological Journal of the Linnean Society* **82**: 607-613.

Lowry DB, Modliszewski JL, Wright KM, Wu CA, Willis JH. 2008. The strength and genetic basis of reproductive isolating barriers in flowering plants. *Philosophical Transactions of the Royal Society B: Biological Sciences* **363**: 3009-3021.

Mari-Ordonez A, Marchais A, Etcheverry M, Martin A, Colot V, Voinnet O. 2013. Reconstructing de novo silencing of an active plant retrotransposon. *Nature Genetics* **45**: 1029-1039.

McClintock B. 1984. The significance of responses of the genome to challenge. *Science* **226**: 792-801.

Middleton C, Stein N, Keller B, Kilian B, Wicker T. 2012. Comparative analysis of genome composition in *Triticeae* reveals strong variation in transposable element dynamics and nucleotide diversity. *Plant journal*: doi: 10.1111/tpj.12048.

Mirouze M, Reinders J, Bucher E, Nishimura T, Schneeberger K, Ossowski S, Cao J, Weigel D, Paszkowski J, Mathieu O. 2009. Selective epigenetic control of retrotransposition in Arabidopsis. *Nature* **461**: 427-430.

Nakagawa S, Schielzeth H. 2013. A general and simple method for obtaining R^2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution* **4**: 133-142.

Parisod C, Alix K, Just J, Petit M, Sarilar V, Mhiri C, Ainouche M, Chalhoub B, Grandbastien M-A. 2010a. Impact of transposable elements on the organization and function of allopolyploid genomes. *New Phytologist* **186**: 37-45.

Parisod C, Alix K, Just J, Petit M, Sarilar V, Mhiri C, Ainouche M, Chalhoub B, Grandbastien M-A. 2010b. Impact of transposable elements on the organization and function of allopolyploid genomes. *New Phytologist* **186**: 37-45.

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.

Parisod C, Salmon A, Zerjal T, Tenaillon M, Grandbastien M-A, Ainouche M. 2009. Rapid structural and epigenetic reorganization near transposable elements in hybrid and allopolyploid genomes in *Spartina*. *New Phytologist* **184**: 1003-1015.

Parisod C, Senerchia N 2012. Responses of transposable elements to polyploidy. In: Grandbastien M-A, Casacuberta JM eds. *Plant Transposable Elements*: Springer-Verlag Berlin Heidelberg, 147-148.

Rieseberg LH. 2001. Polyploid evolution: Keeping the peace at genomic reunions. *Current Biology* **11**: R925-R928.

Senerchia N, Felber F, Parisod C. 2014. Contrasting evolutionary trajectories of multiple retrotransposons following independent allopolyploidy in wild wheats. *New Phytologist*: doi: 10.1111/nph.12731.

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

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

Tiffin P, Olson MS, Moyle LC. 2001. Asymmetrical crossing barriers in angiosperms. *Proceedings of the Royal Society B:-Biological Sciences* **268**: 861-867.

van Slageren MW. 1994. Wild wheats: a monograph of *Aegilops* L. and *Amblyopyrum* (Jaub. & Spach) Eig (*Poaceae*). *Wageningen Agricultural University Papers* **97**.

Vega M, Abbo S, Feldman M, Levy AA. 1994. Chromosome painting in plants: in situ hybridization with a DNA probe from a specific microdissected chromosome arm of common wheat. *Proceedings of the National Academy of Sciences* **91**: 12041-12045.

Wilson P, Driscoll CJ 1983. Hybrid Wheat. In: Frankel R ed. *Heterosis* Berlin Heidelberg: Springer, 94-123.

Supplementary Table S1 Female and paternal donor accessions of the different F1 hybrids genotyped in this study (AE: accessions from the Institutes für Planzengenetik und Kulturplanzenforschung,TA: accessions from the United States Department of Agriculture, IP: accessions from the International Center for Agricultural Research in the Dry Areas and from the Wheat Genetic, NE: accessions and collections from the botanical garden of Neuchâtel. Switzerland.

Cross	female accession number	male accession number	Number of F1 hybrid genotyped
Ae. cylindrica x *Ae. geniculata*	AE870	NE1207	1
Ae. cylindrica x *Ae. geniculata*	PI486244	TA1800	3
Ae. cylindrica x *Ae. geniculata*	PI486244	TA2899	1
Ae. cylindrica x *Ae. geniculata*	PI542181	TA1800	3
Ae. cylindrica x *Ae. geniculata*	TA2204	TA2899	1
Ae. geniculata x *Ae. cylindrica*	PI542181	TA2204	1
Ae. geniculata x *Ae. cylindrica*	PI287737	AE1042	1
Ae. geniculata x *Ae. cylindrica*	NE1207	AE1042	2
Ae. triuncialis x *Ae. cylindrica*	NE0612B	AE1042	2
Ae. triuncialis x *Ae. cylindrica*	PI491442	AE1042	1
Ae. triuncialis x *Ae. cylindrica*	PI491442	NE0001	2
Ae. triuncialis x *Ae. cylindrica*	PI524957	NE0002	1
Ae. cylindrica x *Ae. triuncialis*	PI486244	PI173615	6
Ae. cylindrica x *Ae. triuncialis*	PI486244	PI524957	1
Ae. cylindrica x *Ae. triuncialis*	PI542181	PI542345	1
Ae. cylindrica x *Ae. triuncialis*	TA2204	PI173615	1
Ae. cylindrica x *Ae. triuncialis*	TA2204	PI487246	1
Ae. triuncialis x *Ae. geniculata*	NE0612B	PI287737	1
Ae. triuncialis x *Ae. geniculata*	PI173615	TA1800	1
Ae. triuncialis x *Ae. geniculata*	PI491442	PI287737	5
Ae. triuncialis x *Ae. geniculata*	PI524957	PI491428	2

Supplementary Table S2 Selective primer combinations of amplified fragment length polymorphism (AFLP), sequence specific amplified polymorphism (SSAP), methyl sensitive amplified polymorphism (MSAP) and methyl sensitive transposon display (MSTD) techniques. Adaptors, pre-selective and selective *Eco*RI primers in Parisod *et al.* (2009) and Senerchia *et al.* (2014). Error rates were evaluated on five replicates following Bonin *et al* (2004) on the final binary matrix.

TE family	Primer combination	Error rate %[a/b]	number of loci[a/b]
Barbara	M-CAC*, M-CCA, M-CCG,M-CTG*	3.6/8.7	344/146
BARE1	M-CCA*, M-CAC, M-CTA*, M-CCA	7.0/6.3	434/187
Claudia	M-CGA*, M-CTA, M-CCG, M-CAA*	5.6/7.6	365/202
Egug	M-CCA, M-CGA*, M-CTA*, M-CAA	6.6/6.8	265/217
Fatima	M-CAC, M-CTG*, M-CCG, M-CAA*	4.5/8.6	441/264
Nusif	M-CAC*, M-CTG, M-CTA*, M-CGC	5.5/8.6	380/263
Romani	M-CAC*, M-CCA, M-CGA, M-CGC*	4.6/6.5	332/216
Sabine	M-CAC*, M-CGA, M-CTA*, M-CGC	5.5/6.5	322/191
Xalax	M-CAC*, M-CCA, M-CGA*, M-CTG	3.3/6.0	252/202
AFLP/MSAP	FAM-AAG/M-CGA, NED-E-AAC/M_M-CGA, FAM-E-AGG/M-CCG, NED-E-ACG/M_CAC, NED-E-ACG/M_CCA, FAM-E-ATC/M-CTA, VIC-E-ATT/M-CCG, VIC-E-ATT/M-CAA	5.5/5.6	1400/637

[a] error rate for AFLP / SSAP

[b] error rate for MSAP / MSTD

* combination performed in SSAP and MSTD

Supplementary Table S3 Genome restructuring in F1 hybrids between *Aegilops cylindrica* (CY), *Ae. triuncialis* (TR) and *Ae. geniculata* (GE). Mean proportions and standard deviation for new loci (i.e. detected in F1 hybrids, although unexpected from the additivity of parental profiles) and for lost loci (i.e. expected from the additivity of parental profiles, but not detected in F1 hybrids) in each of the reciprocal hybrid type. Mean proportions of loci (i.e. random sequences, AFLP, and insertions from each of the nine TE families, SSAP) that are common between parents is presented as parental shared loci TE families with significantly different restructuring proportions than random loci as tested by Tukey HSD do not share small letters. NA: not available, because no hybrid survived.

Loci type	number of loci		% parental shared loci		% new loci		% lost loci	
	CY* x *GE	***GE* x *CY***	***CY* x *GE***	***GE* x *CY***	***CY* x *GE***	***GE* x *CY***	***CY* x *GE***	***GE* x *CY***
AFLP	792	701	34.76±0.98	22.21±4.99	8.25±2.71^{a}	5.3±1.89^{a}	33.68±5.64^{ab}	31.79±3.73^{ab}
Barbara	246	228	40.18±1.51	26.88±6.71	6.67±2.61^{ab}	6.35±1.98^{a}	33.67±4.03^{ab}	36.79±8.46^{ab}
BARE1	304	277	41.26±2.7	30.15±1.48	6.61±1.3^{ab}	5.73±2.47^{a}	32.7±4.77^{abc}	34.23±3.24^{ab}
Claudia	246	218	42.82±2	24.06±2.77	5.41±2.23^{ab}	4.58±0.83^{a}	28.15±4.41^{bc}	31.04±7.40^{ab}
Egug	203	216	46.78±2.37	34.67±9.85	4.81±1.76^{ab}	5.32±0.94^{a}	32.47±5.21^{abc}	29.64±4.42^{ab}
Fatima	307	277	44.27±2.05	15.74±0.24	4.69±1.7^{b}	3.09±1.45^{a}	25.84±3.68c	24.87±2.99^{b}
Nusif	280	285	51.75±2.03	23.46±1.1	5.36±1.52^{ab}	3.94±0.49^{a}	27.18±4.07^{bc}	30.37±2.29^{ab}
Romani	191	159	23.94±0.83	28.05±10.2	6.58±2.97^{ab}	5.85±1.84^{a}	35.63±5.64^{a}	32.14±1.97^{ab}
Sabine	256	228	47.79±1.95	25.78±1.06	5.49±1.73^{ab}	5.22±2.21^{a}	32.11±3.5^{abc}	27.15±0.68^{ab}
Xalax	152	136	41.05±3.35	35.05±2.77	6.87±2.72^{ab}	6.37±2.26^{a}	32.96±5.64^{abc}	39.12±1.01^{a}
	CY* x *TR	***TR* x *CY***	***CY* x *TR***	***TR* x *CY***	***CY* x *TR***	***TR* x *CY***	***CY* x *TR***	***TR* x *CY***
AFLP	750	723	41.79±2.03	41.5±1.13	8.78±2.74^{ab}	4.6±1.16^{ab}	32.11±3.11^{abc}	27.94±1.89^{bc}
Barbara	243	231	43.91±3.75	46.92±1.98	7.49±1.7^{abc}	4.75±0.92^{ab}	30.89±2.56^{abcd}	26.96±3.38^{bc}
BARE1	304	291	42.08±2.22	43.73±1.21	7.04±1.81^{abcd}	5.74±1^{ab}	30.47±3.4^{bcd}	26.33±1.96abc
Claudia	252	241	44.91±2.08	45.5±3.65	5.8±1.53c^{de}	4.51±0.87^{ab}	25.77±3.61^{de}	24.16±3.09^{abc}
Egug	195	184	48.52±3.6	46.66±3.37	7.42±2.39^{abc}	6.52±2.2^{ab}	34.78±3.7^{ab}	29.15±2.42^{cd}
Fatima	290	284	56.77±1.4	58.34±1.26	4.47±1.22^{de}	3.5±1.31^{b}	23.21±5.76^{e}	21.38±1.89^{a}
Nusif	286	278	51.06±1.97	55.1±0.99	4.16±0.98^{e}	3.46±1.18^{b}	25.62±4.02^{de}	22.9±1.74^{de}
Romani	185	184	30.67±1.38	29.59±2.58	9.07±3.21^{a}	7.35±3.07^{a}	36.46±6.82^{a}	33.93±2.95^{de}
Sabine	245	240	48.63±2.81	50.12±3.01	6.07±1.09^{bcde}	4.9±1.55^{ab}	26.59±3.21^{cde}	26.45±1.07^{abc}
Xalax	160	148	40.06±3.38	38.51±1.76	4.29±1.85^{de}	3.45±1.78^{b}	35.31±2.78^{ab}	34.89±5.86^{e}
	TR* x *GE	***GE* x *TR***	***TR* x *GE***	***GE* x *TR***	***TR* x *GE***	***GE* x *TR***	***TR* x *GE***	***GE* x *TR***
AFLP	751	NA	40.46 ±0.86	NA	6.07 ±3.09^{ab}	NA	30.7±2.31^{ab}	NA
Barbara	228	NA	47.78 ±2.34	NA	5.96 ±2.71^{ab}	NA	32.95±2.55^{a}	NA
BARE1	294	NA	44.17 ±2.61	NA	7.56 ±2.87^{ab}	NA	31.21±3.36^{ab}	NA
Claudia	238	NA	47.16 ±3.84	NA	5.17 ±2.96^{b}	NA	28.02±5.29^{ab}	NA
Egug	186	NA	49.93 ±2.62	NA	6.70 ±2.20^{ab}	NA	28.38±2.67^{ab}	NA
Fatima	316	NA	54.43 ±1.73	NA	5.23 ±2.02^{ab}	NA	27.89±3.8^{ab}	NA
Nusif	274	NA	38.39 ±1.30	NA	4.51 ±1.10^{b}	NA	26.65±4.36^{b}	NA
Romani	182	NA	40.53 ±3.83	NA	9.65 ±4.73^{a}	NA	29.69±4.6^{ab}	NA
Sabine	242	NA	54.91 ±4.98	NA	5.63 ±2.10^{ab}	NA	28.44±2.29^{ab}	NA
Xalax	147	NA	41.96 ±1.68	NA	6.68 ±2.43^{ab}	NA	32.73±6.67^{ab}	NA

Supplementary Table S4 Reorganization of CG and CHG methylation in F1 hybrids between *Aegilops cylindrica* (CY), *Ae. triuncialis* (TR) and *Ae. geniculata* (GE). Mean proportions and standard deviation of loci showing CG methylation and CHG methylation in each of the reciprocal hybrid type as assessed by the comparison of methylation states in F1 hybrids to their expected additivity in the parents. Parental shared methylation presents mean proportions of random sequences (MSAP) and of insertions from nine TE families (MSTD) showing similar methylation states in both parents. TE families with significantly different proportions of methylation changes than random loci as tested by Tukey HSD do not share small letters. NA: not available, because no hybrid survived.

Loci type	number of loci		% parental shared methylation		% CG methylation		% CHG methylation	
	CY x GE	***GE x CY***	***CY x GE***	***GE x CY***	***CY x GE***	***GE x CY***	***CY x GE***	***GE x CY***
MSAP	591	605	31.68±4.02	31.89±3.91	6.66±1.05^{a}	8.76±2.22^{ab}	8.23±2.12^{a}	9.93±1.83^{ab}
Barbara	149	149	23.96±3.14	21.68±3.26	11.97±3.18^{bc}	12.8±1.17^{ab}	8.85±1.69^{a}	7.32±2.2^{ab}
BARE1	181	171	19.07±3.92	28.75±4.55	10.9±3.14^{abc}	10.61±1.76^{ab}	10.68±2.86^{a}	9.23±2.34^{ab}
Claudia	186	195	20.18±5.08	27.09±1.67	11.61±3^{bc}	9.69±2.46^{ab}	9.63±8.42^{a}	8.84±1.67^{ab}
Egug	221	224	31.39±1.45	32.01±0.9	9.75±2.59^{ab}	8.74±2.12^{ab}	10.7±2.31^{a}	9.04±1.3^{ab}
Fatima	225	213	17.15±1.24	17.88±0.51	11.9±3.42^{bc}	11.48±2.67^{ab}	6.43±3.96^{a}	7.17±1.98^{ab}
Nusif	217	224	16.87±0.25	17.05±0.58	9.74±3.13^{ab}	10.81±5.01^{ab}	7.9±3.85^{a}	10.26±3.49^{ab}
Romani	179	186	24.64±3	25.9±1.35	9.27±1.81^{ab}	7.25±1.76^{b}	10.68±3.35^{a}	11.49±3.69^{a}
Sabine	197	197	15.08±2.62	20.49±1.49	14.97±3.39^{c}	14.78±3.59^{a}	9.53±3.65^{a}	8.39±1.12^{ab}
Xalax	184	192	17.9±2.51	23.26±1.97	8.31±2.54^{ab}	13.06±2.83^{ab}	7.27±5.25^{a}	5.33±2^{b}
	CY x TR	***TR x CY***	***CY x TR***	***TR x CY***	***CY x TR***	***TR x CY***	***CY x TR***	***TR x CY***
MSAP	588	584	32.87±6.04	38.79±2.07	7.18±1.81^{a}	7±1.06^{a}	7.46±1.45^{abcd}	6.05±1.74^{a}
Barbara	154	156	21.32±5.09	18.79±1.4	11.07±4.58^{abc}	15.86±5.41^{a}	7.42±1.62^{abcd}	9.31±6^{a}
BARE1	175	165	22.26±1.99	21.2±2.69	12.26±2.1^{bc}	13.18±4.89^{a}	9.37±2.76^{ab}	9.36±3^{a}
Claudia	183	186	23.65±3.04	24.23±3.9	11.81±3.12^{abc}	11.37±8.71^{a}	5.26±2.29^{cd}	8.32±5.22^{a}
Egug	220	215	31.26±4.17	31.76±1.34	11.1±3.81^{abc}	9.95±2.86^{a}	9.46±2.13^{ab}	10.03±1.78^{a}
Fatima	225	222	17.76±4.86	18.7±2.47	15.03±4.26^{c}	17.75±6.99^{a}	5.05±2.92^{cd}	4.73±3.21^{a}
Nusif	216	217	22±4.4	21.45±2.1	11.59±3.33^{abc}	15.31±9.05^{a}	4.44±1.48^{d}	6.07±3.41^{a}
Romani	185	194	27.08±5.22	30.98±3.76	11.95±3.35^{abc}	10.91±1.49^{a}	10.26±3.46^{a}	8.77±1.37^{a}
Sabine	198	194	18.64±2.71	14.8±3.11	13.8±2.67^{bc}	16.68±10.05^{a}	7.97±2.5^{abc}	6.13±4.07^{a}
Xalax	189	181	21.15±3.37	22.15±3.82	9.23±3.33a^{b}	10.1±6.16^{a}	6.46±1.83^{bcd}	8.63±7.65^{a}
	TR x GE	***GE x TR***	***TR x GE***	***GE x TR***	***TR x GE***	***GE x TR***	***TR x GE***	***GE x TR***
MSAP	584	NA	24.31±14.77	NA	6.76±1.35^{a}	NA	5.45±1.99^{a}	NA
Barbara	151	NA	21.85±1.32	NA	12.19±2.99^{bc}	NA	6.92±2.18a^{b}	NA
BARE1	158	NA	19.91±1.79	NA	8.78±3.30a^{b}	NA	6.93±2.42a^{b}	NA
Claudia	183	NA	28.54±2.36	NA	9.97±2.62^{abc}	NA	8.16±2.39^{abc}	NA
Egug	216	NA	34.64±3.34	NA	8.51±1.65b^{a}	NA	9.36±3.59^{abc}	NA
Fatima	216	NA	17.53±4.47	NA	12.65±4.11^{bc}	NA	5.70±3.78^{a}	NA
Nusif	225	NA	20.26±1.28	NA	8.66±2.94a^{b}	NA	6.98±1.54^{ab}	NA
Romani	187	NA	25.80±1.97	NA	7.60±1.28a^{b}	NA	11.50±2.36^{c}	NA
Sabine	212	NA	19.59±3.97	NA	15.98±3.35^{c}	NA	10.41±3.34^{ba}	NA
Xalax	191	NA	21.56±2.49	NA	11.45±3.97^{abc}	NA	6.01±1.54^{a}	NA

Supplementary Table S5 Proportion and origin of nonrandom lost loci or CG methylated loci for AFLP as in the nine TE family among all loci (in brackets the number of loci) identified in *Aegilops cylindrica* (CY), *Ae. triuncialis* (TR) and *Ae. geniculata* (GE).* identified reciprocal crosses with significant nonrandom loci assessed by tests of proportion on R cran. The origin is from maternald/paternal origin.

	CYxGE		GExCY			CYxTR		TRxCY			TRxGE	
	nonrandom lost loci	origin nonrandom lost loci	nonrandom lost loci	origin nonrandom lost loci		nonrandom lost loci	origin nonrandom lost loci	nonrandom lost loci	origin nonrandom lost loci		nonrandom lost loci	origin nonrandom lost loci
AFLP	28.90% (229)	29%/71%(67/162)	19.41% (136)	29%/71%(39/97)	**	30.40% (228)	53%/47%(121/107)	27.11% (196)	44%/56%(86/110)		28.50% (214)	48%/52%(102/112)
Barbara	25.65% (63)	37%/63%(23/40)	22.78% (52)	21%/79%(11/41)	***	25.51% (62)	63%/37%(39/23)	14.72% (34)	26%/74%(9/25)	**	25.44% (58)	59%/41%(34/24)
BARE1	18.44% (56)	43%/57%(24/32)	19.86% (55)	29%/71%(16/39)		29.28% (89)	48%/52%(43/46)	14.43% (42)	48%/52%(20/22)	***	20.41% (60)	43%/57%(26/34)
Claudia	20.33% (50)	42%/58%(21/29)	19.29% (42)	24%/76%(10/32)		24.21% (61)	52%/48%(32/29)	17.01% (41)	34%/66%(14/27)	*	18.45% (44)	43%/57%(19/25)
Egug	24.69% (50)	48%/52%(24/26)	16.71% (36)	19%/81%(7/29)		30.77% (60)	65%/35%(39/21)	19.57% (36)	47%/53%(17/19)	**	23.09% (43)	37%/63%(16/27)
Fatima	22.13% (68)	37%/63%(25/43)	11.93% (33)	18%/82%(6/27)	**	20.34% (59)	39%/61%(23/36)	20.42% (58)	53%/47%(31/27)		20.27% (64)	42%/58%(27/37)
Nusif	18.25% (51)	35%/65%(18/33)	16.83% (48)	35%/65%(17/31)		22.73% (65)	48%/52%(31/34)	15.47% (43)	40%/60%(17/26)	*	15.68% (43)	44%/56%(19/24)
Romani	23.6% (45)	27%/73%(12/33)	17.67% (28)	29%/71%(8/20)	*	33.51% (62)	45%/55%(28/34)	20.11% (37)	41%/59%(15/22)	**	26.42% (48)	56%/44%(27/21)
Sabine	21.48% (55)	25%/75%(14/41)	16.7% (38)	47%/53%(18/20)		21.22% (52)	44%/56%(23/29)	17.08% (41)	51%/49%(21/20)		21.48% (52)	54%/46%(28/24)
Xalax	21.02% (32)	44%/56%(14/18)	28.78% (39)	36%/72%(14/28)		34.38% (55)	51%/49%(28/27)	20.95% (31)	39%/61%(12/19)	**	21.06% (31)	55%/45%(17/14)
	CYxGE		GExCY			CYxTR		TRxCY			TRxGE	
	nonrandom CG meth loci	origin CG meth loci	nonrandom CG meth loci	origin CG meth loci		nonrandom CG meth loci	origin CG meth loci	nonrandom CG meth loci	origin CG meth loci		nonrandom CG meth loci	origin CG meth loci
AFLP	6.60% (39)	46%/49%/(18/19)	3.31% (20)	50%/50%/(10/10)	*	6.97% (41)	56%/44%/(23/18)	4.45% (26)	46%/54%/(12/14)		6.50% (38)	82%/18%/(31/7)
Barbara	10.74% (16)	56%/44%/(9/7)	6.02% (9)	50%/50%/(4/5)		8.41% (13)	62%/31%/(8/4)	8.95% (14)	64%/36%/(9/5)		9.94% (15)	53%/47%/(8/7)
BARE1	10.49% (19)	42%/42%/(8/8)	9.91% (17)	53%/47%/(9/8)		10.29% (18)	33%/67%/(6/12)	8.47% (14)	43%/57%/(6/8)		6.94% (11)	45%/55%/(5/6)
Claudia	11.85% (22)	55%/45%/(12/10)	7.70% (15)	47%/53%/(7/8)		9.82% (18)	67%/33%/(12/6)	3.23% (6)	50%/50%/(3/3)	*	9.30% (17)	53%/47%/(9/8)
Egug	10.40% (23)	48%/52%/(11/12)	5.81% (13)	62%/38%/(8/5)		9.52% (21)	52%/57%/(11/12)	5.12% (11)	64%/36%/(7/4)		8.77% (19)	42%/58%/(8/11)
Fatima	11.57% (26)	73%/27%/(19/7)	9.91% (21)	43%/57%/(9/12)		12.01% (27)	78%/22%/(21/6)	10.80% (24)	54%/46%/(13/11)		12.97% (28)	54%/46%/(15/13)
Nusif	16.56% (36)	61%/39%/(22/14)	6.69% (15)	47%/53%/(7/8)	**	12.48% (27)	41%/59%/(11/16)	8.76% (19)	42%/58%/(8/11)		8.88% (20)	35%/65%/(7/13)
Romani	8.96% (16)	67%/33%/(10/6)	5.37% (10)	70%/30%/(7/3)		11.37% (21)	43%/57%/(9/12)	8.77% (17)	65%/35%/(11/6)		6.97% (13)	46%/38%/(6/5)
Sabine	9.15% (18)	39%/61%/(7/11)	3.04% (6)	67%/33%/(4/2)		11.09% (22)	55%/36%/(12/8)	10.29% (20)	55%/45%/(11/9)		10.35% (22)	45%/55%/(10/12)
Xalax	17.39% (32)	56%/44%/(18/14)	20.30% (39)	54%/46%/(21/18)		30.21% (57)	54%/46%/(31/26)	17.14% (31)	43%/55%/(14/17)	**	11.81% (30)	53%/47%/(16/14)

Introgression and genome restructuring in natural hybrid zones between two wild wheat species

Natacha Senerchia[1], François Felber[1,2], Christian Parisod[1]

[1]Laboratory of Evolutionary Botany, Institute of biology, University of Neuchâtel, Switzerland

[2]Musée et Jardins botaniques cantonaux, Lausanne, Switzerland

This chapter is in preparation as a brief communication for Evolution

Abstract

Mechanisms and evolution of inter-species barriers through hybridization is of main interest to understand reproductive isolation patterns and hybrids formation and selection. Despite the complexity of working in natural populations, natural hybrid zones allow to study surviving hybrids having bypassed species barriers. Four natural populations of northern Israel composed of the species *Ae. geniculata* and *Ae. triuncialis* in sympatry were investigated, assessing hybridization and recurrent backcrossing in natural conditions. Rather than stable hybrid lineages, inferred hybrids were probably mostly backcrossed with *Ae. triuncialis*, supporting asymmetrical reproductive isolation due to conflicts among transposable elements (TEs) revealed by a previous study on experimental F1 hybrids. Genetic admixture was identified for mostly quiescent TEs among the six investigated TEs as compared to genome-wide loci, indicating that introgression of active TEs is counterselected to probably avoid conflicts and abnormal development of hybrids. In addition, important restructuring involving predominant loss of TE loci was assessed in hybrids, supporting the purging of TE sequences as a crucial process to produce viable hybrids in natural populations.

Introduction

Hybridization, important process among plants is postulated as crucial process in species speciation and evolution (Abbott et al. 2013). Hybrid zones are areas where divergent lineages meet and cross-fertilize. Such specific areas allow to explore the genetics and evolutionary forces underlying species boundaries in natural conditions (Anderson 1949; Barton and Hewitt 1985). Many closely related plants hybridize and backcross spontaneously, having important evolutionary consequences on natural populations (Ellstrand et al. ; Rieseberg and Wendel 1993). Depending on the fitness of hybrids, factors such as habitat preferences, hybridization may result in introgression, stabilization of the hybrid zone, hybrid speciation or extinction (Carney et al. 2000). Riesberg and Wendel *et al.* (1993) argue that the most frequent outcome of hybridization is introgression (i.e. interspecific genetic material exchange).

Hybridization is by no means possible between all species, even close ones. Reproductive isolation is central to the maintenance of sympatric species inhibiting gene-flow. Pre-zygotic reproductive isolation includes the effects of ecological and genetic factors influencing divergence mating between species such as flowering time or pollen-pistil compatibility, whereas post-zygotic reproductive isolation includes hybrid viability and fertility (Tiffin et al. 2001; Rieseberg and Willis 2007). Reproductive barriers may act differentially depending on the direction of the cross revealing asymmetrical hybridization success. Chromosomal rearrangements or genetic incompatibilities of few nuclear loci (Dobzhansky-Muller model) are thought as the main drivers of post-zygotic barriers to gene exchange and will results in symmetrical incompatibilities (Stebbins 1971; Rieseberg and Carney 1998; Tiffin et al. 2001; Coyne and Orr 2004). Asymmetrical reproductive isolation is commonly reported among plant species, suggesting the implication of other(s) mechanism(s)

such as cytonuclear or gametophyte-sporophyte interactions (Lowry et al. 2008). Parisod and Senerchia (2012) suggested that, among others, asymmetrical reproductive isolation may be explained by conflicts among transposable element (TE) following hybridization (Parisod and Senerchia 2012; Senerchia et al. 2014b). Introgression through hybridization and backcrossing reveals new interactions among divergent genomes and may expose genetic incompatibilities accumulated among species. Important genome reorganization was assessed following hybridization in part through the activation of transposable elements (Rieseberg 2001), suggesting a crucial role of TEs in host genome evolution. TEs, specifically LTR retrotransposon composed the majority of many plant genomes (Sabot and Schulman 2009) and showed, at the family level, specific dynamics following hybridization (Senerchia et al. 2014a).

Species of the Triticeae clade evolved through homoploid divergence, hybridization and allopolyploidy and presented a great diversity of di-, tetra- and hexaploid species that leaved in intermingled populations (van Slageren 1994). In particular, the genus *Aegilops*, wild relatives of cultivated wheat independently and naturally evolved multiple species of various ploidy levels that mostly co-occur in the Middle East (van Slageren 1994; Kilian et al. 2011) fostering gene exchange. Interspecific genetic exchanges are documented between many species (Kilian et al. 2011) and to a certain extent are responsible for the badly dated reticulate evolutionary relationships among *Aegilops* taxa (reviewed in Baum et al. 2012). Cytogenetic analyses of Triticeae genomes, such as *Aegilops*, highlighted large genome with gene-poor and gene-rich regions (Feuillet and Keller 1999; Sandhu and Gill 2002).

Natural hybrid zones may provide material to better understand mechanisms and forces acting spontaneously in individual that manage to bypass species boundaries and are viable. Here, we analyzed four hybrid zones from North Israel encompassing the two species

Ae. geniculata and *Ae. triuncialis*. Backcrossing was detected using hybrids index based on AFLP data. Comparison between assignment score based on random sequences (AFLP) and of six TE families (SSAP) indicated that most TEs were differently exchanged between the two tetraploid species. Important genome restructuring was assessed among hybrids of which no sign of proliferation was assessed, but significantly more eliminated fragments than random sequences was shown.

Material and methods

Hybrid zones and sampling

Sampling area was located in the center of origin of both species, representing the local northern distribution area of *Ae. geniculata* and southern distribution area of *Ae. triuncialis*. The two species are morphologically dissimilar and share similarly dry and disturbed habitat (van Slageren 1994), but are rarely observed in sympatry in the Middle East *Ae. geniculata* was identified from 300 to 1750 m (rarely up to 2100m) whereas *Ae. triuncialis* showed larger altitude ranges, from the sea level to 1900 with rare indications up to 2700m.

A total of 75 individuals from four mixed populations of *Ae. geniculata* and *Ae. triuncialis* were sampled in Northern Israel (Golan, during June 2011) (Table 1). Species were intermingled and hybrids were not morphologically identifiable. Leaves of individuals located at least one meter away from each other were sampled following short transects and dried in silica gel.

DNA was extracted from mechanically disrupted leaves following a standard DNeasy plant extraction mini kit protocol from Qiagen AG, Switzerland. Three samples from pure *Ae. geniculata* and *Ae. triuncialis* natural populations were provided by Prof M. Feldman and considered here as parental references.

Table 1 Investigated populations along natural hybrid zones between *Ae. geniculata* and *Ae. triuncialis* in Golan, North Israel. The number of sampled individuals (N) as well as the population composition based on c-fuzzy means clustering of AFLP data.

Location	Coordinate	N	altitude (m)	dominant species	Ae. geniculata (%)	Ae. triuncialis (%)	introgressed ind. (%)
Hermon Mt, Govta creek	33°18'1.7''N 35°46'46.8''E	27	1831	*Ae. geniculata*	62.96	18.52	18.52
Masada	33°12'8.04''N 35°46'2.19''E	8	1092	*mixed*	50.00	25.00	25.00
Nimrod	33°15'10.32''N 35°45'33.1''E	14	1065	mixed	35.71	28.57	35.71
Bukata	33°9'48.2''N 35°46'39.48''E	18	986	*Ae. triuncialis*	16.67	61.11	22.22

Molecular fingerprint techniques and TE assayed

Digestion, ligation, preselective and selective PCR amplifications from amplified fragment length polymorphism (AFLP) and sequence specific amplified polymorphism (SSAP) were carried out in 96-well plates with randomly positioned individual DNA and were performed using the exact same protocol described in Senerchia *et al.* (2014). Selective amplifications were performed with three primer combinations for AFLP and two combinations for each SSAP as described in Table S1. Genotyping error rates were estimated based on the replication of the whole procedure of three samples (i.e. ca. 3% of the dataset) following (Bonin et al. 2004). Fluorescently labeled PCR products were separated with an ABI 3500 capillary sequencer, using GeneScan™ 500 LIZ™. Resulting electropherograms were visualized and scored with GeneMapper 3.5 (Applied Biosystems) using AFLP default peak

detection parameters. The scoring was manually checked and loci were recorded as present (1) or absent (0) in binary matrices.

SSAP was performed independently for six selected LTR retrotransposon families having differentially evolved in the genomes of *Ae. geniculata* and *Ae. triuncialis* (*BARE1*, *Claudia*, *Egug*, *Fatima*, *Romani*, *Sabine*). Following Senerchia *et al.* (2014) *BARE1* and *Romani* showed evidence of proliferation in both *Ae. geniculata* and *Ae. triuncialis*. In contrast, *Claudia* and *Fatima* showed evidence of proliferation in *Ae. geniculata*, *whereas Sabine* showed proliferation in *Ae. triuncialis* but massive elimination in *Ae. geniculata.* In contrast to those TE families expected to get into conflictual interactions in interspecific hybrids, *Egug* was reported as quiescent on both *Ae. geniculata* and *Ae. triuncialis*.

Statistical analyses of the hybrid zones

Following a genomic cline approach (Nolte et al. 2009; Teeter et al. 2010), Individuals were assigned to two clusters based on their AFLP genotype using the fuzzy c-means algorithm (Dunn 1973) implemented in the package 'e1071' in R cran (degree of fuzzification =1.2, verbose=F, Euclidean distance, methods cmeans) (Meyer et al. 2012). The fuzzy c-means clustering is an extension of the multivariable k-means algorithm that iteratively allocate individuals within a predefined number of groups (here two), while minimizing the intragroup variance, and then estimate an assignment score of individuals to each groups (Bezdek 1981). The assignment score is related to the inverse of the Euclidean distance between the focal individual and the centroid of the groups, and can be considered as an admixture score (i.e. hybrid index) estimated without biological assumptions (Gompert et al. 2010; Arrigo et al. 2011).

Whether specific loci were driving the assignment of individuals, the score was estimated 10’000 times after random removal of one locus (with replacement). The mean, 0.05 and 0.95 quantile were assessed for each individual (table S2). This procedure was followed for AFLP and each TE family independently. Based on their AFLP assignment score, individuals were first ranked and inflections points allocated individuals to *Ae. geniculata* (i.e. assignment scores close to zero), to *Ae. triuncialis* (i.e. assignment scores close to one) and hybrids (i.e. intermediate assignment score values between 015 and 0.85). For each individual, the difference in absolute value between the assignment score of each SSAP and AFLP one was estimated by deducing the mean, the 0.05 and 0.95 quantile of each SSAP to the respective AFLP one. Within *Ae. geniculata*, *Ae. triuncialis* and hybrids individuals, the mean and the standard deviations mentioned above calculated were assessed.

Hybrids were further assigned using AFLPOP version 1.1 (Duchesne and Bernatchez 2001), by estimating the likelihood that an individual belongs to each population, to first generation hybrids (F1), or first generation backcrosses *Ae. geniculata*, (BC1GE), or *Ae. triuncialis* (BC1TR) and backcrosses of second or more generations (BC2GE, BC2TR)

Genome restructuring in hybrids

For each locus j, frequencies of band presence (1) and absence (0) were estimated among individuals previously assigned to either *Ae. geniculata* ($D_1(1)_j$ and $D_1(0)_j$) or *Ae. triuncialis* ($D_2(1)_j$ and $D_2(0)_j$) by the c-means based on AFLP. The expected additivity of the progenitors in hybrids was estimated as probabilities of band presence and absence based on frequencies among potential progenitors. The expected probability of band presence in the hybrids was estimated as $E(1)_j = D_1(1)_j \times D_2(1)_j + D_1(1)_j \times D_2(0)_j + D_1(0)_j \times D_2(1)_j$, whereas the expected probability of band absence was estimated as $E(0)_j = D_1(0)_j \times D_2(0)_j$. For each individual

assessed as hybrid based on AFLP data, the observed profile (A(1), A(0)) was compared to the expected one, estimating the probability that the locus j was (i) additive (i.e. identity between expected and observed hybrid profiles) as $E(1)_j \times A(1)_j + E(0)_j \times A(0)_j$, (ii) a new locus (i.e. presence of a band in the observed hybrid profile despite absence in the expected profile) as $E(0)_j \times A(1)_j$, or (iii) a lost locus (i.e. band absence in the observed hybrid profile despite predicted presence in the expected profile) as $(E(1)_j \times A(0)_j)$. Probabilities were estimated for each locus, summed across loci and divided by the total number to obtain proportions of additive, new and lost loci for AFLP and SSAP of each TE family. Proportions for each hybrid individuals were then averaged.

Significant differences between proportions of (i) new loci and (ii) lost loci for AFLP loci and loci for each SSAP (i.e. TE families) were assessed by a one-way analysis of variance (ANOVA) followed by multiple comparisons with Tukey's honest significant differences post-hoc, using the package 'agricolae' on R cran (de Mendiburu 2013).

Results

Hybrid zones assessed with random sequences

Genotyping of 388 AFLP loci (with an error rate of 8.63%, Table S1) assigned individuals to genetic clusters corresponding to either *Ae. geniculata* or *Ae. triuncialis* using fuzzy c-means. Individual assignment scores after resampling followed a relatively smooth sigmoid curve, typical of genomic clines in hybrid zones, with two abrupt inflections defining three groups (Figure 1): individuals belonging to either pure *Ae. geniculata* genetic background (assignments score from 0.003 to 0.07), or pure *Ae. triuncialis* (assignments score from 0.94

to 0.99), as well as intermediate individuals interpreted as hybrids (assignments score from 0.15 to 0.87) (Table S2).

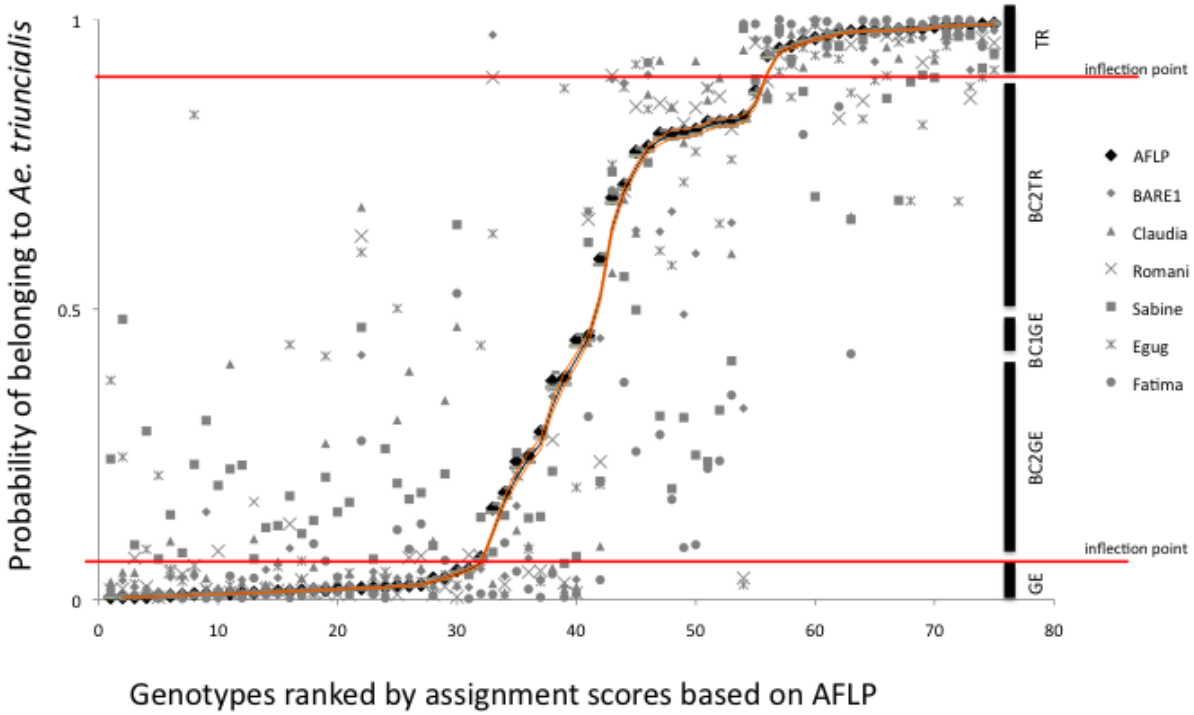


Figure 1 Genomic cline across the natural hybrid zone between *Ae. geniculata* and *Ae. triuncialis*. Assignment of samples to *Ae. geniculata* (0) or to *Ae. triuncialis* (1) by c-fuzzy means on genome-wide AFLP data is represented as black dots with continuous lines show 0.25 and 0.95 quantile. The six TE families are shown according to the panel. Individuals were classified as pure *Ae. geniculata* (GE), pure *Ae. triuncialis* (TR), first generation backcross hybrids with the parental species *Ae. geniculata* (BC1GE), second or more generation backcross hybrids with the parental species *Ae. geniculata* (BC2GE) and second or more generation backcross hybrids with the parental species *Ae. triuncialis* (BC2TR). Red lines reflected inflexion points.

Dispersion of assignment scores based on random AFLP loci evaluated by the 0.05 and 0.95 quantiles was low, indicating that several loci discriminate clusters. Among the 75 analyzed individuals from four natural populations, 32 belonged to *Ae. geniculata*, 20 to *Ae. triuncialis* and 23 were assessed as hybrids (Table S2). The four populations showed between 18.52% and 35.71% of hybrids (Table 1). Population situated at Mt Hermon and in Masada showed more individuals assigned to *Ae. geniculata* and equal proportion of individuals

representing *Ae. triuncialis* and hybrids, whereas population of Bukata showed *Ae. triuncialis* as the dominant species. Population situated at Nimrod showed similar proportions of individuals corresponding to the different groups.

Among the 23 individuals assigned as hybrids, AFLPOP 1.1 allocated two individuals as first generation backcrosses with *Ae. geniculata*, 7 individuals as second or more generations backcrosses with *Ae. geniculata* and 14 individuals as second or more generations backcrosses with *Ae. triuncialis*.

TE sequences across the hybrid zones

The genotyping of SSAP produced from 133 to 250 SSAP loci for each of the six TE families (with an error rate ranging from 3.9 to 6.02%, Table S1). Comparing individual assignment scores based on AFLP to the score assessed of each TE family (i.e. SSAPs) showed that TE sequences were generally congruent with AFLP (Figure 1). Mean deviation between AFLP and SSAP was of 0.09 for *Ae. geniculata*, 0.053 for individuals of *Ae. triuncialis* and 0.236, the highest difference, for hybrid individuals was assessed (Table 2). Considering all TEs, *Sabine* displayed higher difference as compared to AFLP among *Ae. geniculata*, *Ae. triuncialis* and hybrid individuals. *Claudia* and *Egug* among *Ae. geniculata* and hybrids, *Fatima* among *Ae. triuncialis* individuals and hybrids, and *Claudia* among *Ae. geniculata* individuals. *BARE1* and *Romani* showed the lowest differences as compared to AFLP.

Table 2 Assignment scores. Difference between each TE family assignment score assessed by c-fuzzy means analyses and the AFLP assignment score among *Ae. geniculata*, *Ae. triuncialis* and introgressed individuals separately.

	BARE1	*Claudia*	*Egug*	*Fatima*	*Romani*	*Sabine*
Ae. geniculata	0.035±0.002	0.100±0.020	0.140±0.020	0.050±0.009	0.040±0.010	0.180±0.010
Ae. triuncialis	0.018±0.002	0.040±0.006	0.090±0.008	0.060±0.001	0.030±0.004	0.080±0.007
introgressed ind.	0.210±0.020	0.160±0.010	0.200±0.020	0.360±0.020	0.190±0.020	0.260±0.020

Genome changes of introgressed individuals

Genome changes in hybrids was estimated through observed deviation of genetic profiles as compared to expectations based on combinational probabilities of loci frequencies among individuals assessed as pure progenitors across the whole hybrid zone (Table 3).

Proportion of additive AFLP bands among the hybrids showed mean of 68.05% of AFLP loci revealed additive among hybrids, whereas mean proportion of 22.39% were lost and 9.30% were assessed as new loci (table 3). Proportions of additive SSAP bands varied among TE families, ranging from 60.02% for *Sabine* to 66.52% for *Egug*. Proportion of lost bands varied from 25.15% for *Egug* to 30.98% for *Sabine*, and new bands proportions were from 7.20% for *Fatima* to 9.44% for *BARE1*. All TE families presented a significantly higher proportion of lost bands than AFLP, expect *Egug*. No significant differences between AFLP and SSAP proportions of new loci were reported.

Table 3 Genome restructuring proportions in natural introgressed individuals between *Ae. geniculata* and *Ae. triuncialis*. Mean proportions of additive loci (A), lost loci (L) and new loci (N) for random sequences (AFLP) and for the six TE families (SSAP) assessed from the observed deviation from the expected additivity of pure parental individuals. TE families differing significantly from AFLP do not share small letters, as assessed by Tukey tests.

	A	L	N
AFLP	68.05±6.36	22.39±5.48^{a}	9.30±3.3^{a}
BARE1	60.21±4.71	29.95±5.18^{c}	9.44±2.38^{a}
Claudia	63.36±6.63	27.50±5.35^{bc}	8.41±2.77^{a}
Egug	66.52±8.85	25.19±9.38^{ab}	7.54±3.51^{a}
Fatima	62.61±7.92	29.88±7.32^{c}	7.10±2.21^{a}
Romani	61.82±8.13	28.37±6.64^{bc}	9.11±4.01^{a}
Sabine	60.02±5.04	30.97±5.11^{c}	8.23±2.06^{a}

Discussion

Introgression between Ae. geniculata and Ae. triuncialis

Important introgression between *Ae. geniculata* and *Ae. triuncialis* was assessed across natural populations of northern Israel, where both species reach their respective range limits and co-occur. Using reliable fingerprint techniques (AFLP, (Bensch et al. 2002; Bensch and Akesson 2005)), between 18.52% and 35.71% of hybrids were identified, supporting permeable species boundaries across the hybrid zones with recurrent inter-species crossing having circumvented both pre and post zygotic reproductive barriers. All identified 23 hybrids were assessed as backcrosses. Nine hybrids (39%) were backcrossed with *Ae. geniculata*, whereas 14 (61%) involved *Ae. triuncialis*. The absence of inferred F1s or stable hybrid lineage suggests that early hybrids were unable to self-fertilize. Male sterility was indeed reported in artificial F1 hybrids between *Ae. geniculata* and *Ae. triuncialis* (Senerchia et al. 2014b), supporting backcrossing as the main reproductive process for hybrids.

In contrast to Kilian *et al.* (2011), who emphasized hybridization with *Ae. geniculata* as the pollen receiver, natural hybrids here showed preferential backcrossing with *Ae. triuncialis*. Artificial hybridization showed F1 seed production in both directions, but only germination of seeds from the cross *Ae. triuncialis* x *Ae. geniculata* (Senerchia et al. 2014b). Such asymmetrical reproductive isolation is coherent with lower nucleo-cytoplasmic conflicts (Tiffin et al. 2001) following backcrosses with *Ae. triuncialis*. Accordingly, mostly hybrids having a *Ae. triuncialis* cytoplasm may have survived in the natural hybrid zones, but the minor proportion of inferred backcrosses with *Ae. geniculata* deserve explanations. Further analyses using plastid markers are necessary to shade light on the processes ruling gene-flow in these hybrid zones. Asymmetrical hybridization success among *Aegilops* species was associated with conflicts among TEs in F1 hybrids (Senerchia *et al.* 2014b), prompting investigation of TE dynamics in natural hybrid zones.

Inter-species TE-specific loci exchange

Comparing genetic admixture assessed for each of the six TE families (i.e. SSAPs) to random fragments (i.e. AFLP), exchanges of TE loci among species were coherent with genome-wide sequences. Highly active TEs such as *BARE1* and *Romani* showed similar assignment scores than genome-wide loci, sustaining limited inter-species gene-flow in corresponding genome regions. In stark contrast, the quiescent *Egug* (Senerchia et al. 2013; Senerchia et al. 2014a) revealed different assignment scores coherent with large inter-species exchanges as compared random loci. *Sabine* that was quiescent in *Ae. triuncialis* and massively eliminated from *Ae. geniculata* showed a similar pattern than *Egug*. *Claudia* and *Fatima* displayed evidence of recent activation in *Ae. geniculata* but not in *Ae. triuncialis* showed a pattern in between these

extremes. Such patterns are coherent with the hypothesis that divergent arrangements of insertions between species restrict recombination and thus gene-flow. However, the pattern of asymmetrical gene flow, with TE loci from *Ae. geniculata* introgressing more frequently into *Ae. triuncialis* than the opposite, seems to favor the non-mutually exclusive hypothesis that introgression of active TEs is counterselected to avoid conflicts and dysgenesis in hybrids.

Genome changes in natural hybrids

Introgressed individuals presented important deviation from the expected addition of the progenitors in random sequences as well as in each TE fractions, suggesting that hybridization and/or backcrossing events lead to genome changes though restructuring and/or transgressive segregation. Artificial F1 hybrids between the two species revealed important and repeatable structural changes, particularly in TE fractions (Senerchia et al. 2014b), suggesting restructuring of active TE insertions as a necessary process to produce viable hybrids. All here investigated TE families, except *Egug*, were assessed as active in the parental species and were expected to be involved in conflictual interactions in hybrids between *Ae. geniculata* and *Ae. triuncialis* (Senerchia et al. 2013; Senerchia et al. 2014a). However, no evidence of proliferation was identified and loss of TE loci was significantly more frequent than loss of random sequences for all TEs except *Egug*. Hybrids that underwent important proliferation of active TEs may have been eliminated in the natural hybrid zone and the prevalent deletion of TE sequences during early hybrid development suggests that purging of active or deleterious insertions may be essential for hybrid survival (Rieseberg 2001).

Conclusion

Asymmetrical introgression across the hybrid zones between *Ae. geniculata* and *Ae. triuncialis*, with preferential backcrossing towards *Ae. triuncialis*, suggests that this latter species may be advantaged and colonize *Ae. geniculata* range. This asymmetry further supports a TE-specific selection associated with conflicts and restructuring early after hybridization, highlighting the crucial impact of TEs on host genome evolution in natural populations of wild wheats. Long-term monitoring of such hybrid zones will provide a deeper understanding of patterns and processes underlying species integrity.

Acknowledgements

We sincerely thank M. Feldman and E. Millet for the sampling. This work was supported by the Swiss National Science Foundation through the National Centre of Competence in Research 'Plant Survival' and a grant (PZ00P3-131950 to CP).

References

Abbott, R., D. Albach, *et al.* 2013. Hybridization and speciation. *J Evol Biol* **26**:229-246.

Anderson, E. 1949. Introgressive hybridization. John Wiley. *New York, New York, USA.*

Arrigo, N., R. Guadagnuolo, *et al.* 2011. Gene flow between wheat and wild relatives: empirical evidence from *Aegilops geniculata*, *Ae. neglecta* and *Ae. triuncialis*. *Evol.Applic.* **4**:685-695.

Barton, N. H. and G. M. Hewitt. 1985. Analysis of hybrid zones. *Ann. Rev. Ecol. Syst.s* **16**:113-148.

Baum, B. R., T. Edwards, *et al.* 2012. Phylogenetic relationships among the polyploid and diploid *Aegilops* species inferred from the nuclear 5S rDNA sequences (Poaceae: *Triticeae*). *Genome* **55**:177-193.

Bensch, S. and M. Akesson. 2005. Ten years of AFLP in ecology and evolution: why so few animals?. *Mol Ecol* **14**:2899-2914.

Bensch, S., A. Helbig, *et al.* 2002. Amplified fragment length polymorphism analysis identifies hybrids between two subspecies of warblers. *Mol Ecol* **11**:473-481.

Bezdek, J. C. 1981. Pattern recognition with fuzzy objective function algorithms. Kluwer Academic Publishers.

Bonin, A., E. Bellemain, *et al.* 2004. How to track and assess genotyping errors in population genetics studies. *Mol Ecol* **13**:3261-3273.

Carney, S. E., K. A. Gardner, *et al.* 2000. Evolutionary changes over the fifty-year history of a hybrid population of sunflowers. *Evolution* **54**:462-474.

Coyne, J. and H. Orr. 2004. Speciation, Sunderland.

de Mendiburu, F. 2013. agricolae: statistical procedures for agricultural research. R package version 1.1-4.

Duchesne, P. and L. Bernatchez. 2001. AFLPOP : A computer program for simulated and real population allocation based on AFLP data. *Mol Ecol* **3**:380-383.

Dunn, J. C. 1973. A fuzzy relative of the ISODATA process and its use in detecting compact, well-separated clusters. *J. Cybern.* **3**:32-57.

Ellstrand, N. C., H. C. Prentice, *et al.* Gene flow and introgression from domesticated plants into their wild relatives. *Horizontal gene transfer*:217-234.

Feuillet, C. and B. Keller. 1999. High gene density is conserved at syntenic loci of small and large grass genomes. *P.N.A.S.* **96**:8265-8270.

Gompert, Z., L. K. Lucas, *et al.* 2010. Secondary contact between *Lycaeides idas* and *L. melissa* in the Rocky Mountains: extensive admixture and a patchy hybrid zone. *Mol Ecol* **19**:3171-3192.

Kilian, B., K. Mammen, *et al.* 2011. *Aegilops*. Pp. 1-17 *in* C. Kole, ed. Wild Crop Relatives: Genomic and Breeding Resources, Cereals. Springer-Verlag Berlin Heidelberg.

Lowry, D. B., J. L. Modliszewski, *et al.* 2008. The strength and genetic basis of reproductive isolating barriers in flowering plants. *Philos. Trans. R. Soc. Lond., B* **363**:3009-3021.

Meyer, D., E. Dimitriadou, *et al.* 2012. e1071: misc functions of the department of statistics (e1071), TU Wien. R package version 1.6-1.

Nolte, A., Z. Gompert, *et al.* 2009. Variable patterns of introgression in two sculpin hybrid zones suggest that genomic isolation differs among populations. *Mol Ecol* **18**:2615-2627.

Parisod, C. and N. Senerchia. 2012. Responses of transposable elements to polyploidy. Pp. 147-148 *in* M.-A. Grandbastien, and J. M. Casacuberta, eds. Plant Transposable Elements. Springer-Verlag Berlin Heidelberg.

Rieseberg, L. H. 2001. Polyploid evolution: Keeping the peace at genomic reunions. *Curr. Biol* **11**:R925-R928.

Rieseberg, L. H. and S. E. Carney. 1998. Plant hybridization. *New phytol***140**:599-624.

Rieseberg, L. H. and J. F. Wendel. 1993. Introgression and its consequences in plants. Pp. 70-1114 *in* R. Harrison, ed. Hybrid Zones and the Evolutionary Process. Oxford University Press, New York.

Rieseberg, L. H. and J. H. Willis. 2007. Plant speciation. *Science* **317**:910-914.

Sabot, F. and A. H. Schulman. 2009. Genomics of transposable elements in the *Triticeae*. Pp. 387-405 *in* G. J. Muehlbauer, and C. Feuille, eds. Genetics and genomics of the *Triticeae*. Springer US.

Sandhu, D. and K. S. Gill. 2002. Gene-containing regions of wheat and the other grass genomes. *Plant Physiol.* **128**:803-811.

Senerchia, N., F. Felber, *et al.* 2014a. Contrasting evolutionary trajectories of multiple retrotransposons following independent allopolyploidy in wild wheats. *New Phytol*:doi: 10.1111/nph.12731.

Senerchia, N., F. Felber, *et al.* 2014b. Nonrandom genome reorganisation after hybridization reveals a major role fo transposable elements in reproductive isolation. *in prep.*

Senerchia, N., T. Wicker, *et al.* 2013. Evolutionary dynamics of retrotransposons assessed by next generation sequencing in wild relatives of wheat. *Genome Biol Evol* **5**:1010-1020.

Stebbins, G. 1971. Chromosomal evolution in higher plants *in* E. Arnold, ed, London.

Teeter, K. C., L. M. Thibodeau, *et al.* 2010. The variable genomic architecture of isolation between hybridizing species of house mice. *Evolution* **64**:472-485.

Tiffin, P., M. S. Olson, *et al.* 2001. Asymmetrical crossing barriers in angiosperms. *Proc. R. Soc. B-Biol. Sci.* **268**:861-867.

van Slageren, M. W. 1994. Wild wheats: a monograph of *Aegilops* L. and *Amblyopyrum* (Jaub. & Spach) Eig (*Poaceae*). *Wageningen Agricultural University Papers* **97**.

Supplementary material

Supplementary Table S1 Sequences (5'−> 3') selective primers combinations of amplified fragment length polymorphism (AFLP) and sequence specific amplified polymorphism (SSAP). Adaptors, pre-selective and selective *Eco*RI primers in Parisod (2009). Error rates were evaluated on three replicates following Bonin et al (2004) on the final binary matrix.

TE family	Primer combination	Error rate %	number of loci
BARE1	M-CCA, M-CTA	4.4	250
Claudia	M-CAC, M-CTA	6	138
Egug	M-CGA, M-CTA	4.1	133
Fatima	M-CAA, M-CTG	5.1	244
Romani	M-CAC, M-CGC	3.9	141
Sabine	M-CAC, M-CTA	5	138
AFLP	FAM-AAG/M-CGA, VIC-ATT/M-CAA, NED-ACG/M-CAC	8.6	388

Supplementary Table S2 Mean, quantile 0.05 and quantile 0.95 assignment score calculated on AFLP data using C-fuzzy means algorithm.

Population	Assignment score	quantile 0.05	quantile 0.95
Hermon Mt	0.9963	0.9964	0.9961
Hermon Mt	0.9901	0.9906	0.9896
Hermon Mt	0.9503	0.9529	0.9481
Hermon Mt	0.9963	0.9965	0.9961
Hermon Mt	0.9627	0.9644	0.9613
Hermon Mt	0.9865	0.9871	0.9859
Hermon Mt	0.9906	0.9911	0.9902
Hermon Mt	0.9912	0.9917	0.9908
Hermon Mt	0.9888	0.9893	0.9882
Hermon Mt	0.9937	0.9940	0.9934
Hermon Mt	0.9819	0.9829	0.9811
Hermon Mt	0.8433	0.8503	0.8373
Hermon Mt	0.0206	0.0214	0.0197
Hermon Mt	0.0332	0.0347	0.0318
Hermon Mt	0.9867	0.9875	0.9861
Hermon Mt	0.7114	0.7222	0.7023
Hermon Mt	0.9909	0.9913	0.9905
Hermon Mt	0.6238	0.6372	0.6129
Hermon Mt	0.9639	0.9659	0.9625
Hermon Mt	0.9814	0.9826	0.9806
Hermon Mt	0.0379	0.0395	0.0364
Hermon Mt	0.9850	0.9859	0.9843
Hermon Mt	0.9761	0.9772	0.9750

Hermon Mt	0.2856	0.2917	0.2789
Hermon Mt	0.0517	0.0543	0.0493
Hermon Mt	0.0199	0.0206	0.0190
Hermon Mt	0.1240	0.1270	0.1201
Nimrod	0.6183	0.6282	0.6092
Nimrod	0.0113	0.0117	0.0109
Nimrod	0.7626	0.7717	0.7531
Nimrod	0.9838	0.9844	0.9831
Nimrod	0.0098	0.0100	0.0094
Nimrod	0.0135	0.0139	0.0130
Nimrod	0.0231	0.0239	0.0222
Nimrod	0.9960	0.9962	0.9958
Nimrod	0.9480	0.9506	0.9459
Nimrod	0.9888	0.9894	0.9883
Nimrod	0.5534	0.5639	0.5440
Nimrod	0.0184	0.0190	0.0177
Nimrod	0.9808	0.9819	0.9799
Nimrod	0.4135	0.4221	0.4045
Masada	0.0184	0.0192	0.0177
Masada	0.9791	0.9804	0.9782
Masada	0.1771	0.1827	0.1693
Masada	0.0081	0.0084	0.0078
Masada	0.9927	0.9931	0.9924
Masada	0.9937	0.9940	0.9935
Masada	0.5462	0.5538	0.5393
Masada	0.9966	0.9967	0.9964
Bukata	0.3083	0.3152	0.3002
Bukata	0.1775	0.1835	0.1694
Bukata	0.2201	0.2246	0.2140
Bukata	0.1758	0.1811	0.1683
Bukata	0.7532	0.7640	0.7441
Bukata	0.1906	0.1959	0.1831
Bukata	0.0099	0.0103	0.0095
Bukata	0.0204	0.0214	0.0196
Bukata	0.9840	0.9849	0.9833
Bukata	0.2289	0.2340	0.2223
Bukata	0.9747	0.9760	0.9737
Bukata	0.1947	0.2007	0.1868
Bukata	0.0163	0.0170	0.0155
Bukata	0.1977	0.2029	0.1903
Bukata	0.8164	0.8249	0.8088
Bukata	0.1975	0.2040	0.1887
Bukata	0.9860	0.9866	0.9853
Bukata	0.0112	0.0115	0.0107
Ae. geniculata	0.9280	0.9310	0.9257
Ae. geniculata	0.9760	0.9773	0.9750
Ae. geniculata	0.9798	0.9807	0.9791
Ae. triuncialis	0.0465	0.0484	0.0446
Ae. triuncialis	0.0297	0.0309	0.0286
Ae. triuncialis	0.0631	0.0656	0.0600

Conclusions and perspectives

Conclusions

TEs as central players of genome reorganization following polyploidy

Hybridization is a prevalent evolutionary process, particularly among angiosperms (Jiao *et al.*, 2011), and is associated with genome shocks (Genome Shock hypothesis, McClintock, 1984). The merging of divergent genomes was shown to result in revolutionary events through which genomes are reorganized and may lead to speciation (Soltis & Soltis, 1999; Rieseberg, 2001; Doyle *et al.*, 2008; Abbott *et al.*, 2013). Rapid, non-random and reproducible responses following hybridization were suggested among species of the wheat group (Ozkan *et al.*, 2001; Levy & Feldman, 2002).

The evolutionary trajectories of multiples TE families were investigated following various hybridization events, assessing specific reorganization as compared to whole genome sequences. Investigated TE families were selected based on evidence of recent activity in our model species (chapter 2). The selected wild wheat species have interconnected genome composition: four diploid *Aegilops* genomes (Figure 1, Chapter 'Introduction') that (i) differentially combined to produce derived tetraploids with consequences for genome evolution (investigated in chapter 3) and (ii) that, in turn, naturally hybridize as mimicked with artificial F1 hybrids (studied in chapter 4) and surveyed in natural hybrid zones (studied in chapter 5). Overall, evidence of drastic genome reorganization was consistently highlighted following hybridization, with multiple LTR retrotransposon families clearly playing a central role in host genome evolution. In particular, restructuring of selected TE family fractions was assessed following each investigated hybridization event. It suggests the need for hybrid genomes to undergo structural changes even in F1 hybrids.

An important observation was that the divergence between TE families in progenitors impacts on the reorganization level following hybridization. Data from spontaneous tetraploids (established polyploids, chapter 3) and F1 hybrids between tetraploid species (chapter 4), supported TE family specific conflicts revealed at genome merging and thus supported the 'Genome Shock Hypothesis' (McClintock, 1984). Noticeably, the arrangement of TE insertions between progenitors was associated with restructuring of corresponding TEs in both F1 hybrids (short-term restructuring, chapter 4) and established allopolyploids (long-term restructuring, Chapter 3). Despite differences in ploidy levels (i.e. diploids vs tetraploids) and underlying processes (i.e. hybridization vs hybridization associated with whole genome doubling), this correlation was repeatedly identified, which is congruent with hybridization being the main trigger of TE-induced genome reorganization (chapter 1). This result may further offer useful prediction of genome reorganization, with potential application in agronomy. In particular, it indicates that focusing on coding sequences or on some specific genome fraction may not be sufficient to predict reorganization, pointing out the importance of having knowledge of all genomic fractions.

In addition, this work indicated that TEs, at the family level, are highly implicated in genome restructuring of which lost of fragments was dominant and displayed species-specific and TE family-specific dynamic following allopolyploidy. Elimination of fragments was detected in chapter 3 and in chapter 5 (allopolyploidy) but also in chapter 4 (F1 hybrids). It was often more important in TE fractions than in random sequences, hypothesizing purging as a crucial mechanism to repress TE or allowing neo-hybrid to survive. Important reorganization was already assessed in F1 hybrids, indicating that genome restructuring occurs between homeologous chromosomes before meiosis. One hypothesis is that during neo-hybrid gametogenesis and early in their development, merging of divergent genomes in

one cell reveals qualitative or quantitative conflicts, correlated to progenitors TE family divergence. These conflicts would lead to TE activation that may drive genome restructuring and explain why restructuring is observed already in young F1 hybrids.

In addition to restructuring, methylation changes, when investigated in chapter 4, were also identified. They were associated to hybridization success and were especially important in TE fractions. The F1 hybrids (chapter 4) show methylation changes and are also the only one showing no TE specific restructuring.

Methodological considerations of genome evolution of species having complex genomes

Species of the wheat group, in particular *Aegilops* species, offer an ideal system to study genome evolution following hybridization. The main difficulty was to deal with the high diversity of TE families. In fact, complex genomes of *Aegilops* species had to be first described, as carried out here with the analyses of 454 data on *Ae. cylindrica* and *Ae. geniculata*. The need of an adapted method to select recently active TE families and then to develop molecular tools was a crucial step. The approach implemented here was successful to differentiate recently active from quiescent TE families and give tools to collect knowledge about candidate TE families from complex genomes. Most often the selected TE families revealed different evolutionary trajectories among species, stressing on the importance of considering TEs at the family level and on the need of an overall knowledge of the different TE families. Few TE family showed the same dynamics following all investigated hybridization events: *BARE1*, a family known to be active among Poaceae (Manninen & Schulman, 1993), and *Romani* consistently showed evidence of activation, whereas *Egug* and *Nusif* were mostly quiescent. In contrast, the majority of TE families (*Barbara*, *Cereba*,

Claudia, *Danae*, *Daniela*, *Derami*, *Fatima*, *Lila*, *Maximus*, *Quinta*, *WHAM*) showed hybrid-specific and family-specific evidence of proliferation. Two families, *Sabine* and *Xalax*, showed particularly contrasted dynamics depending on the hybrid considered with evidence of proliferation in some species and elimination in others. By investigating multiple TE families, our work showed that TEs does not automatically match genomic background dynamics and that each TE may display intrinsic dynamics. In addition, genome may contain a high diversity of TE families, consequently, knowledge about the dynamic of several TE provides a better overview of TEs as actors of species genomic evolution.

Patterns and processes underlying genome shock

Based on the evolutionary dynamics of multiple TE families in tetraploid species, as evaluated in chapters 2 and 3, conflicts between specific TE family was predicted based on the available model outlined in chapter 1, and tested through patterns of genome reorganization in artificial F1 hybrids studied in chapter 4. The results did not entirely corroborate our hypotheses. In particular, we expected to identify proliferation or evidence of repression (i.e. important fragment loss or de novo methylation) for conflicting families in specific F1 hybrids. Several new TE loci indicative of transposition events were identified in TEs such as *BARE1* or *Romani* for which conflicts were expected, but their proportions were not significantly different from random sequences. Consequently, it was not proved that TEs had independent and different evolutionary trajectories than genome-wide sequences. Concerning marks of repression, deletion of TE sequences was not assessed in the expected TE families in F1 hybrids, but novel methylation occurred specifically in the vicinity of TEs. Important novel CG methylation was assessed around some TEs in specific hybrids,

suggesting resolution of conflicts though methylation. Nevertheless, comparing which TE families displayed such pattern did not show entire consistency. Focusing for example on TEs that have similar pattern in the parents did not show systematically identical pattern in F1 hybrids. In addition, *Sabine*, which showed proliferation in *Ae. cylindrica* but massive elimination in *Ae. geniculata*, was expected to display conflicts when *Ae. geniculata* would have been pollen receiver (cytoplasm having few specific repressors). Surprising enough, it was not the case, but signs of specific methylation were assessed in the reciprocal hybrids and not in the expected hybrids. Discrepancies between predictions and observations may find origin in :

(i) Our predictions of conflicts were based only on restructuring data. Knowledge of patterns of methylation and siRNA loads in the progenitors may be necessary to precisely predict TE dynamics and repression strategies implemented by the host following selected hybridization events.

(ii) F1 hybrids were analyzed only at a specific time of their development (i.e. two weeks old seedlings), whereas conflicts may be revealed and have a strong impact at another development stage (embryo, radicle, seedling, adult plant). In addition, the two surmised regulatory processes (i.e. TE sequence elimination and methylation) may not occur at the same time and may represent complementary repressing mechanisms over evolutionary time. Analyzing F1 hybrids at different stages and investigating their endosperm may help to better understand how the processes controlling TE activity affect genomes and more precisely evaluate conflict predictions.

(iii) Our method based on fingerprinting focuses on proportions of TE loci being significantly higher than those of random sequences in order to

reflect TE-specific changes. Nevertheless, proliferating TE may only generate few new loci that our conservative approach fails to highlight.

(iv) Most crosses produced viable F1 hybrids that were then analyzed. Some seeds did not germinate and were thus not included. Such developmental failure may be due to important TE activation leading to detrimental genome reorganization, causing hybrid death. In addition, the endosperm is known to have unbalanced inheritance (2x maternal: 1x paternal) which may reinforce TE conflicts. Including undeveloped F1 seeds and their endosperms may provide additional pattern, matching or not, conflict predictions.

Perspectives

Overall, this work indicated hybrid-induced genome shocks associated with TE specific reorganization driving the evolution of host genomes to a large extent, and thus having a potential impact on hybrid evolution. Even if gradual changes occur along the hybrid lines, the initial genome shock is determinant to genome reorganization. Futures research directions will help to investigated more precisely conflicts between TEs and their impacts on their host genomes:

1. The focal species, with four genomes C, D, M and U having differentially combined into four natural tetraploids and in hybrids between the different tetraploid species, represent an excellent model to assess genome reorganization following hybridization and to test conflict predictions. First, to complete chapter 3, synthesized neo-tetraploid hybrids (with and without polyploidization) may precise patterns in timing of the

known genome reorganization and the selected TE fractions. In addition, only few artificial crosses with the tetraploid *Ae. crassa* were achieved in chapter 4, because most parental plants died during the extremely cold winter 2011-2012. It would be worthwhile to produce and analyze F1 hybrids made by crossing *Ae. crassa* with *Ae. cylindrica* as well as *Ae. geniculata* and *Ae. triuncialis* in order to compare all possible genome combinations, to fully assess links between differential hybridization and genome reorganization, and to determine if some genomes are more reorganized than others.

2. Methylation changes directed by siRNA repressor occurr following hybridization and are probably an important mechanism controlling TEs (Slotkin & Martienssen, 2007). More focus on methylation changes and their underlying triggers may help to better understand TE conflicts and subsequent TE evolutionary trajectories. Further research should analyze methylation changes in tetraploid species in accessions and in natural populations, and this in random sequences as for multiple TE fractions. Thus, in complement to methylation changes, analyzing siRNA may help to understand TE evolutionary conflicts and the timing of TE controlling. These additional investigations may also allow understanding respective maternal and paternal implication in TE repressing following hybridization.

3. Only surviving F1 hybrids were analyzed. Further research using embryo rescue would clarify whether embryos, but not endosperms, are viable and are specifically affected by TE conflicts and TE proliferation. Not only viable embryo should be analyzed, but also the ones that failed to develop a radicle. This would allow to further link TE dynamics, genome reorganization and hybrid viability at different

developmental stages. In particular, evaluating genome reorganization earlier or later than two weeks seedling may highlight the consequences of conflicts that are potentially already resolved or not yet visible.

4. Given limitations of fingerprinting methods described in chapter 3 and 4, using emerging technique such as retrotransposon capture array (Baillie *et al.*, 2011) may prove more sensitive, potentially allowing to follow active TE insertions rather than insertions of active TE families. This comparison may provide better understandings of whether TE families have followed the master gene model (only few active TE insertions) vs. transposon model (Brookfield & Johnson, 2006). In addition, if TEs follow the master gene model, it would be necessary to determine if few active TE insertions may explain host genome reorganization and abortive hybridization.

5. Only natural hybrid zone between *Ae. geniculata* and *Ae. triuncialis* was analyzed. Perspectives are to include several hybrid zones including more species in order to assess whether hybridization is common among *Aegilops* tetraploid species in natural conditions and how genome-wide and the multiple TE fractions are similarly or differentially reorganized. Possible hybrid zones may be sampled in the center of origin: *Ae. cylindrica* and *Ae. crassa*, *Ae. cylindrica* and *Ae. triuncialis*, *Ae. crassa* and *Ae. triuncialis*. Further investigations on possible hybrid zones between *Ae. cylindrica* and *Ae. geniculata, Ae. crassa* x *Ae. geniculata* have to be carried out. It would also be interesting to assess hybridization rate and reorganization among hybrid zone in the newly colonized areas (marginal areas) and to compare them with those in center of origin. Possible differential gene flow between species and genetic exchanges between central and marginal populations could be assessed.

References

Abbott R, Albach D, Ansell S, Arntzen JW, Baird SJE, Bierne N, Boughman J, Brelsford A, Buerkle CA, Buggs R, *et al.* 2013. Hybridization and speciation. *Journal of Evolutionary Biology* **26**: 229-246.

Baillie JK, Barnett MW, Upton KR, Gerhardt DJ, Richmond TA, De Sapio F, Brennan PM, Rizzu P, Smith S, Fell M, *et al.* 2011. Somatic retrotransposition alters the genetic landscape of the human brain. *Nature* **479**: 534-537.

Brookfield JFY, Johnson L. 2006. The evolution of mobile DNAs: when will transposons create phylogenies that look as if there is a master gene? *Genetics* **173**: 1115-1123.

Doyle J, Flagel L, Paterson A, Rapp R, Soltis D. 2008. Evolutionary genetics of genome merger and doubling in plants. *Annual Review of Genetics* **42**: 443-461.

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**: 97-113.

Levy AA, Feldman M. 2002. The impact of polyploidy on grass genome evolution. *Plant Physiology* **130**: 1587-1593.

Manninen I, Schulman AH. 1993. BARE-1, a copia-like retroelement in barley (Hordeum vulgare L.). *Plant Molecular Biology* **22**: 829-846.

McClintock B. 1984. The significance of responses of the genome to challenge. *Science* **226**: 792-801.

Ozkan H, Levy AA, Feldman M. 2001. Allopolyploidy-induced rapid genome evolution in the wheat (*Aegilops-Triticum*) group. *Plant Cell* **13**: 1735-1747.

Rieseberg LH. 2001. Polyploid evolution: Keeping the peace at genomic reunions. *Current Biology* **11**: R925-R928.

Slotkin RK, Martienssen R. 2007. Transposable elements and the epigenetic regulation of the genome. *Nature Reviews Genetics* **8**: 272-285.

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

Natacha Senerchia

Avenue de France 86
1004 Lausanne
☏ +41 76 532 78 71
@ : natacha.senerchia@unine.ch

Née le 4 août 1982 - Nationalité suisse

FORMATION ET DIPOMES

2010 à ce jour	**Thèse de doctorat,** Laboratoire de Botanique évolutive, Université de Neuchâtel.
2008-2009	**Master** en écologie et physiologie des plantes, Université de Neuchâtel.
2005-2008	**Bachelor en biologie** Université of Lausanne.
1998-2001	**Baccalauréat-Maturité** Gymnase de Beaulieu, Lausanne.
1989-1998	**Diplôme primaire et secondaire** Collège de Cossonay.

PUBLICATIONS

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

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

Senerchia N, Wicher 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.

Parisod & Senerchia. 2013. Responses of transposable elements to polyploidy. In: Plant Transposable Elements (Eds. Grandbastien & Casacuberta), Topics in Current Genetics, Springer 24: 147-168.

EXPERIENCES SCIENTIFIQUES

Février 2013	**Intervention orale**, Biology13 : Evolutionary dynamics of LTR retrotransposons in response to polyploidy"
Février-mai 2012	**Programme de mentoring StartingDoc** : Mentee.
Mai 2012	**Présentation d'un poster** : International Congree on Transposable Element (ICTE), St-Malo, France. « Evolutionary dynamics of LTR retrotransposons in complex genomes».
Août 2012	**Intervention orale** : European Society for Evolutionary Botany (ESEB), Tübingen, Allemagne. « Transposable elements in allopolyploid wild wheats and potential genomic conflicts at hybridization ».
Février 2010	**Intervention orale** : Meeting Wheat Consortium, Zürich, présentation des résultats 2009 du groupe de recherche.
Décembre 2009	**Intervention orale** : Biology 2009, Neuchâtel, Suisse. « Ecological consequences of introgression of transgenic wheat in a wild relative, *Aegilops cylindrica*».
Mars 2009	**Stage de laboratoire**: Agroscope ACW (Changins), Suisse. Tests de laboratoire (ELISA) et analyse de résultats sur le sujet de la sélection variétale.
Juillet-août 2008	**Botaniste:** Université de Lausanne (Prof. Antoine Guisan), Suisse. Relevé botanique non exhaustif dans les Alpes françaises, italiennes et suisses, recherche sur le modèle de niche écologique.