bioRxiv preprint doi: https://doi.org/10.1101/042325; this version posted June 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Transposable Element Evolution in the Allotetraploid *Capsella bursa-pastoris*

J. Arvid Ågren[1,2], Hui-Run Huang[3], Stephen I. Wright[1]

[1] Department of Ecology and Evolutionary Biology, University of Toronto

[2] Present address: Department of Molecular Biology and Genetics, Cornell University

[3] Key Laboratory of Plant Resources Conservation and Sustainable Utilization, South China Botanical Garden, the Chinese Academy of Sciences

Key words: C-value; genome size; mating system; polyploidy

Running head: Transposable element evolution in tetraploid *Capsella*

Word count: 2759

Abstract word count: 236

Number of figures: 5

Number of tables: 1

bioRxiv preprint doi: https://doi.org/10.1101/042325; this version posted June 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Abstract

Premise of the study

Shifts in ploidy affect the evolutionary dynamics of genomes in a myriad of ways. Population genetic theory predicts that transposable element (TE) proliferation may follow because the genome wide efficacy of selection should be reduced and the increase in gene copies may mask the deleterious effects of TE insertions. Moreover, in allopolyploids TEs may further accumulate because of hybrid breakdown of TE silencing. However, to date the evidence of TE proliferation following an increase in ploidy is mixed, and the relative importance of relaxed selection vs. silencing breakdown remains unclear.

Methods

We used high-coverage whole genome sequence data to evaluate the abundance, genomic distribution, and population frequencies of TEs in the self-fertilizing recent allotetraploid *Capsella bursa-pastoris* (Brassicaceae). We then compared the *C. bursa-pastoris* TE profile with that of its two parental diploid species, outcrossing *C. grandiflora* and self-fertilizing *C. orientalis*.

Key results

We found no evidence that *C. bursa-pastoris* has experienced a large genome wide proliferation of TEs relative to its parental species. However, when centromeric regions are excluded, we find evidence of significantly higher abundance of retrotransposons in *C. bursa-pastoris* along the gene-rich chromosome arms, compared to *C. grandiflora* and *C. orientalis*.

Conclusions

bioRxiv preprint doi: https://doi.org/10.1101/042325; this version posted June 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

The lack of a genome-wide effect of allopolyploidy on TE abundance, combined with the increases TE abundance in gene-rich regions suggest that relaxed selection rather than hybrid breakdown of host silencing explains the TE accumulation in *C. bursa-pastoris*.

bioRxiv preprint doi: https://doi.org/10.1101/042325; this version posted June 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Introduction

A central goal of population and comparative genomics research is to understand what factors drive the evolution of genome size and structure (Gregory, 2005; Lynch, 2007; Alföldi and Lindblad-Toh, 2013; Koenig and Weigel, 2015). Coupled with decades of information on genome size from flow cytometry across diverse lineages, the growing wealth of whole genome sequence data has revealed how extensive and rapidly genome size and structure can evolve, even among close relatives (Ungerer et al., 2006; Hawkins et al., 2009; Wright and Ågren, 2011; Tenaillon et al., 2011; Leitch and Leitch, 2013; Ågren and Wright 2015; Ågren et al., 2015). Yet our ability to explain this variation remains in its infancy.

Whole genome duplication via polyploidization has long been considered to be a major contributor to genome evolution (Adams and Wendel, 2005; Soltis and Soltis, 2012; Hollister 2015; Soltis et al. 2015). Most obviously, polyploidization will cause a direct increase in the total amount of DNA per cell. This initial doubling of genome size has often been followed by a process of diploidization, leading to a pattern of DNA loss over time (Leitch and Bennet, 2004; Lysak et al., 2009; Renny-Byfield et al., 2013; Vu et al., 2015). Furthermore, although the direct role of recent polyploidy on genome size evolution can be investigated and controlled for, most plant species have experienced a history of whole genome duplication in their evolutionary history (Vision et al. 2000; Jaillon et al., 2007; Jiao et al., 2011; Vanneste et al., 2014; Li et al. 2015). A lack of complete information on this history can therefore make it difficult to fully investigate the importance of whole genome duplication events on the evolution of genome size and structure.

bioRxiv preprint doi: https://doi.org/10.1101/042325; this version posted June 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

In addition to the direct effect of ploidy on genome size, transposable element (TE) proliferation may follow whole-genome duplications due to the masking of deleterious insertions and a reduction of the efficacy of selection across the genome caused by genome redundancy (recently reviewed in Parisod and Senerchia, 2012; Tayalé and Parisod, 2013). Furthermore, host-mediated silencing of TEs may be disrupted in allopolyploids (where an increase in ploidy is due to interspecific hybridization; Madlung et al., 2002; 2005; Kraitshtein et al., 2010; Yaakov et al., 2011). This combination of relaxed selection and a breakdown of silencing mechanisms could potentially drive dramatic evolution of genome structure following whole genome duplication. In particular, gene-dense euchromatic regions with very low TE content may experience major accumulation of TEs in genic regions. Such a mechanism may explain, for example, the dramatic transposable element expansion in the maize genome following whole genome duplication (Schnable et al., 2009; Baucom et al., 2009; Diez et al., 2014).

On the other hand, genome downsizing in polyploids may lead to a net loss of transposable elements during the process of diploidization (Parisod et al., 2010). The empirical evidence of changes in TE abundance following an increase in ploidy is equivocal. An increase in TE copy number was reported in *Nicotiana tabacum* (Petit et al., 2007; 2010; but see Renny-Byfield et al., 2011), whereas *Orobanche gracilis* appears to have experienced TE loss (Kraitschtein et al., 2010). Thus, overall, the factors driving proliferation vs. loss of transposable elements in polyploids remain poorly understood, and the relative importance of relaxed selection vs. silencing breakdown in TE accumulation is not clear.

bioRxiv preprint doi: https://doi.org/10.1101/042325; this version posted June 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Here, we use high-coverage whole genome sequence data to evaluate the abundance, genomic distribution, and population frequencies of TEs in the self-fertilizing allotetraploid *Capsella bursa-pastoris*. *Capsella bursa-pastoris* is a recently derived allotetraploid, with population genomic evidence for genome-wide reduction in the strength of selection on point mutations due to both gene redundancy and its selfing mating system (Douglas et al. 2015), making it an interesting model to examine the early fate of transposable elements following allopolyploidization. We look for evidence of TE proliferation beyond what would be expected from additivity of its two parental diploid species and use patterns of genomic distribution of TE insertions to distinguish the relative importance of relaxed selection vs. silencing breakdown. We discuss our results in light of the literature on the association between polyploidization, mating system, and TE abundance across plant species.

Material and Methods

Study system

The genus *Capsella* of the Brassicaceae family consists of four species with varying mating system, ploidy, and geographical distribution (Hurka et al. 2012; Figure 1). Selfing is thought to have evolved multiple times from an ancient progenitor of the diploid ($2n = 2x = 16$) *Capsella grandiflora*, a self-incompatible species restricted to Albania and northwestern Greece. Most recently, *Capsella rubella* diverged from *C. grandiflora* within the last 100,000 years (Foxe et al. 2009; Guo et al. 2009; Slotte et al. 2013; Brandvain et al. 2013). *Capsella orientalis* is thought to have evolved selfing prior to *C. rubella*, also from a *C. grandiflora*-like ancestor. *C. orientalis* and *C. grandiflora*

bioRxiv preprint doi: https://doi.org/10.1101/042325; this version posted June 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

have been inferred to have diverged approximately 930,000 years ago, providing the potential for a longer time period of mating system divergence (Douglas et al. 2015). Whereas *C. rubella* has expanded to a larger Mediterranean distribution, *C. orientalis* is now found in an area spanning Eastern Europe to Central Asia (Hurka et al. 2012). The origin of the world-wide distributed *Capsella bursa-pastoris* long remained elusive, but was recently determined to be an allotetraploid ($2n = 4x = 32$) following a hybridization event between *C. grandiflora* and *C. orientalis* within the last 100,000-300,000 years (Douglas et al. 2015). Consistent with this hybrid origin, a principal component analysis of the types of TEs found in *C. bursa-pastoris* puts it as an intermediate between *C. grandiflora* and *C. orientalis* and of all shared insertions found in two of the three species, the majority are between *C. bursa-pastoris* and either *C. grandiflora* or *C. orientalis*, with very few shared between *C. grandiflora* and *C. orientalis* to the exclusion of *C. bursa-pastoris* (Douglas et al. 2015). In this study, we expand the TE analysis of *C. bursa-pastoris*, to test for an accumulation of TEs following allopolyploid origins.

Identification and quantification of transposable elements

To compare the abundance, genomic locations, and population frequencies of TEs in the three species we combined the TE datasets generated by Ågren et al. (2014) and Douglas et al. (2015). These studies applied the PoPoolationTE pipeline of Kofler et al. (2012) on 108-bp paired-end Illumina reads on 7 C. *bursa-pastoris*, as well as 8 C. *grandiflora* and 10 *C. orientalis* individuals sampled across their respective geographical distributions, which provides the most comprehensive picture of TE evolution in each species. Ågren et al. (2014) also analysed TE distributions in *C. rubella*, but for this study we focus on the

bioRxiv preprint doi: https://doi.org/10.1101/042325; this version posted June 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

two direct progenitor species of *C. bursa-pastoris*, *C. grandiflora* and *C. orientalis*. We used the *C. rubella* reference genome (Slotte et al. 2013) and the TE database generated as part of the same study, which consisted of sequence data from seven Brassicaceae species, including *Capsella rubella*, *Brassica rapa*, *Arabis alpina*, *Arabidopsis thaliana* (accessions Col-0, Ler, Kro-0, Bur-0, and C24 from the 1001 *Arabidopsis* genomes project), *Arabidopsis lyrata*, *Eutrema halophila*, and *Schrenkiella parvulum.* Since the PoPoolation TE approach is designed for pooled population data, the original output is an estimate of the population frequency of each TE insertion. We adjusted the pipeline to use population frequencies to infer insertions as homo- or heterozygous. We ignored as spurious insertions with an estimated frequency of <0.2 and considered insertions with a frequency of >0.8 as homozygous. Intermediate frequency insertions were treated as heterozygous. Note that for our highly selfing tetraploid species, insertions identified as 'heterozygous' when mapping to the diploid *C. rubella* reference genome are in fact likely to be homozygous in one of the two homeologous genomes, and so for *C. bursa-pastoris*, intermediate frequency insertions were treated as homozygous at one of the two homeologues, while fixed insertions were treated as present in both homeologues. To avoid falsely inferring independent insertions due to the uncertainty in the method in the precise genomic position, we treated insertions as the same if the distance of the inferred location of two or more insertions across individuals was < 200 bp and the inferred TE family was identical. In previous work, we performed extensive tests to ensure that this approach could generally distinguish homo- and heterozygous insertions (Ågren et al. 2014). We used this approach to determine the abundance, genomic locations, and population frequencies of TEs in the three species.

bioRxiv preprint doi: https://doi.org/10.1101/042325; this version posted June 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Results

We quantified the abundance of four major categories of TEs: DNA, Helitrons, long terminal repeat (LTR) retrotransposons and non-LTR retrotransposons. The three species differ in their mean number of TEs (Kruskal-Wallis chi-squared = 21.342, df = 2, p < 0.00001), but genome-wide all species show similar relative abundance across elements, with LTR elements making up the bulk of the insertions (Figure 2).

To test whether *C. bursa-pastoris* has experienced an accumulation or loss of TEs following its origin, we calculated the expected diploid TE copy number from a *C. orientalis* × *C. grandiflora* hybrid and compared this number to the observed *C. bursa-pastoris* abundance. We randomly paired up *C. orientalis* and *C. grandiflora* chromosomes and calculated the average TE copy number of such a cross (where heterozygous insertions were given a copy number of 0.5). We performed 1,000 replicates of *in silico* crosses, sampling with replacement, based on the present-day TE abundances. We then compared the expected copy number to the observed abundance in *C. bursa-pastoris* and found that *C. bursa-pastoris* harbours slightly but significantly more insertions genome-wide than what would be expected under strict additivity (Figure 2; Wilcoxon rank sum test, $p = 0.01297$). Thus, overall we do not observe evidence for a major reduction in host silencing driving high rates of transposition.

Since TE insertions near genes will likely disrupt gene function, population genetic theory predicts that selection will rapidly remove such insertions (Dolgin and Charlesworth 2008). Following a whole-genome doubling event, a tetraploid like *C. bursa-pastoris* will carry twice as many gene copies as its diploid progenitors and the

bioRxiv preprint doi: https://doi.org/10.1101/042325; this version posted June 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

fitness cost of an insertion should therefore be less. As a consequence, tetraploids may be expected to accumulate more TEs near genes than diploids. To test this prediction, we first excluded the centromeric regions of the genome, following the annotation of Slotte et al. (2013), and considered TE abundance in the gene-rich chromosome arms only. Restricting our attention to these regions, we find that *C. bursa-pastoris* has a considerably higher TE abundance than expected from additivity (Figure 3; Wilcoxon rank sum test, $p = 0.0005828$), particularly for retrotransposons. Second, we used the gene annotation from the reference genome of *C. rubella* (Slotte et al. 2013) to calculate the distance to the closest gene for all TE insertions, in all three species. Again, just like the overall abundance, we are interested in whether *C. bursa-pastoris* has more insertions near genes than what would be expected by additivity from a *C. orientalis* × *C. grandiflora* cross. Using the approach outlined above, we calculated the expected TE copy number within 1000 bp of the closest gene from such a hybrid and compared it to the observed abundance in *C. bursa-pastoris*. We find that *C. bursa-pastoris* harbours significantly more insertions near genes, compared to what would be expected under strict additivity (Figure 4; Wilcoxon rank sum test, $p = 0.000126$).

We used the presence/absence of all TE insertions, across all individuals in the three species to categorize insertions as either singletons (present in only one individual) or non-singletons (present in more than one individual). We find that *C. grandiflora* has the highest proportion of singletons, potentially suggesting a higher TE activity and/or stronger purifying selection against insertions than *C. orientalis* and *C. bursa-pastoris*, which both show similar proportions of singletons (Table 1). However, differences in demographic history between the species are likely also contributing to the frequency

bioRxiv preprint doi: https://doi.org/10.1101/042325; this version posted June 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

spectrum, and the overall count of rare insertions is highest in *C. bursa-pastoris*. Overall, the combination of elevated copy number and a lower proportion of singletons in the tetraploid species is consistent with relaxed purifying selection following the transition to tetraploidy.

To investigate in more detail the relative importance of inherited TE insertions from parental species vs. new ongoing transposition events, we calculated the percentage of *C. bursa-pastoris* insertions found in each diploid progenitor species, separated by insertion frequency class (Figure 5). Note that for insertions found in counts greater than 7 (our sample size), this would imply that the insertion is found in both homeologous genomes at that position. As expected, low-frequency insertions are rarely found in the progenitor diploid species, suggesting a significant number of new insertions in the tetraploid, although clearly some of these cases may have been unsampled in the population but present in an ancestral diploid genome. In contrast, common insertions and those found on both homeologues tend much more often to be found in one or both diploid parental species, reflecting the 'parental legacy' (Buggs et al. 2014) of some insertions inherited in the tetraploid. Also as expected, more insertions are generally shared with *C. grandiflora*, reflecting their greater abundance, however a number of intermediate frequency insertions likely reflect fixed insertions found on the C. orientalis homeologue (Figure 5, insertion frequencies of '7' reflect fixed insertions on one homeologue). Even intermediate-frequency insertions however have a large fraction unsampled from diploid progenitors, suggesting relaxed selection on de novo insertions is allowing TEs to spread since polyploid origins. Overall, these patterns highlight both the

bioRxiv preprint doi: https://doi.org/10.1101/042325; this version posted June 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

additive, inherited contribution and the role of new de novo insertions in gene-rich regions in the TE complement of a recent tetraploid.

Discussion

Overall, we found no evidence that *C. bursa-pastoris* is experiencing a large-scale genome wide proliferation of TEs, as would be expected if there were a genome-wide breakdown of host silencing mechanisms. This is consistent with genome size estimates of the species and its prognitors, which does not indicate a non-additive increase in genome size (Hurka et al., 2012). However, we did detect a considerably higher abundance of TEs than expected when restricting our analysis to gene rich regions. These results are in line with previous work by Douglas et al. (2015) on genome-wide SNP patterns, which suggested that while this allopolyploid has not experienced a large-scale 'genome shock' since its origin, it is undergoing a global quantitative reduction in the efficacy of selection on amino acid and conserved noncoding mutations. Taken together, our results suggest that long-term relaxation of selective constraints is leading to TE accumulation in gene-rich regions, without a major shift in transposition rate.

One important consideration when predicting the effects of polyploidization on genome evolution may be its association with mating system. Under a number of population genetic models highly outcrossing species are predicted to experience higher rates of transposable element activity and copy number (Wright and Schoen, 1999; Morgan, 2001; Charlesworth and Wright, 2001). Allopolyploidization events that are associated with a retention of high rates of outcrossing could therefore represent a 'perfect storm', whereby TE activity remains high while genome redundancy enables

bioRxiv preprint doi: https://doi.org/10.1101/042325; this version posted June 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

rapid proliferation. On the other hand, polyploidy is often associated with elevated rates of self-fertilization compared with diploid relatives, and both asexual reproduction and high rates of selfing are common in polyploid lineages (reviewed in e.g. Mable, 2004; Husband et al., 2008; Robertson et al., 2011; Ramsey and Ramsey, 2014). Highly selfing lineages such as *C. bursa-pastoris* may thus experience a more modest increase in copy number than outcrossers; it would be of interest to investigate similarly-aged outcrossing allopolyploid lineages to assess whether TE accumulation is more dramatic in these species. On the other hand, selfing tetraploid lineages are more likely to experience severe founder events during polyploid origins, and strong genetic drift may further contribute to relaxed selection following whole-genome duplication as we observed in *C. bursa-pastoris* (Douglas et al., 2015).

It is notable that some of the most well-documented ancient TE expansion events, including maize (Schnable et al., 2009; Baucom et al., 2009; Diez et al., 2014) and the *Brassica* genus (Zhang and Wessler, 2004), are associated with ancient allopolyploidization events involving outcrossing lineages. Whether this is simply circumstantial or causal will require in-depth comparative analyses of the joint and unique effects of polyploidy and mating system on genome size and TE proliferation. Although the current age distribution of retroelements in the maize genomes suggests that TE proliferation was more recent than whole genome duplication (Bennett and Leitch 2005), this does not rule out ongoing TE accumulation due to relaxed selection over millions of years. As we gain increasingly detailed insights into the time since last whole genome duplication event in many lineages, investigating how this interacts with mating system to structure genome evolution is becoming increasingly feasible.

bioRxiv preprint doi: https://doi.org/10.1101/042325; this version posted June 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Acknowledgements

We thank two anonymous reviewers for their helpful comments on the manuscript. J.A.Å was supported by a Junior Fellowship from Massey College, S.I.W. by a Natural Sciences and Engineering Research Council (NSERC) Discovery Grant, and H.R.H by the National Natural Science Foundation of China (grant number 31370005) and a fellowship from the China Scholarship Council.

bioRxiv preprint doi: https://doi.org/10.1101/042325; this version posted June 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

References

Adams, K.L., and J.F. Wendel. 2005. Polyploidy and genome evolution in plants. *Current Opinion in Plant Biology* 8: 135–141.

Ågren, J.A., W. Wang, D. Koenig, B. Neuffer, D. Weigel, and S.I. Wright. 2014. Mating system shifts and transposable element evolution in the plant genus *Capsella*. *BMC Genomics* 15: 602.

Ågren, J.A., and S.I. Wright. 2015. Selfish genetic elements and plant genome size evolution. *Trends in Plant Science* 20: 195–196.

Ågren, J.A., S. Greiner, M.T.J. Johnson, and S.I. Wright. 2015. No evidence that sex and transposable elements drive genome size variation in evening primroses. *Evolution* 69: 1053–1062

Alföldi, J., and K. Lindblad-Toh. 2013.Comparative genomics as a tool to understand evolution and disease. *Genome Research* 23: 1063–1068.

Baucom, R.S., J.C. Estill, C. Chaparro, N. Upshaw, A. Jogi, J. Deragon, R.P. Westerman, et al. Exceptional diversity, non-random distribution, and rapid evolution of retroelements in the B73 Maize genome. *PLoS Genetics* 5: e1000732.

bioRxiv preprint doi: https://doi.org/10.1101/042325; this version posted June 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Bennett, M.D, and I.J. Leitch. 2005. Genome size evolution in plants. In T.R Gregory (ed) *The Evolution of the Genome*. Elsevier, San Diego, California, USA.

Buggs, R.J., Wendel, J.F., Doyle, J.J., Soltis, D.E., Soltis, P.S., and Coate, J.E. 2014. The legacy of diploid progenitors in allopolyploid gene expression patterns. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 369: 20130354.

Brandvain, Y., T. Slotte, K.M. Hazzouri, S.I. Wright, and G. Coop. 2013. Genomic identification of founding haplotypes reveals the history of the selfing species *Capsella rubella*. *PLoS Genetics* 9: e1003754.

Charlesworth, D., and S.I. Wright. 2001. Breeding systems and genome evolution. *Current Opinion in Genetics and Development* 11: 685–690.

Diez, C.M., E. Meca, M.I. Tenaillon, and B.S. Gaut. 2014. Three groups of transposable elements with contrasting copy number dynamics and host responses in the maize (*Zea mays* ssp. *mays*) genome. *PLoS Genetics* 10: e1004298.

Douglas, G., G. Gos, K. Steige, A. Salcedo, K. Holm, J.A. Ågren, K. Hazzouri, et al. 2015. Hybrid origins and the earliest stages of diploidization in the highly successful recent polyploid *Capsella bursa-pastoris*. *Proceedings of the National Academy of Sciences USA* 112: 2806–2811.

bioRxiv preprint doi: https://doi.org/10.1101/042325; this version posted June 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Foxe, J.P., T. Slotte, E.A. Stahl, B. Neuffer, H. Hurka, and S.I Wright. 2009. Recent speciation associated with the evolution of selfing in *Capsella*. *Proceedings of the National Academy of Sciences USA* 106: 5241–5245.

Gregory, T.R. 2005. *The Evolution of the Genome*. Elsevier, San Diego, California, USA.

Guo, Y.L., J.S. Bechsgaard, T. Slotte, B. Neuffer, M. Lascoux, D. Weigel, and M.H. Schierup. 2009. Recent speciation of *Capsella rubella* from *Capsella grandiflora*, associated with loss of self-incompatibility and an extreme bottleneck. *Proceedings of the National Academy of Sciences USA* 106: 5246–5251.

Hawkins, J.S., S.R. Proulx, R.A. Rapp, and J.F. Wendel. 2009. Rapid DNA loss as a counterbalance to genome expansion through retrotransposon proliferation in plants. *Proceedings of the National Academy of Science USA* 106: 17811–17816.

Hollister, J.D. 2015. Polyploidy: adaptation to the genomic environment. *New Phytologist* 205: 1034–1039.

Hurka, H., N. Friesen, D.A: German, A. Franzke, and B. Neuffer. 2012. 'Missing link' species *Capsella orientalis* and *Capsella thracica* elucidate evolution of model plant genus *Capsella* (Brassicaceae). *Molecular Ecology* 21: 1223–1238.

bioRxiv preprint doi: https://doi.org/10.1101/042325; this version posted June 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Husband, B.C., Ozimec, B., Martin, S.L., and L. Pollock. 2008. Mating consequences of polyploid evolution in flowering plants: current trends and insights from synthetic polyploids. *International Journal of Plant Sciences* 169: 195-206.

Koenig, K., and D. Weigel. 2015. Beyond the thale: comparative genomics and genetics of *Arabidopsis* relatives. *Nature Reviews Genetics* 16: 285–298.

Kofler, R., A.J. Betancourt, and C. Schlötterer. 2012. Sequencing of pooled DNA samples (Pool-Seq) uncovers complex dynamics of transposable element insertions in *Drosophila melanogaster*. *PLoS Genetics* 8: e1002487.

Jaillon, O., J.M. Aury, B. Noel, B., A. Policriti, C. Clepet, A. Casagrande, N Choisne, et al. 2007. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature* 449: 463–467.

Jiao, Y., N.J. Wickett, S. Ayyampalayam, A.S. Chanderbali, L. Landherr, P.E. Ralph, L.P. Tomsho, et al. 2011. Ancestral polyploidy in seed plants and angiosperms. *Nature* 473: 97–100.

Kraitshtein, Z., B. Yaakov, V. Khasdan, and K. Kashkush. 2010. Genetic and epigenetic dynamics of a retrotransposon after allopolyploidization of wheat. *Genetics* 186: 801–812.

bioRxiv preprint doi: https://doi.org/10.1101/042325; this version posted June 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Leitch, I.J., and M.D. Bennett. 2004. Genome downsizing in polyploid plants. *Biological Journal of the Linnean Society* 82: 651–663.

Leitch,I.J., and A.R. Leitch. 2013. Genome size diversity and evolution in land plants. In I.J. Leitch, J. Greilhuber, J. Doležel, J.F. Wendel (Eds.). *Plant Genome Diversity*, 307–322. Springer,Vienna, Austria.

Li, Z., A.E. Baniaga, E.B. Sessa, M. Scascitelli, S.W. Graham, L.H. Rieseberg, and M.S. Barker. 2015. Early genome duplications in conifers and other seed plants. *Science Advances* 1: e1501084.

Lynch, M. 2007. *The Origins of Genome Architecture*. Massachusetts, USA.

Lysak, M.A., M.A. Koch, J.M. Beaulieu, A. Meister, and I.J. Leitch. 2009. The dynamic ups and downs of genome size evolution in Brassicaceae. *Molecular Biology and Evolution* 26: 85-98.

Mable, B.K. 2004. Polyploidy and self compatibility: is there an association? 162: 803–811.

Madlung, A., R.W. Masuelli, B. Watson, S.H. Reynolds, J. Davison, and L. Comai. 2002. Remodeling of DNA methylation and phenotypic and transcriptional changes in synthetic *Arabidopsis* allotetraploids. *Plant Physiology* 129: 733–746.

bioRxiv preprint doi: https://doi.org/10.1101/042325; this version posted June 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Madlung, A., A.P. Tyagi, B. Watson, H. Jiang, T. Kagochi, R.W. Doerge, R. Martienssen, et al. 2005. Genomic changes in synthetic *Arabidopsis* polyploids. *The Plant Journal* 41: 221–230.

Morgan, M.T. 2001. Transposable element number in mixed mating populations. *Genetics Research* 77: 261–275.

Parisod C., K. Alix, J. Just, M. Petit, V. Sarilar, C. Mhiri, M. Ainouche, et al. 2010. Impact of transposable elements on the organization and function of allopolyploid genomes. *New Phytologist* 186: 37–45.

Parisod C., and N. Senerchia. 2013. Responses of transposable elements to polyploidy. In M.A. Grandbastien and J. Casacuberta (Eds.). *Plant Transposable Elements*, 147–168. Springer, Heidelberg, Germany.

Petit, M., C. Guidat, J. Daniel, E. Denis, E. Montoriol, Q.T Bui, K.Y. Lim, et al. 2010. Mobilization of retrotransposons in synthetic allotetraploid tobacco. *New Phytologist* 186: 135–147.

Petit, M., K.Y. Lim, E. Julio, C. Poncet, F.D. de Borne, A. Kovarik, A.R. Leitch, et al. 2007. Differential impact of retrotransposon populations on the genome of allotetraploid tobacco (*Nicotiana tabacum*). *Molecular Genetics and Genomics* 278: 1–15.

bioRxiv preprint doi: https://doi.org/10.1101/042325; this version posted June 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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

Renny-Byfield S., M. Chester, A. Kovařík, S.C. Le Comber, and MA Grandbastien, M. Deloger, R.A. Nichols et al. 2011. Next generation sequencing reveals genome downsizing in allotetraploid *Nicotiana tabacum*, predominantly through the elimination of paternally derived repetitive DNAs. *Molecular Biology and Evolution* 28: 2843–2854.

Renny□Byfield, A., Kovarik, L.J. Kelly, J. Macas, P. Novak, M.W. Chase, R.A. Nichols et al. 2013. Diploidization and genome size change in allopolyploids is associated with differential dynamics of low□and high□copy sequences. *The Plant Journal* 74: 829–839.

Robertson, K. A., E.E. Goldberg, and B. Igic. 2011. Comparative evidence for the correlated evolution of polyploidy and self-compatibility in Solanaceae. *Evolution* 65: 139–155.

Schnable, P.S., D. Ware, R.S. Fulton, J.C. Stein, F. Wei, S. Pasternak, C. Liang, et al. 2009. The B73 maize genome: complexity, diversity, and dynamics. *Science* 326: 1112–1115.

bioRxiv preprint doi: https://doi.org/10.1101/042325; this version posted June 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Slotte, T., K.M. Hazzouri, J.A. Ågren, D. Koenig, F. Maumus, Y.L. Guo, K. Steige, et al. 2013. The *Capsella rubella* genome and the genomic consequences of rapid mating system evolution. *Nature Genetics* 45: 831–835.

Soltis, P.S., and D.E. Soltis. 2012. *Polyploidy and Genome Evolution*. Springer, New York, New York.

Soltis, P.S., D. Blaine Marchant, Y. Van de Peer, and D.S. Soltis. 2015. Polyploidy and genome evolution in plants. *Current Opinion in Genetics and Development* 35: 199–125.

Tenaillon, M.I., M.B. Hufford, B.S Gaut, and J. Ross-Ibarra. 2011. Genome size and transposable element content as determined by high-thoughput sequencing in maize and *Zea luxurians*. *Genome Biology and Evolution* 3: 219–229.

Ungerer, M.C., S. Strakosh, and Y. Zhen. 2006. Genome expansion in three hybrid sunflower species is associated with retrotransposon proliferation. *Current Biology* 16: r872–r873.

Vanneste, K., G. Baele, S. Maere, and Y. Van de Peer. 2014. Analysis of 41 plant genomes supports a wave of successful genome duplications in association with the Cretaceous–Paleogene boundary. *Genome Research* 24: 1334–1347.

bioRxiv preprint doi: https://doi.org/10.1101/042325; this version posted June 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Vision, T.J., D.G. Brown, and S.D. Tanksley. 2000. The origins of genomic duplications in *Arabidopsis*. *Science* 290: 2114–2117.

Vu, G.T.H., T. Schmutzer, F. Bull, H.X. Cao, J. Fuchs, T.D. Tran, G. Jovtchev, et al. 2015. Comparative genome analysis revealed divergent genome size evolution in a carnivorous plant genus. *Plant Genome* 8: 3.

Wright, S.I, and D.J. Schoen DJ. 1999. Transposon dynamics and the breeding system. *Genetica* 107: 139–148.

Wright, S.I., and J.A. Ågren. 2011. Sizing up *Arabidopsis* genome evolution. *Heredity* 107: 509–510.

Zhang, X., and S.R. Wessler. 2004. Genome-wide comparative analysis of the transposable elements in the related species *Arabidopsis thaliana* and *Brassica oleracea*. *Proceedings of the National Academy of Sciences USA* 101: 5589–5594.

bioRxiv preprint doi: https://doi.org/10.1101/042325; this version posted June 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table 1 Number of singleton and non-singleton transposon insertions in three *Capsella* species

	Singletons	Non-singletons
C. orientalis	524 (20%)	2095 (80%)
C. grandiflora	3060 (29%)	7493 (71%)
C. bursa-pastoris	3768 (22%)	13358 (78%)

bioRxiv preprint doi: https://doi.org/10.1101/042325; this version posted June 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 1. Phylogenetic relationships within the *Capsella* genus. The evolution and natural history of the genus are thoroughly discussed in Hurka et al. (2012), Slotte et al. (2013), and Douglas et al. (2015).

Figure 2. Boxplots of the genome wide abundance of transposable elements (TEs) in the three *Capsella* species. The expected *C. bursa-pastoris* value was generated by performing 1,000 replicates of in silico crosses between *C. orientalis* × *C. grandiflora*, sampling with replacement.

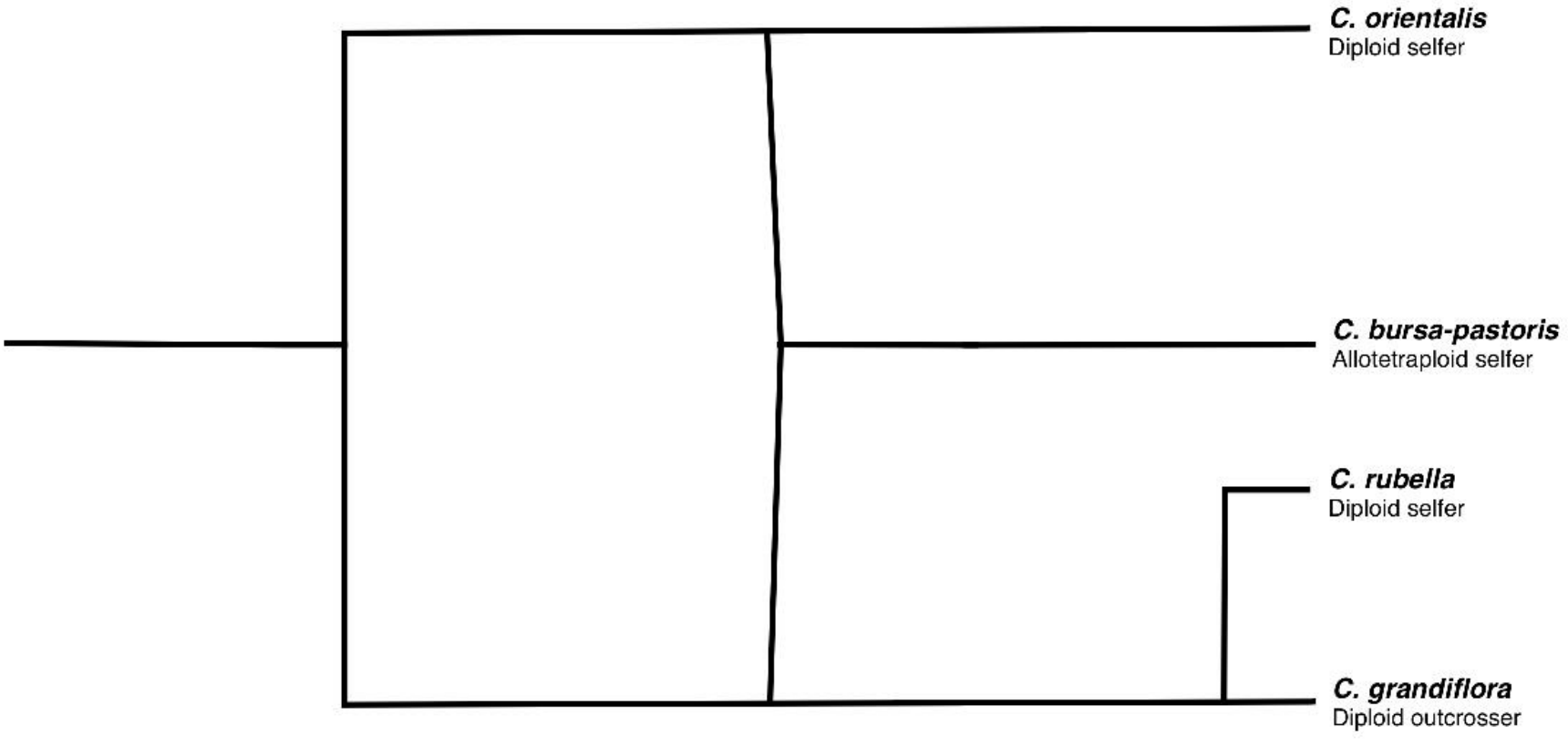
Figure 3. Boxplots of the abundance of transposable elements (TEs) in the three *Capsella* species in the gene rich chromosome arms (centromeric regions excluded). The expected *C. bursa-pastoris* value was generated by performing 1,000 replicates of in silico crosses between *C. orientalis* × *C. grandiflora*, sampling with replacement.

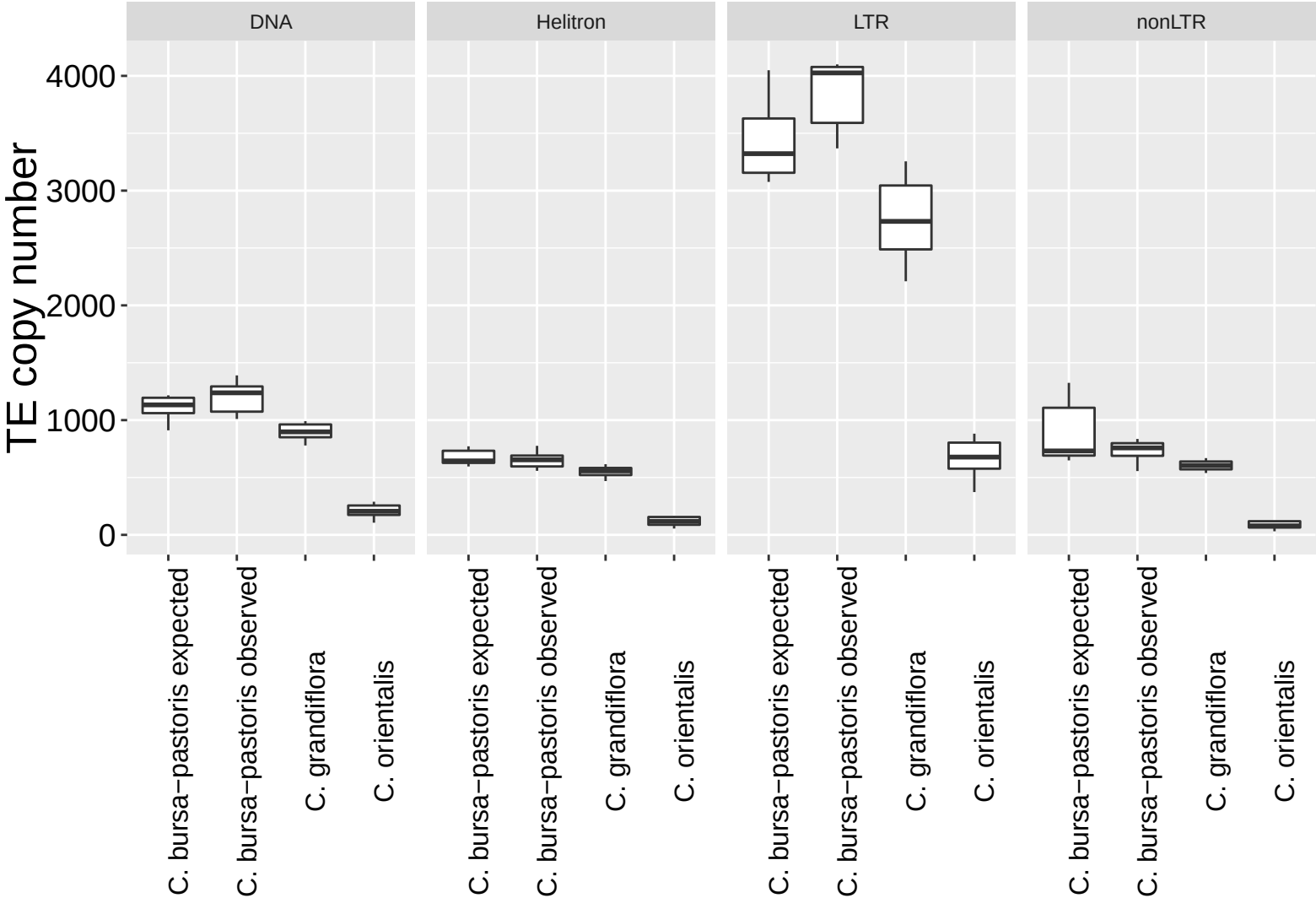
Figure 4. Average abundance of transposable elements (TEs) in 100 bp bins near their closest gene in the three *Capsella* species. Error bars are ± 1 standard error. The expected *C. bursa-pastoris* value was generated by performing 1,000 replicates of in silico crosses between *C. orientalis* × *C. grandiflora*, sampling with replacement.

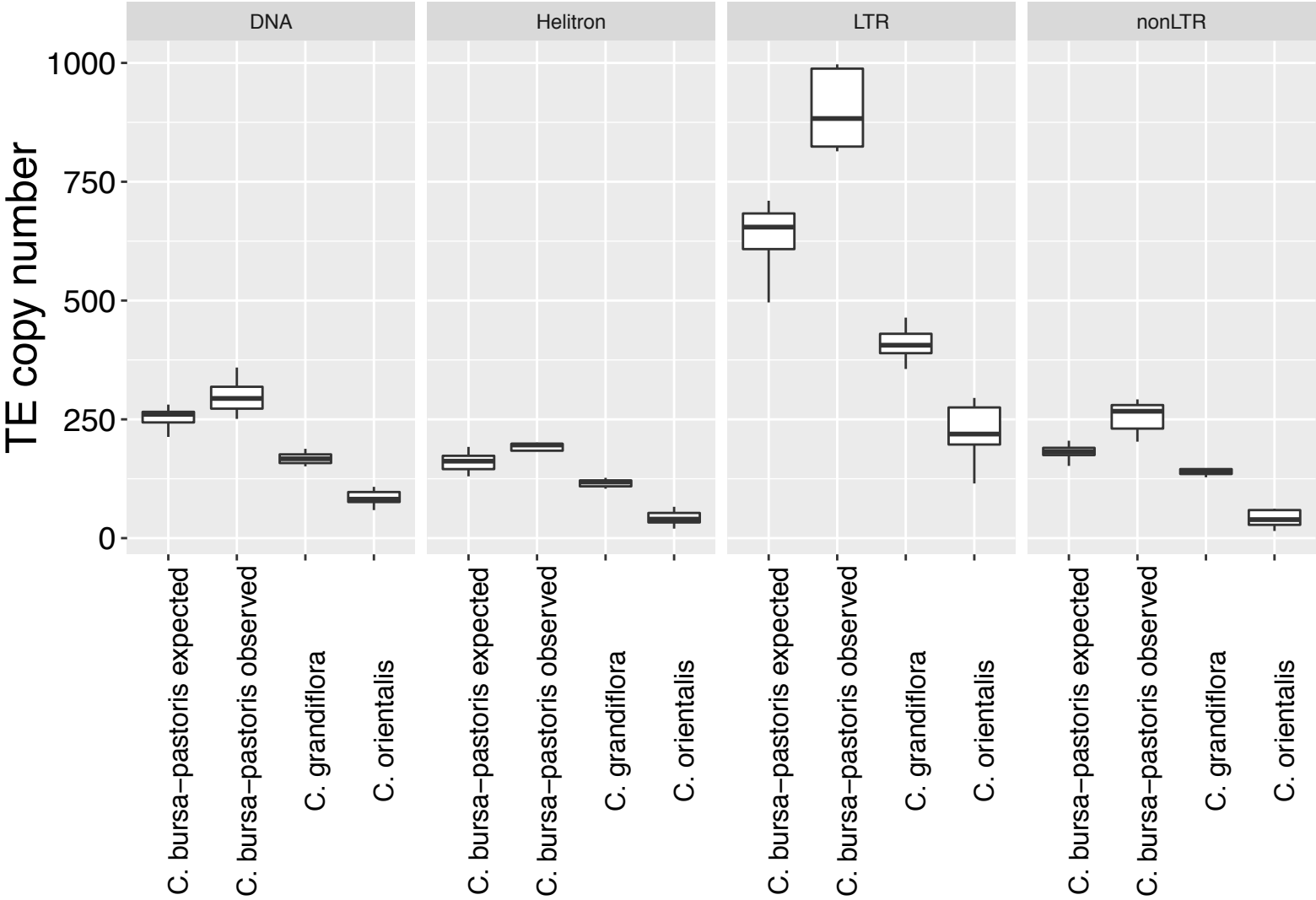
Figure 5. Proportion of *C. bursa-pastoris* transposable element (TE) insertions found in the diploid progenitors *C. orientalis*, *C. grandiflora*, or both. Proportions are separated by TE frequency class in *C. bursa-pastoris*, where the counts represent the number of copies

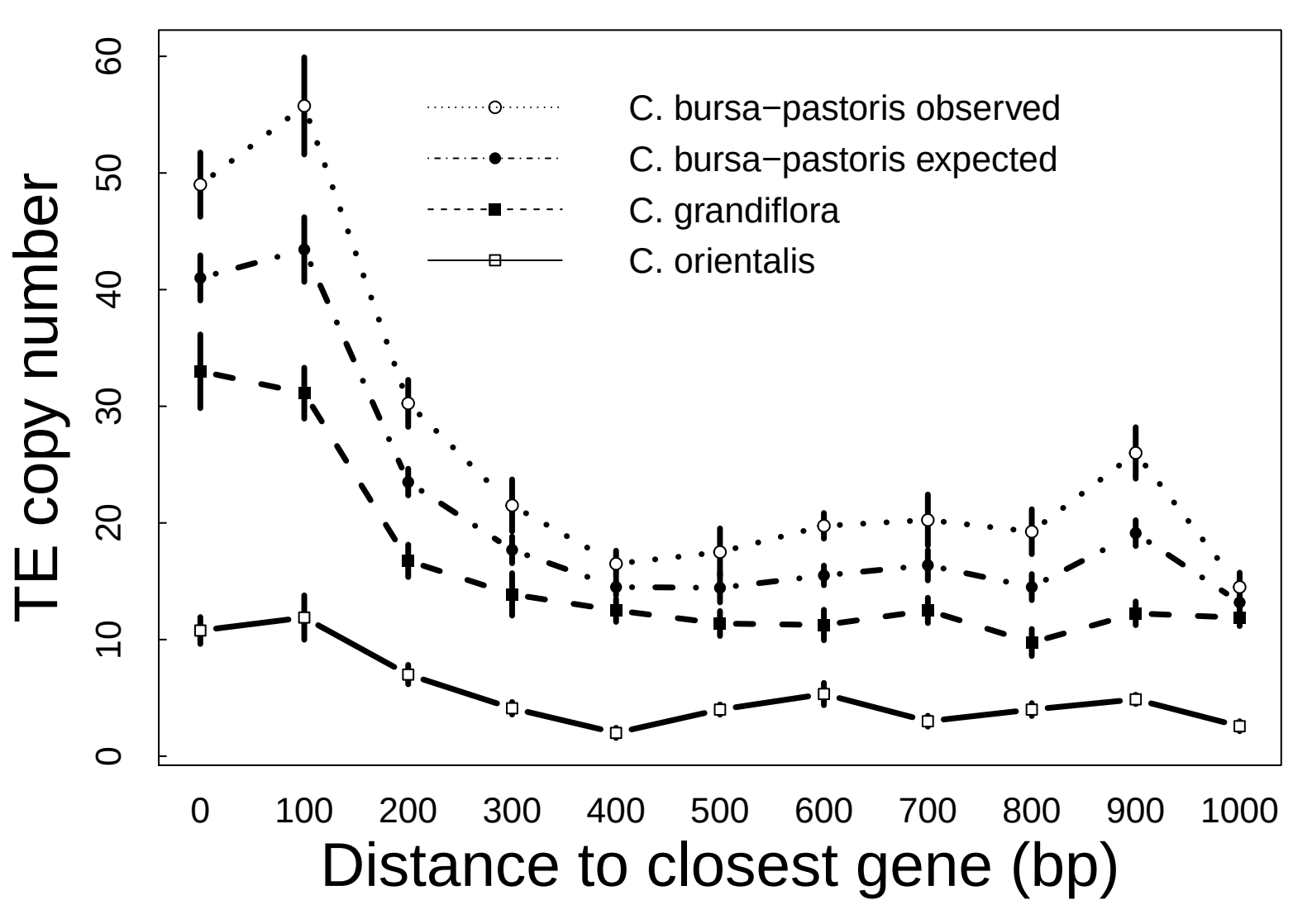
bioRxiv preprint doi: https://doi.org/10.1101/042325; this version posted June 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

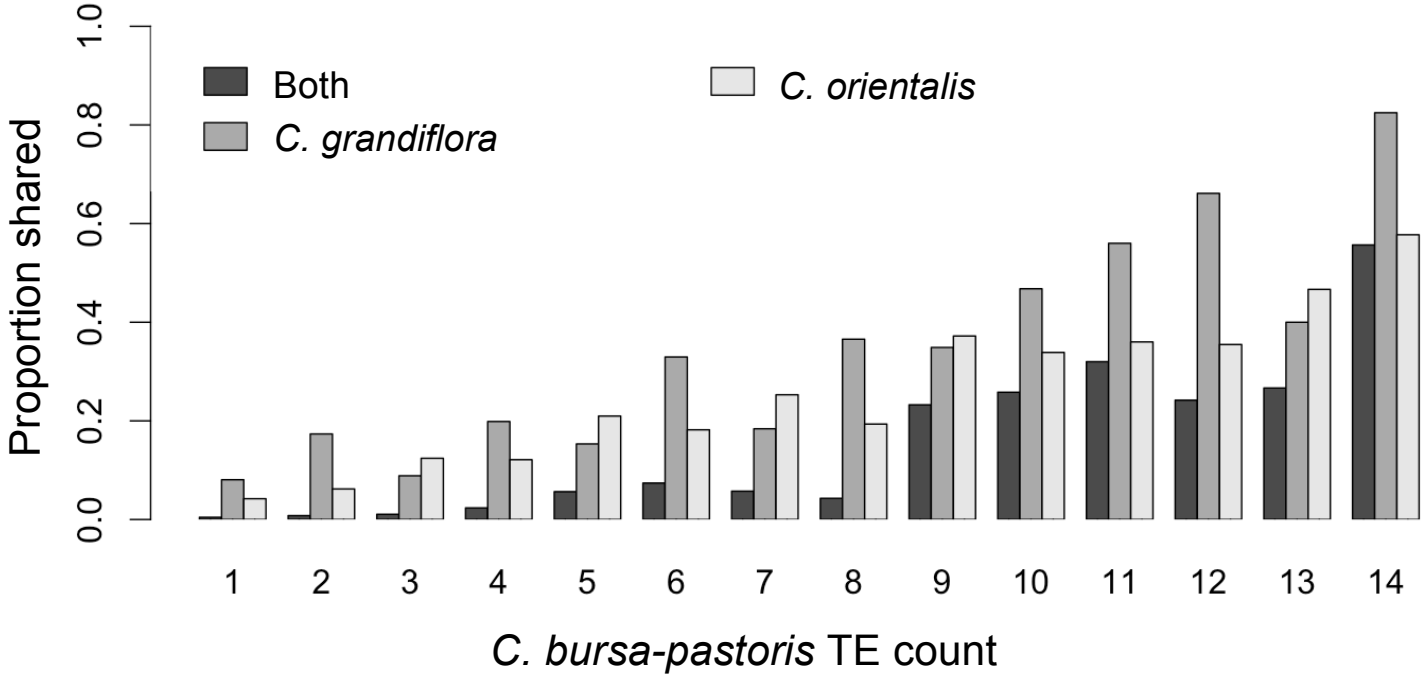
of the insertion in a sample size of 7 *C. bursa-pastoris* individuals. Note that counts greater than 7 imply the insertion is found on both homeologous copies of the genome, and counts of 7 are likely to be cases of fixation events in one of the two homeologues.


C. orientalis
Diploid selfer
C. bursa-pastoris
Allotetraploid selfer
C. rubella
Diploid selfer
C. grandiflora
Diploid outcrosser


DNA
Helitron
LTR
nonLTR
TE copy number
4000
3000
2000
1000
0
C. bursa−pastoris expected
C. bursa−pastoris observed
C. grandiflora
C. orientalis


DNA
Helitron
LTR
nonLTR
TE copy number
1000
750
500
250
0
C. bursa−pastoris expected
C. bursa−pastoris observed
C. grandiflora
C. orientalis


C. bursa−pastoris observed
C. bursa−pastoris expected
C. grandiflora
C. orientalis
TE copy number
0
10
20
30
40
50
60
0
100
200
300
400
500
600
700
800
900
1000
Distance to closest gene (bp)


Proportion shared
0.0
0.2
0.4
0.6
0.8
1.0
Both
C. grandiflora
C. orientalis
1
2
3
4
5
6
7
8
9
10
11
12
13
14
C. bursa-pastoris TE count